The Synaptic Weight Matrix:
Dynamics, Symmetry Breaking, and Disorder

Francesco Fumarola

Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY
2018
ABSTRACT

The Synaptic Weight Matrix: Dynamics, Symmetry Breaking, and Disorder

Francesco Fumarola

A key role in simplified models of neural circuitry (Wilson and Cowan, 1972) is played by the matrix of synaptic weights, also called connectivity matrix, whose elements describe the amount of influence the firing of one neuron has on another.

Biologically, this matrix evolves over time whether or not sensory inputs are present, and symmetries possessed by the internal dynamics of the network may break up spontaneously, as found in the development of the visual cortex (Hubel and Wiesel, 1977). In this thesis, a full analytical treatment is provided for the simplest case of such a biological phenomenon, a single dendritic arbor driven by correlation-based dynamics (Linsker, 1988). Borrowing methods from the theory of Schrödinger operators, a complete study of the model is performed, analytically describing the break-up of rotational symmetry that leads to the functional specialization of cells. The structure of the eigenfunctions is calculated, lower bounds are derived on the critical value of the correlation length, and explicit expressions are obtained for the long-term profile of receptive fields, i.e. the dependence of neural activity on external inputs.

The emergence of a functional architecture of orientation preferences in the cortex is another crucial feature of visual information processing. This is discussed through a model consisting of large neural layers connected by an infinite number of Hebb-evolved arbors. Ohshiro and Weilky (2006), in their study of developing ferrets, found correlation profiles of neural activity in contradiction with previous theories of the phenomenon (Miller, 1994;
Wimbauer, 1998). The theory proposed herein, based upon the type of correlations they measured, leads to the emergence of three different symmetry classes. The contours of a highly structured phase diagram are traced analytically, and observables concerning the various phases are estimated in every phase by means of perturbative, asymptotic and variational methods. The proper modeling of axonal competition proves to be key to reproducing basic phenomenological features.

While these models describe the long-term effect of synaptic plasticity, plasticity itself makes the connectivity matrix highly dependent on particular histories, hence its stochasticity cannot be considered perturbatively. The problem is tackled by carrying out a detailed treatment of the spectral properties of synaptic-weight matrices with an arbitrary distribution of disorder. Results include a proof of the asymptotic compactness of random spectra, calculations of the shape of supports and of the density profiles, a fresh analysis of the problem of spectral outliers, a study of the link between eigenvalue density and the pseudospectrum of the mean connectivity, and applications of these general results to a variety of biologically relevant examples.

The strong non-normality of synaptic-weight matrices (biologically engineered through Dale’s law) is believed to play important functional roles in cortical operations (Murphy and Miller, 2009; Goldman, 2009). Accordingly, a comprehensive study is dedicated to its effect on the transient dynamics of large disordered networks. This is done by adapting standard field-theoretical methods (such as the summation of ladder diagrams) to the non-Hermitian case. Transient amplification of activity can be measured from the average norm squared; this is calculated explicitly for a number of cases, showing that transients are typically amplified by disorder. Explicit expressions for the power spectrum of response
are derived and applied to a number of biologically plausible networks, yielding insights into the interplay between disorder and non-normality. The fluctuations of the covariance of noisy neural activity are also briefly discussed.

Recent optogenetic measurements have raised questions on the link between synaptic structure and the response of disordered networks to targeted perturbations. Answering to these developments, formulae are derived that establish the operational regime of networks through their response to specific perturbations, and a minimal threshold is found to exist for counterintuitive responses of an inhibitory-stabilized circuit such as have been reported in Ozeki et al. (2016), Adesnik (2016), Kato et al. (2017). Experimental advances are also bringing to light unsuspected differences between various neuron types, which appear to translate into different roles in network function (Pfeffer et al., 2013; Tremblay et al., 2016). Accordingly, the last part of the thesis focuses on networks with an arbitrary number of neuronal types, and predictions are provided for the response of networks with a multipopulation structure to targeted input perturbations.
Contents

Acknowledgments xiii

1 Introduction 1

2 Biological foundations 6
   2.1 Fundamentals of neural electrodynamics 6
   2.2 Derivation of rate models 8
   2.3 Range of validity 13
   2.4 Equivalence of alternative rate models 13

3 Hebbian dynamics and the Linsker cell 16
   3.1 Hebbian dynamics in the visual cortex 16
      3.1.1 Hebbian dynamics 16
      3.1.2 Visual cortex development 17
      3.1.3 Hebbian theories of orientation selectivity 20
   3.2 Diagonalization of the Linsker operator 25
      3.2.1 Diagonalization in polar coordinates 25
      3.2.2 Diagonalization in Cartesian coordinates 30
      3.2.3 Connecting the polar and Cartesian representations 32
3.3 Homeostatic constraints .................................................. 36
  3.3.1 Positive semidefiniteness of the constrained operator .......... 38
  3.3.2 Matrix elements of the constrained operator ....................... 39
3.4 Diagonalization of the constrained model ............................... 41
  3.4.1 Long-range solution .................................................. 41
  3.4.2 Moveable node theory ............................................... 42
  3.4.3 A lower bound for the transition point .............................. 46
3.5 Conclusions ..................................................................... 49

4 A multilayer model of orientation development ................................ 52
  4.1 The Hebbian theory of orientation maps .............................. 52
    4.1.1 Review of the experimental literature ............................. 52
    4.1.2 Unconstrained dynamics of the model ......................... 56
    4.1.3 Homeostatic constraint ............................................. 60
    4.1.4 Projection operators ................................................. 63
  4.2 Properties of the time-evolution operator ............................ 64
    4.2.1 Matrix elements ...................................................... 65
    4.2.2 Fourier Transform ................................................... 66
    4.2.3 Positive semidefiniteness ......................................... 68
    4.2.4 Long-term dynamics ................................................ 69
    4.2.5 Symmetries of the system: translations and rotations .......... 70
    4.2.6 Symmetries of the system: parity and CP symmetry .......... 71
    4.2.7 Diagonalization of the unconstrained dynamics ............... 74
5.1 The open problem of synaptic stochasticity ........................................ 125
  5.1.1 Background and motivation ...................................................... 125
  5.1.2 Random matrices beyond quantum mechanics .............................. 126
  5.1.3 Outline of the treatment .......................................................... 129

5.2 General results on spectral properties ........................................... 130
  5.2.1 Preliminary definitions ............................................................. 130
  5.2.2 The shape of supports and the density profiles ............................. 133
  5.2.3 Formulation in terms of singular values ..................................... 135
  5.2.4 Purely random synaptic weight matrices ................................... 136
  5.2.5 Ordering of limits ................................................................. 137
  5.2.6 Relationship to pseudospectra ................................................ 141

5.3 Calculation of spectral properties ................................................ 144
  5.3.1 The shape of supports and density profiles ................................. 144
  5.3.2 Singular value formulation ....................................................... 157
  5.3.3 Validity of the non-crossing approximation ................................ 158
  5.3.4 Proving the compactness of the supports ................................... 169
  5.3.5 The case of zero mean connectivity ......................................... 175

5.4 Applications to specific networks .................................................. 180
  5.4.1 Schur structure of the mean connectivity .................................... 181
  5.4.2 Networks with a single feedforward chain .................................. 182
  5.4.3 Examples motivated by Dale’s law ............................................ 185
  5.4.4 Linearizations of nonlinear networks ....................................... 201

5.5 Conclusions .................................................................................... 206
6 Disordered neural dynamics

6.1 Dynamical functions of non-normality ........................................... 208

6.2 General results ................................................................. 211

6.2.1 Preliminary remarks ...................................................... 211

6.2.2 Result for the average squared norm ...................................... 212

6.2.3 Results for the power spectrum ............................................ 213

6.2.4 The amplifying effects of disorder ......................................... 216

6.2.5 Conditions for applicability .............................................. 217

6.3 Derivation of the general results ............................................. 219

6.3.1 Formulation in terms of diffusive kernel .................................. 219

6.3.2 Summing up the diffuson .................................................. 224

6.4 Applications to specific networks ........................................... 232

6.4.1 Feedforward chain with alternating eigenvalues ...................... 233

6.4.2 Feedforward chain with null eigenvalues ............................... 235

6.4.3 Network with two-state feedforward components .................... 239

6.5 Beyond the non-crossing approximation .................................. 242

6.6 Disorder fluctuations of the covariance .................................... 244

7 Response to targeted input perturbations .................................. 249

7.1 Measuring the response to input perturbations ........................... 249

7.2 Modeling paradoxical responses ............................................ 252

7.2.1 Are excitatory subnetworks independently stable? ................. 252

7.2.2 A uniform excitatory-inhibitory network ............................... 257
Appendix C: Applications of the random-matrix results

.6  C1: Single feedforward chain of length $N$ ........................................ 323
.7  C2: $N/2$ feedforward chains of length 2 ............................................. 330
.8  C3: Composite network with factorizable weights ............................... 341

Appendix D: Moments of the same-frequency covariance ............................. 346
Acknowledgments

I owe an important debt of gratitude to the following people: Professor Robert D. Mawhinney, who has made it possible for me to carry on research outside the physics department; Professor Kenneth D. Miller, who has been my doctoral advisor in the department of neuroscience; Dr. Yashar Ahmadian, with whom Prof. Miller and I worked on the random-matrix projects discussed in chapters 5 and 6; and Dr. Rainer J. Engelken, who has proofread parts of this thesis and has provided advice throughout its writing.
Chapter 1

Introduction

Information is processed in the brain by the coordinated activity of tens of billions of neural cells. However, between the dynamics of a single cell and the emerging behavior of the full high-dimensional complex system (capable of such tasks as processing sensory information, making decisions, and controlling behavior) there is a chasm that involves multiple operational scales, which no single theory is able to bridge.

Depending on the specific question, a diversity of theoretical concepts and perspectives are needed, corresponding to various levels of coarse-graining. At one extreme, we have ”reductionist” models that attempt to account for the detailed biophysics of each cell in a network; at the other extreme, there is the top-down approach of theories that start from information-theoretic, geometric, or evolutionary constraints to infer how computations ought to be performed by a functioning neural circuit.

Both extremes may be essential to the progress of neuroscience, but so can an intermediate approach, neither normative nor microscopic, aimed at building what may be called ”minimal mechanistic models” of brain function. In this spirit, one assembles network models by starting from well-documented mechanistic features of neural units, but simplifies their description in an attempt to single out only the relevant degrees of freedom, i.e. those that are needed to reproduce the observed functional behavior of the network under
This approach has some possible disadvantages. Only a few biophysical variables can be accounted for in every such model, hence one has to proceed by trial and error to identify mechanistic features that are relevant to functional behavior. Moreover, details in the observed neural statistics often lie beyond the predictive power of such simplified models.

The approach has also some important advantages. First, it yields mathematically tractable models, i.e. models that have a manageable parameter space and to which known techniques may be applied, so that the operational regimes of the network under study can be often mapped out. Second, this approach is suitable to answering questions as to what details matter in the recorded behavior of a neurobiological system; this refers, in particular, to identifying the mechanisms through which properties of single cells (and of the connections between them) can affect the network dynamics. The information thus garnered is mathematically transparent and can used as guidance for further, more complex models, as well as for data analysis, and for the design of new experiments.

Among the approaches of this type, the most common one characterizes each neuron-neuron connection with a single variable, known as "synaptic weight". Studying the functional capabilities of the network amounts, then, to computing properties of the matrix formed by these quantities. This concerns in particular its symmetry properties, its dynamical properties, and its spectral properties.

It is definitely not obvious that synaptic transmission can be simplified as involving a single degree of freedom. However, this is not without biological justification, and we will devote the first part of this thesis to demonstrating it. In Chapter 2, we will review the fundamentals of neural circuitry, will mention the models commonly employed, and will
describe the simplifications we choose to adopt.

We will then carry out our program for one of the best-documented skills of cortical circuits – the extraction of higher-level information from visual input. A typical task the brain performs daily consists in determining the orientation of objects observed in the environment. In Chapter 3, we will analyze a minimal mechanistic model of how a single neural cell can learn to recognize a certain visual orientation. We will thus derive equations for the evolution of the synaptic weight matrix and, exploiting an analogy with quantum-mechanical problems, we will fully characterize their solution, arriving at a number of results previously obtained only through simulations.

Cortical cells are known to be arranged on the basis of their functional specialization. For instance, the input features that make a visual-cortex cell fire (color, direction of movement, orientation) depend continuously on the cell’s position within the cortex. In Chapter 4, we will perform a comprehensive analysis of how such functional organization, as concerns orientations, can be achieved in the presence of plausible biological constraints. This will provide an explanation for outstanding questions posed by recent experimental results, and we will discuss the reasons why previous theories have failed to explain them.

In spite of the underlying simplifications, a model based on synaptic weights can be difficult to analyze because of the large number of degrees of freedom involved. Indeed, the weight matrix has as many entries as the square of the number of neurons, and the neural population involved in any functioning cortical circuit is typically very large. This poses problems even for such basic properties of networks as the eigenvalue spectrum of the matrix, all the more so because these matrices are non-normal, and the physics literature has focused mostly on Hermitian operators.
In Chapter 5, we will provide a comprehensive characterization of the spectral properties of large synaptic weight matrices possessing a stochastic component. As in the previous chapters, techniques from physics will be extensively borrowed. We will prove the compactness of the support of eigenvalues in the thermodynamic limit, we will be able to trace its shape analytically, and we will calculate explicitly the profile of eigenvalue density within it. The potential applications of these results to neurobiology are numerous; we will work out in detail those that concern network prototypes most often used these days in neuronal modeling.

How information is encoded, processed and transmitted by neural circuits is related to collective dynamics that evolves on multiple time scales and that may be too complex to be exhaustively described by the mere eigenvalues of weight matrices. Moreover, the strong non-normality of the synaptic weight matrix produces large transients in neural dynamics, which can be relevant in multiple ways to the functionality of cortex, and hence cannot be neglected by theory.

A full chapter will be devoted, therefore, to deriving expressions for strictly dynamical observables (Chapter 6). These include the frequency power spectrum and the magnitude of transients as measured by the average norm squared, which have never been computed analytically for stochastic networks. The technique, which consists of a summation of ladder diagrams, will be adapted from standard field theory. Applications to several biologically plausible weight matrices will then be analyzed. We will also briefly focus on signatures of the linear instability in noisy dynamics.

Technological advances such as optogenetics provide unprecedented tools to measure and manipulate the activity of neurons in a specifically targeted fashion. To respond to these
developments, one needs to know how a network would respond to changes in the input to a given subpopulation or fractions thereof, in any of its operational regimes. In Chapter 7, we will calculate steady-state response functions for an elementary network and, as a result, we will obtain possible constraints on the interpretation of experimental findings that may serve to guide optogenetic experiments. We will then consider the response of a network with a larger number of populations.

Finally, the significance of the results and their connection to the rest of the literature will be reviewed in Chapter 8. We will then discuss in some detail the avenues opened, for future research, both by the results obtained herein and by the techniques we developed. A number of original calculations that supplement the work presented in this thesis can be found in Appendices A-D.
Chapter 2

Biological foundations

2.1 Fundamentals of neural electrodynamics

Neurons, like all other cells in our body, are packed with a vast number and variety of ions and molecules, many of which carry charges.

Most of the time, there is an excess concentration of negative charge inside a neuron. Excess charges that are mobile, like ions, repel each other and build up on the inside surface of the cell membrane. Electrostatic forces attract an equal density of positive ions from the extracellular medium to the outside surface of the membrane, amounting to a total excess charge $Q$.

The semi-permeable cell membrane creates thus a capacitance $C$, and the voltage across the membrane is $v = Q/C$. From this we can determine how much current is required to change the membrane potential at a given rate, namely

$$C_m \frac{dV}{dt} = \frac{dQ}{dt}.$$  \hspace{1cm} (2.1)

Since the membrane is a lipid bilayer, essentially impermeable to most charged molecules, the membrane current $I_m = \frac{dQ}{dt}$ would be zero were it not for the ion channels embedded in the membrane.
Ions are driven through ion channels by a combination of electric forces and diffusion. On one hand, negative membrane potentials attract positive charges into the neuron; on the other hand, ions diffuse through channels because the ion concentrations differ inside and outside the neuron, and these differences are maintained by ion pumps within the cell membrane.

Let the index $\alpha$ label the various types of ion channels. We can define the reversal potential $E_\alpha$ as the membrane potential at which the current flow due to electric forces cancels the diffusive flow through channels of type $\alpha$.

A conductance with a reversal potential $E_\alpha$ tends to move the membrane potential of the neuron toward the value $E_\alpha$. When $V > E_\alpha$ this means that positive current will flow outward, and when $V < E_\alpha$ positive current will flow inward. The difference $V - E_\alpha$ is called the driving force, and the membrane current due to a given channel type is $g_\alpha(V - E_\alpha)$, where $g_\alpha$ is the conductance due to these channels.

Summing over different types of channels, we obtain the total membrane current

$$I_m = \sum_{\alpha=1}^{N_t} g_\alpha (v - E_\alpha), \quad (2.2)$$

where $N_t$ is the number of channel types, and $I_m$ depends on time both through $g_\alpha$ and through $v$.

Piecing together Eq. (2.1) with Eq. (2.2) gives, for the membrane voltage of the $i^{th}$ neuron in a network:

$$C_i \frac{dv_i}{dt} = \sum_{j=1}^{N_c} g_{ij} (E_{ij} - v_i), \quad (2.3)$$
where $C_i$ is the capacitance of the $i^{th}$ neuron, $N_c$ is the number of conductances, $g_{ij}$ is the $j^{th}$ conductance onto the neuron, with reversal potential $E_{ij} = E_{\alpha_j}$, where $\alpha_j$ is the type of the $j$-th channel.

A current of positively charged ions flowing out of the cell through open channels makes the membrane potential more negative (hyperpolarization); current flowing into the cell changes the membrane potential to less negative or even positive value (depolarization). If a neuron is depolarized sufficiently to raise the membrane potential above a threshold level, a positive feedback process is initiated, and the neuron spikes, i.e. it generates an action potential that will propagate along axons, carrying signal to other neurons.

### 2.2 Derivation of rate models

The capacity of channels for conducting ions across the cell membrane can be modified by many factors including the membrane potential (for voltage-dependent channels), the internal concentration of various intracellular messengers (e.g. for Ca$^{2+}$-dependent channels), and the extracellular concentration of neurotransmitters or neuromodulators (e.g. for synaptic receptor channels).

Synaptic transmission at a spike-mediated chemical synapse begins when an action potential invades the presynaptic terminal and activates voltage dependent Ca$^{2+}$ channels leading to a rise in the concentration of Ca$^{2+}$ within the terminal. This causes vesicles containing transmitter molecules to fuse with the cell membrane and release their contents into the synaptic cleft between the pre- and postsynaptic sides of the synapse. The transmitter molecules then diffuse across the cleft and bind to receptors on the postsynaptic neuron.
Binding of transmitter molecules leads to the opening of ion channels that modify the conductance of the postsynaptic neuron, completing the transmission of the signal from one neuron to the other.

Hence, the $g_{ij}$ terms in Eq. (2.3) are really of three types: terms contributing to an intrinsic conductance $g_i^L$ with reversal potential $E_i^L$, terms contributing to an extrinsic input $g_i^{ext}$ with reversal potential $E_i^{ext}$, and within-network synaptic conductances, described by conductances $\tilde{g}_{ij}$ that represent input from neuron $j$ with reversal potential $\tilde{E}_{ij}$.

Dividing Eq. (2.3) by $\sum_k g_{ij}$ and introducing the neuron-specific time factor $\tau_i(t) = C_i/\sum_{k=1}^{N_c} g_{ij}$ gives

$$
\tau_i(t) \frac{dv_i}{dt} = -v_i + \frac{g_i^L E_i^L + g_i^{ext} E_i^{ext} + \sum_j \tilde{g}_{ij} \tilde{E}_{ij}}{g_i^L + g_i^{ext} + \sum_{k=1}^D \tilde{g}_{ik}}, \quad (2.4)
$$

where $D$ is the number of neurons in the network.

In general, we call a synapse "excitatory" (or "inhibitory) if an action potential in the presynaptic neuron increases (or decreases) the probability of an action potential occurring in a postsynaptic cell$^1$.

We will now make two heuristic assumptions:

1. We assume that $\tilde{g}_{ij}$ is proportional to the instantaneous firing rate $r_j$ of neuron $j$, i.e. its probability of spiking per unit time$^2$. We call the proportionality constant $\tilde{W}_{ij}$, so that

$$
\tilde{g}_{ij} = \tilde{W}_{ij} r_j.
$$

---

$^1$This may be regarded, to begin with, as a black-box definition. A neuron's connection to another neuron may be complex, but if the total of excitatory influences exceeds that of the inhibitory influences, the neuron will generate a new action potential at its axon hillock, thus transmitting the information to yet another cell.

$^2$This assumes, among other things, that synaptic time courses are sufficiently rapid compared to the time scale under consideration, so that in the first approximation they can be ignored.
(2) We assume that the firing rate $r_j$ is given by the static nonlinearity $r_j = f(v_j)$. For the aims of this thesis, the functional form of this dependence is not crucial. Several authors, such as Hansel and van Vreeswijk (2002), Miller and Troyer (2002), and Priebe et al. (2004), have derived such a relationship between firing rate averaged over a few tens of milliseconds and voltage.

We model synapses as being either excitatory, with reversal potential $E_E$, or inhibitory with reversal potential $E_I$. Clearly, it is possible to linearly transform the units of voltage so that $E_E = 1$ and $E_I = -1$. Moreover, we define $\tilde{W}_{ij} = \tilde{W}_{ij} E_j$. This is now a synaptic weight that is positive for excitatory synapses and negative for inhibitory synapses. We also define

$$I_i^v \equiv g_i^L E_i^L + g_i^{ext} E_i^{ext} \quad g_i \equiv g_i^L + g_i^{ext}$$

from which we obtain the following conductance-based rate equation:

$$\tau_i(t) \frac{dv_i}{dt} = -v_i + \frac{I_i^v + \sum_{j=1}^{D} W_{ij} f(v_j)}{g_i + \sum_{k=1}^{D} |W_{ik}| f(v_k)}, \quad (2.6)$$

where now $\tau_i(t) = \frac{C_i}{g_i + \sum_{k=1}^{D} |W_{ik}| f(v_k)}$, and the first term on the RHS of the equation is known as the leak term.

Finally, the total conductance represented by the denominator in the last term of Eq. (2.6) may be taken to be constant. This is indeed correct if $g_i^L$ is much larger than synaptic and external conductances, or if inputs tend to be ”push-pull”, with withdrawal of some inputs compensating for addition of others.

We absorb the constant denominator into the definitions of $I_i^v$ and $W_{ik}$, and note that
this also implies that $\tau_i$ is constant, to arrive finally at the ”rate equation”

$$\tau_i \frac{dv_i}{dt} = -v_i + \sum_{j=1}^{D} W_{ij} f(v_j) + I_i^v(t). \quad (2.7)$$

or compactly,

$$T \frac{dv(t)}{dt} = -v(t) + \tilde{W} f(v(t)) + I^v(t). \quad (2.8)$$

Here $v(t)$ is the vector of state variables of all neurons at time $t$; $f(\cdot)$ is usually known as the neuronal input-output function$^3$, which is imposed element-by-element on its vector argument, with $f(v)_i \equiv f(v_i)$ giving the output, i.e. the firing rate, of neuron $i$; $I^v(t)$ is the external input vector; $T = \text{diag}(\tau_1, \tau_2, \cdots, \tau_N)$ is a $N \times N$ diagonal matrix whose diagonal elements are the positive time constants of the neurons (hence $T$ is invertible); and the bar sign has been removed from $W$ to indicate rescaling.

In surprisingly many applications, a linearization of Eq. (2.8) is sufficient to gain insight into the structure of the network over time. This is always true, of course, for studying of the system’s evolution starting from perturbation around a fixed point of Eq. (2.8), or the stability of the fixed point itself.

Suppose that for a constant external input, $I^v_0$, Eq. (2.8) has a fixed point $v_\ast$. Then, given a small perturbation in the input, $I^v(t) = I^v_0 + \delta I^v(t)$, we can write $v(t) = v_\ast + x(t)$, and linearize the dynamics around the fixed point by expanding Eq. (2.8) to first order in $x(t)$

$^3$It typically takes such forms as an exponential, a power law, or a sigmoid function. A favorite choice for analytical investigation is $f(v) = \tanh \left( \frac{v}{v_0} \right)$ for some voltage scale $v_0$. Several properties of the system appear to only depend on the slope $g = [f'']_0$ of the input-output function at the origin and on its saturation values (see e.g. Molgedey et al., 1992).
and $\delta I^v(t)$. This yields the set of linear equations

$$ T \frac{dx(t)}{dt} = -x(t) + W\Phi x(t) + \delta I^v(t), \quad (2.9) $$

for the (small) deviations, where we defined the diagonal Jacobian

$$ \Phi = \text{diag}(f'(v_*)). \quad (2.10) $$

Through much of this thesis, we will also write Eq. (2.9) in a form that assumes $T$ to be proportional to the identity matrix, namely:

$$ \frac{dv_i}{dt} = -\gamma v_i + \sum_{j=1}^{\tilde{W}} W_{ij} f(v_j) + I_i(t). \quad (2.11) $$

whose linearization takes the form

$$ \frac{dv}{dt} = -\gamma v + \sum_{j=1}^{A} Av + I(t). \quad (2.12) $$

In Eq. (2.12), $\gamma = 1/T_{ii}$ will be called through the thesis ”leak parameter”, while the matrix $A = \tilde{W}f^*/T_{ii}$ is what will be mostly refer to as the ”synaptic weight matrix” or ”connectivity” of the system$^4$. The term ”wiring diagram” is also sometimes employed for the connectivity structure of a network.

2.3 Range of validity

While the simplifications we have made so far are substantial, in practice the Eqs. (2.11) and more detailed spiking models have very similar behavior in a number of situations.

Firing rate models can be an ideal tool, in particular, to investigate sufficiently large neuronal systems. Indeed, some authors call them ”population models” because firing rates may be understood as representing the averaged behavior of spike rates of groups of neurons within a circuit (Ermentrout and Terman, 2010).

Even for large populations, however, firing rates fail to describing some properties of spiking models. Most notably, they do not properly reproduce the phenomenon of spike synchrony in networks of neurons. In particular, they fail to capture the onset of synchronous oscillations in networks of inhibitory neurons\(^5\), for which there is both numerical and experimental evidence.

Nonetheless, firing rate models can be a starting point to access a number of different regimes, including that of chaotic dynamics. To overcome the limitations connected to the emergence of synchrony, explicit fixed time delays have been introduced as a heuristic proxy for the combined effects of synaptic and subthreshold integration (Roxin et al., 2006). In this thesis, we will only be concerned with asynchronous spiking regimes.

2.4 Equivalence of alternative rate models

Eq. (2.9), written with a uniform time constant, takes the vector form

\(^5\)In Devalle et al. (2017), it was shown how that this limitation is due to a voltage-dependent synchronization mechanism which is naturally present in spiking neuron models but not captured by standard firing rate equations.
\[ \frac{d\mathbf{v}}{dt} = -\mathbf{v} + \mathbf{I} + Wf(\mathbf{v}). \] (2.13)

Here each element represents the membrane potential of one neuron in the modeled circuit; \( \mathbf{I} \) is the time-dependent vector of external inputs to the neurons in the network; and we have defined the vector function \( f(\mathbf{x}) \) as acting element-by-element on the elements of \( \mathbf{x} \), its \( i^{th} \) element being \( [f(\mathbf{x})]_i = f(x_i) \).

Nonetheless, the model that appears in much of the neuroscientific literature is slightly different from Eq. (2.13), and is written as (see Miller and Fumarola, 2012)

\[ \frac{d\mathbf{r}}{dt} = -\mathbf{r} + f(W\mathbf{r} + \mathbf{I}), \] (2.14)

where \( \mathbf{r} \) is thought of as representing the instantaneous firing rate and \( \mathbf{I} \) represents the vector of external inputs to the neurons.

It is easy to check that Eq. (2.13) can be obtained from Eq. (2.14) by setting \( \mathbf{I} = \mathbf{I} + \tau \frac{d\mathbf{I}}{dt} \) and \( \mathbf{v} = W\mathbf{r} + \mathbf{I} \). This means that the input \( \mathbf{I} \) is a low-pass filtered version of the input \( \mathbf{I} \).

If \( \mathbf{I} \) is given at all times, together with all the initial conditions, then going from an \( \mathbf{r} \)-model to a \( \mathbf{v} \)-model is straightforward, as we simply need to set \( \mathbf{v}(0) = W\mathbf{r}(0) + \mathbf{I}(0) \).

Going from a \( \mathbf{v} \)-model to an \( \mathbf{r} \)-model is also straightforward as long as \( W \) is invertible (which is almost never the case). In the general case, let us define \( \mathcal{N}_W \) to be the nullspace of \( W \), \( \mathcal{N}_W^\perp \) to be the subspace perpendicular to \( \mathcal{N}_W \), \( \mathcal{R}_W \) to be the range of \( W \), and \( \mathcal{R}_W^\perp \) the subspace perpendicular to \( \mathcal{R}_W \).

Given a \( \mathbf{v} \)-model, the equation \( \mathbf{v}(0) = W\mathbf{r}(0) + \mathbf{I}(0) \) has a solution if and only if
\( \nu(0) - I(0) \in \mathcal{R}^W \), so we must choose \( I_{R\perp}(0) = \nu_{R\perp}(0) \). Letting \( D_R \) be the dimension of \( \mathcal{R}^W \) and \( D_N \) the dimension of \( N^W \), we have \( D_R + D_N = D \). So \( I_{R\perp}(0) \) has dimension \( D_N \). This leaves unspecified \( I_R(0) \), which has dimension \( D_R \), as well as \( r(0) \).

To solve for \( r_{N\perp}(0) \), we introduce the Moore-Penrose inverse of the synaptic weight matrix, i.e. the matrix \( W^{-1} \) such that \( WW^{-1} = P_R \) (the projection into the range of \( W \)), while \( W^{-1}W = P_{N\perp} \) (the projection into \( N_{\perp} \)). The matrix \( W^{-1} \) defines a one-to-one mapping from \( \mathcal{R}^W \) to \( N_{\perp}^W \), with \( \mathcal{R}_{\perp}^W \) mapped to 0. We can then solve for \( r_{N\perp}(0) \) as

\[
    r_{N\perp}(0) = W^{-1} [\nu_{R}(0) - I_{R}(0)].
\]

This is a \( D_R \)-dimensional equation for the \( 2D_R \)-dimensional set of unknowns \( \{r_{N\perp}(0), I_{R}(0)\} \), so it determines \( D_R \) of these parameters. The other \( D_R \) parameters remain free, so that a \( D_R \) family of r-models can be mapped into every v-model for a given input \( I(t) \).

---

6 Notice also that the condition \( \nu - I \in \mathcal{R}^W \) is true for all time if it is true in the initial condition.

7 Moreover, since \( r_N \) is also free, and has \( D_N \) dimension, it follows that there are \( D_R + D_N = D \) free parameters.
3.1 Hebbian dynamics in the visual cortex

3.1.1 Hebbian dynamics

The synaptic weight matrix of a network evolves over time in ways that depend, among other things, on the activity of the neurons involved. Such evolution can serve multiple functions. Depending on context, it may be the spontaneous process of internal evolution of the network or the process of ”learning” some information from sensory input.

Mechanisms underlying this evolution are known under the generic name of ”plasticity”. Among them, the best-known one was first proposed by D.O. Hebb, who, in a famous passage of (Hebb, 1949), he made the following hypothesis:

When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased.

In other words, according to Hebb’s rule, synapses are strengthened or stabilized if there is temporal correlation between their pre- and postsynaptic patterns of activity. For small
time variations, this can be linearized into the statement that

\[
\frac{d}{d\tau} W_{ij} \propto r_ir_j, 
\]  

(3.1)

with \( \tau \) as the time variable, while nonlinearities must be introduced to prevent the synaptic weight matrix from diverging in the long term. An important application of this principle was to be found, some thirty years later, in the field of visual cortex studies.

### 3.1.2 Visual cortex development

The visual cortex of mammals is located in the occipital lobe in the back of the head. It is the first receiving area in cerebral cortex for visual sensory information; it receives signals from the lateral geniculate nucleus of the thalamus (hereinafter, LGN), which in turn receives signals directly from the two eyes. The part of the visual cortex that receives directly the sensory inputs from the thalamus is the primary visual cortex, also known as visual area 1 (V1). The functionally defined primary visual cortex is approximately equivalent to the anatomically defined striate cortex. The extrastriate cortex encompasses, in turn, several functional areas (Kandel, 1981.).

The visual cortex extends in each of two dimensions along the cortical surface. These two dimensions contain a continuous map of the world as seen through the two eyes, so that neighboring areas of retina are represented by neighboring areas of cortex. It is said that the cortex contains a "topographic" map of the retinal surface, or is arranged "retinotopically". Visual cortical cells at any given point in the cortex respond to light stimulation from only a small area in the visual world, and this area shifts continuously across these two dimensions
of cortex to yield a map of the visual world.

In the third dimension, the cortex consists of multiple layers. But the area of the world represented by visual cortical cells (as well as many other visual response properties) remain essentially constant through this depth. Such organization of cortical properties in a manner that is invariant through the depth of cortex is known as *vertical* or *columnar* organization (Mountcastle, 1978).

In early research (Hartline, 1938), the "receptive field" (RF) of a cell was defined as the restricted geometric region of the retina where light stimulation evoked a response of the cell (Hartline, 1938). The term is used today, more subtly, for the linear kernel that determines the activity of a visual cortex neuron as a function of retinotopic space. Conversely, the portion of the synaptic weight matrix that describes feedforward input from one LGN neuron to V1 is sometimes called, in visual-cortex context studies, a "projective field" (hereinafter, PF).

In 1958, Hubel and Wiesel performed a number of measurements on the primary visual cortex of cats, and discovered cells in the visual cortex whose RFs tended to align with some given direction. They found this through an experiment by giving a cat specific visual stimuli and measuring the corresponding excitation of the neurons in V1. This breaking of rotational symmetry in the LGN-to-V1 synaptic weights became known as "orientation selectivity". A given distribution of the directions of selectivity over a sheet of the visual cortex is known as an "orientation map".

It was later found that orientation maps exist in the striate cortex of monkeys, kittens, and ferrets before they first open their eyes (Hubel and Wiesel, 1974; Blakemore and van Sluyters, 1975; Chapman and Stryker, 1993). Recent advances in experimental imaging
technologies have made it possible to measure the full map of orientation preferences in young animals. Such experiments show that large-scale orientation maps exist prior to visual experience, and that these maps have many of the same features found in adults (Chapman et al., 1996; Crair et al. 1998; Goedecke et al., 1997).

This gives rise to the question whether such a basic aspect of visual cortex organization is mostly determined by cues directly controlled by patterns of genetic expression, or whether it is more correct to regard development as a flexible, self-organizing process influenced mainly by neural activity and patterns of sensory stimulation. In fact, this is a special instance of the general, outstanding question regarding to what extent brain structures are genetically programmed.

Some authors opt for a view of developing cortex as a plastic and self-organizing system, mostly independent of genetic control (e.g. Farley, 2007). The argument is that it would be costly to include all the specifics of cortex development in our load of genetic information. Indeed, the task of constructing the vertebrate central nervous system is tremendously complex. Perhaps $10^{11}$ neurons must migrate to their proper locations, properly differentiate, send axons to the proper targets (Schmidt and Edwards, 1983; Shatz and Stryker, 1988), and make hundreds of thousands of precise synaptic connections in the correct numbers (Hayes and Meyer, 1989).

In fact, the early stages of this process are known in many instances to occur properly even when all electrical activity of neurons is blocked (Harris, 1981). However, later stages during which inputs sort (usually on a finer scale than their initial innervation) by retinotopic location, modality, or other attributes related to the information that they carry, often appear to depend on patterns of neuronal electrical activity to achieve the final precision of synaptic
connections. In particular, the development of orientation selectivity occurs without vision but requires spontaneous activity (Chapman, 1993; Coppola and Fitzpatrick, 2001).

This is why, as we will now see in detail, the development of orientation columns in the visual cortex has proved to be one of the first testing grounds for Hebb’s rule.

### 3.1.3 Hebbian theories of orientation selectivity

A benchmark application of Hebb’s principle to the visual cortex was performed first by von der Malsburg (in 1973, with a specific set of inputs) and then by Linsker (in 1986 with an ensemble average over inputs). Both simulated the behavior of cells receiving feedforward input through Hebbian synapses from LGN. The dynamics of one such cell was analyzed in more detail by McKay and Miller (1990), who studied numerically the dominant eigenmodes of the time evolution operator, besides presenting some general theorems on their proprieties. A full analytical study of this model cell, however, is still lacking.

In the rest of this chapter, we will re-derive the model Linsker used for a single cortical cell, and we will offer a full mathematical study of its properties, proving the emergence of symmetry breaking and exploring the phase diagram.

The network is shown in Fig. 3.1, panel (a). A double layer of presynaptic cells is connected to a postsynaptic layer. The presynaptic layers can be understood to represent LGN, receiving sensory input from the retina; the postsynaptic cell may be interpreted as a generic cell in V1. LGN contains, in this conceptualization, a layer of ON cells (firing when the corresponding point on the retina receives input), and a layer of OFF cells (firing in a fashion anticorrelated to corresponding retinal input).
We will consider here a single postsynaptic cell receiving input from the presynaptic layers. The connections are purely feedforward, as no recurrent interactions (lateral connectivity) in LGN are considered.

The position of cells within LGN can be referred to with a two dimensional vectors \( r \), having direct retinotopic significance. Let \( x^{\text{ON}}(r, \tau) \) and \( x^{\text{OFF}}(r, \tau) \) be the activity of the ON or OFF cell at position \( r \) at time \( \tau \). We call \( s^{\text{ON}}(r, \tau) \) and \( s^{\text{OFF}}(r, \tau) \) the strength of that input’s connection to the postsynaptic cell we are considering.

The linearization of Eq. (2.8) into Eq. (2.9) allows to write

\[
T \frac{dy}{d\tau} = -\gamma y(\tau) + \int dr A(r) s^{\text{ON}}(r) x^{\text{ON}}(r) + \int dr A(r) s^{\text{OFF}}(r) x^{\text{OFF}}(r),
\]

where we have written separately the input from ON and OFF cells.

The matrix \( T \) on the LHS of eq. (3.2) is diagonal, with entries that are typically much smaller than the time scales of synaptic development. Thus, the activity \( y(\tau) \) can be assumed to be given by the steady state of Eq. (3.2), which, rescaling the synaptic weight matrices by a factor \( \gamma \), may be written by the following linear combination of the inputs:

\[
y(\tau) = \int dr A(r) s^{\text{ON}}(r) x^{\text{ON}}(r) + \int dr A(r) s^{\text{OFF}}(r) x^{\text{OFF}}(r),
\]

where \( A(r) \) is the density of synapses within an arbor. The variables \( x^{\text{ON}} \) and \( x^{\text{OFF}} \) may be understood as stochastic processes, reflecting the stochasticity of visual input\(^1\).

A Hebbian rule for synaptic plasticity is one in which a synaptic strength is increased

\(^1\)However, we are taking the synaptic weight matrices \( s^{\text{ON}} \) and \( s^{\text{OFF}} \) to be deterministic. Considering the matrices themselves as random variables leads to a rich scenario that will be explored in Chapter 5.
when pre- and post-synaptic firing are correlated (and possibly decreased when they are anticorrelated). Following Eq. (3.1), we can write

\[
\frac{d}{d\tau} s^{ON,OFF}(r) \propto x^{ON,OFF}(r)y
\]

(3.4)

and we can set the multiplicative constant to unity by rescaling units. Notice that we are also omitting possible constant terms, included by Linsker but not essential to the development of selectivity.

Substituting (3.3) into (3.4), we obtain

\[
\frac{d}{d\tau} s^{ON,OFF}(r) \propto \sum_{\alpha=ON, OFF} x^{ON,OFF}(r) \int ds A(s) s^\alpha(s) x^\alpha(s)
\]

(3.5)

Assuming that \(S\) changes on a much longer time scale than the random variations in the inputs, we average over the ensemble of patterns \(x^{ON,OFF}(r)\) in the LGN layers. This singles out the long-term correlation functions between LGN activity at different points, which by symmetry we can take to have the form

\[
\langle x^{ON}(r)x^{ON}(s) \rangle = \langle x^{OFF}(r)x^{OFF}(s) \rangle \equiv C^{\text{same}}(r,s)
\]

(3.6)

\[
\langle x^{ON}(r)x^{OFF}(s) \rangle = \langle x^{OFF}(r)x^{ON}(s) \rangle \equiv C^{\text{opp}}(r,s).
\]

(3.7)

Averaging Eq. (3.5) and using the definitions Eqs. (3.6-3.7), we find that the net synap-
tic weight \( s(r) = s^{ON}(r) - s^{OFF}(r) \) evolves according to

\[
\frac{\partial}{\partial \tau} s(r, \tau) = A(r) \int ds \ C(r, s) s(s, \tau),
\]

(3.8)

where \( C = C^{\text{diff}} = C^{\text{same}} - C^{\text{opp}} \). We will take this function to be Gaussian: \( C(r, s) \propto e^{-\frac{(r-s)^2}{2\eta^2}} \), where \( \eta \) is the typical length scale for the decay of correlations, directly depending on the nature of visual input.

It is characteristic of Hebbian rules that synaptic strengths tend to increase without limit. Linsker modeled the biological mechanisms for saturation by including simply upper and lower bounds for all synaptic strengths: \(-s_{\text{max}} \leq s \leq s_{\text{max}}\). This turns Eq. (3.8) into a nonlinear equation; however, as he showed numerically, the theory can rely on these hard limits being large enough so that the principal features of the dynamics are established before the bounds are saturated. Once the hypercube is reached, it will simply capture and preserve the existing weight structure with little subsequent change.

It is therefore possible to extract the long term behavior of the synaptic function simply from analyzing the property of the time-evolution operator in the linear regime. As per Eq. (3.8), this is a Hilbert-Schmidt operator characterized by the kernel

\[
K(r, s; \eta) = A(r) e^{-\frac{(r-s)^2}{2\eta^2}}
\]

(3.9)

and is ostensibly non-Hermitian. However one can make it Hermitian by simply switching
to the new variable \( t(r) = s(r)/\sqrt{A(r)} \). The new time-evolution operator is then:

\[
L^{(0)}(r, s; \eta) = \sqrt{A(r)A(s)} e^{-\frac{(r-s)^2}{2\eta^2}}. 
\]

(3.10)

In addition, we will take the arbor density to be also Gaussian. We write \( A(r) = \frac{1}{2\pi\rho^2} \exp\left(-\frac{r^2}{2\rho^2}\right) \); and neglecting an overall factor (which can be re-absorbed in the units for time), the generic matrix element of \( L^{(0)} \) becomes

\[
L^{(0)}(r, s; \eta) = e^{-\frac{(r-s)^2}{2\eta^2}} e^{-\frac{s^2}{4\eta^2}}. 
\]

(3.11)

Given the linearity of the system at times small enough that the hypercube has not been reached, the solution to (3.8) can be written as

\[
t(r, \tau) = \sum_{\alpha} e^{\lambda_{\alpha}(\tau-\tau_0)} \psi_\alpha(r) \int dr' \psi^*_\alpha(r') t(r', \tau_0)
\]

(3.12)

where the generic index \( \alpha \) labels a complete set of eigenfunctions \( \psi_\alpha \) for the operator with kernel Eq. (3.11), and \( \lambda_{\alpha} \) is the corresponding eigenvalue.

The operator defined by Eq. (3.11) is a bounded operator with a discrete spectrum. Its diagonalization, discussed in the next section, yields complete information on the dynamics. In particular, as the synaptic weight matrix evolves, it becomes exponentially overlapping with the principal eigenspace of \( L^{(0)} \). Mapping out this eigenspace is therefore of importance to understanding the behavior of the model at long times.
Figure 3.1: Panel a: stylized representation of LGN and V1. The two LGN layers model ON and OFF cells, an arbor of radius $\rho$ is shown, while the ellipses stand for a receptive field (RF) and a projective field (PF). Panel b: width $\gamma$ of the rescaled receptive field $T(r)$, Eq. (3.19), as a function of the correlation length of presynaptic activity, $\eta$. The asymptotic value is proportional to the radius $\rho$ of the arbors.

3.2 Diagonalization of the Linsker operator

3.2.1 Diagonalization in polar coordinates

We will refer to the Hilbert-Schmidt operator with kernel (3.10) as the Linsker operator. It is clearly invariant under rotations of the pre-synaptic plane, and its rotational symmetry entails that the eigenfunctions can be taken to have the form $\psi(r, \phi) \propto e^{im\phi}R_{N,m}(r)$, where $N$ (which we will refer to as a "quantum" number) indexes all the radial eigenfunctions corresponding to the same angular number $m$.

From the Perron-Frobenius theorem, given that all the matrix elements are positive, it can be proven that the principal eigenfunction is everywhere positive. From this, we infer that the principal eigenfunction must be one of the functions corresponding to $m = 0$.

We also notice that the operator is invariant under parity (the transformation $x \rightarrow -x$, $y \rightarrow -y$, or $\phi \rightarrow \phi + \pi$). This, and the assumed analyticity of the eigenfunctions, entails
that the Taylor expansion $R_{N,m}(r) = \sum_n a_n r^n$ contains only even powers of $r$ if $m$ is even, and only odd powers of $m$ if $m$ is odd. Thus, all coefficients $a_n$ in the expansion of a radial eigenfunction are zero if $(n + m)_{\text{mod} 2} = 1$.

If we integrate any function of the type “polynomial $\times$ Gaussian” in arbitrary dimensions, the result will be (by Wick’s theorem) identical in structure to the integrand – a ”polynomial $\times$ Gaussian” with a polynomial of the same degree. Here, the matrix element of the operator is a convolution of Gaussian functions, and integrating a convolution of Gaussians will yield yet more Gaussian. We diagonalize the operator in polar coordinates by adopting the following ansatz:

$$\psi(r, \phi) \propto \exp\left[\pm im\phi - \frac{r^2}{2\gamma^2}\right] \sum_{n=0}^N a_n r^{m+2n}$$  \hspace{1cm} (3.13)

where the parameter $\gamma$ has a value as yet unspecified; the coefficients $a_n$, and the non-negative numbers $m$ and $N$ are in principle different for each eigenfunction.

The eigenvalue equation for the operator of Eq. (3.11) becomes

$$e^{-\frac{r^2}{4\mu^2} - \frac{s^2}{2\eta^2}} \sum_{n=0}^N a_n \int_0^\infty ds \ s^{m+2n+1} e^{-\alpha s^2} \int_0^{2\pi} d\phi' e^{\pm im\phi'} + \frac{r^2}{\pi^2} \cos(\phi - \phi') = \lambda_{N,m} e^{\pm im\phi - \frac{r^2}{2\gamma^2}} \sum_{k=0}^N a_k r^{m+2k},$$  \hspace{1cm} (3.14)

where $\lambda_{N,m}$ is the eigenvalue corresponding to the given choice for $N$ and $m$, and the parameter $\alpha$ in Eq. (3.14) is defined as

$$\alpha = \frac{1}{4\mu^2} + \frac{1}{2\eta^2} + \frac{1}{2\gamma^2}.$$  \hspace{1cm} (3.15)
We now shift the angular integration variable by an amount $\phi$, thus pulling a factor $e^{\pm i m \phi}$ out of the integral, which cancels with the identical factor in the RHS. The resulting integral (aside from a factor $2\pi$) is nothing but the definition of the modified Bessel function of the first kind of order $m$. We thus rewrite Eq. (3.14) as

$$2\pi e^{-\frac{r^2}{4\eta^2} - \frac{s^2}{2\eta^2}} \sum_{n=0}^{N} a_n \int_{0}^{\infty} ds \, s^{m+2n+1} e^{-\alpha s^2} I_m \left( \frac{rs}{\eta^2} \right) = \lambda_{N,m} e^{-\frac{r^2}{2\eta^2}} \sum_{k=0}^{N} a_k r^{m+2k}. \quad (3.16)$$

All dependency on the sign of the imaginary exponent has vanished by symmetry, signifying degeneracy of the plus and minus solutions.

To integrate over $s$ in Eq. (3.16), we may use formula 6.631/10 of Gradshteyn and Ryzhik (1996) (Sec. 6.61, p. 738), which can be written as

$$\int_{0}^{\infty} ds \, s^{m+2n+1} I_m (bs) e^{-\alpha s^2} = \frac{n! b^m}{2^{m+1} \alpha^m (m+n+1)} e^{\frac{b^2}{4\alpha}} L_n^{(m)} \left( -\frac{b^2}{4\alpha} \right), \quad (3.17)$$

where $L_n^{(m)}$ is the generalized Laguerre polynomial.

Collecting all the exponential factors on the LHS of Eq. (3.16), including the one coming from radial integration, and equating them to the exponential on the RHS with exponent $\propto \gamma^{-2}$, we arrive at an equation for the width $\gamma$ of the receptive fields:

$$-\frac{1}{4\rho^2} - \frac{1}{2\eta^2} + \frac{1}{4\eta^4 \alpha} = -\frac{1}{2\gamma^2}. \quad (3.18)$$
Plugging (3.18) into (3.15), we obtain

\[ \gamma = \sqrt{2}\rho \left( 1 + \frac{4\rho^2}{\eta^2} \right)^{-1/4}, \]  

(3.19)

which describes how the arbor radius \( \rho \) is renormalized by the correlations.

For fixed \( \rho \), the width \( \gamma \) of the eigenfunctions is a monotonically decreasing function of the ratio \( \rho/\eta \). In this representation (having multiplied the function \( f \) by \( \sqrt{A(r)} \)) the unrenormalized arbor radius is represented by \( \sqrt{2}\rho \). So we can say that, if the ratio \( \rho/\eta \) is very small, no renormalization occurs: \( \gamma = \sqrt{2}\rho \). As the ratio goes to infinity, i.e. the correlations become very local, the range of the eigenfunction is restricted by the correlation length scale, and becomes equal to the geometric mean of the two length scales; namely, \( \gamma \sim \sqrt{\rho \eta} \) (Fig. 1b).

By our choice Eq. (3.19) for \( \gamma \), we have canceled out the exponential terms on the two sides of Eq. (3.16). The rest can be written as

\[
\pi \left( \frac{r}{2\eta^2} \right)^m \sum_{n=0}^{N} \frac{n!a_n}{\alpha^{m+n+1}} L_n^{(m)} \left( -\frac{r^2}{4\eta^4\alpha} \right) = \lambda_{N,m} \sum_{k=0}^{N} a_k r^{m+2k}.\]

(3.20)

Substituting the closed form for the generalized Laguerre polynomials, \( L_n^{(m)}(x) = \sum_{k=0}^{n} (-1)^k \binom{n+m}{n-k} \frac{x^k}{k!} \), and canceling powers of the radius \( r \) from both sides, we obtain

\[
\frac{\pi}{(2\eta^2)^m} \sum_{n=k}^{N} \frac{n!}{k! \binom{n+m}{n-k}} \frac{a_n}{\alpha^{m+n+1}} (4\eta^4\alpha)^{-k} = \lambda_{N,m} a_k. \]

(3.21)

This is a set of \( N + 1 \) equation labeled by \( k = 0, 1, \ldots, N \).
By setting now \( k = N \) in Eq. (3.21), one finds the general formula for the eigenvalues:

\[
\lambda_{N,m} = \frac{\pi}{\alpha} \left( 2\eta^2 \alpha \right)^{-2N-m}.
\] (3.22)

or more conveniently,

\[
\lambda_{N,m} = 2\pi\eta^2 \beta^{-2N-m-1},
\] (3.23)

where \( \beta \equiv 1/q = 1 + \frac{\eta^2}{2\rho^2} + \frac{\eta}{\rho} \sqrt{1 + \frac{\eta^2}{4\rho^2}} \).

The recursive equations (3.21) for the coefficients may be written, for \( k < N \), as

\[
a_k = \frac{(4\eta^4 \alpha)^{N-k}}{1 - (4\eta^4 \alpha^2)^{N-k}} \sum_{n=k+1}^{N} \frac{n!}{k!} \binom{n+m}{n-k} \alpha^{N-n} a_n,
\] (3.24)

where, writing Eq. (3.15) explicitly,

\[
\alpha = \frac{1}{4\rho^2} \left( 1 + \frac{2\rho^2}{\eta^2} + \sqrt{1 + \frac{4\rho^2}{\eta^2}} \right),
\] (3.25)

from which we immediately find the first three eigenfunctions of the operator (3.11):

\[
\begin{align*}
\psi_{0,0}(\vec{r}) &= \frac{1}{\sqrt{\pi \gamma}} \exp \left( -\frac{r^2}{2\gamma^2} \right); \\
\psi_{0,1}^{\pm}(\vec{r}) &= \frac{r}{\sqrt{\pi \gamma^2}} \exp \left( -\frac{r^2}{2\gamma^2} \pm i\phi \right); \\
\psi_{1,0}(\vec{r}) &= \frac{(\gamma^2 - r^2)}{\sqrt{\pi \gamma^3}} \exp \left( -\frac{r^2}{2\gamma^2} \right);
\end{align*}
\] (3.26-3.28)

while the corresponding eigenvalues must be found from Eq. (3.23).

Here, \( \psi_{0,0} \) is a RF that selects no orientation (a non-selective field) while \( \psi_{0,1}^{\pm} \) is a RF
Figure 3.2: The first four radial eigenfunctions corresponding to angular momentum $m = 0$ and $m = 1$, plotted in units of $\gamma$. The eigenfunctions can be enumerated by their number of nodes.

that selects a definite orientation (along the x-axis), hence it displays selectivity. A closed-form solution to Eq. (3.21) is given in Appendix A for the asymptotic regimes of short- and long-range correlations.

### 3.2.2 Diagonalization in Cartesian coordinates

It is instructive to diagonalize the Linsker operator of Eq. (3.11) in Cartesian coordinates, again starting from the eigenvalue equation

$$\lambda \chi(r) = \exp \left( -\frac{r^2}{4\rho^2} \right) \int ds \exp \left( -\frac{|r-s|^2}{2\eta^2} - \frac{s^2}{4\rho^2} \right) \chi(s). \quad (3.29)$$

The action of this operator on any separable function $\chi$ factorizes into the product of two separate integrals involving $r_x$ and $s_x$ respectively. We can thus write Eq. (3.29) as

$$\lambda \chi(r) = \int ds_x L_1(r_x, s_x) \int ds_y L_1(r_y, s_y) \chi(s_x, s_y), \quad (3.30)$$

where $L_1(\xi, \zeta') = \exp \left[ -\frac{\xi^2}{4\rho^2} - \frac{(\xi-\zeta')^2}{2\eta^2} - \frac{\eta^2}{4\rho^2} \right].$
For diagonalizing a 1D integral operator, moreover, a practical starting point is to map it into some known differential operator (Wimbauer et al., 1998). In this case, note that the operator $\hat{H}_1$ defined by

$$\hat{H}_1[f]_\xi = \int d\xi' L_1(\xi, \xi') f(\xi')$$

(3.31)

commutes with the Hamiltonian of the 1D harmonic oscillator $\hat{H}_{h.o.} = \frac{\partial^2}{\partial \xi^2} - \frac{\gamma^2}{2}$, with the same $\gamma$ we computed in polar coordinates$^2$.

This leads us to writing Eq. (3.30) in the basis composed of the eigenfunctions of the harmonic oscillator Eq. (3.31), which are given by

$$\chi_n(\xi) = \frac{1}{\sqrt{2^n n!}} \left( \frac{1}{\pi \gamma^2} \right)^{1/4} H_n \left( \frac{\xi}{\gamma} \right) \exp \left( -\frac{\xi^2}{2 \gamma^2} \right),$$

(3.32)

where $H_n$ is the $n$th Hermite polynomial.

Looking then at the action of the kernel (3.11) on the functions $\chi_{n_x n_y}(r_x, r_y) = \chi_{n_x}(r_x)\chi_{n_y}(r_y)$ defined through Eq. (3.32), we find that $\hat{L}^{(0)}\chi_{n_x n_y}_{\vert_r} = I_{n_x}(r_x)I_{n_y}(r_y)$, where

$$I_n(r) = e^{-r^2/4\sigma^2} \int ds e^{-\frac{(r-s)^2}{2\sigma^2}} e^{-\frac{r^2}{4\gamma^2}} H_n \left( \frac{s}{\gamma} \right).$$

(3.33)

The integral (3.33) can be done by invoking formula 7.374/10 of Gradshteyn and Ryzhik (1996) (Sec. 7.37, p. 843), which can be reworked into

$$\int_{-\infty}^{\infty} e^{-ax^2+bx} H_n(x) = e^{\frac{b^2}{4a}} \sqrt{\frac{\pi}{a}} \left( 1 - \frac{1}{a} \right)^{n/2} H_n \left( \frac{b}{2\sqrt{a(a-1)}} \right).$$

(3.34)

$^2$Here the width $\gamma$ of receptive fields plays the role of the quantum length scale $\sqrt{\frac{h}{m \omega}}$. 

31
Switching to \( x = s/\gamma \), and applying formula (3.34) with \( a = \gamma^2 \alpha, b = \frac{r^2}{\gamma^2} \), we finally arrive at \( I_n(r) = \Lambda_n \chi_n \left( \frac{r}{\gamma} \right) \), where \( \Lambda_n = \sqrt{2\pi \eta \beta^{-n-1/2}} \). Hence, the full eigenvalue of the two-dimensional problem is \( \Lambda_{n_x n_y} = 2\pi \eta^2 \beta^{-n_x-n_y-1} \).

From Eq. (3.32), we then see that the full eigenfunctions of the 2D variables are

\[
\chi_{n_x n_y}(r) = \left( 2^{n_x+n_y} \sqrt{\pi n_x! n_y!} \gamma \right)^{-1} e^{-\frac{r^2}{\gamma^2}} H_{n_x} \left( \frac{r_x}{\gamma} \right) H_{n_y} \left( \frac{r_y}{\gamma} \right).
\]

and may be written out by expressing the Hermite polynomials explicitly, which gives

\[
\chi_{n_x n_y}(r) = \frac{\sqrt{n_x! n_y!}}{2 \frac{n_x+n_y}{2} \sqrt{\pi \gamma}} e^{-\frac{r^2}{\gamma^2}} \sum_{k=0}^{\lfloor n_x/2 \rfloor} \sum_{l=0}^{\lfloor n_y/2 \rfloor} \frac{(-1)^{k+l} (2r_x/\gamma)^{n_x-2k} (2r_y/\gamma)^{n_y-2l}}{k! l! (n_x-2k)! (n_y-2l)!},
\]

so that the three dominant eigenfunctions are

\[
\chi_{0,0}(r) = \frac{1}{\sqrt{\pi \gamma}} \exp \left( -\frac{r^2}{2\gamma^2} \right)
\]

\[
\chi_{1,0}(r) = \frac{r_x}{\sqrt{\pi \gamma^2}} \exp \left( -\frac{r^2}{2\gamma^2} \right)
\]

\[
\chi_{0,1}(r) = \frac{r_y}{\sqrt{\pi \gamma^2}} \exp \left( -\frac{r^2}{2\gamma^2} \right).
\]

### 3.2.3 Connecting the polar and Cartesian representations

As we have showed, the eigenvalues of \( L^{(0)} \) have the form \( \lambda = 2\pi \eta^2 q^{n\lambda} \), where the number \( n_{\lambda} \) is an integer. We computed this number in two different representations. In polar coordinates it is written as \( n_{\lambda} = m + 2N \) while in Cartesian coordinates as \( n_x + n_y \). Of
course the subspace spanned by each eigenvalue will be the same in either representation\(^3\).

Let us now briefly consider the correspondence between eigenfunctions in the two representations. Any polar eigenfunction with quantum number \(n_\lambda\) will be a linear superposition of Cartesian eigenfunctions with quantum numbers \(n_x, n_y\) such that \(n_x + n_y = n_\lambda\). Among those eigenfunctions, however, only some will need to be included, their choice depending on the other quantum number for the polar representation – the angular momentum \(m\).

The terms to be included are those corresponding to a nonzero value of the overlap

\[
W_n^{(m)} = \int d\phi \chi_{n_x n_y}(r, \phi) e^{im\phi},
\]

with \(n = (n_x, n_y)\) and \(\chi_{n_x n_y}\) given by Eq. (3.36). Conversely, in decomposing a certain Cartesian eigenfunction, a certain value of \(m\) has to be considered only if \(W_n^{(m)}(r) \neq 0\) for at least one choice of \(n_x, n_y\) such that \(n_x + n_y = n_\lambda\).

\(^3\)To check that the degeneracy is indeed the same, we note that in polar coordinates it is given by

\[
\mathcal{D}(n_\lambda) = \sum_{0 \leq m_X \leq n_\lambda} \mathcal{D}_m,
\]

where \(\mathcal{D}_m = 2 - \delta_{0m}\) is the degeneracy associated to angular momentum \(m\). Computing this sum, we have

\[
\mathcal{D}(n_\lambda) = 1(n_\lambda \text{ even}) + 2 \left\lceil \frac{n_\lambda}{2} \right\rceil = n_\lambda + 1,
\]

In Cartesian coordinates, on the other hand, the degeneracy is

\[
\mathcal{D}(n_\lambda) = \sum_{0 \leq n_x \leq n_\lambda} \sum_{n_y} \delta_{n_x + n_y, n_\lambda} = \sum_{0 \leq n_x \leq n_\lambda} 1 = n_\lambda + 1.
\]
Writing formula (3.40) explicitly, we have

\[
W^{(m')}_{n_x, n_y}(r) \propto \int_0^{2\pi} d\phi \, H_{n_x} \left( \frac{r}{\gamma} \cos \phi \right) H_{n_y} \left( \frac{r}{\gamma} \sin \phi \right) e^{im\phi}.
\] (3.41)

where the sign of the imaginary exponent is immaterial.

Ignoring for the time being our polar-coordinate diagonalization of the operator, we can derive its main result \((n_x = m + 2N)\) from studying Eq. (3.41):

1) A first conclusion comes from the symmetries of the integral. Hermite polynomials of even degree \((H_n \text{ with odd } n)\) are even functions, and those of odd order are odd functions. The transformation \(\phi \rightarrow \phi + \pi\) inverts \(e^{im\phi}\) in Eq. (3.41) as long as \(m\) is odd, yielding

\[
(-1)^{n_x} H_{n_x} \left( \frac{r}{\gamma} \cos \phi \right) (-1)^{n_y} H_{n_y} \left( \frac{r}{\gamma} \sin \phi \right) e^{2\pi mi} e^{im\phi}.
\] (3.42)

For even values of \(m\), therefore, \(W^{(m')}_{n_x, n_y}\) survives if \(n_x, n_y\) have the same parity; for odd values of \(m\), it survives if they have different parity. It follows that there exists always an integer \(N \geq 0\) such that \(m = n_x + n_y \pm 2N\).

2) The second conclusion comes from the degree of the polynomials, which makes the integral (3.41) vanish if \(m > n_x + n_y\). To see this, rewrite the definite integral in Eq. (3.41) as a contour integral in the complex plane:

\[
W^{(m')}_{n_x, n_y} = -i \left( \frac{\gamma}{r} \right)^{m-1} \oint_{|z|=r/\gamma} dz \, H_{n_x} \left( \frac{z + z^{-1}}{2} \right) H_{n_y} \left( \frac{z - z^{-1}}{2i} \right) z^{m-1}.
\] (3.43)

The Hermite function \(H_{n_x}\) is a sum of powers with its maximum degree being \(n_x\); so the highest-order term in \(z^{-1}\) is \(z^{-n_x}\); similarly, for \(H_{n_y}\) the highest-order term is \(z^{-n_y}\).
Therefore, if $m > n_x + n_y$, the integrand is a polynomial – and the integral Eq. (3.43) is zero by Cauchy’s theorem. We deduce that a second requirement for the survival of the overlap is $m \leq n_x + n_y$, i.e. that $m \leq n\lambda$.

Combining these two requirements on the quantum numbers, one finds that $n\lambda$ can always be written in the form $n_x + n_y = m + 2N$, where $N$ is an integer equal or larger than zero.

Table 3.1: Polar-Cartesian correspondence of quantum numbers. An additional degeneracy, not specified in the table, is associated to the sign of the angular momentum.

<table>
<thead>
<tr>
<th>$(N, m)$</th>
<th>$(n_x, n_y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0,0</td>
<td>0,0</td>
</tr>
<tr>
<td>0,1</td>
<td>0,1, 1,0</td>
</tr>
<tr>
<td>1,0</td>
<td>0,2, 2,0</td>
</tr>
</tbody>
</table>

The three dominant eigenfunctions of the operator (3.11), mapped in between the representations (3.13) and (3.36), are given by $^4$:

$$|N = 0, m = 0\rangle = |n_x = 0, n_y = 0\rangle;$$  \hspace{1cm} (3.44)

$$|N = 0, m = 1\rangle = \frac{1}{\sqrt{2}} \left( |n_x = 1, n_y = 0\rangle \pm i |n_x = 0, n_y = 1\rangle \right);$$  \hspace{1cm} (3.45)

$$|N = 1, m = 0\rangle = \frac{1}{\sqrt{2}} \left( |n_x = 2, n_y = 0\rangle + |n_x = 0, n_y = 2\rangle \right).$$  \hspace{1cm} (3.46)

Fig. (3.3) shows the structure of the mapping for larger values of the quantum numbers.

$^4$Notice that the rules we stated above allow in principle the state $|N = 1, m = 0\rangle$ to have an overlap with $|n_x = 1, n_y = 1\rangle$, but this overlap is null for a further symmetry reason, as can seen by doing $\phi \to 2\pi - \phi$ in (3.41).
3.3 Homeostatic constraints

Eq. (3.1) may be sufficient as a description of Hebb’s rule when it is applied to learning; in the context of development, however, it neglects a widely observed aspect of plasticity. Development under a correlation-based rule is necessarily competitive; the growth of one input is determined not simply by its own activity but by its activity relative to that of competing inputs (Wiesel and Hubel, 1965; Guillery, 1972; Stryker and Strickland, 1984).

This must be kept into account through the use of conservation rules, or constraints added to Hebbian models. A conservation rule ensures that, if one set of cooperating inputs gains in synaptic strength, competitors must lose synaptic strength. In other words, it ensures that the synaptic changes are competitive: some synapses onto a cortical cell can become stronger only at the expense of other synapses ultimately becoming weaker, and
vice versa.

While the existence of such a homeostatic mechanism is generally accepted, it is a subject of debate which would be its most appropriate mathematical modeling (Miller and Stryker 1990, Miller 1992). Linsker used a constraint fixing the percentages of positive and negative synapses in the final receptive field of a cell. Orientation-selective receptive fields developed in his model only in a narrow, tightly tuned regime of the constraint parameter (MacKay and Miller, 1990).

Here, we follow the option originally adopted in von der Malsburg (1973), using a constraint that consists in conserving the total synaptic weight over the postsynaptic cell. The rule is not intended to model the biological mechanism by which competition is achieved, which we will try to mimic more realistically in the next chapter.

Eq. (3.8) must then be modified by the addition of a term ensuring that

$$\frac{\partial}{\partial t} \int s(r, \tau)d\tau = 0.$$  \hspace{1cm} (3.47)

In terms of the rescaled variable $t = s/\sqrt{A}$, Eq. (3.47) becomes

$$\frac{\partial}{\partial \tau} \int t(r, \tau)\sqrt{A(r)}d\tau = 0,$$  \hspace{1cm} (3.48)

i.e., the rescaled synaptic weight must have an invariant projection on the vector $|a\rangle$ defined by $\langle r|a\rangle = \sqrt{A(r)}$. To ensure this, we define the projection operator $P = 1 - |a\rangle\langle a|$ and replace the time evolution operator $L^{(0)}$ of Eq. (3.11) with the ”constrained” operator
\[ L^c = PL^{(0)}P, \text{ whose matrix elements are} \]

\[
L^c(r, s) = \int ds_1 ds_2 \left[ \delta(r - s_1) - \sqrt{A(r)A(s_1)} \right] L^{(0)}(s_1, s_2)
\times \left[ \delta(s_2 - s) - \sqrt{A(s_2)A(s)} \right], \quad (3.49)
\]

and rewriting Eq. (3.49) compactly gives:

\[
\hat{L}^c = L^{(0)} + |a\rangle\langle a|\hat{L}^0|a\rangle\langle a| - 2 \text{ HP} \left[ \hat{L}^0|a\rangle\langle a| \right], \quad (3.50)
\]

where "HP" is the Hermitian part of an operator. This is the constrained form of the Linsker operator that we will now analyze.

### 3.3.1 Positive semidefiniteness of the constrained operator

First, notice that the operator \( \hat{L}^{(0)} \) is positive definite. This follows from the fact that it is Hermitian with an all-positive kernel. Alternatively, we have also seen that its eigenvalues are all positive, though dense in a neighborhood of zero.

Recall now the following lemma: if \( \hat{O} \) is a positive-definite operator in a linear space and \( \hat{P} \) a projection operator on some subspace, then \( \hat{P}\hat{O}\hat{P} \) is positive-semidefinite.

The proof is straightforward. The expectation value of \( \hat{P}\hat{O}\hat{P} \) in an arbitrary state \(|\psi\rangle\) is \( \langle \psi|\hat{P}\hat{O}\hat{P}|\psi\rangle = (\langle \hat{P}^\dagger \psi| \rangle \hat{O} | \hat{P} \psi \rangle) \), but in order to be a projection operator \( P \) must be self-adjoint. Then \( \langle \psi|\hat{P}\hat{O}\hat{P}|\psi\rangle = (\langle \hat{P} \psi| \rangle \hat{O} | \hat{P} \psi \rangle) \), which is the expectation value of \( \hat{O} \) in a state \(|\hat{P} \psi \rangle\). There are two possibilities: (1) \(|\hat{P} \psi \rangle\) is the null state, in which case the expectation value of \( \hat{O} \) is zero, \( \langle \psi|\hat{P}\hat{O}\hat{P}|\psi\rangle = 0 \); or (2) \(|\hat{P} \psi \rangle\) is not the null state, in which
case we can invoke the positive definiteness of $\hat{\mathcal{O}}$ and conclude that $\langle \psi | \hat{P} \hat{\mathcal{O}} \hat{P} | \psi \rangle > 0$. Since this is true for an arbitrary state, $\hat{P} \hat{\mathcal{O}} \hat{P}$ is positive-semidefinite, and in particular, so is $L^c$.

### 3.3.2 Matrix elements of the constrained operator

Both $L^{(0)}$ and $L^c$ commute with the rotation operator, so they can be diagonalized simultaneously with it. Moreover, all the eigenfunctions of $L^c$ with angular momentum $m > 0$ are nothing but the corresponding eigenfunctions of $L^{(0)}$. This follows from the fact that the vector $|a\rangle$ is rotationally invariant.

Hence, we will split the Hilbert space into a zero-momentum sector characterized by $m = 0$ (functions that do not depend on the angular variable, which we also refer to as ”s-waves”) and a finite-momentum sector (spanned by the circular waves $e^{im\phi}$). We already possess the exact diagonalization of the constrained operator in the finite-momentum sector; in the zero-momentum sector, the operator must be diagonalized anew.

We thus have to solve the eigenvalue equation for eigenfunctions $\psi(r)$ that have no angular dependence, and we can write the operator as acting on functions of a single variable.

To do so, we perform first the integrals in Eq. (3.49), finding

$$L^c(r, s) = e^{-\frac{\rho^2 + \eta^2}{4\rho^2}} - \frac{(r-s)^2}{2\rho^2} + \frac{\eta^2}{\eta^2 + 2\rho^2} e^{-\frac{\rho^2 + \eta^2}{4\rho^2}}$$

and then integrate Eq. (3.51) over the angular variable, yielding the matrix elements $L^s$ of the restriction of $L^c$ to the $m = 0$ subspace.
\[ L^s(r, s) = I_0 \left( \frac{rs}{\eta^2} \right) e^{-\left( \frac{1}{2\rho^2} + \frac{1}{\eta^2} \right) \frac{r^2 + s^2}{2}} + \frac{\eta^2}{\eta^2 + 2\rho^2} e^{-\frac{r^2 + s^2}{4\rho^2}} - \frac{\eta^2}{\eta^2 + \rho^2} \left( e^{-\frac{r^2}{4\rho^2} - \frac{s^2}{4\rho^2}} + e^{-\frac{r^2}{4\rho^2} - \frac{s^2}{4\rho^2}} \right), \]  

(3.52)

where we defined \( \frac{1}{\rho^2} = \frac{1}{\rho^2} + \frac{2}{\eta^2 + \rho^2} \) and \( I_0 \) is the modified Bessel function of the first kind.

With Eq. (3.52), we have just shifted the problem from one Hilbert space (functions of two variables with the standard scalar product), to another Hilbert space, that of normalizable functions of one variable with the scalar product defined by \( \langle f | g \rangle = 2\pi \int_0^\infty dr \ r f(r) g(r) \). Notice that in this Hilbert space the operator \( L^s \) is Hermitian, hence it has all-real eigenvalues.

Information on the nodes of the eigenfunctions of (3.52) was derived in McKay and Miller (1990). They started from the assumption, numerically confirmed, that the s-waves eigenfunctions of \( L^{(0)} \) were ordered by their number of nodes. (This assumption is now proven by the formulas for the eigenfunction given above; see Appendix A). As stated, \( L^c \) is simply \( L^{(0)} \) but with an s-wave component (parallel to the arbor density \( A \)) projected away. Using this fact, a series of deductions about eigenlevel shifts were made in McKay and Miller (1990) to show that \( L^c \) must also have its eigenfunctions ordered by their number of nodes, the top eigenfunction having one node (instead of zero) and the others following suit.

In the rest of this chapter, we will proceed to solve the eigenvalue equation for the operator (3.52):

\[ \int_0^\infty ds \ s L^s(r, s) \psi(s) = \frac{\Lambda}{2\pi} \psi(r), \]  

(3.53)
and we will try to estimate in various ways the ratio \( \Xi(\frac{2}{\rho}) = \Lambda/\lambda_{1,0} \) between the principal eigenvalue of \( L^s \) and the principal p-wave eigenvalue of \( \hat{L}^c \), which is \( \lambda_{1,0} = 2\pi\eta^2/\beta^2 \).

It follows from the above that this ratio bears the essential information about the long term symmetries of the synaptic weight. If the ratio \( \Xi \) is larger than unity, then the principal eigenstate of \( L^c \) is an s-wave with radial eigenfunction given by the principal eigenfunction of \( L^s \). Otherwise, the principal eigenfunction of \( L^c \) is the p-wave function \( \psi_{1,0}(r, s) \) computed above.

### 3.4 Diagonalization of the constrained model

#### 3.4.1 Long-range solution

Although the constrained model is not amenable to exact diagonalization, it is easy to show that, in the regime of long-range input correlations (\( \eta \gg \rho \)), rotational symmetry is broken, leading to the development of orientation selectivity.

To see this, we start by noticing that all the radial variables in the eigenvalue equation for \( L^c \) will be confined to a region of order \( \rho \). Hence, the unconstrained operator can be expanded as

\[
L^{(0)}(r, s; \eta) \approx e^{-\frac{r^2+s^2}{4\eta^2}} \left[ 1 - \frac{(r-s)^2}{2\eta^2} \right](3.54)
\]

where further corrections inside the square brackets are of order \( O(\frac{\rho^4}{\eta^4}) \).

Substituting (3.54) into (3.49) leads to the asymptotic matrix element

\[
L^c(r, s; \eta) \approx \frac{r \cdot s}{2\eta^2} \sqrt{A(r)A(s)} = \frac{rs}{2\eta^2} \sqrt{A(r)A(s)} \cos(\phi_r - \phi_s) (3.55)
\]
again with corrections of order $O \left( \frac{1}{\eta^4} \right)$.

It is clear that the only positive eigenvalue of the operator defined by the kernel (3.55) is

$$\psi(r) \propto r \sqrt{A(r)} \cos(\phi - \phi_0)$$

(3.56)

with a positive eigenvalue given by the p-wave eigenvalue $\lambda_{0,1}$ of the unconstrained model.

On the other hand, all the s-wave have zero eigenvalue to this order in the expansion. It follows that for sufficiently long range the principal eigenspace is given by the p-waves.

### 3.4.2 Moveable node theory

We have seen that p-waves dominate the long term dynamics in the long-range regime. We would like to inquire whether there exists a region of parameters where this is not the case, i.e. where rotational symmetry is not broken and the s-waves dominate. These s-waves would describe RFs that are unable to discriminate among the possible orientations of visual input.

If that is the case, there can be no smooth crossover between the two regimes. A linear combination of an s-wave ($m = 0$) and of a p-wave ($m = 1$) could not be an eigenfunction of $L^s$ other than at special points of degeneracy. Let us call $\theta_c = \eta_c / \rho$ the largest value of $h = \eta / \rho_c$ where the principal eigenfunction is non-selective. We would like to find if $\theta_c > 0$ and, if so, to compute the structure of the receptive field for $h < \theta_c$.

A good tool to address this question it is the variational method of quantum mechanics, which may as well be applied to integral operators. The program is carried out as follows: we assume a functional form (trial function) for the principal eigenfunction; we normalize
it; we find the expectation value of our operator in that state; and we maximize it with respect to variational parameters. This leads to the best available approximation of the principal eigenvalue within the given Hilbert subspace\(^5\).

The expectation value of the operator \( \hat{L}^c \) in the trial state \( |\psi\rangle \) is defined as

\[
E[\psi] = \frac{\langle \psi | \hat{L}^c | \psi \rangle}{\langle \psi | \psi \rangle}.
\]  
(3.57)

We have mentioned above that the principal eigenfunction in the \( s\)-sector will be of the \( 2s \) type, i.e. with one radial node. We will choose as our trial function a receptive field with the same functional form as the \( 2s \) eingefungtion of the unconstrained model, only with the position of the node unspecified.

The unconstrained \( 2s \) wave function is, as per Eq. (3.28), a Gaussian of width \( \gamma \) multiplied by the polynomial \((\gamma^2 - r^2)\). We now replace the nodal radius \( \gamma \) with an unspecified value \( R \), and will optimize the expectation value of \( L^c \) with respect to \( R \) all over Hilbert space. Our trial function is therefore

\[
\psi_T(r) = \frac{N}{\sqrt{\pi\gamma}} (R^2 - r^2) \exp\left(-\frac{\gamma^2}{2\gamma^2}\right),
\]  
(3.58)

where the normalization factor is \( N = (2\gamma^4 - 2\gamma^2 R^2 + R^4)^{-1/2} \); this is simply a generalization of Eq. (3.28).

Let us consider the expectation value Eq. (3.57) of the unconstrained operator of Eq.

\(^5\)The analogy enabling this procedure lies in the fact that a Hebbian operator is always bounded from above, just as a Hamiltonian is always bounded from below. Thus the tools customarily applied to Hamiltonians may be applied "upside down" to study Hebbian dynamics.
(3.11) in the state (3.58). This is given by

\[ \mathcal{E}_0 \equiv \langle \psi_T | L^{(0)} | \psi_T \rangle \]

\[ = \frac{N^2}{\pi \gamma^2} \times 4\pi^2 \int_0^\infty \int_0^\infty dr \ r (R^2 - r^2) \ ds \ s \ (R^2 - s^2) I_0 \left( \frac{r s}{\eta^2} \right) e^{- \left( \frac{1}{2\eta^2} + \frac{1}{2\gamma^2} \right) \frac{r^2 + s^2}{2}}. \]

Integration is aided by the following identities, provable from Eq. (3.15):

\[ \alpha = \frac{\beta}{2\eta^2} \]

\[ 1 - 4\alpha^2 \eta^4 + 4\alpha \eta^4 / \gamma^2 = 0, \]

turning Eq. (3.60) into

\[ \mathcal{E}_0(R) = \frac{2\pi \eta^2}{\beta^3} \frac{1 + \beta^2 (1 - R^2 / \gamma^2)^2}{1 + (1 - R^2 / \gamma^2)^2}. \]

To optimize this expectation value, we need to maximize the function \( f(x) = \frac{1 + \beta^2 x}{1 + x} \), where \( x = (R^2 / \gamma^2 - 1)^2 \). The derivative is \( f'(x) = \frac{\beta^2 - 1}{(1 + x)^2} \), always nonnegative because \( \beta \geq 1 \); hence it will be sufficient to maximize \( x \), which is done by choosing the limit \( R \to \infty \). This result is not surprising; in the limit \( R \to \infty \), the moveable node function becomes in fact nodeless, and it is nothing but the 1s Gaussian of width \( \gamma \) which we know as the principal eigenfunction of \( L^{(0)} \).

Let us now consider the expectation value Eq. (3.57) of the full operator \( L^* \) (described in radial coordinates by the matrix element Eq. 3.52) calculated in the moveable-node state
of Eq. (3.58). This can be written as

$$\mathcal{E}(R) = \mathcal{E}_0(R) + \frac{\eta^2 A^2}{\eta^2 + 2\rho^2} - \frac{2\eta^2 AB}{\eta^2 + \rho^2},$$  \hspace{1cm} (3.62)

where

$$A = \frac{N}{\sqrt{\pi\gamma}} \times 2\pi \int_0^\infty (R^2 - r^2) \exp \left[ - \left( \frac{1}{\gamma^2} + \frac{1}{2\rho^2} \right) \frac{r^2}{2} \right] r \, dr \quad (3.63)$$

$$B = \frac{N}{\sqrt{\pi\gamma}} \times 2\pi \int_0^\infty (R^2 - r^2) \exp \left[ - \left( \frac{1}{\gamma^2} + \frac{1}{2\rho^2} + \frac{1}{\rho^2 + \eta^2} \right) \frac{r^2}{2} \right] r \, dr \quad (3.64)$$

or, upon integration,

$$A = 4\sqrt{\pi} N \gamma \rho^3 \frac{(2\rho^2 + \gamma^2)R^2 - 4\gamma^2 \rho^2}{(2\rho^2 + \gamma^2)^2} \quad (3.65)$$

$$B = 4\sqrt{\pi} N \gamma \rho^2 (\rho^2 + \eta^2) \frac{2\rho^2 (R^2 - 2\gamma^2) (\rho^2 + \eta^2) + R^2 \gamma^2 (3\rho^2 + \eta^2)}{[(\rho^2 + \eta^2)(2\rho^2 + \gamma^2) + 2\gamma^2 \rho^2]^2}. \quad (3.66)$$

While expression (3.62) with the substitution of (3.65-3.66) is somewhat intricate, we are ultimately interested only in its maximal value over all the range of nodal radii $R$. We thus expand $\mathcal{E}$ in $h = \eta/\rho$ with the self-consistent ansatz $R^2 = \rho^2 (k^2 h + O(h^2))$, yielding

$$\frac{\mathcal{E}(\rho k \sqrt{h})}{2\pi \eta^2} = 1 - 2 f(k) h + O(h^2),$$  \hspace{1cm} (3.67)

where $f(k) = \frac{8 - 7k^2 + 2k^4}{2 - 2k^2 + k^4}$. The requirements $f'(\bar{k}) = 0$, $f''(\bar{k}) > 0$ lead to

$$\bar{k} = \sqrt{\frac{4 + \sqrt{10}}{3}}, \quad (3.68)$$
which means that the node behaves like as \( R \sim \left( \frac{4+\sqrt{10}}{3} \beta \rho \right)^{1/2} \).

Substituting, we find that the optimal expectation value is

\[
\mathcal{E} \equiv \mathcal{E}(\tilde{r}, \tilde{\eta}) = 2\pi \eta^2 \left[ 1 - (5 - \sqrt{10})h \right].
\]

We can now compare \( \mathcal{E} \) with the exact eigenvalue of the dominant p-wave, which is given by Eq. (3.23) as \( \lambda_{0,1} = 2\pi \eta^2 / \beta^2 \sim 1 - 2h \). Since \( (5 - \sqrt{10}) \sim 1.83 < 2 \), we conclude that the principal s-wave eigenvalue approximated by Eq. (3.58) lies higher. Therefore, the s-waves do indeed dominate for small \( h = \frac{1}{2} \).

This means that, if the visual input decorrelates at sufficiently short distances, the system will never learn to recognize orientations\(^6\).

### 3.4.3 A lower bound for the transition point

We will now show that symmetry is broken for sufficiently large correlation lengths, and we will compute a lower bound on the critical value of the correlation length \( \eta_c = \theta_c \rho \). We can do so by equating our expectation value with the exact 2p eigenvalue, and finding where they cross. The fact that this yields a lower bound on the actual value of the transition point can be proven as follows.

If the variational method reveals the transition at a point \( \theta_c \), it means we have found an s-wave state whose expectation value is larger than the exact eigenvalue of the principal p-\( ^6\)To understand this intuitively, we may think of real-life situation in which scientists train an animal by displaying bars oriented in various directions (e.g. to associate them to rewards); if the bars are too short, their most relevant feature will be their position in space, and not their orientation. The animal, therefore, will be trained to recognize the location of small objects, and not the orientation of lines.
Figure 3.4: Expectation values of the constrained Linsker operator (in units of $2\pi\eta^2$) plotted as a function of $R/\rho$. The three curves refer to: (1) $\lambda_{0,1}$ (expectation value of $L^{(0)}$ or $L^c$ in the exact eigenfunction $\psi_{0,1}$), plotted in green; (2) $\mathcal{E}_0$, expectation value of the unconstrained operator $L^{(0)}$ in the moveable-node state $\psi_T$, plotted in blue; (3) $\mathcal{E}$, expectation value of the constrained operator $L^c$ in the state $\psi_T$, plotted in purple. The figure refers to $h = \frac{2}{\rho} = .01$. Values of $R$ for which the moveable-node state is preferred to the orientation-selective state are different for the two operators: $L^{(0)}$ opts for $\psi_T$ for sufficiently high values of the node radius, $R > \tilde{R}$; $L^c$, for values of $R$ in a narrow window $R \gtrsim \tilde{R}$.

wave for all $h < \theta_c$. Now call $\Theta_c$ the actual critical value of $\eta/\rho$, and suppose ad absurdum that the actual critical point $\Theta_c$ is $\Theta_c < \Theta_c$. That means in the region $\Theta_c < h < \theta_c$ the actual principal state of the operator is a orientation-selective, i.e. $m > 0$. And since the $m > 0$ sector is exactly diagonalizable, this principal p-wave must be the one we already calculated, with eigenvalue $\lambda_{0,1}$.

But if that was true, all the s-wave function would yield expectation values lower than that eigenvalue. Then it would not be possible create a linear combination of them (our trial function) that yields an expectation value $> \lambda_{0,1}$, as we have done. We deduce that we must have $\Theta_c \geq \theta_c$. That is, the variational method provides a lower bound on the actual critical point.
Let us proceed with the calculation. We must first expand $R$ to a higher order, as
\[ R^2 / \rho^2 = c_1 h + c_2 h^2 + O(h^3). \]
The coefficient $c_1$ can be determined by maximizing Eq. (3.62) to the order $O(h)$, which gives
\[ c_1 = \bar{k}^2 = \frac{4 + \sqrt{10}}{3}. \]
Then we calculate the second term in the expansion of $\mathcal{E}$, plug in the value of $c_1$ we found, and maximize with respect to $c_2$. This second order correction, computed at the optimal value of $c_2$, is then included in the expectation value, and the whole thing is compared to the eigenvalue of the leading p-waves, to see which is dominating. One obtains
\[
\mathcal{E} \approx 1 - \left(5 - \sqrt{10}\right) h + \left(\frac{33}{2} - \frac{51}{\sqrt{10}}\right) h^2.
\]
(3.70)

The critical point $\theta_c$ is found where this s-wave expectation value intersects the p-wave eigenvalue given by (Eq. 3.23), that is,
\[
\frac{\lambda_{0,1}}{2\pi \eta^2} \sim 1 - 2h + 2h^2; \quad \text{(3.71)}
\]
setting $\mathcal{E} = \lambda_{0,1}$ yields
\[
\theta_c = \frac{2(75 - 8\sqrt{10})}{997} \approx 0.1.
\]
(3.72)

This is a rigorous lower bound to the critical point, which indeed in the numerical optimization of the moveable node-theory is found to be $\approx 0.3$ (Fig. 3.5), and a full numerical diagonalization of the full s-sector gives close to $\theta_c \approx 0.5$.  

48
Figure 3.5: The ratio $\Xi = \Lambda_s/\lambda_{1,0}$ plotted as a function of $h = \eta/\rho$, as computed by diagonalizing the constrained operator in the sector spanned by $\psi_{0,0}$ and $\psi_{1,0}$ (a procedure equivalent to the moveable node theory). Above, 2D plots of the receptive fields corresponding to the resulting principal eigenfunction. Below, plots of the highest-laying nonselective eigenfunction (2s) at two representative point in parameter space.

### 3.5 Conclusions

We had begun this chapter with the statement that the matrix of synaptic weights is ”plastic”, that is, synaptic weights should be modeled as varying over time in a way that depends on neural activity, reflecting the observed behavior of real synapses. We have then introduced a simple paradigm used to model this time evolution of the synaptic weight matrix – the
Hebbian paradigm, according to which the dynamical changes in a synapse depend directly on the correlation between the activities of the cells it connects.

We proceeded to show analytically how such a mechanism can lead to a functionally important possibility – the spontaneous breaking of symmetries possessed by the synaptic dynamics. The example we focused on concerns the breaking of rotational symmetry in the feedforward weights to cells in the striate visual cortex cells. This is what provides mammals with the ability to recognize orientations, without any need for a genetically prescribed mapping of synaptic connections.

The model of one cortical cell we treated is due to Linsker (1986), who explored its properties numerically. The analytic solution of the model, which was still missing, has been provided in this study. Crucial to symmetry breaking is the presence of homeostatic constrains. We encoded this, following the work of von der Malsburg (1973) and Linsker (1986), by enforcing a conservation rule on the total synaptic input to the cortical cell.

To understand analytically the behavior of the model, we began by considering its unconstrained version (i.e. without any conservation rule) where it was possible to provide a complete diagonalization (further studied in appendix A). In the presence of constraints, the model could only be studied by approximate methods.

The key dimensionless parameter is the decay range of presynaptic correlations, measured in terms of the arbor radius. We drew inspiration from the variational method of quantum mechanics, which allowed to optimize the RF in the "nonselective" sector. We thus obtained lower bounds on the eigenvalues and an analytical estimate of the critical correlation length. Moreover, we derived estimates for the position of the first node as a function of the correlation length.
These results\textsuperscript{7} place on a somewhat firmer mathematical footing what had been known so far only through numerics. More importantly, they form the basis for tackling analytically theories of the interconnected cortical layer that involve Linsker-type cells as primary components.

In the next chapter, after discussing some recent experimental results, we will formulate a more realistic model for orientation selectivity, whose construction will require a deeper look into the biology of visual cortex. In the process of solving it analytically, results from this chapter will prove necessary as intermediate steps.

\textsuperscript{7}Which have partly already appeared in Fumarola and Miller, 2017
Chapter 4

A multilayer model of orientation development

4.1 The Hebbian theory of orientation maps

4.1.1 Review of the experimental literature

In the previous chapter, we discussed how a single cortical cell (considered apart from all the others) could develop a certain functional property (orientation preference) through Hebbian plasticity. Here, we turn to modeling the cortex as a largely recurrent network. This is a statement amply justified by phenomenology; indeed, synapses internal to cortical layers and carrying signal in between cortical cells make up most of the synapses in visual cortex, and are considerably more abundant than feedforward synapses (Binzegger et al. 2004, Stepanyants et al., 2008, Thomson and Lamy, 2007). The problem of modeling Hebbian development in a largely recurrent network will be tackled, again, in a minimal mechanistic fashion. The development of preferred orientation will be used as the testing ground for our approach.

It has been found experimentally that cells’ orientation preferences are locally continuous; moreover, as a recording electrode is moved tangentially through the cortex (Hubel & Wiesel, 1974) preferred orientation changes, either clockwise or counter-clockwise at a roughly constant rate. Measurements of the corresponding angular observable result in
plots such as those of Fig. (4.1). We will refer to these as "orientation maps", although the name is usually applied to those animals in which the arrangement of orientation is quasiperiodical, as with cats and ferrets. In many species, (mice, rats, squirrels, rabbits, hares) these maps show what is called a salt-and-pepper organization of selectivity (for a review, Kaschube, 2014).

Figure 4.1: Maps of 4.32×3.22 mm regions in monkey V1 and cat area 18. Preferred orientation (represented by hue) ranges linearly in 16 steps from vertical (purple, see the color scale under the images), counterclockwise to 11.25 degrees clockwise from vertical (reddish color). Images have been scaled so that white and black represent, respectively, the maximum and the minimum difference D between ON- and OFF- center strengths among illustrated receptive fields. Points of high orientation gradient are shown as white. Map of monkey V1 (area 17) comes from Ts’o et al., 1990; map of cat area 18 from Bonhoeffer and Grinvald, 1991.

A classic result in theoretical neuroscience is that this functional organization of orientation preference across the cortex can arise through Hebbian dynamics from a Mexican-hat structure of input correlations, with same-center-type inputs being more correlated than opposite-center-type at short separations and the converse at longer separations.

The rationale for the assumption of Mexican-hat correlations comes from a general phenomenon known as surround suppression. This phenomenon was first observed in gan-
glion cells (neurons located near the inner surface of the retina) in cats by Kufler (1953) who found that the RFs of these cells consist of a center surrounded by an antagonistic annulus. A similar conclusion was drawn by Wiesel and Hubel (1966) in their study of LGN cells in the rhesus monkey primary visual cortex. It soon appeared, moreover, that while ON cells have a subtractive surround, OFF cells an additive surround (Barlow, 1953; for a historical review, Spillmann, 2014).

If LGN cells are modeled in this fashion, the correlation between the activity of two cells, plotted as a function of distance, is bound to behave nonmonotonically. Indeed, the correlation will be null at distances so long that the receptive fields do not overlap. In the thin (possibly negligible) range of distances such that the cells’ surrounds overlap, correlations will be positive for cells of the same type, negative otherwise. At distances such that the surround of one cell overlaps with the core of the other one, correlations will be negative for same-type cells, positive otherwise. And for distances short enough as to involve a substantial overlap of the two cores, correlations are be positive for same-type cell, negative otherwise.

This scenario, depicted in Fig. 4.2 leads naturally to a Mexican-hat shaped correlation function. Moreover, the profiles of the same-type and opposite-type correlation functions, having opposite signs, will sum up into a profile \( C_{\text{diff}} = C_{\text{same}} - C_{\text{opp}} \) that is also a Mexican hat.

The Mexican-hat assumption was used in the simulations of Hebbian development performed by Miller (1994) and in the analytic study of Wimbauer et al. (1998). In Miller (1994), an extensive numerical study was carried out, leading to the conclusion that "the development of simple-cell receptive fields and cortical maps can result from a Mexican hat.
correlation structure in the dark activity of ON- and OFF-center inputs.” Wimbauer et al. arrived at the more radical conclusion that “in order to get orientation-selective receptive fields, the spatial correlation function of the inputs that drive the development must have a zero crossing.”

The reason why such input correlations facilitate a spatial modulation in the orientation of the receptive fields is easy to grasp. Due to the sign change, an additional length scale appears in the correlation function as compared to a simple fall-off. A Mexican-hat profile makes cortical-space Fourier modes of finite wavelength grow fastest, and the system adapts to the profile of correlations by rotating the RF orientation across the cortex.

Figure 4.2: Argument for the emergence of Mexican-hat correlations between LGN cells. Upper row: correlation between the activity of two cells of the same type (ON or OFF) as a function of distance. Lower row: correlation between the activity of the cells of opposite type (ON-OFF) as a function of distance

This scenario has remained a piece of common wisdom on Hebbian models until 2006, when the first direct measurements of correlations in LGN activity as a function of distance were published. Experiments on ferrets by Ohshiro and Weliky (2006) revealed that no Mexican hat is actually observed in layers A/A1 of LGN. Same-center-type pairs are best correlated at all separations, and the decay with distance is monotonic, as we assumed it to
Measurements were taken at various developmental times and in the presence of various visual inputs (Fig 4.3). The only hint to the existence a zero-crossing in the correlation functions appear with noisy inputs in the 10th week after birth, when orientation maps have already developed.

While such measurements have only performed on ferrets so far, and not yet on cats, they pose a problem for existing Hebbian models, and raise the issue of whether orientation maps can somehow be constructed in the absence of the above-mentioned mechanism.

In this chapter, we are going to show that this is indeed the case if the role of homeostatic constraint is properly taken into account. To do so, in the rest of this section we will set up a two-layer model that improves on the model we used in Chapter 3 through the introduction of lateral connections in the cortex. In section 4.2 we will proceed to analyze its general properties; in Sections 4.3-4.7 we will compute its full phase diagram, adopting for each regime the best-suited tool from a palette of perturbative, variational, and asymptotic methods. We will conclude in Sec. 4.8 with an overview of numerical results.

### 4.1.2 Unconstrained dynamics of the model

We are going to employ the following notation. Presynaptic positions will be called with Greek indices ($\alpha, \beta, \ldots$), postsynaptic positions with the last letters of the Latin alphabet ($x, y, \ldots$), and differences between pre- and post-synaptic positions with letters from the middle of the Latin alphabet (e.g. $r = \alpha - x$). We call $r(x, \tau), r^{(\text{ON})}(\alpha, \tau), r^{(\text{OFF})}(\alpha, \tau)$ the firing rate of neurons in the cortical and of ON or OFF neurons in the LGN layer at time
Figure 4.3: Developmental changes in the spatial pattern of stimulus-evoked and spontaneous LGN neural activity in awake ferrets, as measured by Ohshiro and Weliky (2006). (a) Correlation functions of pairs with the same center sign ($C_{\text{same}}$) in four different age groups. (b) Correlation functions of pairs with the opposite center sign ($C_{\text{opp}}$). (c) Difference of correlation function $C_{\text{diff}}$. Correlation functions for spontaneous activity, activity evoked by natural scene images and activity evoked by white noise are represented by bold, stippled and dotted lines, respectively.

$\tau$ (omitting the time variables wherever obvious). The ON-center and OFF-center synaptic strengths will be $s_{\text{ON}}(x, \alpha, \tau)$ and $s_{\text{OFF}}(x, \alpha, \tau)$.

In the linear regime, applying Eq. (2.9) to this model, we find that the cortical firing
Figure 4.4: Detailed depiction of the two-layer model, including lateral connections in the cortex that were neglected in Chapter 3.

Rates evolve according to

\[ T(x) \frac{dr(x)}{d\tau} = -\gamma r(x) + \sum_{i=ON, OFF} \int d\alpha \ s^{(i)}(x, \alpha) r^{(i)}(\alpha) + \int dy \ W(x, y) r(y), \tag{4.1} \]

where \( W \) is the matrix of synaptic weights for lateral connections in the cortex, and \( T(x) \) is a diagonal matrix whose entries are the time scales of neural activity, which are typically much faster than the typical time scale of synaptic plasticity.

The smallness of the entries of \( T \) allows us, as was done before with Eq. (3.3), to describe synaptic development by relying on the steady state of the fast dynamics Eq. (4.1). The latter is given by

\[ r(x) = \sum_{i=ON, OFF} \int dy d\alpha \ I(x, y) s^{(i)}(y, \alpha)r^{(i)}(\alpha), \tag{4.2} \]
where \( I_{x,y} = [\gamma(1 - W)^{-1}]_{x,y} \).

Hebbian dynamics (Eq. 3.1) implies that the synaptic weights evolve according to

\[
\frac{d s^{\text{ON/OFF}}(x, \alpha)}{dt} \propto r(x) r^{\text{ON/OFF}}(\alpha). \tag{4.3}
\]

Replacing Eq. (4.2) into Eq. (4.4),

\[
\frac{d s^{(i)}(x, \alpha)}{dt} \sim \sum_{j=\text{ON,OFF}} \int dy d\beta \ I(x, y) s^{(j)}(y, \beta) r^{(i)}(\alpha) r^{(j)}(\beta). \tag{4.4}
\]

where the index \((i)\) distinguishes ON and OFF cells.

We will now average Eq. (4.4) over a time scale much longer than the typical time scale of firing-rate dynamics but much shorter than the typical time scale of the synaptic evolution. This separation of time scales leads to

\[
\frac{d s^{(i)}(x, \alpha)}{dt} \sim \sum_{j=\text{ON,OFF}} \int dy d\beta \ I(x, y) s^{(j)}(y, \beta) C^{(i,j)}(\alpha - \beta), \tag{4.5}
\]

where \( C^{(i,j)}_{\alpha - \beta} = \langle r^{(i)}_\alpha r^{(j)}_\beta \rangle \).

Again, the quantity of interest is the net synaptic weight \( s(x, \alpha) = s^{\text{ON}}(x, \alpha) - s^{\text{OFF}}(x, \alpha) \). We will now see how to describe its long-term dynamics in the presence of homeostasis.
4.1.3 Homeostatic constraint

In Chapter 3, we discussed that correlation-based development entails competition between the various components of a neural system, and that this competition is usually taken into account through conservation rules.

We followed classic papers in the field by adopting a rule that conserved the total incoming synaptic strength to a generic cortical cell. This was enough to allow for a rigorous derivation of various results on the emergence of selectivity.

Now that we are considering a larger set of cortical cells interacting with each other, we may question the plausibility of the traditional conservation rule. From the physiology of the problem, indeed, there are reasons to believe that the competition affecting $S^{(ON)} - S^{(OFF)}$ happens at the level of the arbors, not at the level of the receptive fields. Research has repeatedly shown that axonal arbors compete with each other throughout development (Elliott et al. 1997; Gan and Macagno, 1997). If they lose overall synaptic strength, they compete more effectively to retain it; if they have too much, they compete less effectively, so that arbor retraction takes place (for a review, Luo and O’Leary, 2005; Jacobson, 2013 – Chapter 8). While little is known about the mechanisms underlying this homeostatic balancing, it appears that no single LGN input can increase its innervation strength into the cortex beyond its share (nor will it be competed away by others).

Dendritic competition occurring in the cortex may also take place, but it will most likely perceive the sum of incoming synaptic input from ON and OFF cells, rather than their difference; axonal competition in LGN, on the other hand, concerns ON and OFF cells individually, and therefore will affect the difference between their projections strengths. In can
therefore be argued that the physiology of the problem, rather than justifying the traditional constraint, suggests a constraint mechanism conserving the total projection strength from each presynaptic cell, i.e. the net synaptic output from each LGN neuron.

This amounts to fixing

$$\frac{d}{dT} \sum_x s^{ON}_{x,\alpha} = \frac{d}{dT} \sum_x s^{OFF}_{x,\alpha} = 0 \ \forall \alpha. \quad (4.6)$$

This can be satisfied through the addition of some yet-unspecified terms to Eqs. (4.5), turning them into

$$\frac{d}{dT} s^{ON}_{x,\alpha} = A_{x-\alpha} \sum_{y,\beta} I_{x-y} \left[ C^{ON,ON}_{\alpha-\beta} s^{ON}_{y,\beta} + C^{ON,OFF}_{\alpha-\beta} s^{OFF}_{y,\beta} \right] - \epsilon^{ON}_{\alpha} A_{x-\alpha}; \quad (4.7)$$

$$\frac{d}{dT} s^{OFF}_{x,\alpha} = A_{x-\alpha} \sum_{y,\beta} I_{x-y} \left[ C^{OFF,OFF}_{\alpha-\beta} s^{OFF}_{y,\beta} + C^{OFF,ON}_{\alpha-\beta} s^{ON}_{y,\beta} \right] - \epsilon^{OFF}_{\alpha} A_{x-\alpha}; \quad (4.8)$$

where $\epsilon^{ON}$ and $\epsilon^{OFF}$ will be defined in such a way as to implement the conservation constraints Eq. (4.6) of the model.

Define now $t^{ON}_{x,\alpha} = \frac{s^{ON}_{x,\alpha}}{\sqrt{A_{x-\alpha}}}$ and $t^{OFF}_{x,\alpha} = \frac{s^{OFF}_{x,\alpha}}{\sqrt{A_{x-\alpha}}}$, so that Eqs. (4.7-4.8) become

$$\frac{d}{dT} t^{ON}_{x,\alpha} = \sqrt{A_{x-\alpha}} \sum_{y,\beta} I_{x-y} \left[ C^{ON,ON}_{\alpha-\beta} \sqrt{A_{y-\beta}} t^{ON}_{y,\beta} + C^{ON,OFF}_{\alpha-\beta} \sqrt{A_{y-\beta}} t^{OFF}_{y,\beta} \right] - \epsilon^{ON}_{\alpha} \sqrt{A_{x-\alpha}}; \quad (4.9)$$

$$\frac{d}{dT} t^{OFF}_{x,\alpha} = \sqrt{A_{x-\alpha}} \sum_{y,\beta} I_{x-y} \left[ C^{OFF,OFF}_{\alpha-\beta} \sqrt{A_{y-\beta}} t^{OFF}_{y,\beta} + C^{OFF,ON}_{\alpha-\beta} \sqrt{A_{y-\beta}} t^{ON}_{y,\beta} \right] - \epsilon^{OFF}_{\alpha} \sqrt{A_{x-\alpha}}. \quad (4.10)$$

It will be convenient to regard the functions $t^{ON}$ and $t^{OFF}$ as vectors in a Hilbert space, so that (in bra/ket notation) $t^{ON}_{x,\alpha} = \langle x, \alpha | t^{ON} \rangle$ and $t^{OFF}_{x,\alpha} = \langle x, \alpha | t^{OFF} \rangle$. 
Again using, as in the previous section, \( C_{ON,ON} = C_{OFF,OFF} \) and \( C_{ON,OFF} = C_{OFF,ON} \), we can rewrite Eqs. (4.9-4.10) in terms of the sole operators \( L^s \) and \( L^d \), defined by:

\[
\langle x, \alpha | L^s | y, \beta \rangle = \sqrt{A_{x-\alpha}} I_{x-y} C_{\alpha-\beta}^{ON,ON} \sqrt{A_{y-\beta}}, \quad \langle x, \alpha | L^d | y, \beta \rangle = \sqrt{A_{x-\alpha}} I_{x-y} C_{\alpha-\beta}^{ON,OFF} \sqrt{A_{y-\beta}}. \tag{4.11}
\]

Namely, we have

\[
\frac{d}{d\tau} \langle \alpha | t^{ON} \rangle = L^s \langle \alpha | t^{ON} \rangle + L^d \langle \alpha | t^{OFF} \rangle - \mathcal{E}^{ON} | \alpha \rangle; \tag{4.13}
\]
\[
\frac{d}{d\tau} \langle \alpha | t^{OFF} \rangle = L^s \langle \alpha | t^{OFF} \rangle + L^d \langle \alpha | t^{ON} \rangle - \mathcal{E}^{OFF} | \alpha \rangle. \tag{4.14}
\]

Here, the vector \( |\alpha\rangle \) is defined by \( \langle x, \alpha | \alpha \rangle = \sqrt{A_{x-\alpha}} \). The operators \( \mathcal{E}^{ON}, \mathcal{E}^{OFF} \) that appear in Eqs. (4.13-4.14) have the form

\[
\mathcal{E}^{ON} = \sum_{\beta} \epsilon^{ON}_{\beta} P_{\beta}, \quad \mathcal{E}^{OFF} = \sum_{\beta} \epsilon^{OFF}_{\beta} P_{\beta}; \tag{4.15}
\]

where \( P_{\beta} \) is the operator that effects projection into the subspace with basis \( \{|y, \beta\}_y\); that is,

\[
P_{\beta} = \sum_{y} |y, \beta\rangle \langle y, \beta|. \tag{4.16}
\]

The expressions for \( \epsilon^{ON}, \epsilon^{OFF} \) can be found from the conservation laws Eq. (4.6), which may be rewritten in the form

\[
\frac{d}{d\tau} \langle \alpha | P_{\beta} | t^{ON} \rangle = 0 \quad \forall \beta, \quad \frac{d}{d\tau} \langle \alpha | P_{\beta} | t^{OFF} \rangle = 0 \quad \forall \beta. \tag{4.17}
\]
Substituting Eqs. (4.13-4.14) into the two constraints (4.17), and using the expressions (4.15) for $\mathcal{E}^{\text{ON}}$, $\mathcal{E}^{\text{OFF}}$, we find:

\[
\epsilon_{\beta}^{\text{ON}} = \frac{\langle a | P_\beta L^s | t^{\text{ON}} \rangle + \langle a | P_\beta L^d | t^{\text{OFF}} \rangle}{\langle a | P_\beta | a \rangle}, \quad \epsilon_{\beta}^{\text{OFF}} = \frac{\langle a | P_\beta L^s | t^{\text{OFF}} \rangle + \langle a | P_\beta L^d | t^{\text{ON}} \rangle}{\langle a | P_\beta | a \rangle} \quad (4.18)
\]

Since the quantity of interest is $s_{x,\alpha} = s_{x,\alpha}^{\text{ON}} - s_{x,\alpha}^{\text{OFF}} = \sqrt{A_{x,\alpha}} (t_{x,\alpha}^{\text{ON}} - t_{x,\alpha}^{\text{OFF}})$, we must proceed to compute the time evolution of $|t\rangle \equiv |t^{\text{ON}} - t^{\text{OFF}}\rangle$. From Eqs. (4.13-4.14), we find

\[
\frac{d}{dt} |t\rangle = \left(1 - \sum_\beta \frac{P_\beta |a\rangle \langle a | P_\beta}{\langle a | P_\beta | a \rangle}\right) L |t\rangle, \quad (4.19)
\]

where $1$ stands for the identity operator and we have defined $L \equiv L^s - L^d$.

The matrix elements of the operator $L$ in the $|x, \alpha\rangle$ basis are

\[
\langle x, \alpha | L | y, \beta \rangle = \sqrt{A_{x,\alpha}} I_{x-y} C_{\alpha-\beta} \sqrt{A_{y,\beta}}, \quad (4.20)
\]

where $C_{\alpha-\beta} = C_{\alpha-\beta}^{\text{ON,ON}} - C_{\alpha-\beta}^{\text{ON,OFF}}$.

### 4.1.4 Projection operators

To rewrite Eq. (4.19) in a more transparent form, we may define the single-arbor ket $|a_\beta\rangle = P_\beta |a\rangle$, with elements $\langle x, \alpha | a_\beta \rangle = \delta_{\alpha,\beta} \sqrt{A_{x,\alpha}}$. Let us also notice that the projection operator $P_\beta$ of Eq. (4.16) is orthogonal, hence self-adjoint. Using this fact, as well as the
idempotence of \( P_\beta \), we obtain

\[
\frac{d}{dt}|t\rangle = \left[ 1 - \sum_\beta \frac{|a_\beta\rangle\langle a_\beta|}{\langle a_\beta|a_\beta\rangle} \right] L|t\rangle. \tag{4.21}
\]

Defining \( P = 1 - \sum_\beta \frac{|a_\beta\rangle\langle a_\beta|}{\langle a_\beta|a_\beta\rangle} \), we have from Eq. (4.21)

\[
\frac{d}{dt}|t\rangle = PL|t\rangle = P|t\rangle + P(L(1 - P)|t\rangle. \tag{4.22}
\]

Notice that if the initial condition is such that \( P|t_0\rangle = |t_0\rangle \), this property will be conserved by Eq. (4.21)\(^1\). In this case, we can drop the last term in Eq. (4.22) and write simply

\[
\frac{d}{dt}|t\rangle = P(L|t\rangle \equiv L^p|t\rangle \tag{4.23}
\]

In the following, we will be interested in exploring the principal eigenspace of the operator \( L^p = PLP \), as this determines the fastest growing modes. The biologically relevant questions will be: (1) whether this eigenspace contains orientation-selective RFs; (2) if these are uniform in cortical space or they break translational invariance in the cortical variable, thus giving rise to a cortical modulation of orientation preferences.

### 4.2 Properties of the time-evolution operator

In this section, we will analyze the basic properties of the time-evolution operator \( L^p \) of Eq. (4.23), namely: 1) the form of its matrix elements, in real space and after a Fourier transform

\(^1\)This would be equivalent to having the same number of OFF and ON cells.
in the cortical variables, (2) its semipositivity, (2) the general structure of its spectrum, (3) its commutation properties with translation and rotation operators, (4) its symmetry with respect to parity, complex conjugation, and their combination. Finally, we will diagonalize exactly the unconstrained operator $L$, which will serve as a starting point for studying the properties of $L^p$ in greater detail.

4.2.1 Matrix elements

Written explicitly, the generic matrix element of the operator $L^p$ of Eq. (4.23) is

$$L^p(x, \alpha; y, \beta) = \int dx_1 d\alpha_1 dx_2 d\alpha_2 \left[ \delta(x - x_1)\delta(\alpha - \alpha_1) 
- \sqrt{A(x - \alpha)}\delta(\alpha - \alpha_1)\sqrt{A(x_1 - \alpha_1)} \right] L(x_1, \alpha_1; x_2, \alpha_2) \times
\left[ \delta(x_2 - y)\delta(\alpha_2 - \beta) - \sqrt{A(x_2 - \alpha_2)}\delta(\alpha_2 - \beta)\sqrt{A(y - \beta)} \right].$$  (4.24)

Switching to the relative coordinates $r = \alpha - x$ and integrating out the delta functions, from Eq. (4.24) we obtain

$$L^p(x, r; y, s) = L(x, r; y, s) + S(x, r; y, s) + T(x, r; y, s) + \tilde{T}(x, r; y, s),$$  (4.25)

where

$$\frac{S(x, r; y, s)}{\sqrt{A(r)A(s)}} = \int ds_1 ds_2 \sqrt{A(s_1)A(s_2)} L(x + r - s_1, s_1; y + s - s_2, s_2)$$  (4.26)

$$T(x, r; y, s) = -\sqrt{A(r)} \int du \sqrt{A(u)} L(x + r - u, u; y, s)$$  (4.27)

$$\tilde{T}(x, r; y, s) = -\sqrt{A(s)} \int du \sqrt{A(u)} L(x, r; y + s - u, u).$$  (4.28)
Comparing Eqs. (4.9-4.10) with Eqs. (4.13-4.14), we see that the matrix elements of 
$L = L^d - L$ is

$$L(x, r; y, s) = \sqrt{A(r)} \sqrt{A(s)} I(x - y)C(x - y + r - s).$$  \tag{4.29}

For the functions $I$ and $C$, the simplest assumption coherent with experiments (see the discussion in Sec.4.1.1) is a Gaussian dependence on distances:

$$C(\alpha, \beta) = \exp \left[ -\frac{(\alpha - \beta)^2}{2\zeta^2} \right], \quad I(x, y) = \exp \left[ -\frac{(x - y)^2}{2\eta^2} \right],$$  \tag{4.30}

where $\eta$ and $\zeta$ are the two characteristic length-scales\(^2\). These, and the arbor radius $\rho$, are the coordinates of the parameter space of the model.

### 4.2.2 Fourier Transform

Let us now obtain the Fourier transform of the matrix element (4.25) in the cortical variable $x$. We can use the translational invariance of $L$ and (Eq. 4.29) and of $L^p$ (see Eq. 4.24) to define

$$L(x, r; y, s) \equiv L(x - y; \Delta = x - y) = \int \frac{d\omega}{(2\pi)^2} e^{-i\omega\Delta} L(r, s; \omega).$$  \tag{4.31}

We will refer to the spatial frequency vector $\omega$ as the cortical wavevector and to its modulus as the wavenumber; analogously, the Fourier transform of the constrained operator, Eq.\(^2\)Notice that, differing from the previous chapter, we are calling $\eta$ the length scale of recurrent cortical interaction.
\[
L^p(r, s; \omega) = L(r, s; \omega) - \sqrt{A(r)} \int du \sqrt{A(u)} L(u, s; \omega) e^{-i\omega(r-u)}
- \sqrt{A(s)} \int du \sqrt{A(u)} L(r, u; \omega) e^{-i\omega(u-s)} + \sqrt{A(r)A(s)} \int ds_1 ds_2 \sqrt{A(s_1)A(s_2)} L(s_1, s_2; \omega) e^{-i\omega(r-s-s_1+s_2)}. \tag{4.32}
\]

With the choice (4.30) for the interaction and correlation functions, the Fourier transform of Eq. (4.29) reads
\[
L(r, s; \omega) \sim L^{(0)}(r, s; \mu) \exp \left[ -\frac{\omega^2}{2\Omega^2} - i \frac{\eta^2}{\mu^2} \omega(r - s) \right], \tag{4.33}
\]
where we have neglected an overall prefactor that can be absorbed in the definition of time.

The constants \( \mu \) and \( \Omega \) in Eq. (4.33) are given by
\[
\mu^2 = \eta^2 + \zeta^2 \quad \quad \Omega^2 = \frac{1}{\eta^2} + \frac{1}{\zeta^2}, \tag{4.34}
\]
and \( L^{(0)}(r, s; \mu) = \sqrt{A(r)A(s)} e^{-\frac{(r-s)^2}{2\mu^2}} \) is the time-evolution operator of the model studied in the previous chapter (see Eq. (3.10) of Sec. 3.1.3), with its correlation parameter replaced here by a Pythagorean combination of the two length scales in the problem.

We now: (1) substitute the Fourier-transformed matrix element of \( L \) Eq. (4.33) into the expression (4.32) that we derived for the matrix element of \( L^p \); (2) insert the Gaussian form of the arbors \( A(r) = \frac{1}{2\pi \rho^2} \exp \left( -\frac{r^2}{2\rho^2} \right) \); (3) perform the integral over all the intermediate space variables.
We thus arrive at decomposing the constrained operator of Eq. (4.24) into

\[ \hat{L}^p = \hat{L} + \hat{S} + \hat{T} + \hat{T}^t, \] (4.35)

where

\begin{align*}
L(r,s;\omega) &= \exp\left[-\frac{\omega^2}{2\Omega^2} - \frac{i}{\mu^2} \omega(r - s) - \frac{r^2 + s^2}{4\rho^2} - \frac{(r - s)^2}{2\mu^2}\right]; \quad (4.36) \\
S(r,s;\omega) &= \frac{\mu^2}{\mu^2 + 2\rho^2} \exp\left[-\frac{\omega^2}{2\Omega^2} - \frac{\rho^2\xi^4\omega^2}{2\mu^2(\mu^2 + 2\rho^2)} - i\omega(r - s) - \frac{r^2 + s^2}{4\rho^2}\right]; \quad (4.37) \\
T(r,s;\omega) &= -\frac{\rho^2}{\mu^2 + \rho^2} \exp\left[-\frac{\omega^2}{2\Omega^2} - \frac{\rho^2\xi^4\omega^2}{2\mu^2(\rho^2 + \mu^2)} - \frac{1}{2} \left(\frac{1}{2\rho^2} + \frac{1}{\rho^2 + \mu^2}\right)r^2 \\
&\quad - \frac{s^2}{4\rho^2} - i\omega\left(\frac{\rho^2 + \eta^2}{\rho^2 + \mu^2}r - s\right)\right]; \quad (4.38)
\end{align*}

and, just as in the one-cell problem (see Eq. 3.49 of Sec. 3.3), the additive constraint operators \( \hat{S} \) and \( \hat{T} \) have fully separable matrix elements.

### 4.2.3 Positive semidefiniteness

It will be useful to rely on the positive semidefiniteness of the Fourier-transformed operator \( L^p(\omega) \) for any given \( \omega \).

Consider Eq. (4.32), and supposed we regard the direction of the wavevector \( \omega \) as fixed (in the following, we will take it be parallel to the x-axis). We can then write

\[ L^p(\omega) = \left(1 - |a_\omega\rangle\langle a_\omega|\right)L(\omega)\left(1 - |a_\omega\rangle\langle a_\omega|\right), \]

(4.39)

where \( \langle r|a_\omega\rangle = \sqrt{A(r)}e^{-i\omega r_x} \).
Thus, even for a given wavenumber \( \omega \), the constrained operator is nothing but the unconstrained operator sandwiched between two identical projection operators. Accordingly, one can proceed as we did in the previous chapter (Sec. 3.3.1) to show that the operator is positive semidefinite.

4.2.4 Long-term dynamics

A notable difference between this model and the one-cell model of Chapter 3 is that here, because of translational invariance, the spectrum of the time-evolution operator is continuous.

Once the operator is diagonalized at an arbitrary spatial frequency, we expect to find a series of possibly overlapping bands, corresponding for low wavenumbers to different rotational states. Call \( \Lambda_M(\omega) \) the eigenvalue structure of the principal band, and \( \omega_M \) the position of its (possibly broad) maximum, corresponding to the fastest-growing eigenspace.

The developmental process has a similar effect as filtering with a spatially isotropic bandpass. The function \( \Lambda_M(\omega) \) can be interpreted as the corresponding filter profile, and the location of the maximum of this filter may depend in nontrivial ways on the parameters of the model.

If the system lies in parameter space at a point such that \( \omega_M = 0 \), the dynamics will tend to flatten out any inhomogeneity in the initial condition. If \( \omega_M > 0 \), on the other hand, the long-term RF will be heterogeneous on a scale \( \sim 1/\omega_M \). Of course, because we have a broad maximum of nearly optimal wavevector, we may expect local but no long-range periodicity.
The evolution of the RF at any given point in the cortex, finally, may cancel or emphasize whatever degree of orientation selectivity is possessed by the initial condition, depending on the structure of the eigenspace associated to the principal mode.

Hence, we can gain insight on long-term behavior by focusing on the dominant band of the spectrum, and on the RFs it represents.

### 4.2.5 Symmetries of the system: translations and rotations

Since LGN activity reflects retinal input, we expect no physical changes from simultaneously rotating both the cortical layer and the ON/OFF LGN sheets by the same angle. This would be equivalent, in practice, to rotating the brain while equally rotating also the outside world. The same is of course true about translations.

The time-evolution operator $L_p(\omega)$ has thus two symmetries: (1) Simultaneous shifts of the cortex and of LGN do not affect the matrix elements; (2) If the wavevector $\omega$ is rotated, and the relative coordinates $(r, s)$ are rotated by the same angle, the matrix elements are also unchanged. If we consider the null wave-number $\omega = 0$, this operation reduces to rotating the $r$-coordinates only, and this will not affect the matrix elements.

It follows from the discussion of Sec. (4.2.4) that the problem can largely be treated as a study of zero-temperature phase transitions. Translational symmetry is broken if the principal eigenstate of the system corresponds to a nonzero wavenumber. Rotational symmetry is broken if the principal eigenstate is not invariant under simultaneous rotation of the wavevector and of the radial coordinate. For instance, this happens if the wavenumber under consideration is $\omega = 0$ and the eigenfunction is a p-wave or has any other angle-

It follows that there are several symmetry classes for the solution, and we may give shorthand names to the three phases that will emerge from the forthcoming analysis. We call "N-phase" (non-selective) the phase in which no invariance is broken. We use the name "R-phase" for the phase where rotational symmetry is broken but translational symmetry is not; and T-phase for the phase where translational symmetry is broken, and so is rotational symmetry. A summary of these labels is given in table (4.1).

<table>
<thead>
<tr>
<th>Phase Label</th>
<th>Receptive field uniform across cortex</th>
<th>Orientation Selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>R</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>T</td>
<td>×</td>
<td>✓</td>
</tr>
</tbody>
</table>

### 4.2.6 Symmetries of the system: parity and CP symmetry

An important property of the eigenfunctions of $\hat{L}^p$ concerns their behavior under the action of the operators $\hat{P}_x$ and $\hat{P}_y$, defined by

$$
\hat{P}_x \psi(r_x, r_y) = \psi(-r_x, r_y) \quad \quad \hat{P}_y \psi(r_x, r_y) = \psi(r_x, -r_y).
$$

(4.40)

As we take the wavevector $\omega$ to be aligned with the x-axis, the commutation rule $[\hat{L}^p, \hat{P}_y] = 0$ is immediately verified from Eq. (4.24), hence $\hat{L}^p$ and $\hat{P}_x$ can be diagonalized simultaneously, and the eigenfunctions of $\hat{L}^p$ may be chosen as either symmetric or antisymmetric under inversion of the $r_y$ coordinate.

On the other hand, the operators $\hat{L}^p$ and $\hat{P}_x$ do not commute, as can be seen from Eq.
However $\hat{P}_x$ does commute with the product $\hat{C}\hat{P}_x$, where $\hat{C}$ is the antilinear operator such that $\hat{C}\psi(r) = \psi^*(r)$.

If we write our complex receptive field $\psi(r) = u(r) + iv(r)$ as the real-valued vector function $\psi(r) = \begin{pmatrix} u(r) \\ w(r) \end{pmatrix}$, this operator takes the form

$$\hat{C}\hat{P} = \begin{pmatrix} \hat{P} & 0 \\ 0 & -\hat{P} \end{pmatrix}.$$ (4.41)

which is visibly an Hermitian operator, hence its eigenvalues must be real. Since $\hat{C}\hat{P}^2 = 1$, it follows that the eigenvalues are $\pm 1$.

In this representation, a generic integral operator $\hat{O}$ takes the matrix form $\hat{O} = \begin{pmatrix} \hat{A} & -\hat{B} \\ \hat{B} & \hat{A} \end{pmatrix}$, where the kernels $A(r, s)$ and $B(r, s)$ of $\hat{A}$ and $\hat{B}$ are the real and imaginary part of the kernel of $O(r, s)$. Such an operator clearly commutes with multiplications of the wave functions by a phase factor $e^{i\theta}$. Indeed, that is now represented by the rotation of the complex plane

$$\hat{R}_\theta = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix},$$ (4.42)

and we have $[\hat{O}, \hat{R}] = 0$.

If $[\hat{O}, \hat{C}\hat{P}] = 0$, it follows that the eigenfunctions $[u(r), w(r)]$ of $\hat{O}$ can be chosen to be eigenvectors of the operator $\hat{C}\hat{P}$, whose eigenvalues we discussed after Eq. (4.41). That
is, they can be chosen to obey the constraint

\[
\begin{pmatrix}
  u(-r) \\
  -w(-r)
\end{pmatrix} = \hat{C} \hat{P} \begin{pmatrix}
  u(r) \\
  w(r)
\end{pmatrix} = \lambda_{\mathcal{CP}} \begin{pmatrix}
  u(r) \\
  w(r)
\end{pmatrix} = \pm \begin{pmatrix}
  u(r) \\
  w(r)
\end{pmatrix},
\]

from which we can see that either \( u(r) \) is symmetric and \( w(r) \) antisymmetric, or vice versa. In both cases, the symmetric and antisymmetric part of the function are separated by a phase shift of magnitude \( \pi \).

Applying this to the constrained Linsker operator, we conclude that the eigenfunctions of \( \hat{L}^p \) will consist of a component \( \psi_S \) that is symmetric in \( \hat{P}_x \) and a component \( \psi_S \) that is antisymmetric, the two components being separated by a phase shift \( \pi \).

We can thus write

\[
\psi(r) \propto \psi_S(r) \pm i\psi_A(r),
\]

where \( \psi_S \) and \( \psi_A \) are real, and \( \psi_S \) (\( \psi_A \)) an even (odd) function in \( r_x \).

Notice that the operator \( \hat{C} \hat{P} \) does not commute with the operator \( \hat{R}_\theta \) defined by Eq. (4.42). This means that by diagonalizing \( \hat{C} \hat{P} \) we have effectively fixed the gauge of the wave functions.

Thus, we have shown that it is always possible to write the eigenfunctions of \( \hat{L}^p \) in the form \( \psi_S(r) + i\psi_A(r) \), where \( \psi_S \) and \( \psi_A \) are real. If we back-transform to real space in the cortical coordinates \( x \), this means that symmetric and antisymmetric RFs will alternate along the direction of cortical modulation, as shown in Fig. (4.5).
4.2.7 Diagonalization of the unconstrained dynamics

In Fourier space, the unconstrained two-layer model is given by Eq. (4.36), whose diagonalization can be obtained, through a gauge transformation, from an eigenbasis of the one-cell operator given by Eq. (3.11) of the previous chapter (Sec. 3.1.3).

Indeed, if we define the gauge transform
\[ \Psi(r_x, r_y) = \exp\left(i \frac{\eta^2}{\mu^2} \Omega r_x\right) \chi(r_x, r_y), \]
then it is clear that \( \Psi(r) \) is an eigenfunction of \( \hat{L} \) if and only if \( \chi(r) \) is an eigenfunction of
\[ L^d(r, s) = \exp\left[-\frac{\omega^2}{2\Omega^2} - \frac{r^2 + s^2}{4\mu^2} - \frac{(r - s)^2}{2\mu^2}\right], \tag{4.45} \]
which is equivalent to the unconstrained one-cell operator of Eq. (3.11), and can be decomposed in terms of the same eigenfunctions, with the three following differences:

1. all matrix elements are being multiplied by an invariant prefactor \( \exp\left(-\frac{\omega^2}{2\Omega^2}\right) \);
2. the length \( \eta \) will have to be replaced by \( \mu \) in all the formulas;
(3) the parameters $\gamma$ and $\beta$ must now be understood with the definitions

$$\gamma = \sqrt{2\rho} \left( 1 + \frac{4\rho^2}{\mu^2} \right)^{-1/4}, \quad \beta = 1 + \frac{\mu^2}{2\rho^2} + \frac{\mu}{\rho} \sqrt{1 + \frac{\mu^2}{4\rho^2}},$$  \hspace{1cm} (4.46)$$

which generalize the results of Sec. 3.2, and in particular Eq. (3.19).

Hence, the eigenfunctions of $\hat{L}$ have the form

$$\Psi_{n_x, n_y}(r; \omega) = e^{in_x\omega r_x/\mu^2} \chi_{n_x, n_y}(r),$$

where the functions $\chi_{n_x, n_y}(r)$ are defined in Eq. (3.35), only with $\gamma$ as in Eq. (4.46).

The corresponding eigenvalues are

$$\Lambda_{n_x, n_y} = 2\pi \mu^2 e^{-\frac{\omega^2}{4\pi^2}} \beta^{-n_x - n_y - 1},$$

with $\beta$ defined as per Eq. (4.46). Since all the dependence on the wavenumber is in the exponential scaling factor, it is apparent that the optimal wavenumber is always $\omega = 0$.

Since $L$ is diagonalizable, and the other operators summing up to $L^p$ have separable matrix elements, each of the four operators summing up to $L^p$ in Eq. (4.35) is diagonalizable exactly. Unfortunately the sum of the four is not. But while no closed-form solution is available in the whole range of parameters, it is possible to study the operator separately in the various regions of the parameter space.

We will characterize the space of parameters in terms of the two dimensionless variables $\eta/\rho$ and $\zeta/\rho$. 

75
4.3 Perturbation theory for short correlations

4.3.1 Solution with uncorrelated input

In the limit $\zeta \ll \min(\rho, \eta)$, from Eq. (4.29) it can be shown that the matrix element of $L^p$ in real space has the form

$$L^p(x, r; y, s) = \delta(x - y + r - s) \tilde{L}(r, s), \quad (4.47)$$

where the operator $\tilde{L}$ is nothing but the one-cell Linsker operator $\hat{L}^c$ defined by Eq. (3.49) of Sec. 3.3, with the role of the correlation function $C(r, s)$ now played by the cortical interaction function $I(r, s)$.

If we look for eigenfunctions of $L^p$ in the form $\psi(r) = f(r)e^{-i\omega(x + r)}$, the characteristic equation reduces to $\lambda f(r) = \int ds \tilde{L}(r, s)f(s)$, which means that $f(r)$ is the corresponding eigenfunction of $\tilde{L}$ and the eigenvalue is independent on the wavenumber. Hence, the formulas we derived in Chapter 3 are directly applicable to this limit.

In particular, we already know that the principal eigenfunction of the operator is orientation selective if $\eta > \Theta_c\rho$, and is non-selective if $\eta < \Theta_c\rho$, with the critical ratio $\Theta_c$ is bounded from below by the value $\theta_c$ of Eq. (3.72).

For $\zeta = 0$ and $\eta \leq \theta_c\rho$ the principal eigenfunction in the variational approximation of Sec. 3.4.2 is

$$\psi_0(x, r) \propto (R^2 - r^2) \exp \left( -\frac{r^2}{2\gamma^2} + i\omega(x + r_x) \right), \quad (4.48)$$

with the nodal radius $R \propto \sqrt{\eta\rho}$ calculated in the previous chapter, and eigenvalue $\lambda \sim$
For $\zeta = 0$ and $\eta > \Theta_c \rho$, the exact principal eigenfunction as per Eq. (3.27) is given by

$$\psi(r, x) = (k_+ e^{i\phi} + k_- e^{-i\phi}) \exp \left[ -i \omega (x + r) - \frac{r^2}{2\sigma^2} \right],$$

(4.49)

for any vector $\omega$ and arbitrary coefficients $k_+$ and $k_-$. Here we are calling $\sigma$ the RF width $\gamma$ of formula (4.46), computed for $\zeta = 0$.

We will refer in particular to the combinations

$$\begin{pmatrix} \psi^x(x, r) \\ \psi^y(x, r) \end{pmatrix} = \begin{pmatrix} r_x \\ r_y \end{pmatrix} \exp \left( - \frac{r_x^2 + r_y^2}{2\gamma^2} + i \omega (x + r) \right),$$

(4.50)

which are the instances of the ”2p” waves $\Psi_{0,1}$ and $\Psi_{1,0}$ corresponding to the R phase. Since the cortical wavevector is aligned along the $x$-axis, $\psi^x$ describes RFs aligned parallel to the cortical wavevector, and $\psi^y$ describes RFs aligned orthogonally to it. Accordingly, we will call the $\psi^x$ the longitudinal eigenfunction and $\psi^y$ the transverse eigenfunction.

These two eigenfunctions share, as per Eq. (3.23), share the exact eigenvalue

$$\Lambda^{x,y} = 2\pi \eta^2 \left( 1 + \frac{\eta^2}{2\rho^2} + \frac{\eta^2}{\rho} \sqrt{1 + \frac{\eta^2}{4\rho^2}} \right)^{-2} ;$$

(4.51)

the double degeneracy we thus find for $\eta > \Theta_c \rho$ adds up to the overall degeneracy we have

in the wavenumber $\omega$ for any value of $\eta$.

The above shows that the point $P_0$ defined by $(\zeta_0, \eta_0) = (0, \Theta_c \rho)$ belongs to a phase boundary. Moreover, this phase boundary cannot stop there because it is a boundary be-
tween two phases that have different symmetries— one that displays orientation selectivity and one that does not. How is this phase boundary continued for $\zeta > 0$? Will it curve up or down?

Since we possess the exact solution for $\zeta = 0$, we will use it as a starting point for a perturbation theory in the small parameter $\zeta/\eta$. The perturbation theory (schematized in Fig. 4.6) will prove the following three facts:

1. the $\omega$ degeneracy is removed for any $\eta$ is removed by an infinitesimal $\zeta > 0$, and this in such a way that $\omega = 0$ is always the principal eigenstate;

2. the phase boundary starting at the point $(0, \Theta_c\rho)$ has a flat slope at that point in the $\zeta/\eta$ plane;

3. the $xy$ degeneracy of the p-wave eigenfunctions survives at finite $\zeta$.

Figure 4.6: Schematic depiction of the phase diagram of the model, with the LGN correlation length on the x-axis and the length scale of recurrent cortical interaction on the y-axis. The shaded area display results from the perturbation theory for small $\zeta/\eta$. Blank regions of this diagram will be explored in the following sections. The question mark refers regions of the diagram with longer-range correlations, which will be studied next.
4.3.2 Optimal wavenumber for oriented eigenfunctions

It is possible to study the two 2p eigenfunctions separately because the full operator \( L^p \) does not couple these two eigenfunctions for any value of the parameters.

Indeed, we have it by symmetry that \( \langle \psi^x | \hat{L}^p | \psi^y \rangle = 0 \) for any \( \zeta \) and \( \omega \). This means that we can study the effect of a small but finite \( \zeta \) separately on the two eigenfunctions. We begin with the \( y \)-oriented wave.

Transverse orientations

While \( \psi^y \) is an eigenfunction of \( \hat{L} \) only for \( \zeta = 0 \), it can be checked that its generalization

\[
\chi_{0,1}(r) = \frac{r_y}{\sqrt{\pi \sigma^2}} \exp \left( -\frac{r^2}{2\sigma^2} \right) e^{-\frac{r^2}{2\gamma^2}} e^{i \frac{r^2}{2\nu^2} \omega \tau}, \tag{4.52}
\]

is an exact eigenfunction of the full operator \( \hat{L}^p \) over the whole phase diagram. Indeed, it is an exact eigenfunction of \( L \) and, being orthogonal to the constraint ket \( |a_\omega \rangle \), it belongs to the kernel of the constraint operators \( \hat{S} \) and \( \hat{T} \).

The eigenvalue is given by

\[
\Lambda^y = \frac{2\pi \mu^2}{\beta^2} \exp \left( -\frac{\omega^2}{2\Omega^2} \right), \tag{4.53}
\]

with \( \beta \) defined as per Eq. (4.46).

For \( \zeta = 0 \), as we knew, this eigenvalue is independent on the wavenumber. However, for any \( \zeta > 0 \), Eq. (4.53) describes an eigenvalue that decreases monotonically with the wavenumber, hence the degeneracy is removed. We can conclude that, in the limit of small
ζ, the principal y-oriented eigenfunction is uniform over the receptive field.

**Longitudinal orientations**

We now turn to considering the x-oriented function ψ^x. It can be checked that ψ^x is orthogonal to the constraint state |a_ω⟩, which entails

\[ \langle ψ^x | ˆS | ψ^x \rangle = \langle ψ^x | ˆT | ψ^x \rangle = 0, \]  \hspace{1cm} (4.54)

This holds true for any value of ζ, and means that we may neglect the constraint operators completely.

We will attempt a perturbation theory in the parameter \( \epsilon = \zeta^2/\eta^2 \). Hence we write the operator \( ˆL^p \) as \( ˆL^p = ˆL^p(\epsilon = 0) + ˆΔ + O(\epsilon^2) \), where \( ˆΔ \) includes the first order in \( \epsilon = \zeta^2/\eta^2 \), and we will treat \( ˆΔ \) as a perturbation. In the shift operator \( ˆΔ = ˆΔ_L + ˆΔ_S + ˆΔ_T + ˆΔ_T^\dagger \), because of what we said, we only have to compute the \( ˆL \)-term.

We begin by expanding to the first order in \( \epsilon \) Eq. (4.36), which yields

\[ \Delta L(r,s;ω) = \epsilon \left[ -\frac{\omega^2\eta^2}{2} + i\omega(r_x - s_x) + \frac{(r - s)^2}{2\eta^2} \right] e^{-iω(r_x - s_x) - \frac{r^2 + s^2}{4\eta^2} - \frac{(r - s)^2}{2\eta^2}}. \]  \hspace{1cm} (4.55)

We then keep only the terms that have a nonvanishing expectation value in ψ^x; in particular, we neglect terms that change sign if we swap the two variables \( r_x \) and \( s_x \), because the integral would be zero. In addition, we may ignore terms whose expectation value in ψ^x will bear no dependence on the wavenumber. This leaves a single first-order term in Eq.
that obeys all these requirements, namely:

\[
\Delta L(r, s; \omega) \sim -\frac{\omega^2 \zeta^2}{2} e^{-i\omega(r_x-s_x)} e^{i^{2}s^2} (r_x-s_x)^2 r_x^2 s^2 . \tag{4.56}
\]

The corresponding expectation value is

\[
\Delta = \langle \psi^x | \Delta L | \psi^x \rangle = -\frac{\omega^2 \zeta^2}{2} \Lambda_2 p (\zeta = 0) , \tag{4.57}
\]

a negative shift in the eigenvalue that is minimized by setting \( \omega = 0 \).

We have thus proven that, for sufficiently small \( \zeta \) and given \( \eta \), the principal eigenstate is always cortically uniform \( (\omega = 0) \), as long as the principal eigenstate for the given \( \eta \) and \( \zeta = 0 \) is an R phase. This entails that in the limit \( \zeta \to 0 \), the principal eigenstate of the system has a zero wavenumber for any \( \eta > \Theta_c \rho \). Hence, the slope of the phase boundary at \( (0, \Theta_c \rho) \) cannot be positive.

While the degeneracy in wavenumber has been removed by first-order perturbation theory, the degeneracy between the \( x \) and \( y \) orientations has not been removed, as seen by comparing Eq. (4.2) and Eq. (delta delta)\(^3\).

### 4.3.3 Optimal wavenumber for non-oriented eigenfunctions

The s-wave (i.e. non-oriented) eigenfunctions of \( \tilde{L}^p \) are also degenerate in the wavenumber for \( \zeta = 0 \), as follows from the divergence of the cutoff wavenumber \( \Omega \).

To see which spatial frequencies effectively dominate, we must build a perturbation

\(^3\)Moreover, it can be seen that this degeneracy would not be removed by carrying this perturbation theory to any further order.
theory in $\epsilon = \zeta^2/\eta^2$ by starting from an arbitrary s-wave eigenfunction $\psi_s$ of the zero-wavenumber operator. Defining the ket $|s\rangle$ by $\langle r|s\rangle = \psi_s(r)e^{-i\omega r_x}$, and using Eq. (4.39), we can write the level shift as

$$\Delta L^p(\omega) = \langle s| (1 - |a_\omega\rangle\langle a_\omega|) \Delta L_\omega (1 - |a_\omega\rangle\langle a_\omega|) |s\rangle$$  \hfill (4.58)

where the matrix elements of $\Delta L_\omega$ have the form given in Eq. (4.55).

We notice that gauge factors of the type $e^{-i\omega r_x}$ will cancel in the integrands of all scalar products that appear in eq. (4.58). As a consequence, the third term in the square brackets of Eq. (4.55) may ignored, as it adds no dependence on the wavenumber.

The first term in the square brackets of Eq. (4.55), on the other hand, yields the level shift

$$\Delta L_e^{(1)} = -\frac{\omega^2 \zeta^2}{2} \langle s| L^p(\zeta = 0)|s\rangle = -\frac{\omega^2 \zeta^2}{2} \int \psi_s(r)L_c(r,s)\psi_s(s)dr$$  \hfill (4.59)

where the expectation value, as we have seen, doesn’t depend on wavenumber. From the positive semidefiniteness of $L^c$, proven in section 3.3.1 we deduce that the shift (4.59) decreases monotonically with the wavenumber.

The only remaining term is the second one in the square brackets of Eq. (4.55), namely

$$\Delta L_e^{(2)}(r,s;\omega) = i\epsilon \omega (r_x - s_x)e^{-i\omega(r_x - s_x)} - \frac{r_s^2}{4\eta^2} - \frac{(r_s - s_s)^2}{2\eta^2},$$  \hfill (4.60)
which also yields a level shift of the form

$$\Delta L_e^{(2)} = \langle s | \Delta L^{(2)} | s \rangle + |\langle s | a_\omega \rangle|^2 \langle a_\omega | \Delta L_\omega | a_\omega \rangle - 2 \text{Re} \left[ \langle s | a_\omega \rangle \langle a_\omega | \Delta L^{(2)} | s \rangle \right] ;$$  \hspace{0.5cm} (4.61)

since the gauge factors of Eq. (4.60) cancel everywhere in Eq. (4.61), the resulting matrix element is effectively antisymmetric in the swapping of the \( r \) and \( s \) coordinates. This makes the first two terms in Eq. (4.61) vanish by symmetry; since the final square bracket is purely imaginary, the third term is also zero.

Hence the full eigenvalue shift is given by Eq. (4.59) and decreases monotonically with the wavenumber. We conclude that the wavenumber degeneracy at \( \zeta = 0 \) is removed by an infinitesimal range of presynaptic correlations, and that the DC cortical modes is favored for any \( \eta \).

Building upon this, it is not difficult to show that to the lowest order in \( \zeta/\eta \) the phase boundary is a flat line. Indeed, to the lowest order the dominant wavenumber remains null, hence the phase boundary overlaps with the phase boundary we may compute (with arbitrary \( \eta \) and \( \zeta \)) for the zero-wavenumber operator \( L^\rho(\omega = 0) \); this phase boundary depends only on \( \eta^2 + \zeta^2 \) (as we will see now more in detail) and therefore has a flat slope and negative curvature at \( \zeta = 0 \).

Hence, the phase boundary starting at the point \( P_0 = (0, \Theta_c \rho) \) has a flat slope in that point (Fig. 4.6).
4.4 Forbidden regions for the N- and R-phases

If we set $\omega = 0$ in the kernel of the operator $L^p$ (as per Eq. 4.35), it can be seen that the resulting operator, while not exactly solvable, becomes equivalent to the one-cell constrained operator defined by Eq. (3.49), only with the substitution $\eta \to \mu$.

The structure of its principal eigenspace can therefore be deduced from our study of Chapter 3. Namely, there must be a critical value $\mu_c$ such that, for $\mu > \mu_c$, the principal eigenspace of the operator has degeneracy $D = 2$ and p-wave basis functions (R phase); for $\mu < \mu_c$, it consists instead of a non-degenerate s-wave eigenfunction $\Psi_s(r)$.

If the T-phase is subdominant everywhere, hence absent from the principal eigenspace, the space of parameters will contain a circular arc of radius $\mu_c$ as its only phase boundary. The N-phase will be found in a quarter-circle of radius $\mu_c$ centered on the origin, and the R-phase everywhere outside that quarter-circle.

This critical radius has the value $\mu_c = \rho \Theta_c$, where $\Theta_c$ is the phase boundary of the one-cell model. Results from our variational analysis of Chapter 3 translate here into a lower bound on the critical radius, $\mu_c \geq \theta_c \rho$, with $\theta_c = \frac{2(75 - 8\sqrt{10})}{997} \approx 0.1$ being the lower bound to $\Theta_c$ given by Eq. (3.72).

Those results also provide a qualitatively satisfactory estimate for $\Psi_s$, the exact N-phase eigenfunction of the model, which from Eqs. (3.68-3.69) can be seen to be

$$\Psi_s(r_x, r_y) \propto \left( R^2 - r_x^2 - r_y^2 \right) \left( -\frac{r_x^2}{2\gamma^2} \right),$$

(4.62)
with \( R \sim \left( \frac{4 + \sqrt{10}}{3} \rho \mu \right)^{1/2} \) and the eigenvalue

\[
\frac{\mathcal{E}}{2\pi \mu^2} \sim 1 - \left( 5 - \sqrt{10} \right) \frac{\mu}{\rho} + \left( \frac{33}{2} - \frac{51}{\sqrt{10}} \right) \left( \frac{\mu}{\rho} \right)^2.
\] (4.63)

The bipartite structure of the phase diagram we have just described (Fig. 4.7) would only hold true if the optimal value \( \omega_M \) of the cortical wavenumber was found to be null everywhere, which may not be the case.

On the other hand, we have shown rigorously (in Sec. 4.3) that the optimal wavenumber is indeed zero for \( \zeta = 0 \) and \( \eta > \Theta_c \rho \). Moreover, the point \( (\zeta, \eta) = (0, \Theta_c \rho) \) has been shown to lie along a phase boundary, and we have also proven that the boundary cannot slope upwards at that point.

The arc-shaped NR boundary may of course be changed beyond recognition by the presence of T-phase regions. These may enter portions of the phase diagram otherwise occupied by the R and N phase, substantially upsetting its topology. However, it follows from the above considerations that no point in parameter space such that \( \mu > \mu_c \) can possibly belong to an N-phase subregion, and no point such that \( \mu < \mu_c \) can belong to an R-phase subregion\(^4\).

The resulting scenario is therefore the following. The optimal wavenumber will vary continuously across parameter space (in ways we are going to calculate) and the corresponding eigenfunction will change in accordance with it. This eigenfunction, in principle, may resemble neither the s-waves nor the p-waves – or it may be a linear combinations of

\(^4\)This may be regarded as following from a variational argument. If the optimal wavenumber is null, the expectation value of the operator must be optimized within the \( \omega_M = 0 \) sector, proving the claim.
these and many more. Yet, at all points of the phase diagram where the cortical wavenumber becomes vanishingly small, the principal eigenfunction will behave in a way we can predict.

In regions such that $\mu > \mu_c$, the eigenfunctions will progressively tend to resemble a p-wave as $\omega_M$ is sufficiently decreased; in regions such that $\mu < \mu_c$, the eigenfunctions will asymptotically lose all but the leading s-wave components as the optimal wavenumber approaches zero$^5$.

Figure 4.7: Schematic description of the variational argument for the NR boundary. The T phase may appear on both sides of the dashed circular arc; the N phase cannot appear above the arc; the R phase is forbidden below the arc.

4.5 The long-range limit

4.5.1 The long-range limit: derivation

We call long-range limit ($\mu \gg \rho$) the case where either cortical interactions are long-range ($\eta \gg \rho$) or LGN interactions are ($\eta \gg \rho$) or both.

$^5$For $\mu \gg \mu_c$, the results of the calculations of Sec. 4.5 may be checked against this prediction. The same is true, with both regimes, for the variational results of Sec. 4.6.
We begin by Taylor expanding Eq. (4.36) into

\[ \hat{L}^p \sim \hat{L}_0 + \hat{R} + \hat{S} + \hat{T} + \hat{T}^\dagger, \]

with

\[
L_0(r, s; \omega) = \exp \left[ -\frac{\omega^2}{2\Omega^2} - i\frac{\eta^2}{\mu^2} \omega (r - s) - \frac{r^2 + s^2}{4\rho^2} \right], \tag{4.65}
\]

\[
R(r, s; \omega) = \left( \frac{r_i s_i}{\mu^2} + \frac{r_i r_j s_i s_j}{2 \mu^4} + \ldots \right) \exp \left[ -\frac{\omega^2}{2\Omega^2} - i\frac{\eta^2}{\mu^2} \omega (r - s) - \frac{r^2 + s^2}{4\rho^2} \right]. \tag{4.66}
\]

where summation over repeated indices is implied.

In every integration where this kernel would play a role, variables representing relative LGN-V1 coordinates are confined by the arbor densities to a radius of order \( \rho \). Using this, we can rely on the smallness of \( \rho/\mu \) to discard all but the lowest term in \( R \).

Simplifying analogously \( \hat{S} \) and \( \hat{T} \), we can turn Eq. (4.64) into

\[
L^p(r, s; \omega) = e^{-\frac{\omega^2}{2\Omega^2}} \left[ a_c^* (r) a_c (s) + A^* (r) A (s) + q^2 a_1^* (r) a_1 (s),\right.

\[
\left. -q \left( a_1^* (r) a_d (s) + a_d^* (r) a_1 (s) \right) \right], \tag{4.67}
\]

where \( c = (\eta/\mu)^2 \), \( d = \frac{\eta^2 + \rho^2}{\mu^2} \), \( q = \exp \left( -\frac{\zeta^4 \rho^2 \omega^2}{2 \mu^4} \right) \), and we have defined the functions

\[
a_\beta (r) = \exp \left( i \beta \omega r_x - \frac{r^2}{4\rho^2} \right) \quad A^* (r) = \frac{r}{\mu} \exp \left[ -\frac{r^2}{4\rho^2} + i \frac{\eta^2}{\mu^2} \omega r_x \right]. \tag{4.68}
\]

We will now treat the \( \eta \gg \rho \) and \( \zeta \gg \rho \) cases separately, even if these assumptions lead to similar mathematical results.
The regime $\eta \gg \rho$

If $\eta \gg \rho$, we have $d \sim c$, so the operator Eq. (4.64) becomes

$$L^p(r, s; \omega) = e^{-\frac{\eta^2}{2\eta^2}} \left[ a_c^*(r)a_c(s) + A^*(r)A(s) + q^2a_1^*(r)a_1(s) - q \left( a_c^*(r)a_c(s) + a_c^*(s)a_1(r) \right) \right]$$  \hspace{1cm} (4.69)

Let us now define the two unknowns $I_\beta = \int a_\beta(r)\psi(r)dr$ for $\beta = 1, c$, and the third unknown $K = \int A_x(r)\psi(r)dr$, and let us use the self-consistent assumption that $\int A_y(r)\psi(r)dr = 0$ (which is checked below in section 4.5.2). The eigenvalue equation for the operator of Eq. (4.69) becomes

$$\lambda e^{\frac{\eta^2}{2\eta^2}}\psi(r) = I_1 \left[ q^2a_1^*(r) - qa_c^*(r) \right] + I_c \left[ -qa_1^*(r) + a_c^*(r) \right] + KA_x^*(r)$$  \hspace{1cm} (4.70)

and computing the three unknown integrals from Eq. (4.70) itself, one obtains:

$$\frac{\lambda}{2\pi\rho^2}e^{\frac{\omega^2}{2\omega^2}}I_1 = i\frac{\rho^2\zeta^2}{\mu^3}\omega K$$  \hspace{1cm} (4.71)

$$\frac{\lambda}{2\pi\rho^2}e^{\frac{\omega^2}{2\omega^2}}I_c = (q^3 - q)I_1 + (1 - q^2)I_c$$  \hspace{1cm} (4.72)

$$\frac{\lambda}{2\pi\rho^2}e^{\frac{\omega^2}{2\omega^2}}K = -q^3\frac{i\rho^2\zeta^2}{\mu^3}\omega I_1 + q^2\frac{i\rho^2\zeta^2}{\mu^3}\omega I_c + \frac{\rho^2}{\mu^2}K$$  \hspace{1cm} (4.73)

from which it follows that we can replace the infinite-dimensional operator of Eq. (4.35)
with the 3x3 matrix \( \hat{L} = 2\pi \rho^2 e^{-\frac{\rho^2}{2\pi}} \hat{M} \), where

\[
\hat{M} = \begin{pmatrix}
0 & 0 & Jq \\
q^3 - q & 1 - q^2 & 0 \\
-q^3 J & q^2 J & (\rho/\mu)^2
\end{pmatrix}
\] (4.74)

for \( J = i\rho^2 \zeta^2 \omega / \mu^3 \).

This matrix has only two nonzero eigenvalues, both positive as we expect from the discussion in Sec. 4.2.3. The larger one is

\[
\lambda = \pi \rho^2 e^{-\frac{\rho^2}{2\pi}} \left( 1 + \frac{\rho^2}{\mu^2} - e^{-\omega^2 \zeta^4 \rho^2 / \mu^4} + \sqrt{(1 - \frac{\rho^2}{\mu^2} - e^{-\omega^2 \zeta^4 \rho^2 / \mu^4})^2 + \frac{4\rho^4 \zeta^4 \omega^2}{\mu^6} e^{-2\omega^2 \zeta^4 \rho^2 / \mu^4}} \right),
\] (4.75)

while the corresponding eigenfunction is obtained from Eq. (4.70) through the principal eigenvector of the matrix \( \hat{M} \).

This is found from Eq. (4.74) to be, before normalization,

\[
I_1 = 2J^2 q^3
\] (4.76)

\[
I_c = (1 - q^2) \left[ 1 - s - q^2 + \sqrt{(1 - s - q^2)^2 - 4J^2 q^4} \right]
\] (4.77)

\[
K = J q^2 \left[ 1 + s - q^2 + \sqrt{(1 - s - q^2)^2 - 4J^2 q^4} \right],
\] (4.78)
with $s = (\rho/\mu)^2$. In the limit $s \to 0$, keeping $x$ fixed, we find

$$
\left( I_1, I_c, K \right) \to \left( 0, 1, 0 \right),
$$

(4.79)

hence the principal eigenfunction for the kernel Eq. (4.69) is found to be

$$
\psi(r) = \psi_1(r; \omega_M \tilde{\omega}) \propto -qa_1^*(r) + a_c^*(r)
$$

$$
= e^{-\frac{s^2}{4\mu^2}} \left[ e^{-i\frac{\omega_0^2}{\mu^2}\omega_M \tilde{\omega} r - i\omega_M \tilde{\omega} r - \frac{1}{2} \left( \frac{\omega_0^2}{\mu^2} \right)^2} \right],
$$

(4.80)

where $\tilde{\omega}$ is an arbitrary unit vector.

**The regime $\zeta \gg \rho$**

For $\zeta \gg \rho$, the eigenvalue equation of (4.67) can be written as

$$
\lambda e^{\frac{2\pi i}{\Omega^2}} \psi(r) = I_1 \left( q^2 a_1^*(r) - qa_d^*(r) \right) + I_c a_c^*(r) - I_d qa_1^*(r) + K A_x^*(r),
$$

(4.81)

where we defined the three unknown quantities $I_\beta = \int a_\beta(r) \psi(r) \, dr$ (for $\beta = 1, c, d$) and the fourth unknown $K = \int A_x(r) \psi(r) \, dr$. Again, we are using the self-consistent assumption that $\int A_y(r) \psi(r) \, dr = 0$, which will be duly checked in Sec. 4.5.2.
Define $J_x = \exp \left( -\frac{\rho^2 x^2 \omega^2}{2} \right)$, so that

\[
q = J_{1-c};
\]

\[
\int a_{\alpha}(r)a_{\alpha}^*(r)dr = 2\pi \rho^2 J_{\alpha-\beta};
\]

\[
\int a_{\beta}(r)A_x^*(r)dr = (2\pi \rho^2) \frac{i \rho^2 \omega}{\mu} (\beta - c) J_{\beta-c};
\]

while $\int dr A_x(r)A_x^*(r) = 2\pi \rho^4 / \mu^2$.

From Eq. (4.81), we obtain

\[
\frac{\lambda}{2\pi \rho^3} e^{\frac{\omega^2}{2\pi^2}} I_1 = (q^2 - q J_{1-d}) I_1 + q I_c - q I_d + i \frac{\rho^2 \omega}{\mu} (1 - c) q K;
\]

\[
\frac{\lambda}{2\pi \rho^3} e^{\frac{\omega^2}{2\pi^2}} I_c = (q^3 - q J_{d-c}) I_1 + I_c - q^2 I_d;
\]

\[
\frac{\lambda}{2\pi \rho^3} e^{\frac{\omega^2}{2\pi^2}} I_d = (q^2 J_{1-d} - q) I_1 + J_{d-c} - q J_{1-d} I_d + i \frac{\omega \rho^4}{\mu^3} q K;
\]

\[
\frac{\lambda}{2\pi \rho^3} e^{\frac{\omega^2}{2\pi^2}} K = \left( -i \frac{\omega \xi_q^2 \rho^2 q^3}{\mu^3} + i q \frac{\omega \rho^4}{\mu^3} J_{d-c} \right) I_1 + i \frac{\omega \xi_q^2 \rho^2 q^3}{\mu^3} q^2 I_d + \frac{\rho^2}{\mu^2} K;
\]

from which it follows that, in this limit, we can replace our infinite-dimensional operator with the $4 \times 4$ matrix $\hat{L} = 2\pi \rho^2 e^{-\frac{\omega^2}{2\pi^2}} \hat{M}$, where

\[
\hat{M} = \begin{pmatrix}
q^2 - q J_{1-d} & q & -q & i \frac{\omega \xi_q^2 \rho^2 q^3}{\mu^3} \\
q^3 - q J_{d-c} & 1 & -q^2 & 0 \\
q^2 J_{1-d} - q & J_{d-c} & -q J_{1-d} & i \frac{\omega \rho^4}{\mu^3} q \\
-i \frac{\omega \xi_q^2 \rho^2 q^3}{\mu^3} + i q \frac{\omega \rho^4}{\mu^3} J_{d-c} & 0 & i \frac{\omega \xi_q^2 \rho^2 q^3}{\mu^3} q^2 & \frac{\rho^2}{\mu^2}
\end{pmatrix}.
\]

Now, we have $d - c = \frac{\rho^2}{\mu^3}$ and $1 - d = \frac{\xi_q^2 \rho^2}{\mu^3}$; since we are considering the regime
where $\zeta \gg \rho$, we can write $1 - d \sim \frac{\zeta^2}{\mu^2}$, so that $J_{1-d} \sim q$. Notice that we are making no assumption on the magnitude of $\eta$. The matrix thus simplifies to

$$
\hat{M} = \begin{pmatrix}
0 & q & -q & \frac{i\omega\rho^2\zeta^2}{\mu^2} q \\
q^3 - qJ_{d-c} & 1 & -q^2 & 0 \\
q^3 - q & J_{d-c} & -q^2 & \frac{i\omega\rho^2}{\mu^3} q \\
-\frac{i\omega^2\rho^2}{\mu^2} q^3 + \frac{i\omega^2\rho^4}{\mu^3} qJ_{d-c} & 0 & \frac{i\omega^2\rho^2}{\mu^2} q^2 & \frac{\rho^2}{\mu^2}
\end{pmatrix} \quad (4.90)
$$

Let us adopt one more self-consistent assumption, concerning the optimal wavenumber, that will be verified immediately once the optimal wavenumber is computed from the resulting eigenvalue. Namely, we assume $\omega \ll \frac{\mu}{\rho^2}$, so that we can write $J_{d-c} \sim 1$ and neglect the terms in $\omega\rho^4/\mu^3$. The matrix Eq. (4.90) becomes

$$
\hat{M} = \begin{pmatrix}
0 & q & -q & \frac{i\omega\rho^2\zeta^2}{\mu^2} q \\
q^3 - q & 1 & -q^2 & 0 \\
q^3 - q & 1 & -q^2 & 0 \\
-\frac{i\omega^2\rho^2}{\mu^2} q^3 & 0 & \frac{i\omega^2\rho^2}{\mu^2} q^2 & \frac{\rho^2}{\mu^2}
\end{pmatrix} \quad (4.91)
$$

We have reduced an infinite dimensional problem to a four-dimensional problem, which we can solve exactly. From Eq. (4.91), we see that

$$
\det \left( \hat{M} - \lambda \right) = \lambda^2 \left[ \rho^2 - \frac{\omega^2\rho^4\zeta^4}{\mu^6} q^4 - \rho^2 q^2 + \lambda \left(q^2 - 1 - \frac{\rho^2}{\mu^2}\right) + \lambda^2 \right] \quad (4.92)
$$

and from Eq. (4.92), it is found that the two non-null eigenvalues correspond to those of Eq. (4.74). Hence formula (4.75) for the eigenvalue still holds true and, in particular, the
optimal wavenumber will be the same in the two regimes.

![Figure 4.8: Summary of results obtained from the perturbative treatment (Sec. 4.3), the variational argument (Sec. 4.4), and the asymptotic rank reduction (Sec. 4.5).](image)

4.5.2 The long-range limit: analysis of results

**Calculation of the phase boundary**

The system is in the T-phase if the wavenumber maximizing the principal eigenvalue is positive, while it is in either the R or N phase if that optimal wavenumber is null. In terms of the dimensionless variable \( x = \omega^2 \zeta^4 \rho^2 / \mu^4 \) (such that \( q = e^{-x/2} \)) we can write the principal eigenvalue (4.75) as

\[
\lambda = \pi \rho^2 f(x),
\]

\[
f(x) = e^{-\alpha x/2} \left( 1 + s - e^{-x} + \sqrt{(1 - s - e^{-x})^2 + 4sx e^{-2x}} \right),
\]

with \( \alpha = \left( \frac{\mu n}{\zeta \rho} \right)^2 \).

It can be seen that the eigenvalue goes always to zero for large spatial frequencies, because \( f(x \to \infty) \propto \exp(-\alpha x/2) \). Hence the optimal wavenumber is necessarily larger
than zero as long as the derivative of the eigenvalue at zero wavenumber is positive.

The derivative $f'(x)$ computed at $x = 0$ is $f'(0) = 2 - s\alpha$. If $s = 0$, which is the case we just considered, this derivative is always positive, so the optimal wavenumber is always nonzero. As soon as $s$ is larger than zero, however, one can find values of $\alpha$ sufficiently large as to make the derivative at the origin negative. Hence, the behavior of the phase boundary far from the origin is described by $\rho^2 \alpha/\mu^2 < 2$, i.e. $\eta/\zeta < \sqrt{2}$.

This may also be written as the T-phase dominance condition

$$\zeta > \zeta_c(\eta) = \sqrt{\frac{1}{2}} \eta.$$ (4.95)

**Form of the eigenfunction**

The assumptions $\eta \gg \rho$ and $\zeta \gg \rho$ lead thus to the same eigenfunction (4.80), which we may write as

$$\psi(r) \propto e^{-\frac{i \rho^2}{4 \mu^2} - \frac{i \rho^2}{2 \mu^2} \omega r_x} \left( 1 - e^{-i \frac{\rho^2}{\mu^2} \omega r_x - \frac{1}{2} \frac{\rho^2}{\mu^2} \omega^2} \right).$$ (4.96)

Notice that the value of $\omega$ to be plugged into Eq. (4.96) is the value that maximizes the eigenvalue (4.75). In regimes where the optimal frequency is null, we must take the $\omega \to 0$ limit in Eq. (4.96), which yields

$$\psi(r) \propto r_x \exp \left( -\frac{\rho^2}{4 \mu^2} \right),$$ (4.97)

equal to the orientation-selective eigenfunction of the one-cell problem, that is, to an R-phase. This is in agreement with the argument of Sec. 4.4 that is depicted in Fig. 4.7.

---

6In those parts of the phase diagram, therefore, the homeostatic constraint is satisfied through the indi-
Normally oriented eigenfunctions

In order to obtain Eq. (4.80), we made at the very outset the self-consistent assumption
\[ \int A_y^*(r) \psi(r) d\mathbf{R} = 0, \]
which we used to write both Eq. (4.70) and (4.81). The subspace we have focused on was indeed orthogonal to \( A_y \), and we found this subspace to be an asymptotic eigenspace of the system.

Nonetheless, the same system may also possess eigenfunctions having a nonzero overlap with \( A_y(r) \). Do these eigenfunctions correspond to a higher eigenvalue than those we calculated?

The self-consistency of our initial assumption is straightforward to check. If we repeat the above by relaxing the assumption \( \int A_y^*(r) \psi(r) d\mathbf{R} = 0 \), we have to diagonalize a \( 5 \times 5 \) matrix instead of a \( 4 \times 4 \) one. However, this matrix is diagonal in its \( A_y \)-sector. The resulting extra eigenvalue is a strictly decreasing function of wavenumber, hence it must be computed at zero wavenumber, where we find \( \Lambda_y = 2\pi \rho^4 / \mu^2 \). This is not surprising because we know from studying the one-cell case that, if we replace the one-cell length-scale parameter with \( \mu \), we have

\[ \lambda_{N=0,m=1} = \frac{2\pi \mu^2}{\beta^2} \sim \frac{2\pi \mu^2}{(\mu^2 / \rho^2)^2} = 2\pi \rho^4 / \mu^2. \]  

(4.98)

Let us compare this eigenvalue with the eigenvalue of the cortically modulated solution \( \Lambda_M \) which we found above. In the regions of the phase diagram where \( \omega_M > 0 \), we have \( \Lambda_y < \Lambda_M \), hence the normally oriented solution is suppressed at long times. In the regions where optimal wavenumber is \( \omega_M = 0 \), on the other hand, it can be seen that \( \Lambda_y = \Lambda_M \).

vidual selectivity of cells, and does not need to be satisfied through variations over cortical space; that is why translational symmetry can be restored.
We thus find that, when the RF varies across the cortex, it tends to be oriented along the
direction of cortical modulation. When it is uniform across the cortex, its direction becomes
immaterial, hence we have degenercy in the direction of the RF.

**Formulas for the optimal wavenumber**

The behavior of the optimal wavenumber over the phase diagram is described by the fol-
lowing:

\[
\omega_M \sim \begin{cases} 
\frac{1}{\rho} \sqrt{\log \left( 1 + \frac{2\rho^2}{\eta^2} \right)} & \text{if } \zeta \gg \rho \sim \eta \\
\frac{\sqrt{2}}{\rho} \sqrt{\log \left( \frac{\rho}{\eta} \right)} & \text{if } \zeta \gg \rho \gg \eta \gg \exp \left( -\frac{\zeta^2}{\rho^2} \right) \\
\frac{\mu}{\sqrt{2} \zeta} \sqrt{\frac{2\zeta^2 - \eta^2}{\eta^2 - \zeta^2}} & \text{if } \eta^2 \gg 2\zeta^2 - \eta^2 > 0, \eta, \zeta \gg \rho 
\end{cases} \tag{4.99}
\]

which are derived and discussed in Appendix B.

In the same Appendix, proof is given for the validity of the self-consistent hypothesis
on \(\omega_M\) that we introduced before Eq. (4.91).

### 4.6 Cartesian-basis expansion

To make further progress, we will need to rotate the operator from its representation in the
real-space LGN coordinates to a representation in the basis of LGN formed by Cartesian
eigenfunctions of the one-cell Linsker operator with an added plane-wave factor.

In other words, we will be using the basis functions

\[
\chi_{n_x,n_y}(\mathbf{r}) = \chi_{n_x}(r_x)\chi_{n_y}(r_y)e^{-\frac{i}{\mu \omega} \mathbf{r}_x}, \tag{4.100}
\]
where \( \chi_n(r) \) is given by Eq. (3.32).

The usefulness of this basis lies in the fact that the unconstrained one-cell operator of Eq. (4.29) is diagonal in the basis defined by Eq. (4.100):

\[
\langle n_x n_y | L | m_x m_y \rangle = \delta_{n_x m_x} \delta_{n_y m_y} \frac{2\pi \mu^2}{\beta^{n_x+n_y+1}} e^{-\frac{r^2}{4\Omega^2}}.
\] (4.101)

(No confusion should arise between the angular momentum number \( m \) of Eq. (3.13) and the Cartesian labels \( m_x, m_y \) of Eq. (4.101).

The full constrained operator at any given wavenumber \( \omega \) has the form \( \hat{L}^p = \hat{L} + \hat{S} + \hat{T} + \hat{T}^\dagger \). In the Cartesian basis defined by Eq. (4.100), the matrix elements of the constraint operators are:

\[
S(n, m; \omega) = \bar{S} \int dr ds \, \chi_{n_x}(r_x) \chi_{n_y}(r_y) \chi_{m_x}(s_x) \chi_{m_y}(s_y) \\
\times e^{-i\omega r_x + i\omega s_x + i\frac{\mu^2}{2\mu^2} \omega (r_x - s_x)} \frac{r_x^2 + r_y^2 + s_x^2 + s_y^2}{4\mu^2};
\] (4.102)

\[
T(n, m; \omega) = -\bar{T} \int dr ds \, \chi_{n_x}(r_x) \chi_{n_y}(r_y) \chi_{m_x}(s_x) \chi_{m_y}(s_y) e^{-i\omega \left( \frac{r_x^2 + r_y^2}{\mu^2 + \rho^2} \right)} \\
\times e^{i\mu^2 \omega (r_x - s_x)} \frac{1}{2} \left( \frac{1}{2\mu^2 + 2\rho^2} \right) (r_x^2 + r_y^2) - \frac{r_x^2 + r_y^2}{4\mu^2};
\] (4.103)

where

\[
\bar{S} = \frac{\mu^2}{\mu^2 + 2\rho^2} \exp \left[ -\frac{\omega^2}{2\Omega^2} - \frac{\rho^2 \zeta^4 \omega^2}{\mu^2 (\mu^2 + 2\rho^2)} \right];
\] (4.104)

\[
\bar{T} = \frac{\mu^2}{\mu^2 + \rho^2} \exp \left[ -\frac{\omega^2}{2\Omega^2} - \frac{\rho^2 \zeta^4 \omega^2}{2\mu^2 (\mu^2 + \rho^2)} \right].
\] (4.105)
Calculating these quantities will be straightforward once we possess knowledge of the function \( f_n(h, \sigma^2) \equiv \int_{-\infty}^{\infty} dr \, \chi_n(r) e^{i \sigma r - \frac{r^2}{2\sigma^2}} \) for arbitrary values of \( h \) and \( \sigma \).

Or, calling \( \kappa = \sqrt{\frac{1}{2} \left( \frac{1}{\gamma^2} + \frac{1}{\sigma^2} \right)} \), relabeling \( x = \kappa r \) and using Eq. (3.32),

\[
f_n = \frac{c_n}{\kappa} \int dx \, H_n \left( \frac{x}{\kappa \gamma} \right) e^{i \kappa x - x^2}. \tag{4.106}
\]

where \( c_n = \frac{1}{\sqrt{2\pi n!}} \left( \frac{1}{\pi \gamma^2} \right)^{1/4} \).

The integral in Eq. (4.106) can be done by applying formula 7.374/8 (p. 843) of Gradshteyn and Ryzhik (1996), yielding

\[
f_n = \sqrt{\frac{2\pi \sigma \gamma}{\pi \gamma}} \, c_n \, e^{n} \left( \frac{\sigma^2 - \gamma^2}{\sigma^2 + \gamma^2} \right)^{n/2} H_n \left( \frac{h \sigma \gamma}{\sqrt{\sigma^4 - \gamma^4}} \right) \exp \left[ - \frac{(h \gamma \sigma \omega)^2}{2(\sigma^2 + \gamma^2)} \right], \tag{4.107}
\]

where we clearly have \( f_n(-h, \sigma^2) = f_n^*(h, \sigma^2) \).

Applying Eq. (4.107) allows us to turn Eqs. (4.103) into

\[
S(n, m) = \bar{S} f_{nx} \left( -\frac{\xi^2}{\mu^2}, 2\rho^2 \right) f_{ny} \left( 0, 2\rho^2 \right) f_{mx} \left( \frac{\xi^2}{\mu^2}, 2\rho^2 \right) f_{my} \left( 0, 2\rho^2 \right); \tag{4.108}
\]

\[
T(n, m) = -\bar{T} f_{nx} \left( \frac{\eta^2}{\mu^2} - \frac{\eta^2 + \rho^2}{\mu^2 + \rho^2}, 2\rho^2 \right) f_{ny} \left( 0, 2\rho^2 \right) f_{mx} \left( \frac{\xi^2}{\mu^2}, 2\rho^2 \right) f_{my} \left( 0, 2\rho^2 \right); \tag{4.109}
\]

where

\[
\frac{1}{\bar{\rho}^2} \equiv \frac{1}{\rho^2} + \frac{2}{\rho^2 + \mu^2} = \frac{3\rho^2 + \mu^2}{\rho^2(\rho^2 + \mu^2)}, \tag{4.110}
\]

and we note for future reference that \( \bar{\rho} \approx \frac{1}{\sqrt{3}} \rho \) if \( \mu \ll \rho \).

Some remarks are in order:
(1) For $\omega = 0$, the constraint $|a_\omega\rangle$ should be orthogonal to all the non-s waves, and in particular to all the choices of $(n, m)$ that make the wave function antisymmetric in either variable. Indeed, this is verified in Eqs. (4.108-4.109) because $H_n(0)$ vanishes if $n$ is odd.

(2) For $\omega \neq 0$, these matrix elements vanish under the condition that either $n_y$ or $m_y$ is odd. This means that the constraint is orthogonal to the whole subspace spanned by odd values of the y-number. Those eigenfunctions are unperturbed, and their eigenvalue is the same in the unconstrained problem, namely $\Lambda_{n_x,n_y} = 2\pi \mu^2 \beta^{-n_x-n_y-1} \exp \left( -\frac{\omega^2}{2\Omega^2} \right)$.

(3) From the expression Eq. (4.107) for $f_n$, considering separately the cases $\sigma^2 - \gamma^2 > 0$ and $\sigma^2 - \gamma^2 < 0$, it is easy to see that the function $f_n$ will be purely imaginary if $n$ is odd, purely real if $n$ is even. But $n_y$ cannot be odd without canceling the whole matrix element. We conclude that all the matrix elements are real except those where $n_x$ is odd, which are purely imaginary.

To sum up, the matrix elements of $L^\rho$ are real for even $n_x$ and $n_y$, imaginary for odd $n_x$ and even $n_y$, and they vanish for arbitrary $n_x$ and odd $n_y$.

4.6.1 Combining s-waves with longitudinal eigenfunctions (0x theory)

We will now restrict the diagonalization of $L^\rho$ to the two-dimensional subspace spanned by $|\psi_0\rangle = |0, 0\rangle$ and $|\psi_x\rangle = |1, 0\rangle$. We will refer to the resulting variational treatment as the ”0x” theory for this model.

The unconstrained operator is diagonal, $L_{0x} = L_{x0} = 0$, and using the properties of the
constraint operators we can write the operator in the matrix form

\[
\mathbf{M} = \begin{pmatrix}
M_{00} & M_{0x} \\
M_{0x}^* & M_{xx}
\end{pmatrix} = \begin{pmatrix}
L_{00} + S_{00} + 2T_{00} & S_{0x} + T_{0x} + T_{x0}^* \\
(S_{0x} + T_{0x} + T_{x0}^*)^* & L_{xx} + S_{xx} + 2T_{xx}
\end{pmatrix}
\] (4.111)

We thus have to compute nine quantities: \(L_{00}, L_{xx}, S_{00}, S_{xx}, S_{0x}, T_{00}, T_{0x}, T_{x0}, T_{xx}\).

The diagonal matrix elements of the unconstrained operator are simply the many-cell eigenvalues

\[
L_{00} = 2\pi \mu^2 \beta^{-1} e^{-\frac{\sigma^2}{2\sigma^2}}; \quad L_{xx} = 2\pi \mu^2 \beta^{-2} e^{-\frac{\rho^2}{2\rho^2}}. \quad (4.112)
\]

For the other matrix elements, we need to write down relevant values of the \(f\)-functions of Eq. (4.106). Using Eq. (4.107) and recalling that the first Hermite polynomials are \(H_0(x) = 1, H_1(x) = 2x\), the functions we need are found to be

\[
f_0(h, \sigma^2) = \left(\frac{4\pi}{\gamma^2}\right)^{1/4} \frac{\sigma \gamma}{\sqrt{\sigma^2 + \gamma^2}} \exp\left[-\frac{(h \gamma \sigma \omega)^2}{2(\sigma^2 + \gamma^2)}\right], \quad (4.113)
\]

\[
f_1(h, \sigma^2) = \sqrt{2i} \frac{\sigma^2 \gamma h \omega}{\sigma^2 + \gamma^2} f_0(h, \sigma^2); \quad (4.114)
\]

Plugging Eqs. (4.113-4.114) into Eqs. (4.108-4.109), we conclude that

\[
S_{00} = \tilde{S} \left[ f_0 \left(\frac{\zeta^2}{\mu^2}, 2\rho^2\right) \right]^2 \left[ f_0 \left(0, 2\rho^2\right) \right]^2;
\] (4.115)

\[
S_{0x} = \tilde{S} f_0 \left(-\frac{\zeta^2}{\mu^2}, 2\rho^2\right) \left[ f_0 \left(0, 2\rho^2\right) \right]^2 f_1 \left(\frac{\zeta^2}{\mu^2}, 2\rho^2\right);
\] (4.116)
\[ S_{xx} = \bar{S} \left| f_1 \left( \frac{\zeta^2}{\mu^2} , 2\rho^2 \right) \right|^2 \left[ f_0 \left( 0, 2\rho^2 \right) \right]^2 ; \]  
(4.117)

\[ T_{0,0} = -\bar{T} f_0 \left( \frac{\eta^2}{\mu^2} - \frac{\rho^2 + \eta^2}{\rho^2 + \mu^2} , 2\rho^2 \right) f_0 \left( 0, 2\rho^2 \right) f_0 \left( \frac{\zeta^2}{\mu^2}, 2\rho^2 \right) f_0 \left( 0, 2\rho^2 \right) ; \]  
(4.118)

\[ T_{0,x} = -\bar{T} f_0 \left( \frac{\eta^2}{\mu^2} - \frac{\rho^2 + \eta^2}{\rho^2 + \mu^2} , 2\rho^2 \right) f_0 \left( 0, 2\rho^2 \right) f_1 \left( \frac{\zeta^2}{\mu^2} , 2\rho^2 \right) f_0 \left( 0, 2\rho^2 \right) ; \]  
(4.119)

\[ T_{x,0} = -\bar{T} f_1 \left( \frac{\eta^2}{\mu^2} - \frac{\rho^2 + \eta^2}{\rho^2 + \mu^2} , 2\rho^2 \right) f_0 \left( 0, 2\rho^2 \right) f_0 \left( \frac{\zeta^2}{\mu^2}, 2\rho^2 \right) f_0 \left( 0, 2\rho^2 \right) ; \]  
(4.120)

\[ T_{x,x} = -\bar{T} f_1 \left( \frac{\eta^2}{\mu^2} - \frac{\rho^2 + \eta^2}{\rho^2 + \mu^2} , 2\rho^2 \right) f_0 \left( 0, 2\rho^2 \right) f_1 \left( \frac{\zeta^2}{\mu^2} , 2\rho^2 \right) f_0 \left( 0, 2\rho^2 \right) . \]  
(4.121)

The principal eigenvalue of matrix (4.111) can be written in terms of the quantities (4.115-4.121) as

\[ \Lambda(\omega) = \frac{1}{2} \left( L_{00} + L_{xx} + S_{00} + S_{xx} \right) + T_{00} + T_{xx} + \]  
\[ + \left[ \left( L_{00} - L_{xx} + S_{00} - S_{xx} + 2T_{00} - 2T_{xx} \right)^2 + 4 \left| S_{0x} + T_{0x} + T_{x0} \right|^2 \right]^{1/2} . \]  
(4.122)

and we will now proceed to study this eigenvalue under the assumption \( \rho/\mu \ll 1 \).

### 4.6.2 Short-range limit

Let us now adopt the assumption \( \mu \ll \rho \), which entails \( \bar{\rho}^2 \sim \rho^2/3 \) and \( \gamma^2 \sim \rho \mu \). Moreover,

\[ \frac{\eta^2}{\mu^2} - \frac{\rho^2 + \eta^2}{\rho^2 + \mu^2} \sim -\frac{\zeta^2}{\mu^2} + \frac{\zeta^2}{\rho^2} + o \left( \frac{\mu^2}{\rho^2} \right) . \]  
(4.123)

Here we may neglect \( \frac{\zeta^2}{\rho^2} \) as long \( \frac{\zeta^2}{\mu^2} = O(1) \) (thus, we are barring ourselves from considering the limit \( \zeta \to 0 \)).
With this simplification, the only values of the $f$-function we need are:

\begin{align*}
f_0(0, 2\rho^2) & \sim f_0(0, 2\bar{\rho}^2) \sim (4\pi \rho \mu)^{1/4}; \quad (4.124) \\
f_0(\pm \frac{\zeta^2}{\mu^2}, 2\rho^2) & \sim f_0(\pm \frac{\zeta^2}{\mu^2}, 2\bar{\rho}^2) \sim (4\pi \rho \mu)^{1/4} \exp \left( -\frac{\rho \zeta^4 \omega^2}{2\mu^3} \right); \quad (4.125) \\
f_1 \left( \frac{\zeta^2}{\mu^2}, 2\rho^2 \right) & \sim f_1 \left( \frac{\zeta^2}{\mu^2}, 2\bar{\rho}^2 \right) \sim \sqrt{2} \frac{\rho^{1/2} \zeta^2 \omega}{\mu^{3/2}} f_0 \left( \frac{\zeta^2}{\mu^2}, 2\rho^2 \right). \quad (4.126)
\end{align*}

Applying this, we can see from Eqs. (4.115-4.121) that that the matrices $\hat{S}$ and $\hat{T}$ may be re-written in the form $\hat{S} \sim \hat{S} \hat{Q}$ and $\hat{T} \sim -\hat{T} \hat{Q}$, where

\begin{align*}
Q_{00} &= 4\pi \rho \mu \exp \left( -\frac{\rho \zeta^4 \omega^2}{\mu^3} \right); \\
Q_{0x} &= \frac{\sqrt{2} \rho^{1/2} \zeta^2 \omega}{\mu^{3/2}} Q_{00} \\
Q_{x0} &= Q_{0x}^*; \\
Q_{xx} &= \frac{2\rho \zeta^4 \omega^2}{\mu^3} Q_{00};
\end{align*}

so that, applying Eq. (4.111), we obtain

\begin{align*}
M_{00} &= L_{00} + (\bar{S} - 2\bar{T}) Q_{00}; \\
M_{0x} &= (\bar{S} - 2\bar{T}) Q_{0x}; \\
M_{xx} &= L_{xx} + (\bar{S} - 2\bar{T}) Q_{xx}.
\end{align*}

(4.131)

We would like now to expand Eqs. (4.131) in the small parameter $u = \mu/\rho$. Notice that this expansion is complicated by the dependence of the various matrix elements on $\omega$, because we do not know of what order of magnitude the optimal wavenumber $\omega$ will turn out to be in terms of $\mu/\rho$. In particular, the off-diagonal matrix element is directly
proportional to \( \omega \), and only a posteriori we will be able to check if the optimal wavenumber makes it of the same order as the diagonal matrix elements.

Approximating \( \beta \sim 1 + u \), we expanding the equations (4.112), one obtains

\[
L_{00} \sim 2\pi \mu^2 e^{-\frac{\omega^2}{\Omega^2}} (1 - u) \quad L_{xx} \sim 2\pi \mu^2 e^{-\frac{\omega^2}{\Omega^2}} (1 - 2u). \tag{4.132}
\]

Using moreover the fact that, by virtue of Eqs. (4.104-4.105), we can write \( \vec{S} - 2\vec{T} \sim -\frac{3}{2} \mu^2 \exp \left(-\frac{\zeta^2 \omega^2}{2}\right) \), from the Eqs. (4.131) we arrive at

\[
M_{00} = e^{-\frac{\omega^2}{\Omega^2}} \left( 2\pi \mu^2 \frac{2\pi \mu^3}{\rho} - 6\pi \frac{\mu^3}{\rho} e^{-x^2 \zeta^2 \omega^2} \right); \tag{4.133}
\]
\[
M_{0x} = e^{-\frac{\omega^2}{\Omega^2}} \left( -3 \times 2^{3/2} \pi i \frac{\mu^{3/2} \zeta^2 \omega}{\rho^{1/2}} e^{-x^2 \zeta^2 \omega^2} \right); \tag{4.134}
\]
\[
M_{xx} = e^{-\frac{\omega^2}{\Omega^2}} \left( 2\pi \mu^2 \frac{4\pi \mu^3}{\rho} - 12\pi \zeta^4 \omega^2 e^{-x^2 \zeta^2 \omega^2} \right); \tag{4.135}
\]

where \( x = \frac{\zeta}{\mu} \sqrt{\frac{1}{2} + \frac{\rho}{\mu}}. \)

In the limit \( \omega \to 0 \) (or also \( \zeta \to 0 \)), Eq. (4.111), via Eqs. (4.133, 4.134, 4.135), takes the form

\[
\hat{M}(\omega = 0) = \begin{pmatrix} 2\pi \mu^2 - 8\pi \frac{\mu^3}{\rho} & 0 \\ 0 & 2\pi \mu^2 - 4\pi \frac{\mu^3}{\rho} \end{pmatrix}. \tag{4.136}
\]

So the two components are decoupled, and the \( |0, 0\rangle \) eigenfunctions lies below the \( |0, 1\rangle \) eigenfunction, which in turn lies below the moveable-node eigenfunction.

We switch to dimensionless units by defining \( z = \zeta / \rho \), \( w = x \zeta \omega \) and \( \alpha = (2x^2 \zeta^2 \Omega^2)^{-1} \). We will additionally write \( x = y / \sqrt{u} \), because \( y \) stays finite in the limit \( u \to 0 \). We write
hence

\[ M_{00} = 2\pi \mu^2 e^{-\alpha w^2} \left( 1 - u - 3u e^{-w^2} \right); \quad (4.137) \]
\[ M_{0x} = 2\pi \mu^2 e^{-\alpha w^2} \left( -\frac{3\sqrt{2i}z}{y} e^{-w^2} \right); \quad (4.138) \]
\[ M_{xx} = 2\pi \mu^2 e^{-\alpha w^2} \left( 1 - 2u - \frac{6uz^2}{y^2} w^2 e^{-w^2} \right); \quad (4.139) \]

and the principal eigenvalue of the matrix 4.111), i.e. of \( L^p \) in the reduced Hilbert space, is

\[
\Lambda(w; z, h) = \pi \mu^2 e^{-\alpha w^2} \left[ 2 - 3u - 3ue^{-w^2} - \frac{6uz^2}{y^2} w^2 e^{-w^2} + \sqrt{(u - 3ue^{-w^2} + \frac{6uz^2}{y^2} w^2 e^{-w^2})^2 + 8 \times \frac{z^2 w^2}{y^2} e^{-2w^2}} \right] , \quad (4.140)
\]

where \( u = h^2 + z^2 = \sqrt{\eta^2 + \zeta^2/\rho^2} \), \( y = \frac{z}{u} \sqrt{1 + \frac{u}{2}} \) and \( \alpha = \frac{y^2}{2\mu^2 w^2} \).

### 4.6.3 Local stability of the R-phase

The local stability of the zero wavenumber state is immediately obtained from formula (4.140), namely

\[
\left. \frac{\partial^2 \Lambda}{\partial w^2} \right|_{w=0} = -\pi \mu^2 \frac{4u}{z^2(u + 2)} \left( 12z^4 - 18z^2 + h^2 + 12z^2h^2 - 2uh^2 \right) \\
\propto - [h^2 - 18z^2 + O(u^3)] , \quad (4.141)
\]

104
which means that the cortically uniform orientation-selective solution $\Psi_{2\eta}$ becomes locally unstable if

$$\frac{h}{z} < \kappa_L \equiv 3\sqrt{2} \sim 4.24.$$ (4.142)

This estimates the slope of the continuous phase boundary due to local instability, which is depicted as a dashed red line in Fig. 4.9).

As a corollary of formula (4.142), the zero-wavenumber solution is stable along the whole $\eta$-axis and unstable along the whole $\zeta$-axis, in agreement with our long-range solution of Sec. 4.5.\footnote{Moreover, as we showed in Sec. 4.3; zero wavenumber eigenfunctions are stable all along the line $\zeta = 0$; in the regime in which we have found the zero wavenumber to be locally stable, numerics will show that it is in fact globally stable.}

To study the vicinity of the transition, we need to expand Eq. (4.140) for small values of $w$, which yields

$$\frac{\Lambda(w)}{2\pi \mu^2} \sim 1 - 2u + \frac{u}{2z^2} (18z^2 - h^2) w^2 - 45 u w^4 + O(w^6),$$ (4.143)

maximized by $w = \sqrt{\frac{18z^2 - h^2}{6\sqrt{5}z}}$ or, switching back to dimensional variables,

$$\omega_M \sim \frac{1}{6\sqrt{5}} \frac{\mu^{3/2}}{\sqrt{\mu \kappa^3}} \sqrt{18\zeta^2 - \eta^2}.$$ (4.144)

The eigenvalue predicted by the 0x theory is the expectation value of Eq. (4.143) computed at the optimal wavenumber in Eq. (4.144). This is given by

$$\frac{\Lambda_{0x}}{2\pi \mu^2} = 1 - 2u + u \frac{(h^2 - 18z^2)^2}{720z^4},$$ (4.145)
an eigenvalue that will be useful once we check the global stability of the 0x solution.

4.6.4 0x eigenfunction

Having computed the 0x eigenvalue, we may turn to looking at the 0x eigenfunction. The principal eigenvector of matrix (4.111) is given by the vector \( v = (v_0, v_x) \), where

\[
\begin{align*}
    v_0 &= M_{xx} - M_{00} - \sqrt{(M_{00} - M_{xx})^2 + 4|M_{0x}|^2} \\
    v_x &= 2M_{0x}^* 
\end{align*}
\]  \hspace{1cm} (4.146)

Just as we proved to be the case in general (Sec. 4.2.6), we see from Eq. (4.146) that there is a phase shift of magnitude equal to \( \pi \) between the symmetric and antisymmetric components of the eigenfunction, and we are at liberty to write it as

\[
\psi(r) = \exp \left( -i \frac{\eta^2}{\mu^2} \omega r_x \right) \left( \Psi_{\text{non-selective}}(r) + i \Psi_{\text{selective}}(r) \right), \quad (4.147)
\]

Here the two \( \Psi \)’s are real, and the gauge factors is negligible wherever the wavenumber is sufficiently small.

4.6.5 The uniform phase

Equation (4.146) shows that \( v_x \propto M_{0x} \). Plugging (4.144) into (4.138), we see that in the vicinity of the local instability boundary we have

\[
M_{0x} \propto \frac{\mu^3}{\rho_\zeta} \sqrt{18\zeta^2 - \eta^2}, \quad (4.148)
\]
thus $v_x$ vanishes near the critical line.

From Eq. (4.146), using the fact that $\lim_{\omega \to 0} (M_{xx} - M_{00}) > 0$, one can see that $v_0$, too, vanishes near the critical line. As $M_{0x} \to 0$, calling $\Delta = M_{xx} - M_{00}$, we can write

$$v_0 = \Delta - \sqrt{\Delta^2 + 4|M_{0x}|^2} \sim -2\frac{|M_{0x}|^2}{\Delta}. \quad (4.149)$$

and this vanishes faster than $v_x$ as $\omega \to 0$: $\frac{v_0}{v_x} \propto M_{0x}$. Therefore, the uniform phase we have been discussing is an R phase, with $v_0 = 0$. Since it is an R-phase, the receptive field is the longitudinal p-wave $\chi_{1,0}(r)$, which is degenerate with the transverse p-wave $\chi_{0,1}(r)$, so degeneracy arises as the T-phase gives way to the R-phase.

It is also interesting to consider the case $\eta = 0$ and $\zeta \to 0$. The optimal wavenumber (4.144) diverges as $(\rho \zeta)^{-1/2}$ for $\eta \equiv 0$ and $\zeta \to 0$. However, the exponential factor $e^{-x^2\zeta^2\omega^2}$ in the matrix elements (4.133-4.135) does not vanish, but becomes of order one. Plugging (4.144) into (4.133-4.135) for $\eta = 0$ we see that $M_{0x}$ and $M_{00} - M_{xx}$ are both of order $\zeta^3/\rho$.

This suggests that along the $\zeta$ axis, for sufficiently small $\zeta$, the 0 and x components are of the same order. Since we know that this is not true if we approach the origin of the phase diagram from above, it follows that the origin of the phase diagram constitutes a singular point for the principal eigenfunction$^8$.

---

$^8$Which is not surprising because at that point the operator is unbounded (Halmos-Sunder, 1978).
Figure 4.9: Structure of the variational theory for short-range interactions. The RT local instability line \((\eta = \kappa_L \zeta)\) is shown in red. The global instability curves of the N phase \((\mu = \theta_c \rho \text{ and } \eta = \kappa_G \zeta)\) are shown in blue. Phase boundaries in black. Regions corresponding to different symmetries have been drawn in different colors (see the legend).

### 4.7 The N-phase domain

#### 4.7.1 Global stability of the 0x solution

Of course the 0x state’s being locally stable at some optimal wavenumber does not by itself imply that it is globally stable. A partial check for global stability consists in comparing the expectation value (4.145) of the T phase with the expectation value of our moving-node function (N-phase), whose approximate eigenvalue is given by Eq. (3.70): \(\frac{\Lambda_{MN}}{2\pi \mu^2} \sim 1 - (5 - \sqrt{10}) u\). The jump from the 1s+2s solution to the 0x solution would then happen when \(\Lambda_{MN} = \Lambda_{0x}\), i.e. when

\[
\frac{(h^2 - 18z^2)^2}{720z^4} = \sqrt{10} - 3 \sim 0.16. \tag{4.150}
\]
Formula (4.150) is justified in a neighborhood of the point $z = 0, h = 0$, and can be rewritten as

$$h^4 - 36z^2h^2 + \kappa z^4 = 0,$$

(4.151)

with $\kappa = 2484 - 720\sqrt{10} \approx 207$. There are two solutions to Eq. (4.151), namely

$$\frac{h^2}{z^2} = 18 \pm \sqrt{324 - \kappa}.$$

(4.152)

Choosing the plus sign yields $h/z \approx 5.4$, while choosing the minus sign yields $h/z \approx 2.7$. Obviously, only the second solution is compatible with condition (4.142); since we have gotten here by using formula (4.144), valid under condition (4.142), it follows that only the minus-sign solution must be kept.

We conclude that there is a discontinuous phase boundary due to global instability and the T phase prevails over the N phase under the condition

$$\frac{h}{z} < \kappa - \sqrt{18 - \sqrt{720\sqrt{10} - 2160}} \approx 2.68.$$

(4.153)

### 4.7.2 Shape of the phase boundary

We have seen that a sharp boundary exists where the the zero-wavenumber 1s-2s function (our approximation to the N phase) goes under the 0x-function (our approximation for the T phase), made out of 1s and 2p waves at finite wavenumber.

For $\zeta = 0$, the point where the phase boundary is located is of course $P_0$, which we
define here by $(\zeta_0, \eta_0) = (0, \theta_c \rho)^9$. At this point (where, incidentally, the weight of the 0x eigenfunction lies all on 2p), the transition is between the 1s-2s and 2p functions – the same transition we studied in the 1-cell case.

Let us observe how the phase boundary (as described by the 0x theory) evolves as $\zeta$ is increased from zero. As long as $\zeta/\eta$ stays sufficiently small ($\zeta/\eta < 1/3 \sqrt{2}$), the R phase will lie above the T phase everywhere in the vicinity of the circumference $\mu = \rho \theta_c$. Since that circumference is the boundary between the R and N phases, and since T can be neglected as being dominated by R, it follows that the phase boundary will proceed simply along the circumference as we increase $\zeta$ from 0.

Now consider approaching the point $P_T$ defined by the intersection of $z^2 + h^2 = \theta_c^2$ with $h/\zeta = 3\sqrt{2}$ (Eq. 4.142), where the T phase comes on a par with the R phase. From this point downward, the phase boundary is crossing a region where the R-phase is no longer stable, therefore the T phase must be reckoned with. As long as the T phase has a lower eigenvalue than the N phase, however, the boundary will continue to run along the circumference. The point where T becomes globally favored over the N phase is the intersection of the two equations:

\[
\begin{align*}
  z^2 + h^2 &= \theta_c^2 \\
  h &= \kappa_{Gz}
\end{align*}
\]

(4.154)

With slight abuse of notation we have used the term $P_0$ both for the point $(0, \Theta_c \rho)$ and for the approximate transition point obtained by replacing the true critical ratio $\Theta_c$ with $\theta_c$. The latter is the termination of the phase boundary in this variational theory, not in the full model, but the difference is inessential in the arguments to come.
We call \( P^* = (z^*, h^*) \) the solution of this system, given by

\[
\begin{align*}
    z^* &= \frac{\theta_c}{\sqrt{1 + \kappa_G^2}} \approx 0.11 \\
    h^* &= \frac{\theta_c \kappa_G}{\sqrt{1 + \kappa_G^2}} \approx 0.30.
\end{align*}
\] (4.155)

From here downward, the phase boundary of the \( N \) phase follows the global stability boundary, which is given by the straight line \( h = \kappa_G z \) as per Eq. (4.153). Hence, the point \( P^* \) defines the corner of a circle-sector inside which the \( N \)-phase is confined (Fig. 4.9).

We conclude that the phase diagram contains three phases: an \( N \) phase with a one-node radial eigenfunction extending within a sector-shaped domain centered on the origin, adjacent to the \( \eta \)-axis and bordered by the circular arc that joins the points \( P_0 \) and \( P^* \); an \( R \) phase that begins on top of the \( N \) phase and extends above an oblique line in the \( (\zeta, \eta) \) plane; the \( T \) phase covering all the rest of the phase diagram.

This also means that there exists a triple point at the location \( P_T \), i.e. a point where the three phases meet. We find that this point is the intersection of \( z^2 + h^2 = \theta_c \) with the local-instability curve \( h/z = 3\sqrt{2} \). Solving the system yields the coordinates of the triple point within the 0x theory:

\[
\begin{align*}
    \frac{\zeta_T}{\rho} &= \frac{\theta_c}{\sqrt{19}} \approx 0.22 \\
    \frac{\eta_T}{\rho} &= \frac{3\sqrt{2}\theta_c}{\sqrt{19}} \approx 0.97.
\end{align*}
\] (4.156)
4.7.3 Corrections to the variational results

Direction of the shifts

The application we have made here of the variational method differs in one important way from its application in the previous chapter.

In Chapter 3, we used the exact eigenfunction as representative of the p-wave phase, and an approximate eigenfunction as representative of the s-wave phase. Hence, we could prove that the variationally computed critical interaction length \( \theta_c \rho \) was a lower bound to the actual value. In the two-layer model we are analyzing, on the other hand, the 0x function is not an exact eigenfunction of the system. Therefore, here we have used approximate eigenfunctions to represent both phases – the nonselective and selective phases. The actual RF is neither exactly a 0x function in the T phase, nor exactly a 1s-2s function in the N phase.

As a consequence, we cannot predict in which direction the NT boundary will be shifted from the calculated shape; whether it will move to the right or to the left. By changing either of the two trial functions, the boundary may be pushed either into the N region or into the T region, depending on the Hilbert-space extension being explored.

The opposite is true, however, as concerns the RT boundary. Additional corrections to the variational theory considered so far could only push the RT boundary further to the left. No corrections are to be made to the R eigenfunction, which is exact, hence only the T eigenvalue will be augmented.
Smoothening of the transitions

As we have seen, the RT phase boundary is a continuous boundary, meaning that the R and T eigenfunctions turn continuously into each other. The NR and NT phase boundaries are discontinuous within the limits of our variational theory, as they signal a discontinuous change of the principal eigenfunction.

A variational theory with a larger basis could in principle smoothen the transition underlying a boundary, by including approximate eigenfunctions that bridge the gap between the eigenfunctions in the two phases. This would turn a discontinuous transition into a continuous one. Symmetry considerations, however, can serve to rule out this possibility.

This is obvious for what concerns the RN boundary. Our approximation for the N phase included the first two s-waves, but including more of them could not possibly create a bridging with the R eigenfunction. The R eigenfunction $|R\rangle$ is known exactly, and $\langle R|L^p(\omega = 0)|N\rangle$ for any s-wave $|N\rangle$.

The same is not true about the NT transition. Indeed, corrections of the 0x eigenfunction can make the transition continuous, localizing the global instability. As we approach the NT boundary from within the T region, the wavenumber will tend to zero and at the same time the principal eigenfunction will lose its non-s components. Thus, corrections to the 0x theory can in principle bridge the transition, making it continuous.

The phase boundary, however, will remain sharp, that is, its eigenvalue will be nonanalytic along the boundary. The reason is that, as we showed through perturbation theory in Sec. (4.3), the zero-wavenumber N phase is dominant in a right neighborhood of the segment $\zeta = 0, \eta \in [0, \eta_c]$. It follows that the transition from N to T, where it happens,
is a transition from a uniformly zero to a non-zero cortical wavenumber, which cannot be analytic\textsuperscript{10}.

**Shift of the triple point**

We consider here how corrections to the variational treatment may possibly affect the the NR boundary, which the 0x theory locates along the circular arc that joins the points $P_0$ and $P^*$. 

One way, of course, will be the upward shift of the point $P_0$, due to the fact that $\theta_c < \Theta_c$, and a simultaneous expansion of the arc. But notice that this is not essential because, if we possessed the exact value of $\Theta_c$, the 0x theory would yield the same geometric structure we just derived.

A bridging of the R and N phase would be a more substantial effect. As we discussed, the only way to bridge the R and N phase will be to go through the $T$ phase, creating an $NTR$ sequence. Hence, it is possible that a T-phase ribbon will appear along the supposed NR boundary once additional basis functions are included in the variational theory. This intermediate T-ribbon will allow for a continuous evolution of the principal eigenfunction in between the N and R phase, which would not be possible at $\omega = 0$. This will also have the effect of moving the triple point further up along the circumference $\mu = \theta_c$\textsuperscript{11}.

On the other hand, we have shown that at the top of the circumference, i.e. in the point $P_0$, the RT transition must be discontinuous. Indeed, we proved in Section (4.3) that the

\textsuperscript{10}In fact, the N solutions have also an additional symmetry – under rotations of the sole presynaptic layer while keeping the postsynaptic layer fixed – that is broken by introducing a finite wavenumber. Hence, a nonanalytic transition is needed for the $T$ phase to begin.

\textsuperscript{11}In agreement with our previous statement concerning the RT boundary, which can only be moved leftwards from its calculated value
preferred wavenumber for \( \zeta = 0 \) is \( \omega = 0 \). We also showed perturbatively that a right neighborhood of the \( \zeta = 0 \) axis contains an R phase for any \( \eta > \theta_c \).

Suppose now that the eigenfunctions sitting on either side of an NR boundary will be bridged continuously by an intermediate T phase. Clearly this intermediate T-phase ribbon cannot touch the \( \eta \) axis, or more precisely it can only touch it at the point \( P_0 \), and the triple point cannot move any further up than that along the \( \eta \) axis. It follows that, if the bridging occurs wherever allowed, the intermediate T-phase will have the shape of an inlet from the larger T region we calculated, tapering as it approaches the point \( P_0 \).

Since no symmetry prevents this full bridging of the variational eigenfunctions, we expect \( P_0 \equiv (0, \rho \Theta_c) \) to be the exact triple point of the system.
Figure 4.10: Cortical wavenumber in units of the inverse arbor radius (see color bar). The two axes correspond to the cortical interaction length and the LGN correlation length, both measured in units of the arbor radius. Clearly visible are the boundaries of the two cortically uniform phases.
Figure 4.11: Principal eigenfunction of the Fourier-transformed model at the optimal wavenumber, computed for three representative points in the phase diagram corresponding to the values of $\eta/\rho$ and $\zeta/\rho$ indicated underneath the plots. Hues represent the value of the receptive fields, normalized wave-function-wise, the offset being blue for the real parts and green for the imaginary parts. LGN is represented by a $25 \times 25$ cell grid with a side equivalent to three arbor radii. Eigenfunctions have been rotated so as to nullify the integral of the imaginary part. In accordance with our analysis, this results in eigenfunction whose real part is symmetric under inversion of the cortical-wavevector axis, and whose imaginary part is antisymmetric under the same.
Figure 4.12: Orientation map obtained for a 100×100-grid cortical layer with pace 0.2ρ, computed by summing the principal eigenfunction of the model (Eq. 4.96, with η = .5ρ, ζ = 3ρ) for 20 equally spaced angular directions of the wavevector, represented by the periodic color map.
Figure 4.13: Result of four dynamical simulations for a 32x32 cortex with 16x16 RFs, an arbor size of 13x13, and periodical boundary conditions. Green region corresponds to ON-prevalence, red regions to OFF prevalence. The parameters are $\zeta = \rho$, $\zeta = 1.5\rho$ respectively for the two rows and $\eta = \rho$, $\eta = 1.5\rho$ respectively for the two columns.
4.8 Numerics and conclusions

4.8.1 Numerical results

We will review the numerical results on the model before presenting a closing discussion to this Chapter.

In Fig. 4.10, the optimal wavenumber $\omega_M$ is plotted as a function of the model’s parameters, revealing the sharp boundaries between translationally invariant and cortically modulated regions. The slope of the NT boundary appears, interestingly, to be lower than predicted by the variational theory. This means that the N-phase becomes more robust when it can be constructed from more complex patterns than just linear combinations of 1s and 2s waves, and suggests that the 0x function may work better in approximating the actual T phase, leaving less room for improvement. As for the triple point, it moves all the way to the $\eta$-axis and coincides with the point $P_0$, as was suggested in Sec. (4.7.3).

Fig. 4.11 shows RFs plotted at three representative points chosen from the N, R, and S regions of the phase diagram. Eigenfunctions have been rotated so that the real part is symmetric under inversion of the cortical-wavevector’s axis and the imaginary part is antisymmetric under the same, which we have seen to be always possible (Sec. 4.2.6). In particular, this makes the N receptive field purely real and the R receptive field purely imaginary.

Fig. 4.12 is the result of summing the T-phase eigenfunction (4.96), computed at the optimal wavenumber, over a large number of different orientation of cortical wavevectors, so as to simulate biologically plausible initial conditions. The phase shifts between differently directed components are taken to be uniformly distributed. The dominant orientation of the
resulting RF is indicated by different hues, just as in the experimental plots of Fig. 4.1.

Finally, Fig. 4.13 displays the result of dynamical simulations performed by evolving random initial conditions with the operator $\hat{L}^p$ for a sufficiently long time to ensure convergence, with parameter values given in the caption. RFs are shown at corresponding points on a $100 \times 100$ grid of cortical cells, and areas of each grid corresponding to opposite signs of the RF are shown in different colors.

### 4.8.2 Conclusions

This chapter has worked out a possible resolution to the enigma posed by Ohshiro and Wei- liky (2006) with their experiments on the development of orientation selectivity in ferrets. Studying visual cortex during development under various visual inputs, these authors found no sign change in LGN correlations in spite of this being the foundation for existing theories on the Hebbian development of selectivity (Miller 1994, Wimbauer et al., 1998).

Here, we have re-examined the problem by devoting a more careful treatment to the biology of axonal competition. By building and solving a minimal model of receptive field development in which the total projection strength from each presynaptic cell is conserved, we have shown that a constrained dynamics keeping axonal competition into account is able to produce a realistic salt-and-pepper organization of selectivity even with a simple fall-off in input correlations.

The multilayer model we employed consists of two layers representing ON/OFF cells in LGN and a layer representing V1. The key dimensionless parameters can be taken to be the widths of LGN correlations and of cortical lateral interaction, in units of the arbor
radius. As the model is translationally and rotationally invariant, possible solutions can break either symmetry, or both, or none. Orientation selectivity requires breaking rotational symmetry, while variation of preferred orientations across cortex requires also the breaking of translational symmetry.

We have studied the phase diagram of the model using different methods in different regimes. For small LGN interaction length, we used a perturbative expansion of a locally exact solution, complementing it with variational arguments. For long-range interaction, we expanded the matrix elements asymptotically in real space and diagonalized the reduced-rank problem exactly. For small LGN interaction we constructed a variational theory that is able to account for the geometric structure of phase boundaries.

The uniformly non-selective phase (neither symmetry broken) prevails for sufficiently small values of the width parameters. Rotational invariance is broken for sufficiently long-range cortical interactions; translational, for sufficiently long-range LGN correlations. Besides computing the phase boundaries, we evaluated explicitly the preferred value of the cortical wavenumber across parameter space and the functional form of receptive fields in the various regimes.

The relevance of this study to current experimental inquiries lies in the proof that spatial organization of orientation selectivity can develop through ON/OFF Hebbian learning without Mexican-hat correlations, provided a constraint fixes the total synaptic strength to cortex from each ON and OFF input. Indeed, this finding may help dispel the common wisdom that long-range inhibition would be necessary to create smooth orientation maps, thus bridging the theory of Hebbian development to advances in our experimental knowledge.
of visual cortex\textsuperscript{12}.

From the theoretical viewpoint, the relevance of this work lies largely in the application of quantum mechanics tools to Hebbian evolution, which exemplifies a general and far-reaching mathematical possibility.

The variational method, in particular, has not been applied before, to the best of our knowledge, to Hebbian models. Although we have only applied it here to the development of orientation selectivity, it may certainly be applied to other existing problems in Hebbian development, such as those concerning the emergence of ocular dominance, direction selectivity, and phase selectivity among others. The fastest-growing patterns emerging during development can be treated as the ground state of a quantum-mechanical system, and where their symmetries are concerned, mapping the parameter space of the system becomes a standard exercise in quantum mechanics.

On a more general note, it should be said that little has been done so far in studying possible plausible nonlinear complications of Hebbian models, beyond the mere aim of ensuring convergence. Yet, nonlinear variants of Linsker-type Hebbian dynamics may be key to creating bottom-up models of orientation development that are able to reproduce not only a salt-and-pepper organization such as we have demonstrated, but also the bandpass spectrum measured in the cortical architecture of several groups of animals, notably cats and monkeys. This would serve to bridge Hebbian theory to Landau-type models of universal behavior such as those of Kaschube et al. (2010).

Finally, it should be remembered that the models we studied both in this chapter and

\textsuperscript{12}Preliminary results from the work we have discussed in this chapter were presented in Fumarola and Miller (2017).
in Chapter 3 relied importantly on the initial conditions being random, or at least not finely tuned to avoid the principle eigenspace of the dynamics. We have also allowed for the activity patterns to be stochastic, and only their correlation functions appeared to influence the synaptic weights on the time scale of development. On the other hand, the emerging connectivity matrix has been regarded as a fully deterministic object.

This approach, however, may fail to be adequate in a number of applications. In particular, that may be the case when one focuses on details of the neural dynamics on a shorter time scale than that of synaptic plasticity. Fluctuations in the synaptic weights can play the role played by quenched disorder in systems like spin glasses, creating thus the random field in which neural trajectories move. This problem will be approached in a systematic fashion starting in the next chapter.
Chapter 5

The Spectral Properties of Weight Matrices

5.1 The open problem of synaptic stochasticity

5.1.1 Background and motivation

The notions on synaptic plasticity reviewed at the beginning of Chapter 3, which form the basis for our discussion of visual cortex, should be sufficient to establish that the entries of the synaptic weight matrix are strongly history dependent.

Plasticity, which underlies learning and memory, naturally gives rise to a synaptic connectivity that is largely dependent on previous experience, that is, on the past trajectories of the system. Since the matrix encodes aspects of the statistical structure of sensory environment over time, it will contain a conspicuous stochastic component reflecting the inherent stochasticity of particular histories of sensory experience.

Even apart from this intrinsic stochasticity, the necessity of stochastic modeling comes from the large number of degrees of freedom involved in any given neural circuit. This adds up, moreover, to the randomness coming from sample-to-sample variability. Thus, in the modeling of any cortical component, no matter how specific and well-documented, the stochasticity of synaptic weights is seldom negligible.

In the attempt to describe plausible neurobiological behavior, therefore, one must gen-
eralize from the dynamics of a system with a fixed, regular connectivity to the expected behavior of a typical sample from an appropriate connectivity ensemble. Moreover, this stochasticity (a) may be too strong to allow for perturbative treatments; (b) may not be i.i.d. across elements of the connectivity matrix; (c) may be superimposed on a deterministic component that is itself non-negligible.

5.1.2 Random matrices beyond quantum mechanics

The theory of random matrices arose in the 1950’s, when physicists began to experiment with heavy nuclei more extensively than they had so far. These nuclei absorb and emit thousands of frequencies, and not only the systems are non-integrable, but writing the Hamiltonian may already be a problem, as there are hundreds of nucleons involved. Even with all the analytical machinery of Schrödinger operators, adequately solving the eigenvalue problem is unfeasible.

Dyson (1962) summarized the motivation behind the use of statistical methods as follows:

The statistical theory will not predict the detailed sequence of levels in any one nucleus. But it will describe the general appearance and the degree of irregularity of the level structure, that is expected to occur in any nucleus which is too complicated to be understood in detail.

This program was first carried out in Wigner (1958), and has later also been applied to quantum systems on the mesoscopic scale, whose Hamiltonian fluctuates with different types of symmetry constraints. The probability distribution of the levels has been calculated analytically in many cases; for a review see Mehta (2004), Guhr et al. (1998), or Bai and
Silverstein (2006). In all these applications, the matrices to be diagonalized are Hermitian, because so is the Hamiltonian of the system.

Applying results from quantum physics directly to neurobiology would not be possible. The situation in neurobiology is made more intricate by two inescapable complications. In the first place, the synaptic weight matrix of a network is never entirely random, but features substantial deterministic structure. For example, connections between neurons typically have restricted spatial range; they show specificity with respect to neuronal type, location, and response properties. In the second place, this base structure (as well as the disorder on top of it) is in general described by non-Hermitian matrices.

In fact, whereas Hermiticity is ensured in the quantum case by a general principle (the conservation of probability), non-Hermiticity is ensured in the neuronal context by an empirical principle, traditionally known as Dale’s law.

Dale’s law is the observation (Dale, 1935; Eccles, 1954) that individual neurons release the same neurotransmitter at all of their synapses (see discussion in Section 2.2). This observation has been found to hold true with very few exceptions (Jonas et al., 1998; Root et al., 2014). In the context of theoretical studies, it refers more specifically to the fact that an individual neuron makes either only excitatory or only inhibitory synapses.

This functional principle constrains the structure of neural connectivity (Strata and Harvey, 1999) and imposes a certain architecture on the synaptic weight matrix, forcing all elements in each column of the matrix (describing the synaptic projections of a certain neuron) to have the same sign. Especially when the typical weight magnitude is much larger than typical differences between the magnitudes of excitatory and inhibitory weights, such a matrix can be extremely nonnormal (cf. footnote 2), much more so than a fully random
Furthermore, nonnormality can lead to important dynamical properties not seen for normal matrices. As will be discussed below, networks with a recurrent connectivity pattern described by a nonnormal matrix can be described as having a hidden feedforward connectivity structure between orthogonal activity patterns, each of which can also excite or inhibit itself (Ganguli et al., 2008).

Besides containing a vast number of results on Hermitian random matrices, the literature contains also some results on fully random non-Hermitian matrices with i.i.d. zero-mean elements (Feinberg and Zee, 1997). However, there is a shortage of results on quantities of interest for nonnormal matrices that fall in between the two extremes of fully random and fully deterministic.

A natural departure from a nonnormal deterministic structure (as described by a fixed connectivity matrix $M$) is to additively perturb it with a fully random matrix $J$ with i.i.d. zero-mean elements. In many important examples, however, the strength of disorder (deviations from the mean structure) is not uniform and itself has some structure (e.g. for each connection it can depend on the types of the connected nodes or neurons). Moreover, the deviations of the strength of different connections or interactions from their average need not be independent. Hence it is important to move beyond a simple i.i.d. deviation from the mean structure.
5.1.3 Outline of the treatment

In this chapter, we will study ensembles of large $N \times N$ random matrices of the form $A = M + LJR$ where $M$, $L$ and $R$ are arbitrary ($M$) or arbitrary invertible ($L$ and $R$) deterministic matrices that are in general nonnormal, and $J$ is a completely random matrix with zero-mean i.i.d. elements of variance $1/N$. The matrix $M$ is thus the average of $A$, and describes average connectivity. When $L$ and $R$ are diagonal, they specify variances that depend separably on the row and column of $A$; when they are not diagonal, the elements of $A$ are not statistically independent. In fact, $L$ and $R$ allow correlations that have separable dependence on row and column indices.

As will be shown in Sec. 5.4.4, this form arises naturally in linearizations of dynamical systems involving simple classes of nonlinearities. This type of ensemble is also natural from the random-matrix-theory viewpoint, as it describes a classical fully random ensemble – an i.i.d. random matrix $J$ – modified by the two basic algebraic operations of matrix multiplication and addition.

Using the Feynman diagram technique in the large $N$ limit, formulae will be derived for the density of the eigenvalues of $A$ in the complex plane, which generalize the well-known "circular law" for fully random matrices (Girko, 1984). Moreover, a correct regularizing procedure will be described for finding the support of the eigenvalue density in the limit $N \to \infty$ in certain highly nonnormal cases of $M$; the naive interpretation of the formulae fails in these cases, which are not yet discussed in the literature.

The results will be summarized in the next section, the derivations will be presented in Sec. 5.3, while Sec. 5.4 will be devoted to explicit results of analytical and numerical
calculations based for some specific examples of $M$, $L$ and $R$.

## 5.2 General results on spectral properties

### 5.2.1 Preliminary definitions

We will study ensembles of large $N \times N$ random matrices of the form

$$ A = M + LJR, \quad (5.1) $$

where $M$, $L$ and $R$ are arbitrary ($M$) or arbitrary invertible ($L$ and $R$) deterministic matrices, and $J$ is a random matrix of i.i.d. elements with zero mean and variance $1/N$. We are interested in results for the limit of the eigenvalue density as $N \to \infty$; therefore, the matrices $M$, $L$, $R$ and $J$ must each be understood, more precisely, as an infinite sequence of matrices dependent on $N$.

Since $J$ and therefore $LJR$ have zero mean, $M$ is the ensemble average of $A$. The random fluctuations of $A$ around its average are given by the matrix $LJR$, which for general $L$ and/or $R$ has dependent and non-identically distributed elements, due to the possible mixing and non-uniform scaling of the rows (columns) of the i.i.d. $J$ by $L$ ($R$).

We will be mostly interested in the statistics of the eigenvalues of Eq. (5.1). But although the statistics of the eigenvalues and eigenvectors of $A$ are of interest in their own right, in the following we will also consider directly certain properties of the linear dynamical system given by Eq. (2.12). In studying this system, it will be generally assumed that Eq. (2.12) is asymptotically stable. This means that $M$, $L$, $R$ and the leak parameter $\gamma$ of Eq. (2.12)
must be chosen so that for any typical realization of $J$, no eigenvalue of $-\gamma I + M + LJR$ has a positive real part; this can normally be achieved, for example, by choosing a large enough $\gamma > 0$.

Universality theorems ensure that, for given $M$, $L$ and $R$, the obtained result for the eigenvalue density in the limit $N \to \infty$ will not depend on the exact choice of the distribution of the elements of $J$, beyond its first two moments. The universality of the eigenvalue density for general $M$, $L$ and $R$ was established in Tao et al. (2010), following earlier work on the universality of the circular law established and successively strengthened in Bai (1997), Tao and Vu (2008), Gotze and Tikhomirov (2010).

A corollary of these theorems is that we are at liberty to calculate the eigenvalue density in any ensemble of our choice; hence, this choice can be dictated by merely pragmatical considerations.

In the derivation given below, we will assume that the random $J$ belongs to the complex Ginibre ensemble (Ginibre, 1965), i.e. the distribution of the elements of $J$ is complex Gaussian. Results obtained for the Ginibre ensemble extend to any i.i.d. $J$ whose elements have the same first two moments, namely, zero mean and variance $1/N$. This includes, for instance, instances of $J$ with binary (spin-like) elements, or with heavy-tailed (e.g. log-normal) distribution of the entries.

The following notation will be adopted. The ”operator norm” of an operator, when mentioned without any further specification, means the maximum singular value of that operator; for any matrix $B$, its operator norm will be denoted by $\|B\|$. On the other hand,
the (normalized) Frobenius norm of an operator is defined as

\[ \|B\|_F^2 \equiv \frac{1}{N} \sum_{ij} |B_{ij}|^2 = \frac{1}{N} \text{Tr}(BB^\dagger); \]  

(5.2)

equivalently, the Frobenius norm \( \|B\|_F \) is the root mean square of the singular values of \( B \).

For general matrices, \( A \) and \( B \),

\[ \text{tr}(A) \equiv \frac{1}{N} \text{Tr}(A), \quad A^{-\dagger} \equiv (A^\dagger)^{-1}, \quad \frac{1}{A} \equiv A^{-1}, \quad \frac{A}{B} \equiv AB^{-1}, \]

and when adding a scalar to a matrix it is implied, as in the previous chapters, that the scalar is multiplied by the appropriate identity matrix. We will denote by \( \mathbf{1} \) the identity matrix in a number of dimensions to be deduced from context. Only when a special point is to be made we will indicate dimensionality, e.g. by writing \( \mathbf{1}_N \).

General nonholomorphic functions of complex variables will be called \( f(z) \) instead of \( f(z, \bar{z}) \) for the sake of simplicity; for a variable \( z = x + iy \), the Dirac delta function is defined as \( \delta^2(z) \equiv \delta(x)\delta(y) \), the Wirtinger derivatives as \( \partial_x \equiv \partial/\partial x + i\partial/\partial y \) and \( \partial_z \equiv \partial/\partial z = (\partial/\partial x - i\partial/\partial y)/2 \).

Standard notations will be used for the relative orders of functions in the large-\( N \) limit: namely, \( O(f(N)) \) is used when for large enough \( N \) the absolute value of that quantity is bounded above by a fixed positive multiple of \( |f(N)| \); \( \Theta(f(N)) \) when, for large enough \( N \), the absolute value of that quantity is bounded above \textit{and} below by a fixed positive multiple of \( |f(N)| \); and \( o(f(N)) \) when its ratio to \( |f(N)| \) vanishes for \( N \to \infty \).

The only conditions to be imposed on \( M, L \) and \( R \) are that \( \|M\|_F, \|L\|_F, \|R\|_F \),
\[ \|L^{-1}MR^{-1}\| \text{ and } \|(LR)^{-1}\| \text{ are bounded as } N \to \infty. \] The bound on \( \|LR\| \) will be used in Sections 5.3.3 and 5.3.4; the Frobenius norm conditions, on the other hand, are assumptions in the universality theorem of Tao et al. (2010), which will be relied upon as just mentioned.

Finally, it will be assumed that, for all \( z \in \mathbb{C} \), the distribution of the eigenvalues of \( M_z M_z^\dagger \) (where \( M_z \) is defined below in Eq. 5.5) tends to a limit distribution as \( N \to \infty \). This last condition makes precise the requirement that \( M, L \) and \( R \) are defined consistently as functions of \( N \), so that a limit spectral density for \( M + LR \) is meaningful; however, it does not impose any further limits on the growth of the eigenvalues of \( M_z M_z^\dagger \) with \( N \), beyond the norm bounds imposed above.

### 5.2.2 The shape of supports and the density profiles

The density of the eigenvalues of \( M + LR \) in the complex plane for a realization of \( J \) (which is known in probability theory as the ”empirical spectral distribution”) is defined by

\[
\rho_J(z) = \frac{1}{N} \sum_\alpha \delta^2(z - \lambda_\alpha),
\]

where \( \lambda_\alpha \) are the eigenvalues of \( M + LR \).

It is known (Tao et al., 2010) that \( \rho_J(z) \) is asymptotically self-averaging. This means that, defining \( \rho(z) \equiv \langle \rho_J(z) \rangle_J \) as the ensemble average of \( \rho_J(z) \), the function \( \rho_J(z) - \rho(z) \) converges to zero (in the distributional sense) with probability one as \( N \to \infty \). Thus for large enough \( N \), any typical realization of \( J \) yields an eigenvalue density \( \rho_J(z) \) that is arbitrarily close to \( \rho(z) \).
It will be proven in Sec. 5.3 that for large $N$, with certain cautions and excluding special cases described below (Eqs. (5.18)–(5.19) and preceding discussion), $\rho(z)$ is nonzero in the region of the complex plane satisfying

$$\text{tr}[(M_z M_z^\dagger)^{-1}] \geq 1$$  \hspace{1cm} (5.4)

where we defined

$$M_z \equiv L^{-1}(z - M)R^{-1}. \hspace{1cm} (5.5)$$

Using the definition Eq. (5.2), one may also express Eq. (5.4) as

$$\left\| R \frac{1}{z - M} L \right\|_F \geq 1. \hspace{1cm} (5.6)$$

Moreover, inside this region, the density profile $\rho(z)$ is given by

$$\rho(z) = \frac{1}{\pi} \frac{\partial}{\partial z} \text{tr} \left[ \frac{(RL)^{-1} M_z^\dagger}{M_z M_z^\dagger + g(z)^2} \right], \hspace{1cm} (5.7)$$

where $g(z)$ is a real, scalar function found by solving

$$\text{tr} \left[ \frac{1}{M_z M_z^\dagger + g^2} \right] = 1,$$  \hspace{1cm} (5.8)

for $g$ for each $z$.

It may be checked that these formulas correctly yield a well-known limit in the case $M = 0$, $L = 1$, and $R = \sigma 1$. We have then $M_z = z/\sigma$; hence, Eq. (5.4) yields $\sigma^2/|z|^2 \geq 1$.  

134
or \(|z| \leq \sigma\) for the support; as for the density profile, Eqs. (5.7)–(5.8) yield the uniform
distribution \(\rho(z) = 1/(\pi \sigma^2)\) within the circle. We thus recover correctly the so-called
"circular law” for fully random matrices\(^1\).

5.2.3 Formulation in terms of singular values

It is possible and illuminating to express Eqs. (5.6)–(5.8) exclusively in terms of the singular
values \(s_i(z)\) of the matrix \(M_z\). Vanishing singular values, if any, will be included among
the \(s_i(z)\) so that we will always have \(N\) of them.

Since the squared singular values of \(M_z\) are the eigenvalues of the Hermitian \(M_z M_z^\dagger\),
we can evaluate the trace in Eq. (5.8) in the eigenbasis of the latter matrix, and rewrite this
equation as

\[
\frac{1}{N} \sum_{i=1}^{N} \frac{1}{s_i(z)^2 + g^2} = 1. \tag{5.9}
\]

Similarly, Eq. (5.4) can be equivalently rewritten as

\[
\frac{1}{N} \sum_{i=1}^{N} s_i(z)^{-2} \geq 1. \tag{5.10}
\]

In Sec. 5.3.2, it will be shown that Eq. (5.7) can also be written in a form that makes it
explicit that the dependence of \(\rho(z)\) on \(M\), \(L\) and \(R\) is \textit{only} through the singular values of

\(^1\)The circular law was discussed in Girko (1984), Bai (1997), Tao and Vu (2008), and Gotze and
Tikhomirov (2010). Formulae (5.4)–(5.8) generalize both the circular law and some results obtained in the lan-
guage and methods of free probability theory (Emery et al., 2000) by Biane and Lehner (2001) for \(L \propto R \propto 1\)
(first appeared in Khoruzhenko, 1996) as well as by Feinberg and Zee (1997) (extended to finite \(N\) by Hikami
and Pnini, 1998) with a \textit{normal} choice of the mean matrix \(M\).
$M_z$ and their derivatives with respect to $z$ and $\bar{z}$; namely, we have

$$\rho(z) = \frac{1}{\pi} \partial_\bar{z} \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{\partial_z (s_i(z)^2)}{s_i(z)^2 + g(z)^2} \right].$$  \hspace{1cm} (5.11)

### 5.2.4 Purely random synaptic weight matrices

For the special case of $M = 0$ and general $L$ and $R$, these formulas can be simplified considerably. The spectrum is then isotropic around the origin in this case, i.e. $\rho(z)$ depends only on $r \equiv |z|$ and its support is a disk centered at the origin with radius

$$r_0 = \|RL\|_F = \left[ \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 \right]^{1/2},$$  \hspace{1cm} (5.12)

where $\sigma_i$ are the singular values of $RL$ (this follows from Eq. (5.10) by noting that for $M = 0$, the singular values of $M_z = z(RL)^{-1}$ are $s_i(z) = |z|/\sigma_i$).

Within this support the spectral density is given by

$$\rho(r) = -\frac{1}{2\pi r^2} \partial_r (g(r)^2),$$  \hspace{1cm} (5.13)

where $g(r)^2 > 0$ is found by solving

$$1 = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_i^2 r^2 + g(r)^2}.$$  \hspace{1cm} (5.14)

Integrating Eq. (5.13), we see that the proportion of eigenvalues lying a distance larger
than \( r \) from the origin is, in this case, given by

\[
\begin{align*}
  n_>(r) &= \begin{cases} 
  g(r)^2 & (r < r_0) \\
  0 & (r \geq r_0) 
\end{cases} 
\end{align*}
\]  
\( (5.15) \)

In Sec. 5.3.5 we will prove that the eigenvalue density, given by Eqs. (5.13)–(5.14), is always a decreasing function of \( r = |z| \), i.e. for \( r > 0 \) its derivative with respect to \( r \) is strictly negative, as long as the limit distribution of the \( \{\sigma_i\} \) as \( N \to \infty \) has nonzero variance (otherwise \( \rho(z) \) is given by the circular law with the radius of Eq. 5.12).

The values of spectral density at \( r = 0 \) and \( r = r_0 \) can be calculated explicitly for general \( L \) and \( R \):

\[
\begin{align*}
  \rho(r = 0) &= \frac{1}{\pi N} \left[ \frac{1}{N} \sum_{i=1}^{N} \sigma_i^{-2} \right]^{-1} \\
  \rho(r = r_0) &= \frac{1}{\pi} r_0^2 \left[ \frac{1}{N} \sum_{i=1}^{N} \sigma_i^4 \right]^{-1} \leq \rho(r = 0). 
\end{align*}
\]  
\( (5.16, 5.17) \)

### 5.2.5 Ordering of limits

As noted above, some crucial cautions apply in using the above formulae for the eigenvalue density and its boundary Eqs. (5.4)–(5.8), or equivalently Eqs. (5.9)–(5.11); and for \( M = 0 \), Eqs. (5.13)–(5.14).

We have written these formulas for finite \( N \) (assuming it is large); however, the non-crossing approximation used in deriving these formulas is only guaranteed to yield the correct result for the eigenvalue density in the limit, i.e. \( \lim_{N \to \infty} \rho(z) \) (as discussed in Sec.
5.3.3); finite-size corrections obtained from Eqs. (5.4)–(5.8) are not in general correct, and \(o(1)\) contributions to \(g(z)^2\) or \(\rho(z)\) obtained from Eqs. (5.8) and (5.7) should be discarded.

Furthermore, in general, the correct way of finding the support of \(\lim_{N \to \infty} \rho(z)\) using Eq. (5.4) is by setting the left side of the inequality (5.4) to \(\lim g^2 \to 0^+ \lim N \to \infty\) of the left side of Eq. (5.8), as shown in Sec. 5.3.3. However, in writing Eq. (5.4) we have simply set \(g = 0\) in Eq. (5.8), and thus implicitly taken the limit \(g^2 \to 0^+\) before the \(N \to \infty\) limit.

To correctly express the support, we must first define the function

\[
\mathcal{K}(g, z) \equiv \lim_{N \to \infty} \frac{1}{M_z M_z^I + g^2} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{1}{s_i(z)^2 + g^2} \right]
\]  

(5.18)

for fixed, strictly positive \(g\), which serves to regularize the denominators in Eq. (5.18) for \(s_i(z)\) which are zero or vanishing in the limit \(N \to \infty\).

The generally correct way of expressing Eq. (5.4) or Eq. (5.10) is then

\[
\mathcal{K}(0^+, z) \equiv \lim_{g \to 0^+} \mathcal{K}(g, z) \geq 1.
\]  

(5.19)

Let us denote the support of \(\lim_{N \to \infty} \rho(z)\), given by Eq. (5.19), by \(S_{0^+}\) and the region specified by the limit \(N \to \infty\) of Eq. (5.4) or Eq. (5.10) by \(S_0\). For many examples of \(M, L\) and \(R\), the limits \(N \to \infty\) and \(g \to 0^+\) commute everywhere and hence \(S_{0^+} = S_0\). However, if there are \(z\)’s at which some of the smallest \(s_i(z)\) are either zero or vanish in the limit \(N \to \infty\), the two limits may fail to commute, and the naive use of Eq. (5.4) can yield a region, \(S_0\), strictly larger than and containing \(S_{0^+}\), the correct support of \(\lim_{N \to \infty} \rho(z)\).
We can write compactly

$$\mathcal{K}(0^+, z) = \lim_{g \to 0^+} \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{s_i^2(z, N) + g^2}$$

(5.20)

where we have included explicitly the dependence of the singular values on $z$ and $N$. While the normalized sum in Eq. (5.20) will converge pointwise as $N \to \infty$, it is seen in the equation that uniform convergence can be impeded by the presence of asymptotically vanishing singular values.

In particular, at points $z$ in the complex plane such that a $\Theta(1)$ number of $s_i(z)$ are zero or $o(1)$, these singular values do not make a contribution to $\mathcal{K}(g, z)$ for $g > 0$. Indeed, their contribution to the sum in Eq. (5.18) is $O(N^{-1})$, hence they make no contribution to $\mathcal{K}(0^+, z)$. But if they vanish sufficiently fast as $N \to \infty$ they can make a nonzero contribution to the left side of Eq. (5.10). Such points $z$ may fall within $S_0$, but not within $S_0^+$.

For finite $N$, the $s_i(z)$ can vanish exactly when $z$ coincides with an eigenvalue of $M$; thus the above situation can, e.g., arise close to eigenvalues of $M$ that are isolated and far from the rest of $M$’s spectrum, so that they fall outside the support of $\lim_{N \to \infty} \rho(z)$. In such cases, the spectrum of $M + LJR$ will nonetheless typically also contain isolated eigenvalues (which do not contribute to $\lim_{N \to \infty} \rho(z)$) with effectively deterministic location, i.e. within $o(1)$ distance of corresponding isolated eigenvalues of $M$. Examples of this phenomenon, for which $S_0 - S_0^+$ is not empty but has zero measure, have been studied in Tao (2013), and O’Rourke and Renfrew (2014). For symmetric matrices, outlier eigenvalues corresponding to eigenvalues of the mean matrix were first studied in Edwards and Jones (1976).
For some choices of $M$, $L$ and $R$, a more interesting case can arise such that for $z$ in a certain region of the complex plane with nonzero measure, all $s_i(z)$ are nonzero at finite $N$ (hence $M$ has no eigenvalue there), but a few $s_i(z)$ are $o(1)$ and vanish sufficiently fast as $N \to \infty$; in particular when $L \propto R \propto I$, this can occur for certain highly nonnormal $M$. In such cases the non-commutation of the two limits can lead to a difference $S_0 - S_0^+$ with nonzero measure.

In the cases we will examine, this signifies that there exists a finite, non-vanishing region outside the support of $\lim_{N \to \infty} \rho(z)$ (typically surrounding it) where, although $\rho(z)$ is $o(1)$, it nonetheless converges to zero sufficiently slowly that a $\Theta(1)$ number of”outlier” eigenvalues lie there. Note that, nonetheless, the vast majority of eigenvalues, i.e. $(1 - o(1))N$ of them, lie within the support of the limit density.

We will discuss examples of this phenomenon in Sec. 5.4 below. In one of the examples (discussed in Sec. 5.4.3), the existence of such outlier eigenvalues was first noted in Rajan and Abbott (2006); a remarkable paper by Terence Tao has later characterized their distribution (Tao, 2013). However, the connection between such outlier eigenvalues and nonzero but $o(1)$ singular values of $M_z$ that arises, e.g., for highly nonnormal $M$, has not been studied (or noticed) anywhere before.

---

2The designation “highly nonnormal” can be motivated, when $L$ and $R$ are proportional to the identity matrix, as follows. Let us denote by $\Sigma_\epsilon(M)$ the (operator-norm based) $\epsilon$-pseudospectrum of $M$, i.e. the region of $z$'s over which $\|(z - M)^{-1}\| > \epsilon^{-1}$. For fixed $N$, the true spectrum of $M$, which we denote by $\Sigma(M)$, is the set of points over which the smallest singular value of $(z - M)$ is exactly zero and hence $\|(z - M)^{-1}\| = \infty$. For finite $N$, $\lim_{\epsilon \to 0^+} \Sigma_\epsilon(M) = \Sigma(M)$ for any $M$. However, for nonnormal $M$ this approach could be much slower than in the normal case (see the discussion in Sec. 5.2.6). Now suppose that, as in the atypical cases under discussion, in a finite region of the complex plane the smallest singular value of $M_z$ is nonzero for finite $N$, but vanishes in the limit $N \to \infty$. This means that the operator norm of $(z - M)^{-1}$ is finite over such a region but goes to infinity as $N \to \infty$. Hence, if we define $\Sigma_\epsilon(M) = \lim_{N \to \infty} \Sigma_\epsilon(M)$ and $\Sigma(M) \equiv \lim_{N \to \infty} \Sigma(M)$, we see that in such cases $\lim_{\epsilon \to 0^+} \Sigma_\epsilon(M) \neq \Sigma(M)$ (or equivalently, $\lim_{\epsilon \to 0^+} \lim_{N \to \infty} \Sigma(M) \neq \lim_{N \to \infty} \lim_{\epsilon \to 0^+} \Sigma(M)$). More generally, this indicates that at finite but large $N$, the $\epsilon$-pseudospectra of such matrices can cover a significantly broader region than the spectrum even for very small $\epsilon$, indicating extreme nonnormality.
Some additional findings obtained through numerical simulations are the following:

(1) the distribution of these outliers remains indeed random as $N \to \infty$;

(2) its average behavior, in many cases, is not correctly given by the non-crossing approximation;

(3) the distribution of these outliers is in general less universal than $\lim_{N \to \infty} \rho(z)$; e.g. it can depend on the choice of real vs. complex ensembles for $J$.

### 5.2.6 Relationship to pseudospectra

We will discuss here a remarkable connection between the result (5.4) for the support of the stochastic spectrum and the notion of pseudospectra. This discussion applies to the case in which the limits $g^2 \to 0^+$ and $N \to \infty$ commute, so that Eq. (5.4) correctly describes the support.

Pseudospectra are generalizations of eigenvalue spectra that are particularly useful in the case of nonnormal matrices. The eigenvalue spectrum of matrix $M$ can be thought of as the set of points, $z$, in the complex plane where $(z - M)^{-1}$ is singular, i.e. it has infinite norm. Given a fixed choice of matrix norm, $\| \cdot \|$, the pseudospectrum of $M$ at level $\sigma$, or its “$\sigma$-pseudospectrum” in the given norm, is the set of points $z$ for which $\|(z - M)^{-1}\| \geq \sigma^{-1}$ (thus as $\sigma \to 0$ we recover the spectrum).

An equivalent definition may also be given. For the choice of the operator norm made in Sec. 5.2.1 (i.e. when $\|A\|$ is taken to be the maximum singular value of $A$), the $\sigma$-pseudospectrum can be equally characterized as the set of points, $z$, for which there exists a matrix perturbation $\Delta M$, with $\|\Delta M\| \leq \sigma$, such that $z$ is in the eigenvalue spectrum of
\( M + \Delta M \). The equivalence between these two definitions is also true for any matrix norm derived from a general vector norm (see Trefethen and Embree, 2005, for a proof). We can state that, in the operator norm, the \( \sigma \)-pseudospectrum of \( M \) is the set to which its spectrum can be perturbed by adding to it arbitrary perturbations of size \( \sigma \) or smaller.

In the present context, we may think of \( LJR \) as a perturbation of \( M \). Let us focus for simplicity on the case where \( L \) and \( R \) are proportional to the identity, i.e., we have \( \Delta M = \sigma J \), with a positive scalar \( \sigma \). The result Eq. (5.6) reads, in this case, as \( \| \sigma(z - M)^{-1} \|_F \geq 1 \) or \( \| (z - M)^{-1} \|_F \geq \sigma^{-1} \). In other words, as \( N \to \infty \), the spectrum of \( M + \sigma J \), for an i.i.d. random \( J \) with zero mean and variance \( 1/N \), is exactly the \( \sigma \)-pseudospectrum of \( M \) in the normalized Frobenius norm defined by Eq. (5.2).

Interestingly, the perturbation, \( \Delta M = \sigma J \), has normalized Frobenius norm \( \sigma \) as \( N \to \infty \); indeed, this norm is \( \sigma \sqrt{\sum_{ij} J_{ij}^2 / N} \), which, by the law of large numbers, converges to \( \sigma \) for large \( N \). As \( N \to \infty \), the spectrum in response to the random perturbation \( \sigma J \), which has size \( \sigma \) (in normalized Frobenius norm), is the \( \sigma \)-pseudospectrum of \( M \) in the normalized Frobenius norm.

This result sounds similar to the equivalence of the two definitions of pseudospectra for the operator norm which we noted above (one based on the norm of \((z - M)^{-1}\), and one based on the spectra of bounded perturbations), but it differs at least for two key reasons:

1. The equivalence between the two definitions of pseudospectra is true for the operator norm (maximal singular value) but \textit{not} for the normalized Frobenius norm (root mean square of singular values) that appears in Eq. (5.6).

2. For the operator norm, it is \textit{not} in general true that the \( \sigma \)-pseudospectrum of \( M \) is equivalent to the spectrum obtained from a single stochastic perturbation of \( M \) of size \( \sigma \).
even in the limit $N \to \infty$. (That is not true although, for numerical purposes, the spectra arising from such random perturbations are sometimes used as a “poor man’s version” or approximation of pseudospectra; see Trefethen and Embree, 2005).

The latter statement can be proven as follows. The operator norm of the random i.i.d. perturbation, $\sigma J$ (i.e. its maximum singular value) converges almost surely to $2\sigma$ as $N \to \infty$ (Yin, 1988). Condition (5.6) for $z$ to be in the spectrum under this random perturbation is $\|(z - M)^{-1}\|_{F} \geq \sigma^{-1}$, or $\text{rms}({s_i(z)^{-1}}) \geq 1$ where the $s_i(z)$ are the singular values of $z^{-M} \sigma$ and $\text{rms}({x_i})$ represents the root-mean-square of the set of values $\{x_i\}$. This is not equivalent to the condition that $z$ be in the $2\sigma$-pseudospectrum of $M$ in the operator norm, i.e. that $\|(z - M)^{-1}\| \geq (2\sigma)^{-1}$ or $s_{\text{min}}(z)^{-1} \geq \frac{1}{2}$, where $s_{\text{min}}(z)$ is the minimum of the $s_i(z)$; in fact, noting that $s_{\text{min}}(z)^{-1} \geq \text{rms}({s_i(z)^{-1}})$, it is easy to see that the spectrum under random i.i.d. perturbations with operator norm $\|\sigma J\| = 2\sigma$ is strictly a proper subset of the $2\sigma$-pseudospectrum in the operator norm.

For example, for $M = 0$, the “poor man’s $2\sigma$-pseudospectrum” in the limit $N \to \infty$ is a ball of radius $\sigma$ about the origin (the circular law), while the true $2\sigma$-pseudospectra of the zero matrix is the ball of radius $2\sigma$ about the origin.

In sum, in the operator norm, the $\sigma$-pseudospectrum of $M$ for any $N$ is equivalent to the set of points $z$ for which some perturbation $\Delta M$ with $\|\Delta M\| \leq \sigma$ can be found such that $z$ is in the spectrum of $M + \Delta M$ (Trefethen and Embree, 2005). In the normalized Frobenius norm in the limit $N \to \infty$, however, the $\sigma$-pseudospectrum of $M$ is equivalent to the spectrum of $M + \Delta M$ where $\Delta M$ is any random perturbation with zero-mean i.i.d. elements with $\|\Delta M\|_F = \sigma$. This statement for the normalized Frobenius norm holds when the two limits $N \to \infty$ and $\gamma \to 0^+$ commute; when the two limits do not commute, the
support of the spectral distribution of \( M + \Delta M \) is a subset of the \( \sigma \)-pseudospectrum of \( M \) in the normalized Frobenius norm.

### 5.3 Calculation of spectral properties

#### 5.3.1 The shape of supports and density profiles

In this section we derive formulae Eqs. (5.4)–(5.7) for the average spectral density \( \rho(z) \) of random matrices of the form \( A = M + LJR \) (where \( M, L \) and \( R \) are deterministic matrices and \( J \) is random with i.i.d. elements of zero mean and variance \( 1/N \)).

As mentioned in Sec. 5.2, the spectral density is self-averaging for large \( N \). Furthermore, as established in Tao and Krishnapur (2010), it is also universal in the large \( N \) limit, in the sense that it is independent of the details of the distribution of the elements of \( J \) as long as its mean and variance are as stated. The same universality theorem also ensures that the real or complex nature of \( J \) does not by itself affect \( \rho(z) \) to leading order. That enables us to carry out calculations in any convenient ensemble.

We will consider therefore the computationally handy case where \( J \) is a zero-mean complex Gaussian random matrix with \( \langle J_{ab}J_{cd} \rangle = 0 \), and

\[
\langle J_{ab}J_{cd}^* \rangle = \frac{1}{N} \delta_{ac} \delta_{bd}.
\]  

(5.21)

Thus \( \langle |J_{ab}|^2 \rangle = \frac{1}{N} \), and all other first and second moments of \( J \) (including \( \langle J_{ab}^2 \rangle \)) vanish.
The measure on $J$ can be written as

$$d\mu(J) \propto e^{-\mathcal{N} \text{Tr}(J J^\dagger)} \prod_{ab} d\text{Im} J_{ab} d\text{Re} J_{ab}. \quad (5.22)$$

In this form, and by the invariance of the trace, it is clear that the measure is symmetric with respect to the group $U(N) \otimes U(N)$, acting on $J$ by $J \mapsto U J V^\dagger$ where $U$ and $V$ are arbitrary $N \times N$ unitary matrices.

For a particular realization of $J$, we define the Green’s function $G(z; J)$ by

$$G(z; J) \equiv \frac{1}{M_z - J}, \quad (5.23)$$

where $M_z = L^{-1}(z - M)R^{-1}$ (Eq. 5.5). In the case $L, R \propto 1$, $G(z; J)$ will be proportional to the resolvent of $A$, $\frac{1}{z - A}$. More generally we have

$$\frac{1}{z - A} = R^{-1}G(z; J)L^{-1}. \quad (5.24)$$

We will also make use of the identity

$$\delta^2(z) = \frac{1}{\pi} \partial_z \partial_{\bar{z}} \ln |z|^2 = \frac{1}{\pi} \partial_z \left( \frac{1}{z} \right); \quad (5.25)$$

the first equal sign in this equation follows by noting that $4\partial_z \partial_{\bar{z}} = \nabla^2$, where $\nabla^2$ is the 2-D Laplacian and recalling from electrostatics (Jackson, 1998) that the solution of Poisson’s equation for a point charge at origin, i.e. $\nabla^2 \phi(z) = 4\pi \delta^2(z)$, in 2-D is given by the potential field $\phi(z) = \ln |z|^2$. The second identity in Eq. (5.25) follows from $\partial_z \ln |z|^2 = \partial_z (\ln z + \cdots)$. 

145
\[ \ln(z) = \frac{1}{z} + 0. \]

Using Eq. (5.25), we can write the empirical spectral density, defined in Eq. (5.3), as

\[ \rho_J(z) = \frac{1}{\pi} \frac{1}{N} \sum_{\alpha} \frac{1}{z - \lambda_\alpha} = \frac{1}{\pi} \partial_z \text{ tr} \frac{1}{z - A}. \quad (5.26) \]

Performing the ensemble average we obtain

\[ \rho(z) \equiv \langle \rho_J(z) \rangle_J = \frac{1}{\pi} \partial_z \text{ tr}[(RL)^{-1} \langle G(z; J) \rangle_J], \quad (5.27) \]

where we used Eq. (5.24), and the linearity and cyclicity of the trace. Thus, to calculate \( \rho(z) \), the task boils down to calculating \( \langle G(z; J) \rangle_J \).

The diagrammatic technique provides a method for calculating averages of products of \( G(z; J) \)'s. However, this method in its standard form relies on \( A \) being a Hermitian matrix. Indeed, it begins with an expansion of \( G(z; J) \) in powers of \( J \), which is only valid when \( z \) is far enough from the spectrum of \( A \), i.e. away from the points we are most interested in. For Hermitian matrices, this is no problem as the spectrum is confined to the real line, and therefore \( G(z; J) \) and \( \langle G(z; J) \rangle_J \) will be analytic outside the real line. Thus, one can start from the expansion for \( z \) far away outside the real line, perform the averaging over \( J \), and sum up the most dominant contributions to obtain a result analytic in \( z \); this result can then be analytically continued to \( z \) arbitrarily close to the spectrum on the real line, yielding information about the spectrum.

All this would seemingly fail in the case of a nonnormal (and in particular non-Hermitian) \( A \), with eigenvalues that in general cover a two dimensional region in the com-
plex plane. Using a trick introduced by Feinberg and Zee (1997), however, we can turn this problem into the auxiliary problem of averaging the Green’s functions for a Hermitian matrix.

This is done at the computationally manageable cost of doubling all the degrees of freedom. One defines a $z$-dependent, $2N \times 2N$ Hermitian “Hamiltonian”

$$H(z) \equiv \begin{pmatrix} 0 & M_z - J \\ M_z^\dagger - J^\dagger & 0 \end{pmatrix},$$

whose corresponding resolvent or Green’s function (which will depend on a new complex variable $\eta$) is the $2N \times 2N$ matrix

$$G(\eta, z; J) \equiv (\eta - H(z))^{-1} = -\begin{pmatrix} \frac{\eta}{(M_z - J)(M_z - J)^\dagger - \eta^2} & \frac{M_z - J}{(M_z - J)^\dagger(M_z - J) - \eta^2} \\ \frac{(M_z - J)^\dagger}{(M_z - J)(M_z - J)^\dagger - \eta^2} & \frac{\eta}{(M_z - J)^\dagger(M_z - J) - \eta^2} \end{pmatrix}.$$ (5.29)

For $\eta \to i0$, we see that

$$G(0, z; J) = -\begin{pmatrix} 0 & (M_z - J)^\dagger \\ (M_z - J)^{-1} & 0 \end{pmatrix},$$

and thus from Eq. (5.23), for any realization of $J$

$$G(z; J) = -\lim_{\eta \to i0} G^{21}(\eta, z; J)$$ (5.31)
Here we have used the notation

\[
G(\eta, z; J) = \begin{pmatrix}
G^{11}(\eta, z; J) & G^{12}(\eta, z; J) \\
G^{21}(\eta, z; J) & G^{22}(\eta, z; J)
\end{pmatrix},
\]  

(5.32)

where \(G^{\alpha\beta}\) (with \(\alpha, \beta \in \{1, 2\}\)) are \(N \times N\) matrices, forming the four blocks of \(G\).

We have written the limit in Eq. (5.31) as \(\eta \to i0\) to emphasize that until the end of the calculations \(\eta\) is to retain a nonzero imaginary part, which serves to regularize the denominators in Eq. (5.29) (c.f. the discussion after Eq 5.54).

Since we will be carrying out a perturbation expansion in powers of \(J\), it is convenient to decompose the Hamiltonian according to

\[
H(z) = H_0(z) - \mathcal{J}, \quad H_0(z) \equiv \begin{pmatrix}
0 & M_z \\
M_z^\dagger & 0
\end{pmatrix}, \quad \mathcal{J} \equiv \begin{pmatrix}
0 & J \\
J^\dagger & 0
\end{pmatrix}.
\]

(5.33)

We will sometimes use a tensor product notation to denote matrices in this doubled-up space, e.g. writing \(\mathcal{J} = \sigma^+ \otimes J + \sigma^- \otimes J^\dagger\), where we defined the \(2 \times 2\) matrices

\[
\sigma^+ = \begin{pmatrix}
0 & 1 \\
0 & 0
\end{pmatrix}, \quad \sigma^- = \begin{pmatrix}
0 & 0 \\
1 & 0
\end{pmatrix}.
\]

(5.34)

By a slight abuse of notation we also denote \(2N \times 2N\) matrices \(\sigma^+ \otimes 1_{N \times N}\) by \(\sigma^+\) (and will denote the identity matrix in any space by \(1\)).
From Eqs. (5.31) we obtain

\[ \text{tr} [(RL)^{-1} G(z; J)] = -\text{tr} \left[ (\sigma^+ \otimes (RL)^{-1}) G(i0^+, z; J) \right], \quad (5.35) \]

and from Eq. (5.27)

\[
\rho(z) = - \lim_{\eta \to i0} \frac{1}{\pi} \partial_z \text{tr} \left( (\sigma^+ \otimes (RL)^{-1}) G(\eta, z) \right) = - \lim_{\eta \to i0} \frac{1}{\pi} \partial_z \text{tr} \left( (RL)^{-1} G^{21}(\eta, z) \right)
\]

(5.36)

where we defined

\[ G(\eta, z) \equiv \langle G(\eta, z; J) \rangle_J. \quad (5.37) \]

Having expressed \( \rho(z) \) in terms of the ensemble average of the Green’s function for a Hermitian matrix, we can now develop a diagrammatic method for calculating ensemble averages of products of \( G(\eta, z; J) \) (including \( G(\eta, z) \)). Being the Green’s function of a Hermitian matrix, \( G(\eta, z; J) \) and hence \( G(\eta, z) = \langle G(\eta, z; J) \rangle_J \) are analytic functions of \( \eta \) for \( \eta \) outside the real line, and therefore analytic continuation can be used to take the limit \( \eta \to i0 \) after obtaining the average over \( J \) for \( \eta \) sufficiently away from the real line.

We will denote the elements of a generic \( 2N \times 2N \) matrix \( A \) by \( A_{ab}^{\alpha\beta} \), where the Greek indices range in \{1, 2\} and the Latin indices range in \{1, \ldots, N\}. Using this notation, the definition (5.33), and Eq. (5.21), we can write the covariance for the components of \( J \) as

\[
\left\langle J_{ab}^{\alpha\beta} J_{cd}^{\gamma\delta} \right\rangle_J = \frac{1}{N} \delta_{ad}\delta_{bc} \left( \sigma_{\alpha\beta}^{+}\sigma_{\gamma\delta}^{-} + \sigma_{\alpha\beta}^{-}\sigma_{\gamma\delta}^{+} \right) \quad (5.38)
\]

The terms proportional to \( \sigma^+ \sigma^+ \) and \( \sigma^- \sigma^- \) involve \( \langle J_{ab} J_{cd} \rangle \), or its complex conjugate,
which vanish for the complex Gaussian ensemble.

The various indices on the RHS of Eq. (5.38) can be decoupled by rewriting the expression in parentheses as $\pi^{1}_{\alpha\delta}\pi^{2}_{\gamma\beta} + \pi^{2}_{\alpha\delta}\pi^{1}_{\gamma\beta}$, where

\[
\pi^{1} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \pi^{2} \equiv \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},
\]

yielding

\[
\langle J_{ab} J_{cd} \rangle = \frac{1}{N} \sum_{r=1}^{2} (\pi^{r}_{\alpha\delta} \delta_{ad})(\pi^{3-r}_{\gamma\beta} \delta_{cb}).
\]

Also, since $J_{ab}$ have zero mean, we have $\langle J \rangle = 0$.

The starting point of the diagrammatic method is the perturbation expansion of

\[
G(\eta, z; J) = (\eta - H_{0}(z) - \mathcal{J})^{-1}
\]

in powers of $\mathcal{J}$

\[
G(\eta, z; J) = G(\eta, z; 0) \sum_{n=0}^{\infty} [\mathcal{J}G(\eta, z; 0)]^{n}
\]

where $G(\eta, z; 0)$ is given by Eq. (5.29) with the $J$’s set to zero. This equation is represented diagrammatically in the third line of Fig. 5.1; the thin arrows defined in the first line of the figure represents $G(\eta, z; 0)$, and the dashed lines represent a power of $\mathcal{J}$ before ensemble averaging.

To obtain the average resolvent, $\mathcal{G}(\eta, z)$, we average Eq. (5.41) term by term with respect to the ensemble Eq. (5.22). Since the measure is Gaussian with zero mean, according to Wick’s theorem the average of each term of Eq. (5.41) involving $n$ factors of $\mathcal{J}$ is given by a sum over the contributions of all possible complete pairings of the $\mathcal{J}$’s in that term (in
Figure 5.1: The first two lines define different elements of Feynman diagrams: the Green’s function for $J = 0$ (zero disorder), $G^\alpha_\beta(\eta, z; 0)$, the covariance of two $J$ elements, the ensemble averaged Green’s function, $G(\eta, z) \equiv \langle G(\eta, z; J) \rangle_J$, and the self-energy $\Sigma(\eta, z)$. Eq. (5.43) (the matrix indices for $G(\eta, z)$ and $\Sigma(\eta, z)$ are arranged as for $G^\alpha_\beta(\eta, z; 0)$).

The third line is the diagrammatic representation of the expansion Eq. (5.41) of $G(\eta, z; J)$ before averaging over $J$, where the $J$’s are represented by dashed lines. Averaging over Eq. (5.22) is performed by pairing all $J$’s and connecting them with the wavy lines representing $\langle J J \rangle$. In the large $N$ limit, the contribution of crossing pairings is suppressed by negative powers of $N$; the sum of all non-crossing diagrams, shown on the fourth line, yields the leading contribution to $G(\eta, z)$ for large $N$. The last line shows the diagrammatic representation of Eq. (5.42), which if iterated generates all the non-crossing diagrams. Alternatively, $G(\eta, z)$ can be found by solving this self-consistent equation directly.

In particular, since $\langle J \rangle_J = 0$, terms in Eq. (5.41) with odd powers of $J$ vanish after averaging). Each pairing can be represented as a Feynman diagram, as shown in Fig. 5.1, the first two lines of which define the diagram elements.

For example, the last diagram in the fourth line of Fig. 5.1 shows one possible pairing of the term in Eq. (5.41) corresponding to $n = 6$. The contribution of each pairing diagram is given by a product of factors, one per each pair, given by Eq. (5.40) (represented by wavy lines) with the right indices for that pair, as well as the factors of $G(\eta, z; 0)$ (represented...
by thin arrows), with all the intervening Greek and Latin matrix indices summed over their proper ranges.

We will show in Section 5.3.3 that for $\text{Im } \eta \neq 0$, and so long as $\|(RL)^{-1}\|$ remains bounded as $N \to \infty$, only non-crossing pairings need to be retained in the large $N$ limit, as crossing pairings are suppressed by inverse powers of $N$ and do not contribute in the limit. A pairing diagram is non-crossing if it can be drawn on a plane with its wavy lines drawn only on the half-plane above the straight arrow line, without any wavy lines crossing.

As the last two lines of Fig. 5.1 demonstrate, all non-crossing diagrams can be generated by iterating the equation

$$
\mathcal{G}(\eta, z) = \mathcal{G}(\eta, z; 0) + \mathcal{G}(\eta, z; 0) \Sigma(\eta, z) \mathcal{G}(\eta, z),
$$

starting from $\mathcal{G}(0, z) = \mathcal{G}(\eta, z; 0)$. This equation is represented diagrammatically in the last line of Fig. 5.1, with the ”self-energy” matrix, $\Sigma(\eta, z)$, defined by the diagram in the second line of that figure, i.e.

$$
\Sigma(\eta, z) \equiv \langle \mathcal{J} \mathcal{G}(\eta, z) \mathcal{J} \rangle.
$$

Using Eq. (5.40) we obtain

$$
\Sigma_{\alpha \delta}^{c} (\eta, z) = \delta_{\alpha \delta} \sum_{r=1}^{2} \pi^{r} \frac{1}{N} \text{Tr} (\pi^{3-r} \mathcal{G}(\eta, z)),
$$

starting from $\mathcal{G}(0, z) = \mathcal{G}(\eta, z; 0)$.
which using Eq. (5.39) we can write as

\[
\Sigma(\eta, z) = \begin{pmatrix}
-i g_2(\eta, z) & 0 \\
0 & -i g_1(\eta, z)
\end{pmatrix},
\] (5.45)

where we have defined the scalar functions

\[
g_\alpha(\eta, z) \equiv i \text{ tr} \mathbb{G}^{\alpha}(\eta, z).
\] (5.46)

Using Eq. (5.45), we can solve Eqs. (5.42)–(5.45) for \( \mathbb{G}(\eta, z) \) at once, in terms of \( g_\alpha(\eta, z) \), and then use Eq. (5.46) to obtain a self-consistency equation, which can be solved for \( g_\alpha(\eta, z) \). To this end, we multiply Eq. (5.42) by \( \mathbb{G}^{-1}(\eta, z; 0) \) on the left, and by \( \mathbb{G}^{-1}(\eta, z) \) on the right, to obtain

\[
\mathbb{G}(\eta, z) = \left[ \mathbb{G}^{-1}(\eta, z; 0) - \Sigma(\eta, z) \right]^{-1} = [\eta - H_0(z) - \Sigma(\eta, z)]^{-1}.
\] (5.47)

Using this expression with Eqs. (5.33) and (5.45), it can be easily checked that

\[
\mathbb{G}(\eta, z) = -\begin{pmatrix}
(\bar{\eta} + ig_1)K_1^{-1} & (z - M)K_2^{-1} \\
(\bar{\bar{\eta}} - M^\dagger)K_1^{-1} & (\bar{\eta} + ig_2)K_2^{-1}
\end{pmatrix},
\] (5.48)

where \( K_1 \equiv M_zM_z^\dagger + (g_1 - i\eta)(g_2 - i\eta) \) and \( K_2 \equiv M_z^\dagger M_z + (g_1 - i\eta)(g_2 - i\eta) \), and we dropped the arguments of \( g_\alpha(\eta, z) \) for succinctness. Imposing Eq. (5.46), one obtains the
self-consistency equations

\begin{align*}
    g_1 &= (g_1 - i\eta) \text{tr}(K_1^{-1}), \\
    g_2 &= (g_2 - i\eta) \text{tr}(K_2^{-1}).
\end{align*}

(5.49)  

(5.50)

Before solving these equations for \( g_1 \) and \( g_2 \), one can easily show that \( \text{tr}(K_1^{-1}) = \text{tr}(K_2^{-1}) \). One way to see this is to use the singular value decomposition (SVD) of \( M_z \) in the form

\[ M_z = U_z S_z V_z^\dagger, \]

(5.51)

where \( S_z \) is a nonnegative diagonal matrix with the singular values of \( M_z \), \( s_i(z) (i = 1, \cdots, N) \), on the diagonal, and \( U_z \) and \( V_z \) are unitary matrices. Just as in Sec. 5.2.3, we are including possibly vanishing singular values among \( s_i(z) \), hence \( S_z, U_z \) and \( V_z \) are always \( N \times N \) matrices.

Using the invariance of trace under similarity transforms, we obtain \( \text{tr}(K_1^{-1}) = \text{tr}(K_2^{-1}) = \text{tr}(S_z^2 + (g_1 - i\eta)(g_2 - i\eta))^{-1} \). From this equality, it is not hard to see that Eqs. (5.49-5.50) cannot be simultaneously satisfied unless \( g_1(\eta, z) = g_2(\eta, z) \equiv g(\eta, z) \), with \( g(\eta, z) \) satisfying

\[ g = (g - i\eta) \text{tr}\left[ \frac{1}{S_z^2 + (g - i\eta)^2} \right], \]

(5.52)

or as written in the original basis

\[ g = (g - i\eta)\text{tr}\left[ \frac{1}{M_z M_z^\dagger + (g - i\eta)^2} \right]. \]

(5.53)
Noting from Eqs. (5.45) that the self-energy is thus proportional to the $2N \times 2N$ identity matrix, from Eqs. (5.47) and (5.29) (for $J = 0$) we obtain

$$G(\eta, z) = G(\eta + ig(\eta, z), z; 0) = -\begin{pmatrix} \frac{i\kappa}{M_z M_z^\dagger + \kappa^2} & \frac{M_z}{M_z^\dagger M_z + \kappa^2} \\ \frac{M_z^\dagger}{M_z M_z^\dagger + \kappa^2} & \frac{i\kappa}{M_z^\dagger M_z + \kappa^2} \end{pmatrix}$$  \quad (5.54)

where $\kappa \equiv g(\eta, z) - i\eta$.

According to Eq. (5.31), for the case of present interest one must solve Eq. (5.53) in the limit $\eta \to i0$. Note, however, that as shown in Sec. 5.3.3, the non-crossing approximation is in general guaranteed to work only for $\text{Im} \eta \neq 0$; hence the limit $\eta \to i0$ must be taken after the limit $N \to \infty$ (as already pointed out in Sec. 5.2, taking the limits in this order is important in cases where some of the singular values in $S_z$ vanish in the limit $N \to \infty$).

For the present purposes, it will suffice to let $\eta = i\epsilon$ for some real positive $\epsilon$ and take the limit $\epsilon \to 0^+$ at the end. In this case one must seek a positive solution for $g(i\epsilon, z)$ in Eq. (5.53); this is because by definition, $g(\eta, z) = \text{tr} G_{11}(\eta, z) = \langle \text{tr} iG_{11}(\eta, z; J) \rangle_J$ and from Eq. (5.29) one obtains $g(i\epsilon, z) = \langle \text{tr} \frac{\epsilon}{(M_z - J)(M_z - J) + \epsilon^2} \rangle_J$, which for $\epsilon > 0$, is the ensemble average of the trace of a positive definite matrix and hence positive.

Taking the limit $N \to \infty$ while keeping $\epsilon$ (and hence $\epsilon + g$) positive and nonzero, we will define

$$\mathcal{K}(\kappa, z) \equiv \lim_{N \to \infty} \text{tr} \left[ \frac{1}{M_z M_z^\dagger + \kappa^2} \right] = \lim_{N \to \infty} \text{tr} \left[ \frac{1}{S_z^2 + \kappa^2} \right]$$  \quad (5.55)

for $\kappa = g + \epsilon > 0$. We can then rewrite Eq. (5.53) in terms of $\kappa$ rather than $g$:

$$\kappa [1 - \mathcal{K}(\kappa, z)] = \epsilon.$$  \quad (5.56)
Since $\epsilon$ and $\kappa = g + \epsilon$ are positive, it follows that $1 - \mathcal{K}(\kappa, z)$ must also be positive. In the limit $\epsilon \to 0^+$, two possible situations can occur:

1. It may be that $g, \kappa \to 0^+$, in which case we must have either

$$\lim_{\kappa \to 0^+} \mathcal{K}(\kappa, z) < 1, \quad (5.57)$$

or $\lim_{\kappa \to 0^+} \mathcal{K}(\kappa, z) = 1$;

2. Alternatively, the solution for $g$ stays finite and positive in the limit while $\mathcal{K}(\kappa, z) \to 1^-$ as $\kappa \to g^+$. In this case, $g(z) \equiv \lim_{\epsilon \to 0^+} g(\epsilon, z)$ must satisfy $\mathcal{K}(g(z), z) = 1$, i.e.

$$1 = \lim_{N \to \infty} \text{tr} \left[ \frac{1}{S^2_z + g(z)^2} \right]. \quad (5.58)$$

Note further that, since $\mathcal{K}(\kappa, z)$ is a decreasing function of $\kappa$, in the second case we have $\mathcal{K}(0^+, z) \geq \mathcal{K}(g(z), z) = 1$, i.e.

$$\lim_{\kappa \to 0^+} \mathcal{K}(\kappa, z) \geq 1. \quad (5.59)$$

Thus the two possible solutions are realized in complementary regions (with a shared boundary) of the complex $z$-plane, respectively given by Eqs. (5.57) and (5.59).

Let us substitute the $g(z) = 0$ solution for the case (5.57) in Eq. (5.54), and naively set $\eta = i\epsilon$ (and thus $\kappa$) to zero, to obtain

$$\mathcal{G}(\eta = i0^+, z) = G(\eta = i0^+, z; 0). \quad (5.60)$$
From Eqs. (5.30)–(5.31), this solution yields

\[
\langle G(z; J) \rangle_J = -G^{21}(\eta = i0^+, z) = M_z^{-1} = R(z - M)^{-1} L
\] (5.61)

which is analytic outside the spectrum of \( M \). Hence from Eq. (5.27), it yields \( \rho(z) = 0 \), at least outside the spectrum of \( M \). The more careful analysis presented in Sec. 5.3.4, in which we correctly take the limit \( N \to \infty \) in Eq. (5.36) before taking \( \epsilon \to 0^+ \), will confirm that in the region Eq. (5.57), \( \lim_{N \to \infty} \rho(z) \) always vanishes.

We conclude that the support of \( \lim_{N \to \infty} \rho(z) \) is where Eq. (5.59) holds (which is Eq. (5.19) of Sec. 5.2); here \( g(z) \) is to be found by solving Eq. (5.58), or equivalently Eq. (5.8) or Eq. (5.9). In this region, we obtain \( \rho(z) \) by substituting Eq. (5.54), with the solution of Eq. (5.58), into Eqs. (5.36). This yields Eq. (5.7), which we rewrite here as

\[
\begin{align*}
\rho(z) &= \frac{1}{\pi} \partial_z E(z), \\
E(z) &= \text{tr} \left[ \frac{(RL)^{-1}M_z^\dagger}{M_z^2 M_z^\dagger + g(z)^2} \right],
\end{align*}
\] (5.62) (5.63)

with \( g(z) \) given by Eq. (5.58), or equivalently Eq. (5.8).

### 5.3.2 Singular value formulation

We will obtain here an alternative expression for \( \rho(z) \) that is equivalent to Eqs. (5.62)–(5.63), and that explicitly shows how the spectral density depends only on the singular values of \( M_z \).
Noting that \( \partial_z \left( M_z M_\dagger_z \right) = (RL)^{-1} M_\dagger_z \), we can write Eq. (5.63) as

\[
E(z) = \text{tr} \left[ \frac{\partial_z \left( M_z M_\dagger_z \right)}{M_z M_\dagger_z + g(z)^2} \right];
\]  
(5.64)
on the other hand, we have

\[
\partial_z \text{tr} \ln \left[ M_z M_\dagger_z + g(z)^2 \right] = \text{tr} \left[ \frac{\partial_z \left( M_z M_\dagger_z + g^2(z) \right)}{M_z M_\dagger_z + g(z)^2} \right] = E(z) + \partial_z \left( g^2(z) \right),
\]  
(5.65)
where to write the last term we used Eq. (5.8).

Thus we obtain

\[
E(z) = \partial_z \varphi(z),
\]  
(5.66)
\[
\varphi(z) \equiv -g^2(z) + \text{tr} \ln \left[ M_z M_\dagger_z + g(z)^2 \right].
\]  
(5.67)
or using the SVD, Eq. (5.51),

\[
\varphi(z) = -g^2(z) + \text{tr} \ln \left[ S_z^2 + g(z)^2 \right] = -g(z)^2 + \frac{1}{N} \sum_{i=1}^{N} \ln \left[ s_i(z)^2 + g(z)^2 \right].
\]  
(5.68)
Finally, substituting Eq. (5.68) in Eq. (5.66), and using Eq. (5.9), one arrives at Eq. (5.11).

### 5.3.3 Validity of the non-crossing approximation

In this section we will give the justification for the non-crossing approximation used in Sec. (5.3.1). That is, we will show that the only diagrams not suppressed by inverse powers of \( N \) are the non-crossing diagrams.
Figure 5.2: The orbits (shown by thin red paths) for two diagrams for the spectral density in a complex $J$ ensemble. The non-crossing diagram on top has three orbits: orbit (1) is the external orbit connecting the two ends of the Green’s function, while orbits (2) and (3) are the internal orbits. As in Eqs. (5.69) and (5.70), they contribute $\text{Tr} (\sigma^+ G(\eta, z; 0) \pi^r G(\eta, z; 0))$, $\text{Tr} (\pi^{3-r_1} G(\eta, z; 0) \pi^{r_2} G(\eta, z; 0))$ and $\text{Tr} (\pi^{3-r_2} G(\eta, z; 0))$, respectively, with $r_1$ and $r_2$ summed over 1 and 2 (cf. Eq. 5.40). The trace contributed by each of the three orbits is $O(N)$, which when combined with the three factors of $1/N$ accounting for the two wavy lines and the normalization of the external orbit’s trace, yield an $O(1)$ expression for this diagram. By contrast, the crossing diagram on the right has no internal orbits. Its only external orbit contributes $\text{Tr}(\sigma^+ G(\eta, z; 0) \pi^{r_2} G(\eta, z; 0) \pi^{3-r_1} G(\eta, z; 0) \pi^{3-r_2} G(\eta, z; 0) \pi^{r_1} G(\eta, z; 0))$ which after the trace normalization is $O(1)$. Accounting for two factors of $1/N$ coming from the wavy lines, we then see that this crossing diagram is $O(N^{-2})$ and hence is suppressed as $N \to \infty$.

As explained after Eq. (5.41), averaging of $G(\eta, z; J)$ over the disorder $J$ involves summing over all complete pairings of the factors of $J$ in every term of the expansion Eq. (5.41), with each pairing of each term represented by a diagram as shown in Fig. 5.1. Each such diagram is composed of a solid directed line (each segment of which represents a factor of $G_{ab}^{\alpha\beta}(\eta, z; 0)$), with a number of wavy lines (each representing the expression Eq. (5.40), with different indices) connecting different points on the solid arrow line, and all the internal matrix indices summed over.

For the purpose of calculating the eigenvalue density, according to Eqs. (5.36)–(5.37), what we need to calculate is actually $\text{tr}(\sigma^+ (RL)^{-1} \langle G(\eta, z; J) \rangle_J)$; thus we can imagine the solid arrow making a loop by closing in on itself and sandwiching $\sigma^+ \otimes (RL)^{-1}$ (see Fig. 5.2).
Given the structure of the Kronecker deltas in Eq. (5.40), it is more convenient for the present purposes, however, to think of each diagram as consisting in a certain number of "orbits". Each orbit is formed as follows: we start somewhere on the solid line and we move upon it always along its arrow until the next wavy line is encountered; at such points we leave the solid line, and we continue on the wavy line without crossing it, because Eq. (5.40) is composed of two Kronecker deltas (one for each side of the wavy line) enforcing index identification at the corresponding ends on each side\(^3\); we will thus return somewhere else on the solid line, and we continue as before until we reach the initial point (see Fig. 5.2).

As we go around this orbit, for each solid line traversed we write down, from right to left, a \(G(\eta, z; 0)\) and for each wavy line a \(\pi^r_i\) (see Eqs. 5.39) where \(i\) is the index of the wavy line. Because all matrix indices are summed over, such adjacent factors multiply like matrices, and since the orbit forms a loop, in the end we obtain the trace of the matrix product thus obtained. Notice that this recipe for assigning the contribution of each orbit accounts for the Kronecker deltas and \(\pi^r_i\)'s in Eq. (5.40), but not for the factor \(\frac{1}{N}\) and the sum over \(r\)'s; we will account for these, at the end, after Eq. (5.70).

A generic orbit, which we will refer to as internal, closes on itself after traversing, say, \(m\) wavy lines sandwiching \(m\) Green's functions (e.g. the orbits labeled 2 and 3 in panel (a) of Fig. 5.2), and thus contributes a trace of the form

\[
I_{m,r} \equiv \text{Tr}(G(\eta, z; 0)\pi^{r_1} \cdots G(\eta, z; 0)\pi^{r_m}),
\]

\(^3\)This structure is a consequence of using a complex ensemble for \(J\), for which the covariances \(\langle J_{ab}J_{cd} \rangle\) vanishes. For the real Gaussian ensemble, by contrast, the latter do not vanish; in this case Eq. (5.40) becomes \(\langle J_{ab}^{\alpha\beta}J_{cd}^{\gamma\delta} \rangle = \frac{1}{N} \left[ \delta_{ad}\delta_{bc} \left(\sigma_{\alpha\beta}^{++}\sigma_{\gamma\delta}^{++} + \sigma_{\alpha\beta}^{+-}\sigma_{\gamma\delta}^{+-} \right) + \delta_{ac}\delta_{bd} \left(\sigma_{\alpha\beta}^{++}\sigma_{\gamma\delta}^{-+} + \sigma_{\alpha\beta}^{+-}\sigma_{\gamma\delta}^{+-} \right) \right] = \frac{1}{N} \sum_{r=1}^{2} \left[ (\pi^r_{\alpha\delta}\delta_{\alpha\beta}) (\pi^{3-r}_{\gamma\delta}\delta_{\gamma\delta}) + (\pi^r_{\alpha\gamma}\delta_{\alpha\delta}) (\pi^{3-r}_{\beta\delta}\delta_{\beta\delta}) \right].\)
where \( r \) is short-hand for \( \{r_{i_1}, \ldots, r_{i_m}\} \), and \( i_k \) are the indices of the wavy lines traversed in the orbit.

In every diagram, there is also exactly one orbit (e.g. the orbits labeled 1 in both panels of Fig. 5.2) that includes, in addition, the factor \( \sigma^+(RL)^{-1} \) sandwiched between the two external Green’s functions. This orbit, which we will call the external orbit, contributes a trace of the form

\[
E_{n, \tilde{r}} \equiv \text{Tr}(\sigma^+(RL)^{-1}G(\eta, z; 0) \pi^{r_{i_1}} \cdots \pi^{r_{i_m}} G(\eta, z; 0))
\]

where \( n \) is the number wavy lines the orbit traverses and \( \tilde{r} \) is short for \( \{r_{j_1}, \ldots, r_{j_n}\} \), and \( j_k \) are the indices of the wavy lines traversed in this orbit\(^4\).

The full expression for the diagram is obtained by multiplying all such trace factors contributed by every orbit in the diagram, as well as a factor of \( N^{-w-1} \) where \( w \) is the number of wavy lines in the diagrams, to account for the \( N^{-1} \) in Eq. (5.40) for each wavy line, as well as the extra \( N^{-1} \) which normalizes the trace in the external orbit Eq. (5.70) as dictated by Eq. (5.36). The obtained expression is finally summed over all the \( r \)-indices corresponding to each wavy line, as required by Eq. (5.40).

The justification for the non-crossing approximation is based on the claim (proven further below) that each trace contributed by a orbit (external or internal) as in Eqs. (5.69)–(5.70) is \( O(N) \), irrespective of \( \eta, z, m \) or \( r \). Once that is accepted as true, it follows that a diagram’s scaling with \( N \) depends solely on the number of orbits and wavy lines it contains.

\(^4\)In writing Eq. (5.70), we dropped the \( \frac{1}{N} \) that normalizes the trace in Eqs. (5.36), but we will account for it below. For succinctness, in Eqs. (5.69)–(5.70) we also suppressed the arguments \((\eta, z)\) for \( I_{m, r} \) and \( E_{n, \tilde{r}} \) on which they depend
Specifically, any diagram will yield an expression that is $O(N^\alpha)$ with

$$
\alpha = f - w - 1, 
$$

(5.71)

where $f$ is the number of orbits in the diagram (the sum over at most $2^w$ possible configurations of $r_i$ does not contribute to the scaling with $N$).

A classic argument can now be used to show that the contributions of crossing diagrams are suppressed by inverse powers of $N$. Though the argument is well known (t’Hooft, 1974; Brezin and Zuber, 1978), we will summarize it here to make the presentation in this thesis self-contained.

Let $V$ denote the total number of vertices in the diagram (i.e. the number of intersections of wavy lines and the solid line, plus an extra one representing the insertion of $\sigma^+(RL)^{-1}$ in the solid line loop) and let $E$ denote its total number of edges, i.e. $E \equiv w + s$, where $s$ is the number of solid line segments ($s = 5$ in both panels of Fig. 5.2). It is easy to see that $V = s$. Thus we have $E - V = w$.

We can define the number of “faces” in the diagram by $F \equiv f + 1$, and its “Euler characteristic” by

$$
\chi \equiv F - E + V, 
$$

(5.72)

We then find that $\chi = F - (E - V) = f + 1 - w$, and from Eq. (5.71) we then obtain

$$
\alpha = \chi - 2; 
$$

(5.73)

thus, the contribution of a diagram is $O(N^\alpha)$, with $\alpha$ determined solely by the diagram’s
formal “Euler characteristic” via Eq. (5.73).

It can be shown that a diagram with \( F \) formal “faces” and a formal “Euler characteristic” \( \chi \) as defined above, can be drawn on (embedded in) a two-dimensional orientable surface with Euler characteristic \( \chi \), such that no edges (solid or wavy) cross to create new vertices, and each face created on the surface by its partitioning by the drawn diagram, a) is topologically a disk, and b) has a one-to-one correspondence with and is encircled by an orbit in the diagram. We now count, among the orbits, also the loop formed by the solid arrow line, thus the number of faces on the surface is indeed \( F = f + 1 \). The \( \chi \) we defined above for the diagram agrees with the Euler characteristic of the surface, as conventionally defined.

Topologically, such a surface is a generalized torus with \( g \) holes, where \( g \) is the number of edge-crossings in the diagram. While the surface with zero holes is the sphere (or after decompactification, the plane), the diagram in the example of Fig. 5.2, panel b, can be drawn in this manner only on a torus, which corresponds to \( g = 1 \). Using the generalized Euler formula \( \chi = 2 - 2g \) (for the proof see Stahl (2005), theorem 4.2.1), we conclude that

\[
\alpha = -2g; \tag{5.74}
\]

therefore the only diagrams that are not suppressed by inverse powers of \( N \) are those that can be drawn, as described above, on the plane.

Since we took the area enclosed by the solid arrow line loop as a face by itself, this means that the diagram should be drawable with the wavy lines remaining outside this area (in order not to partition it into several faces) without crossing each other; this is the
precise definition of the diagram being non-crossing – as was to be proven. It should be mentioned, for the sake of precision, that this is a more restrictive property than planarity of the diagram; for example the graph in panel (b) of Fig. 5.2 is planar, as one of the wavy lines can be drawn inside the solid loop without crossing any other line, but it is not non-crossing as defined here.

Let us now go back to justifying the claim that the traces contributed by the orbits as in Eqs. (5.69)–(5.70) are \( O(N) \). For this purpose we will make use of the singular value decomposition of \( M_z \) introduced in Eq. (5.51). Defining the unitary matrix

\[
U_z \equiv \begin{pmatrix} U_z & 0 \\ 0 & V_z \end{pmatrix},
\]

(5.75)

and using Eq. (5.51), we can write \( H_0(z) \), defined in Eq. (5.33), as

\[
H_0(z) = U_z \tilde{H}_0(z) U_z^\dagger,
\]

(5.76)

where

\[
\tilde{H}_0(z) \equiv \begin{pmatrix} 0 & S_z \\ S_z & 0 \end{pmatrix}.
\]

(5.77)

Let us also define \( \tilde{G}(\eta, z; 0) \equiv U_z^\dagger G(\eta, z; 0) U_z \), such that

\[
G(\eta, z; 0) = U_z \tilde{G}(\eta, z; 0) U_z^\dagger.
\]

(5.78)
Then using the definition $G(\eta, z; 0) = (\eta - H_0(z))^{-1}$, we see that

$$
\tilde{G}(\eta, z; 0) = \frac{1}{\eta - H_0(z)} = \left( \begin{array}{cc}
\frac{\eta}{\eta^2 - S^2_z} & \frac{S_z}{\eta^2 - S^2_z} \\
\frac{S_z}{\eta^2 - S^2_z} & \frac{\eta}{\eta^2 - S^2_z}
\end{array} \right),
$$

where we used Eq. (5.77) to write the last equality. Given the block-diagonal nature of Eq. (5.75) and the definitions Eq. (5.39), we also have

$$
\pi^r = U_z \pi^r U^\dagger_z.
$$

We now substitute $G(\eta, z; 0)$ and $\pi^r$ in Eqs. (5.69)–(5.70) with the right hand sides of Eqs. (5.78) and (5.80), respectively. After canceling the $U_z$’s we obtain

$$
I_{m,r} = \text{Tr}\left(\tilde{G}(\eta, z; 0)\pi^{r_1} \cdots \tilde{G}(\eta, z; 0)\pi^{r_m}\right),
$$

$$
E_{n,\tilde{r}} = \text{Tr}\left(\sigma^+ A(z)\tilde{G}(\eta, z; 0)\pi^{r_1} \cdots \tilde{G}(\eta, z; 0)\pi^{r_n}\right)
$$

where we defined

$$
A(z) \equiv U^\dagger_z (RL)^{-1} V_z,
$$

such that $U^\dagger_z [\sigma^+ \otimes (RL)^{-1}] U_z = \sigma^+ \otimes A(z) \equiv \sigma^+ A(z)$.

For the internal orbits, we see from Eq. (5.79) that each $\tilde{G}(\eta, z; 0)$, depending on whether it is sandwiched between the same projectors $\pi^r$, or between two opposite projectors, $\pi^r$ and $\pi^{3-r}$, contributes a diagonal factor equal to $\eta/(\eta^2 - S^2_z)$ or $S_z/(\eta^2 - S^2_z)$, respectively.

Thus, for any configuration of $r_i$’s, if the number of Green’s functions sandwiched the
second way is \( k (1 \leq k \leq m) \), we obtain

\[
I_{m,r}(\eta, z) = \sum_{i=1}^{N} \frac{\eta^{m-k} s_i(z)^k}{(\eta^2 - s_i(z)^2)^m} \quad (1 \leq k \leq m)
\]  

(5.84)

for the internal orbits (in particular, we see that the sole dependence of \( I_{m,r}(\eta, z) \) on \( r \) is via the number \( k \)). Hence, we have

\[
|I_{m,r}(\eta, z)| \leq N \max_i \left| \frac{\eta^{m-k} s_i(z)^k}{(\eta^2 - s_i(z)^2)^m} \right|. 
\]  

(5.85)

When the imaginary part of \( \eta \) is nonzero, the denominator in the right hand side of Eq. (5.85) cannot vanish for any value of \( s_i(z) \). Assuming \( \text{Im} \, \eta > 0 \), it will be sufficient for the present purposes to substitute Eq. (5.85) with the weaker bound

\[
|I_{m,r}(\eta, z)| \leq N \max_s \left| \frac{\eta^{m-k} s^k}{(\eta^2 - s^2)^m} \right|, \quad (\text{Im} \, \eta > 0)
\]  

(5.86)

where now the maximum is taken for \( s \) ranging over the whole \([0, \infty)\). Since \( \text{Im} \, \eta > 0 \) the expression has no singularities at finite real \( s \), and since \( 2m > k \), it cannot diverge as \( s \to \infty \) either; thus it has a finite maximum independent of \( N \).

More precisely, it is easy to show that \( \max_s \left| \frac{\eta^{m-k} s^k}{(\eta^2 - s^2)^m} \right| \leq \left[ \frac{\sqrt{2}}{\text{Im} \, \eta} \right]^m \), irrespective of \( k \) as long as \( 1 \leq k \leq m \), yielding

\[
|I_{m,r}(\eta, z)| \leq N \left[ \frac{\sqrt{2}}{\text{Im} \, \eta} \right]^m, \quad (\text{Im} \, \eta > 0).
\]  

(5.87)

While as \( \eta \to i0 \), which is the limit we have to take after summing up the relevant diagrammatic series, \( s_i(z) \) that approach zero as \( N \) grows can make this expression unbounded as \( N \to \infty \).
Similarly, the trace for the external orbit can be written in the new basis Eq. (5.79) as

\[ E_{n,\tilde{r}}(\eta, z) = \sum_{i=1}^{N} A_{ii}(z) \frac{\eta^{n-\tilde{k}} s_i(z)^{\tilde{k}}}{(\eta^2 - s_i(z)^2)^n}, \quad (1 \leq \tilde{k} \leq n), \]  

where \( \tilde{k} \) is the number of Green’s functions in Eq. (5.70) sandwiched between two \( \pi^r \)'s with different superscripts; this convention works correctly for the external orbit as well, if we account for the presence of \( \sigma^+ \) by imagining a \( \pi^2 (\pi^1) \) to the left (right) of the leftmost (rightmost) Green’s function.

From Eq. (5.83), we can write \( A_{ii}(z) = u_i(z)^\dagger (RL)^{-1} v_i(z) \), where we defined the vectors \( u_i(z) \) and \( v_i(z) \) to be the \( i \)-th column of \( U_z \) and \( V_z \), respectively. By the Cauchy-Schwartz inequality, we then have

\[ |A_{ii}(z)| \leq \|u_i(z)\| \|(RL)^{-1} v_i(z)\| \leq \|u_i(z)\| \|v_i(z)\| \|(RL)^{-1}\| \]  

where \( \|(RL)^{-1}\| \) is the operator norm, or the maximum singular value, of \( (RL)^{-1} \). But since \( U_z \) and \( V_z \) are unitary matrices, \( u_i(z) \) and \( v_i(z) \) are unit vectors, and we obtain

\[ |A_{ii}(z)| \leq \|(RL)^{-1}\|. \]  

(5.90)

Going back to Eq. (5.88), this yields the bound

\[ |E_{n,r}(\eta, z)| \leq N \|(RL)^{-1}\| \max_{i} \left| \frac{\eta^{n-\tilde{k}} s_i(z)^{\tilde{k}}}{(\eta^2 - s_i(z)^2)^n} \right|. \]  

(5.91)

The only difference with the inequality for \( I_{m,r} \) is the factor \( \|(RL)^{-1}\| \). Repeating the same
as for the internal traces, we therefore see that

$$|E_{n,r}(\eta, z)| \leq N \left[ \frac{\sqrt{2}}{\text{Im} \eta} \right]^n \|RL\|^{-1}, \quad (\text{Im} \eta > 0), \quad (5.92)$$

and thus a sufficient condition for $E_{n,r}$ to be $O(N)$ for $\text{Im} \eta > 0$, is that $\|RL\|^{-1}$ remains bounded as $N \to \infty$, i.e.

$$\|RL\|^{-1} = O(1). \quad (5.93)$$

Combining Eqs. (5.87) and (5.92), and given the prescription in the paragraph following Eq. (5.70), we can bound the absolute value of the contribution of a diagram with genus $g$ (or $g$ crossings), $w$ wavy lines, and $s$ solid lines, by $2^w \left[ \frac{\sqrt{2}}{\text{Im} \eta} \right]^s \|RL\|^{-1} N^{-2g}$. Here we have used the fact that $f - w - 1 = -2g$, as proven above; the power of $s$ is obtained by noting that the powers of $m$ and $n$ in the bounds Eqs. (5.87) and (5.92), when summed over all orbits, must equal $s$, since every Green’s function or solid line appears in exactly one orbit.

Hence for a fixed, nonzero $\text{Im} \eta$, the contribution of crossing diagrams (i.e. those with $g \geq 1$) goes to zero as $N \to \infty$. Thus if we take the limit $N \to \infty$ before the limit $\eta \to i0^+$, ignoring the crossing diagrams is safe, and the expression for $\rho(z)$ obtained from Eq. (5.36) after analytic continuation of $\text{tr}(\sigma^+(RL)^{-1}G(\eta, z))$ to $\eta = i0$, with $G(\eta, z)$ given by the contribution of non-crossing diagrams to $\langle G(\eta, z; J) \rangle_j$, gives the correct result for $\lim_{N \to \infty} \rho(z)$.

It should be mentioned that when the smallest singular value $s_i(z)$ remains bounded away from zero as $N \to \infty$, even at $\eta = 0$ the traces Eqs. (5.84) and (5.88) are $O(N)$, as
is not hard to check, justifying the non-crossing approximation at $\eta = 0$. Thus it is only when some $s_i(z)$ are $o(1)$ that it becomes important to send $\eta$ to $i0^+$ only after the limit $N \to \infty$ has been taken. In particular, in such cases, applying the limit $\eta \to i0^+$ to the results obtained using the non-crossing approximation before taking the limit $N \to \infty$, may yield finite-size contributions to $\lim_{N \to \infty} \rho(z)$, which in general may yield incorrect subleading corrections.

### 5.3.4 Proving the compactness of the supports

In this section we will prove that in the region Eq. (5.57), the eigenvalue density vanishes. More precisely, we prove that $\rho(z) \equiv \lim_{\epsilon \to 0^+} \lim_{N \to \infty} \rho_N(z, \epsilon) = 0$, where

\[
\rho_N(z, \epsilon) = \frac{1}{\pi} \frac{\partial}{\partial \tilde{z}} \text{tr} \left[ \frac{(RL)^{-1}M_z^\dagger}{M_zM_z^\dagger + \kappa^2} \right],
\]

(5.94)
is obtained by substituting Eq. (5.54) into Eqs. (5.36). Here, $\kappa = g(z, \epsilon) + \epsilon$ is the solution of Eq. (5.56), which as we argued in Sec. 5.3.1, vanishes as $\epsilon \to 0^+$ when $z$ is in the region Eq. (5.57). Note that since Eq. (5.56) is defined in the limit $N \to \infty$, $\kappa$ is a function of $z$ and $\tilde{z}$ but has no dependence on $N$.

Recall that for $\epsilon > 0$, $g(z, \epsilon)$ is positive and therefore $\kappa > \epsilon > 0$. Expanding the derivative in Eq. (5.94), we obtain

\[
\pi \rho_N(z, \epsilon) = \text{tr} \left[ \frac{(RL)^{-1}(RL)^{-1}}{M_zM_z^\dagger + \kappa^2} \right] - \text{tr} \left[ \frac{(RL)^{-1}M_z^\dagger}{M_zM_z^\dagger + \kappa^2} \right] \frac{1}{M_zM_z^\dagger + \kappa^2} \frac{1}{M_zR^{-1}L} \frac{1}{M_zM_z^\dagger + \kappa^2}
\]

\[
- \text{tr} \left[ \frac{(RL)^{-1}M_z^\dagger}{M_zM_z^\dagger + \kappa^2} \right] \frac{1}{M_zM_z^\dagger + \kappa^2} \frac{1}{M_zR^{-1}L} \frac{1}{M_zM_z^\dagger + \kappa^2} \right] \frac{1}{M_zM_z^\dagger + \kappa^2}
\]

\[
\frac{1}{M_zM_z^\dagger + \kappa^2} \frac{1}{M_zR^{-1}L} \frac{1}{M_zM_z^\dagger + \kappa^2} \right] \frac{1}{M_zM_z^\dagger + \kappa^2}
\]

\[
\text{tr} \left[ \frac{(RL)^{-1}M_z^\dagger}{M_zM_z^\dagger + \kappa^2} \right] \frac{1}{M_zM_z^\dagger + \kappa^2} \frac{1}{M_zR^{-1}L} \frac{1}{M_zM_z^\dagger + \kappa^2} \right] \frac{1}{M_zM_z^\dagger + \kappa^2}
\]

(5.95)
Defining \( Q = 1 - M_z^\dagger \frac{1}{M_z M_z^\dagger + \kappa^2} M_z \) and suppressing the explicit dependence of \( \kappa \) on \( z \) for simplicity, we can write

\[
\pi \rho_N(z, \epsilon) = \text{tr} \left[ \frac{(RL)^{-1} Q (RL)^{-1}}{M_z M_z^\dagger + \kappa^2} \right] - \text{tr} \left[ \frac{(RL)^{-1} M_z^\dagger}{M_z M_z^\dagger + \kappa^2} \frac{1}{M_z M_z^\dagger + \kappa^2} \right] \partial \kappa^2(\kappa^2). \tag{5.96}
\]

Now, by the Woodbury matrix identity we can see that \( Q = \frac{\kappa^2}{M_z M_z^\dagger + \kappa^2} \), and substituting this into Eq. (5.96) yields

\[
\pi \rho_N(z, \epsilon) = \text{tr} \left[ \frac{(RL)^{-1} (RL)^{-1}}{M_z M_z^\dagger + \kappa^2} \right] \kappa^2 - \text{tr} \left[ \frac{(RL)^{-1} M_z^\dagger}{M_z M_z^\dagger + \kappa^2} \frac{1}{M_z M_z^\dagger + \kappa^2} \right] \partial \kappa^2(\kappa^2). \tag{5.97}
\]

Differentiating Eq. (5.56) with respect to \( \bar{z} \) yields

\[
-\partial \bar{z}(\kappa^2) = \frac{-2\kappa^2 \partial \kappa}{1 - \kappa - 2\kappa^2 \partial \kappa}, \tag{5.98}
\]

with the partial derivatives of \( \mathcal{K}(\kappa, \bar{z}) \) given by

\[
-\partial_{\kappa^2} \mathcal{K} = \mathcal{T}_\infty(\kappa) \equiv \lim_{N \to \infty} \mathcal{T}_N(\kappa) \tag{5.99}
\]

\[
-\partial \kappa \mathcal{K} = \mathcal{V}_\infty(\kappa) \equiv \lim_{N \to \infty} \mathcal{V}_N(\kappa),
\]

where we defined

\[
\mathcal{T}_N(\kappa) \equiv \text{tr} \left[ \frac{1}{(M_z M_z^\dagger + \kappa^2)^2} \right] \tag{5.100}
\]

\[
\mathcal{V}_N(\kappa) \equiv \text{tr} \left[ \frac{1}{M_z M_z^\dagger + \kappa^2 M_z M_z^\dagger + \kappa^2} \right]. \tag{5.101}
\]
we thus obtain

\[ \pi \rho_N(z, \epsilon) = \kappa^2 T_N(\kappa) + \frac{2 \kappa^2 \mathcal{V}_\infty(\kappa) \mathcal{V}_N(\kappa)^\ast}{1 - K(\kappa) + 2 \kappa^2 T_\infty(\kappa)} \] (5.102)

where we defined

\[ T_N(\kappa) \equiv \text{tr} \left[ \frac{(RL)^{-1}}{M_z^\dagger M_z + \kappa^2 M_z M_z^\dagger + \kappa^2} \right]. \] (5.103)

Having eliminated derivatives of \( \kappa \), we now simply need to show that \( \lim_{\kappa \to 0^+} \lim_{N \to \infty} \) of the right side of Eq. (5.102) vanishes for \( z \) is in the region Eq. (5.57) (where \( \kappa = 0^+ \) is the solution of Eq. (5.56) as \( \epsilon \to 0^+ \)).

We will start by bounding the traces \( T_N(\kappa) \) and \( \mathcal{V}_N(\kappa) \) in Eq. (5.102). For \( \mathcal{V}_N(\kappa) \) we use the singular value decomposition Eq. (5.51):

\[ |\mathcal{V}_N(\kappa)| = \text{tr} \left[ \frac{(RL)^{-1} M_z^\dagger}{M_z M_z^\dagger + \kappa^2 M_z M_z^\dagger + \kappa^2} \right] = \text{tr} \left[ U_z^\dagger (RL)^{-1} V_z \frac{S_z}{(S_z^2 + \kappa^2)^2} \right] \leq \|(RL)^{-1}\| \text{tr} \left[ \frac{S_z}{(S_z^2 + \kappa^2)^2} \right] \] (5.104)

(where in the last line we used Eq. 5.90); we will write this compactly as

\[ |\mathcal{V}_N(\kappa)| \leq \|(RL)^{-1}\| \mathcal{V}_N^>(\kappa) \] (5.105)

where we have defined

\[ \mathcal{V}_N^>(\kappa) \equiv \frac{1}{N} \sum_{i=0}^{N} \frac{s_i(z)}{(s_i(z)^2 + \kappa^2)^2}. \] (5.106)
Taking the limit \( N \to \infty \), we obtain from Eq. (5.105)

\[
|V_\infty(k)| \leq CV_\infty^>(k) \tag{5.107}
\]

where \( V_\infty^>(k) \equiv \lim_{N \to \infty} V_N^>(k) \), and \( C \) is an upper bound on \( \|(LR)^{-1}\| \) (which we have assumed is \( O(1) \) as \( N \to \infty \)).

To bound \( T_N(k) \), we use the inequality

\[
|\text{tr}(ABCD)| \leq ||A|| ||C|| \text{tr} (BB^\dagger)^{\frac{1}{2}} \text{tr} (DD^\dagger)^{\frac{1}{2}}. \tag{5.108}
\]

This can be derived by first using the Cauchy-Schwartz inequality, \( |\text{tr} (AB)|^2 \leq \text{tr} (AA^\dagger) \text{tr} (BB^\dagger) \), and then using the inequality \( |\text{tr} (AB)| \leq ||B|| \text{tr} (A) \), valid for positive semi-definite \( A \) (which in turn follows from the definition of \( ||B|| \) after unitary diagonalization of \( A \)).

Using now (5.108) we obtain

\[
\left| \text{tr}\left[ \frac{(RL)^{-1} (RL)^{-\dagger}}{M_2 M_z + \kappa^2 M_z M_2^\dagger + \kappa^2} \right] \right| \leq \|(RL)^{-1}\|^2 \text{tr}\left[ \left( \frac{1}{M_2 M_2^\dagger + \kappa^2} \right)^2 \right] \tag{5.109}
\]

or

\[
|T_N(k)| \leq \|(RL)^{-1}\|^2 T_N^>(k). \tag{5.110}
\]

and using the inequalities (5.105), (5.107) and (5.110) in Eq. (5.102) we arrive at

\[
\pi|\rho_N(z, \epsilon)| \leq C^2 \left[ \kappa^2 T_N^>(k) + \frac{2 \kappa^2 V_\infty^>(k) V_N^>(k)}{1 - K(k) + 2 \kappa^2 T_\infty^>(k)} \right] \tag{5.111}
\]
Taking the $N \to \infty$ limit (while keeping $\kappa$ finite), and defining $\rho(z, \epsilon) \equiv \lim_{N \to \infty} \rho_N(z, \epsilon)$, this becomes

$$\pi \rho(z, \epsilon) \leq C^2 \left[ \kappa^2 \mathcal{T}_\infty^\prec(\kappa) + \frac{2(\kappa \mathcal{V}_\infty^\prec(\kappa))^2}{1 - \mathcal{K}(\kappa) + 2\kappa^2 \mathcal{C}_\infty(\kappa)} \right]$$  \hspace{1cm} (5.112)

where we defined

$$\mathcal{T}_\infty^\prec(\kappa) \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N} \frac{1}{(s_i(z)^2 + \kappa^2)^2} \hspace{1cm} (5.113)$$

$$\mathcal{V}_\infty^\prec(\kappa) \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N} \frac{s_i(z)}{(s_i(z)^2 + \kappa^2)^2} \hspace{1cm} (5.114)$$

Thus, to show that $\lim_{\epsilon \to 0^+} \rho(z, \epsilon) = 0$, it suffices to show that $\kappa^2 \mathcal{T}_\infty^\prec(\kappa)$ and $\kappa \mathcal{V}_\infty^\prec(\kappa)$ vanish as $\kappa \to 0^+$. Indeed, since $z$ is in the region Eq. (5.57), $1 - \mathcal{K}(\kappa)$ and hence the denominator in the last term in Eq. (5.112) remains positive as $\kappa \to 0^+$.

Let us rewrite Eqs. (5.113-5.114) as

$$\mathcal{T}_\infty^\prec(\kappa) = \int_0^\infty \frac{\rho_s(s; z)ds}{(s^2 + \kappa^2)^2} \hspace{1cm} (5.115)$$

$$\mathcal{V}_\infty^\prec(\kappa) = \int_0^\infty \frac{s \rho_s(s; z)ds}{(s^2 + \kappa^2)^2} \hspace{1cm} (5.116)$$

where we are defining

$$\rho_s(s; z) = \lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N} \delta(s - s_i(z)) \hspace{1cm} (5.117)$$

as the limit of the density of the singular values of $M_z^6$.

---

6More precisely, we only need to define the limit Eq. (5.117) in the sense of distributions, i.e. such that
Note that, as seen from Eqs. (5.115-5.116), contributions to $T_\infty^\infty(\kappa)$ and $V_\infty^\infty(\kappa)$ from integration on $[s_0, \infty)$ for any fixed, nonzero $s_0$ remain finite as $\kappa \to 0^+$; only singular contributions arising from the region $s = O(\kappa) \ll 1$ can contribute to $\kappa^2 T_\infty^\infty(\kappa)$ and $\kappa V_\infty^\infty(\kappa)$ as $\kappa \to 0^+$. Thus we only need concern ourselves with the portion of integrals from 0 to some arbitrary small, but fixed $s_0$, and show that $\kappa^2 \int_0^{s_0} \frac{\rho_k(s; z) ds}{(s^2 + \kappa^2)^2}$ and $\kappa \int_0^{s_0} \frac{s \rho_k(s; z) ds}{(s^2 + \kappa^2)^2}$ vanish as $\kappa \to 0^+$.

Let us first consider separately the case in which there is a region of $z$ outside Eq. (5.57), where a single (more generally $O(1)$) singular value $s_i(z)$ vanishes as $N \to \infty$, while all the other $s_i(z)$ remain bounded from below. Instances of this type of situation are the examples of Eqs. (C.1) (Appendix C) and (5.151). An $O(1)$ set of (vanishing) singular values does not contribute to the density Eq. (5.117) and since the other $s_i(z)$ are bounded from below, there is an $s_0$ below which $\rho_k(s; z)$ identically vanishes. So the claim is clearly true for such cases.

More in general, we can exploit the fact that $z$ is in the region Eq. (5.57), so that

\[
\lim_{\kappa \to 0^+} \int_0^{\infty} \frac{\rho_k(s; z) ds}{s^2 + \kappa^2} < 1. \tag{5.118}
\]

and we conclude that as $s \to 0^+$ the density, $\rho_k(s; z)$, must vanish at least as fast as $s^\alpha$, i.e. it must be $O(s^\alpha)$, for some $\alpha > 1$; otherwise the integral in Eq. (5.118) diverges in the limit.

for any regular test function, $f(s^2)$, bounded at infinity and regular everywhere, including at $s^2 \to 0^+$, we have $\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(s_i(z)^2) = \int_0^{\infty} f(s^2) \rho_k(s; z) ds$. We do not assume any smooth form for $\rho_k(s; z)$; in particular, $\rho_k(s; z)$ may have delta function singularities when an $O(N)$ singular values converge to the same value as $N \to \infty$. Also note that this assumption does not forbid the possibility that some $s_i(z)$ diverge as $N \to \infty$; the requirement that $\|M\|$ remain bounded automatically guarantees that these will not be numerous enough to contribute to $\rho_k(s; z)$ at infinity.
Let us therefore choose \( s_0 \) to be small enough such that for \( s \leq s_0, \rho(s; z) < cs^\alpha \) for some constant \( c \) and \( \alpha > 1 \). It is then an elementary exercise to show that \( \kappa^2 \int_0^{s_0} \frac{s^\alpha ds}{(s^2 + \kappa^2)^2} \) and \( \kappa \int_0^{s_0} \frac{s^{\alpha+1} ds}{(s^2 + \kappa^2)^2} \) are \( O(\kappa^{\min(2, \alpha-1)}) \) and \( O(\kappa^{\min(1, \alpha-1)}) \), respectively, as \( \kappa \to 0^+ \), and since \( \alpha > 1 \), they both vanish in the limit, proving the claim.

### 5.3.5 The case of zero mean connectivity

**Deriving Eq. (5.12)**

For the special case of \( M = 0 \), we have \( M_z = z(RL)^{-1} \). If we let \( \sigma_i \) be the singular values of \( RL \), then the singular values of \( M_z \) will be given by \( s_i(z) = |z|\sigma_i^{-1} \). Substituting this in Eq. (5.9) and multiplying both sides by \( r^2 = |z|^2 \), we obtain Eq. (5.14). It is immediately seen that \( g(z) \), \( \varphi(z) \) and \( \rho(z) \) depend only on the radius \( r = |z| \).

Similarly, we can rewrite Eq. (5.68) as

\[
\varphi(r) = -g(r)^2 + \frac{1}{N} \sum_{i=1}^{N} \ln[r^2\sigma_i^{-2} + g(r)^2].
\]  

To find the spectral radius (boundary of the spectrum) \( r_0 \) we have to solve Eq. (5.14) for \( r \), setting \( g(r) = 0 \). This gives

\[
r_0^2 = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^2 = \|RL\|_F^2,
\]  

yielding the boundary of Eq. (5.12).
Deriving Eqs. (5.13-5.15)

Let us define the proportion of eigenvalues lying outside a radius $r$ from the origin by $n_>(r)$. To obtain Eqs. (5.13) and (5.15), first note that

$$\rho(r) = \frac{1}{\pi} \partial_z \partial_z \varphi(z) = \frac{1}{4\pi} \nabla^2 \varphi(z) = \frac{1}{4\pi r} \frac{\partial}{\partial r} (r \partial_r \varphi(r)),$$  \hspace{1cm} (5.121)

where we used the expression of Laplacian, $\nabla^2 = \partial_x^2 + \partial_y^2$, in 2-D polar coordinates in the last equality.

Using this with the definition $n_>(r) = 2\pi \int_r^\infty \rho(r) r dr$, one obtains $n_>(r) = \left[ \frac{r}{2} \partial_r \varphi(r) \right]_r^\infty$. For the limit at $r \to \infty$, note that for $r > r_0$, $g(r) = 0$ and we have

$$\varphi(r) = \frac{1}{N} \sum_{i=1}^N \ln(r^2 \sigma_i^{-2}) = 2 \ln r - \frac{2}{N} \ln \det(RL)$$  \hspace{1cm} (5.122)

hence $\frac{r}{2} \partial_r \varphi(r) \to 1$ as $r \to \infty$. It follows that

$$n_>(r) = 1 - \frac{r}{2} \partial_r \varphi(r).$$  \hspace{1cm} (5.123)

Differentiating Eq. (5.119) and using Eq. (5.14) we obtain

$$\partial_r \varphi(r) = 2r \frac{1}{N} \sum_{i=1}^N \frac{1}{r^2 + \sigma_i^2 g(r)^2},$$  \hspace{1cm} (5.124)
and

\[ n_>(r) = 1 - r^2 \frac{1}{N} \sum_{i=1}^{N} \frac{1}{r^2 + \sigma_i^2 g(r)^2} = g(r)^2 \frac{1}{N} \sum_{i=1}^{N} \frac{\sigma_i^2}{r^2 + \sigma_i^2 g(r)^2}. \]  

(5.125)

Using Eq. (5.14) once again we obtain Eq. (5.15). Finally, using the latter together with Eqs. (5.121)–(5.123) yields Eq. (5.13).

**Extremal values of the spectral density**

We will prove some further general properties for the eigenvalue density for \( M = 0 \). Let us first define

\[ I_{n,k}(g, r) \equiv \left\langle \frac{\sigma^{-k}}{(g^2 + \sigma^{-2} r^2)^n} \right\rangle_\sigma \]  

(5.126)

and

\[ \langle f(\sigma) \rangle_\sigma \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} f(\sigma_i). \]  

(5.127)

We will assume here that the \( \sigma_i \)’s have a limit density, \( \rho_\sigma(\sigma) \), such that \( \langle f(\sigma) \rangle_\sigma = \int_0^\infty f(\sigma)\rho_\sigma(\sigma)\,d\sigma \) is well-defined for \( f(\sigma) \) with sufficiently fast decay at infinity. Note that since we assumed that \( \|(RL)^{-1}\| = (\min_i \sigma_i)^{-1} = O(1) \), this density has no measure at \( \sigma = 0 \) and hence the averages in Eq. (5.126) are non-singular for \( n, k \geq 0 \). Also \( \langle f(\sigma) \rangle_\sigma \) is finite as long as \( f(\sigma) = O(\sigma^2) \) as \( \sigma \to \infty \), as we are assuming that the \( \|RL\|_F = O(1) \) and \( \lim_{N \to \infty} \|RL\|_F^2 = \langle \sigma^2 \rangle_\sigma \).

First, we will obtain general expressions for \( \rho(r = 0) \) and \( \rho(r = r_0) \), with \( r_0 \) given by Eq. (5.12). Using Eq. (5.14), re-expressed as \( I_{1,0}(g, r) = 1 \), we can put Eq. (5.13),
\[ \pi \rho(r) = -\frac{\partial n_{1}(r)}{\partial(r^2)} \] into the form

\[ \pi \rho(r) = \frac{\frac{\partial I_{1,0}}{\partial(r^2)}}{\frac{\partial I_{2,0}}{\partial(r^2)}} = \frac{I_{2,2}(g, r)}{I_{2,0}(g, r)} . \quad (5.128) \]

Using the facts that at \( r = 0 \), \( g = 1 \), and at \( r = r_0 \), \( g = 0 \), one obtains

\[ \rho(r = 0) = \frac{1}{\pi} \frac{I_{2,2}(1, 0)}{I_{2,0}(1, 0)} = \frac{1}{\pi} \langle \sigma^{-2} \rangle_\sigma \quad (5.129) \]

\[ \rho(r = r_0) = \frac{1}{\pi} \frac{I_{2,2}(0, r_0)}{I_{2,0}(0, r_0)} = \frac{1}{\pi} \langle \sigma^2 \rangle_\sigma . \quad (5.130) \]

Using now the fact that \( \sigma^4 \) and \( \sigma^{-2} \) are anti-correlated and that \( \sigma^2 = \sigma^4 \sigma^{-2} \), we see that

\[ \langle \sigma^2 \rangle_\sigma \leq \langle \sigma^4 \rangle_\sigma \langle \sigma^{-2} \rangle_\sigma \] or

\[ \rho(r = r_0) \leq \rho(r = 0), \quad (5.131) \]

In this equation, the equality occurs if and only if \( \rho(\sigma) \) is deterministic, i.e., a delta-function. This can happen if all but an \( o(1) \) fraction of the \( \sigma_i \)'s have the same limit as \( N \to \infty \); in that case the eigenvalue distribution is given by the circular law.

**Slope of the density profile**

More generally, one can prove that \( \rho(r) \) is a decreasing function of \( r \) for any choice of \( L \) and \( R \) (with \( M = 0 \)). Using \( \frac{dp(r)}{dr} = 2r \frac{dp(r)}{d(r^2)} \) and Eq. (5.128), we obtain

\[ \frac{dp(r)}{dr} = 2r \frac{dI_{1,0}}{d(r^2)} \frac{I_{2,2}}{I_{2,0}} \frac{I_{2,0} - I_{2,2}}{I_{2,0}^2} , \quad (5.132) \]

where we are dropping the explicit \( (g, r) \) dependence of \( I_{n,k} \)'s whenever it is convenient.
Applying now the identities

\[
\frac{d}{d(r^2)} = \frac{\partial}{\partial(r^2)} + (g^2) \frac{\partial}{\partial(g^2)} = \frac{\partial}{\partial(r^2)} - \pi \rho(r) \frac{\partial}{\partial(g^2)}
\]

(5.133)

\[
\frac{\partial I_{n,k}}{\partial(g^2)} = -n I_{n+1,k}
\]

(5.134)

\[
\frac{\partial I_{n,k}}{\partial(r^2)} = -n I_{n+1,k+2}
\]

(5.135)

one finds

\[
\frac{d\rho(r)}{dr} = -4r \left[ \frac{I_{2,0}^2 I_{3,1} - 2I_{2,2}I_{2,0}I_{3,2} + I_{2,2}^2 I_{3,0}}{I_{2,0}^3} \right]
\]

(5.136)

It will prove useful to define the expectation operator \(\langle f(\sigma) \rangle'_\sigma\) as

\[
\langle f(\sigma) \rangle'_\sigma = \frac{\langle f(\sigma) \rangle_{\sigma^2}}{\langle \sigma^2 \rangle'_\sigma},
\]

(5.137)

in terms of which we can write

\[
\frac{d\rho(r)}{dr} = -4r \left[ \left( \frac{\sigma^{-4}}{g^2 + \sigma^{-2} r^2} \right)' - 2 \left( \frac{\sigma^{-2}}{g^2 + \sigma^{-2} r^2} \right)' \left( \langle \sigma^{-2} \rangle'_\sigma \right) + \left( \frac{1}{g^2 + \sigma^{-2} r^2} \right)' \left( \langle \sigma^{-2} \rangle'_\sigma \right) \right]
\]

(5.138)

or

\[
\frac{d\rho(r)}{dr} = -4r \left[ \text{Cov}' \left[ \frac{\sigma^{-2}}{g^2 + \sigma^{-2} r^2}, \sigma^{-2} \right] - \langle \sigma^{-2} \rangle'_\sigma \text{Cov}' \left[ \frac{1}{g^2 + \sigma^{-2} r^2}, \sigma^{-2} \right] \right]
\]

(5.139)

where \(\text{Cov}'[f, g] = \langle fg \rangle'_\sigma - \langle f \rangle'_\sigma \langle g \rangle'_\sigma\) is the covariance under \(\langle \cdot \rangle'_\sigma\).

Now since \(\frac{\sigma^{-2}}{g^2 + \sigma^{-2} r^2}\) and \(\sigma^{-2}\) are both strictly decreasing functions of \(\sigma\) (since \(g > 0\) for \(r < r_0\)), while \(\frac{1}{g^2 + \sigma^{-2} r^2}\) is a strictly increasing function of \(\sigma\) (for \(r > 0\)), the first covariance...
on the right hand side of Eq. (5.139) is positive, while the second one is negative, and therefore
\[ \frac{d\rho(r)}{dr} \leq 0. \] (5.140)

This slope is zero at \( r = 0 \) and strictly negative for \( r > 0 \) as long as \( \text{Var}[\sigma] > 0 \). When \( \text{Var}[\sigma] = 0 \), again we obtain the circular law. At \( r = r_0 \), we obtain
\[ \rho'(r_0) = -\frac{4}{r_0} \langle \sigma^{-2} \rangle_{\sigma}' \left( \langle \sigma^{-2} \rangle_{\sigma}' \langle \sigma^2 \rangle_{\sigma}' - 1 \right) = -\frac{4}{r_0} \frac{\langle \sigma^2 \rangle_{\sigma}}{\langle \sigma^4 \rangle_{\sigma}} \left( \langle \sigma^2 \rangle_{\sigma} \langle \sigma^6 \rangle_{\sigma} - \langle \sigma^4 \rangle_{\sigma}^2 \right) \] (5.141)

The curvature of \( \rho(r) \) at zero can also be evaluated by taking the limit \( r \to 0 \) of the bracket in Eq. (5.139), noting that \( g \to 1 \) as \( r \to 0 \). We thus find that
\[ \rho''(r = 0) = -4\text{Var}'[\sigma^{-2}] = -4 \frac{\text{Var}[\sigma^2]}{\langle \sigma^4 \rangle_{\sigma}} \leq 0. \] (5.142)

### 5.4 Applications to specific networks

In this section we present the results of explicit calculations of the eigenvalue density for specific examples of \( M, L \) and \( R \). To make the presentation more readable, we will state the results without proof. However, outlines of all the derivations are provided in Appendix C.

For many of the examples presented here, the matrices \( L \) and \( R \) defined above will be both proportional to the identity matrix; thus, in these examples the full synaptic weight matrix is of the form \( M + \sigma J \) where \( \sigma > 0 \) determines the strength of disorder in the matrix. In Secs. 5.4.3 and 5.4.4, we also present examples with nontrivial \( L \) and/or \( R \).
5.4.1 Schur structure of the mean connectivity

We recall that any mean connectivity matrix $M$ can be turned into an upper-triangular form by a unitary transformation, i.e.

$$M = U T U^\dagger,$$  \hspace{1cm} (5.143)

where $U$ is unitary and $T$ is upper-triangular (i.e. $T_{ij} = 0$ if $i > j$) with its main diagonal consisting of the eigenvalues of $M$. The difference between nonnormal and normal matrices is that for the latter, $T$ can be taken to be strictly diagonal. Equation (5.143) is referred to as a Schur decomposition of $M$ (Horn and Johnson, 1990), and we refer to the orthogonal modes of activity represented by the columns of $U$ as Schur modes.

The Schur decomposition provides an intuitive way of characterizing the dynamical system Eq. (2.12). Rewriting Eq. (2.12), with $J$ and $I(t)$ set to zero, in the Schur basis by defining $y = U^\dagger x$ (i.e. $y_i$ is the activity in the $i$-th Schur mode), we obtain $\frac{dy}{dt} = -\gamma y + Ty$. We see that activity in the $j$-th Schur mode provides an input to the equation for the $i$-th mode only when $i \leq j$ (as $T_{ij} = 0$ for $i > j$). Thus the coupling between modes is feedforward, going only from higher modes to lower ones, without any feedback. We will refer to $T_{ij}$’s for $j > i$ as feedforward weights. As these vanish for normal matrices, we can say a matrix is more nonnormal the stronger its feedforward weights are.

Due to the invariance of the trace, the norm, and the adjoint operation under unitary transforms, the general formula for the spectral density Eq. (5.7) takes the same form in any basis, so in particular we can work in the Schur basis of $M$. This means that $M$ can be replaced by $T$ provided $L$ and $R$ are also expressed in $M$’s Schur basis. The unitary invariance of the results is a consequence of both the invariance of $\rho(z)$ and of the invariance.
Figure 5.3: Eigenvalue spectrum of $M + \sigma J$ for $M$ given by Eq. (5.144) with $w = 1$ and $\lambda_n = \pm i$ (with $+i$ and $-i$ alternating). The red big dots at $\pm i$ indicate the eigenvalues of $M$. The red curve is the outer boundary of the eigenvalue spectrum of $A$ as computed numerically using Eq. (5.4) for disorder with zero mean and variance $\sigma = 0.5$. The black dots are eigenvalues from the numerical diagonalization of a realization of $A$ with real, Gaussian $J$'s of size $N = 700$.

of the statistical ensemble Eq. (5.22) for $J$ (whence the invariance of $LJR$ when $L \propto R \propto 1$) under unitary transforms like Eq. (5.143). Thus, we will use the feedforward structure of the Schur decomposition to characterize the different examples we consider below.

In keeping with the overall approach advocated in the Introduction, the examples will be chosen to demonstrate interesting features of nonnormal matrices in their simplest possible settings.

### 5.4.2 Networks with a single feedforward chain

In this type of network, each and every Schur mode is only connected to its lower adjacent mode, forming a long feedforward chain of length $N$. For simplicity, we take all feedfor-
ward weights in this chain to have the same value \( w \), so that

\[
M = T = \begin{pmatrix}
\lambda_1 & w & 0 & \cdots \\
0 & \lambda_2 & w & \cdots \\
& \vdots & \vdots & \ddots 
\end{pmatrix} \tag{5.144}
\]

or more succinctly \( M_{nm} = w \delta_{n+1,m} + \lambda_n \delta_{nm} \).

As a first example, Figure 5.3 shows the eigenvalue distribution of \( A = M + \sigma J \) for \( M \) of the form Eq. (5.144) with alternating imaginary eigenvalues, \( \lambda_n = (-1)^{n+1}i \). The black dots in Fig. 5.3 are the eigenvalues of \( A \) for one realization of \( J \). As expected, they are scattered around the highly degenerate spectrum of \( M \) at \( \pm i \) (represented by the two red dots). The red curve marks the outer boundary of the eigenvalue spectrum of \( A \) as computed numerically using Eq. (5.4), showing the agreement between theory and numerics.

As a second example, we will consider \( M \) of the form (5.144) with all eigenvalues zero \( (\lambda_n = 0 \forall n) \). This is a case where the ordering of limits, as discussed in Sec. 5.2.5, is of crucial importance.

In fact, using Eq. (5.4) directly would yield \( |z| \leq \sqrt{|w|^2 + \sigma^2} \) for the support of the eigenvalue density. However, using the correct procedure, Eqs. (5.18)–(5.19), one finds that this formula is only correct for \( \sigma \geq |w| \), while for \( \sigma < |w| \), the true support of the eigenvalue density in the limit \( N \to \infty \) is the annulus

\[
\sqrt{|w|^2 - \sigma^2} \leq |z| \leq \sqrt{|w|^2 + \sigma^2}, \tag{5.145}
\]

(this was first obtained in Khoruzhenko, 1996). Within this support, the eigenvalue density
is given in either case by

\[
\rho(z) = \frac{1}{\pi \sigma^2} \left[ 1 - \frac{|w|^2}{\sqrt{4|w|^2|z|^2 + \sigma^2}} \right].
\] (5.146)

Figure 5.4 demonstrates the close agreement of Eqs. (5.145)–(5.146) with the empirical spectrum of \(M + \sigma J\) for a single realization of \(J\), for \(N = 2000\) and two different values of \(\sigma\).

The discrepancy we noted between Eq. (5.145) and the results obtained by the naive use of Eq. (5.4) is due to the fact that for \(|z| < |w|\), \(M_z = (z - M)/\sigma\) has an exponentially small singular value (i.e. an eigenvalue of order \(O(e^{-cN})\), as is shown in Appendix C) which makes the result of Eqs. (5.18)–(5.19) dependent on the order of the two limits \(N \to \infty\) and \(g \to 0^+\). As we discussed after Eq. (5.11), such a discrepancy can signify the existence of an \(O(1)\) number of outlier eigenvalues outside the support of \(\lim_{N \to \infty} \rho(z)\). Simulations show that this is the case for \(|z| < \sqrt{|w|^2 - \sigma^2}\) (see Fig. 5.4).

The most striking aspect of these results is revealed in the limit \(\sigma \to 0\). For \(\sigma = 0\), the spectrum is that of \(M\), which is concentrated at the origin. Remarkably, however, as seen from Eqs. (5.145)–(5.146), for very small but nonzero \(\sigma\) the bulk of the eigenvalues are concentrated in the narrow ring with modulus \(|z| \approx |w|\). Thus in the limit \(N \to \infty\) the spectrum has a discontinuous jump at \(\sigma = 0\). This is a consequence of the extreme nonnormality of \(M\), as nonnormality manifests itself in the sensitivity of the spectrum to small perturbations.

This sensitivity is quantified by the notion of pseudospectra, which we introduced in Sec. 5.2.6. The (operator-norm) \(\epsilon\)-pseudospectrum of \(M\) is the region of complex plane to
which its spectrum can be perturbed by adding to $M$ a matrix of operator norm no larger than $\epsilon$. As we mentioned, this is precisely the set of complex values $z$ for which $\|(z - M)^{-1}\| > \epsilon^{-1}$, and therefore by the definition of the operator norm $\| \cdot \|$, the region in which $\|(z - M)^{-1}\|^{-1} = s_{\min}(z - M) < \epsilon$, where $s_{\min}(z - M)$ is the least singular value of $z - M$.

As noted above, for $|z| < |w|$, $s_{\min}(z - M)$ is exponentially small: $s_{\min}(z - M) \leq |w| \frac{z}{w} |w|^N$ (for a proof see after Eq. (C.17), Appendix C). Thus the $\epsilon$-pseudospectrum of $M$ contains the set of points $z$ satisfying $|w| \frac{z}{w} |w|^N < \epsilon$, i.e. the centered disk with radius $|w| \left( \frac{\epsilon}{|w|} \right)^{1/N}$ which approaches $|w|$ as $N \to \infty$. In other words, for large enough $N$, any point $|z| < |w|$ is in the $\epsilon$-pseudospectrum for any fixed $\epsilon$, no matter how small.

It has been stated (Trefethen and Embree, 2005) that dense random perturbations, of the form $\sigma J$ considered here, tend to trace out the entire $\epsilon$-pseudospectrum (where $\epsilon = \sigma \|J\| \approx 2\sigma$). This forms the basis for a numerical approach to estimating pseudospectra.

However, the present result shows that, for $\epsilon, \sigma \ll |w|$, the spectrum of such perturbations traces out the $\epsilon$-psuedospectrum in a highly uneven fashion. The vast majority ($\Theta(N)$) of the perturbed eigenvalues only trace out the boundary of the pseudospectrum, $|z| \approx |w|$, while only a few ($O(1)$) eigenvalues lie in its interior. Thus, dense random perturbations can fail as a way of visualizing (operator-norm based) pseudospectra.

### 5.4.3 Examples motivated by Dale’s law

In this section we consider examples motivated by Dale’s law in neurobiology. We will first consider two examples of connectivity matrices respecting Dale’s law which take the
Figure 5.4: The eigenvalue spectra of $A = M + \sigma J$ for $N = 2000$ and $M$ given by Eq. (5.144) with $\lambda_n = 0, w = 1$ for single realizations of real Gaussian $J$. $\sigma = 0.95$ and $0.5$ in the left and right panels, respectively. The red circles mark the circular boundaries of the spectral support given by Eq. (5.145). The insets show a comparison of the analytic formula Eq. (5.146) for the spectral density (black smooth trace) and histograms corresponding to the particular realization shown in the main plot (red jagged trace).

form Eq. (5.1) with $L = \sigma^{-1} R = 1$, and a scalar $\sigma$. At the end of this section, we consider an example with nontrivial $L$ and $R$.

$N/2$ feedforward chains of length 2

In the first example, we consider a matrix $M$ that, as we will show, has a Schur form $T$ composed of $N/2$ disjoint feedforward chains, each connecting only two modes (we assume $N$ is even). For simplicity we will focus on the case where all eigenvalues are zero.
Thus in the Schur basis we have

\[ T = \begin{pmatrix} 0 & w_1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & w_2 & \cdots \\ 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = W \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \]  

(5.147)

where we defined \( W \) to be the \( N/2 \times N/2 \) diagonal matrix of Schur weights \( W = \text{diag}(w_1, w_2, \ldots, w_{N/2}) \).

The matrix \( T \) in Eq. (5.147) arises as the Schur form of a mean matrix of the form

\[ M = \frac{1}{2} \begin{pmatrix} K & -K \\ K & -K \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \otimes K \]  

(5.148)

where \( K \) is a normal but otherwise arbitrary \( N/2 \times N/2 \) matrix. Note that \( M \) is nonetheless nonnormal. The feedforward weights in Eq. (5.147) are then the eigenvalue of \( K \).

When \( K \) has only positive entries, matrices of the form Eq. (5.148) satisfy Dale’s principle, and were studied in Murphy (2009), in the context of networks of excitatory and inhibitory neurons. We should imagine here a grid of \( N/2 \) spatial positions, with an excitatory and an inhibitory neuron at each position. \( \frac{1}{2} K \), a matrix with positive entries, describes the mean synaptic strength between spatial positions, which is taken to be identical regardless of whether the projecting, or receiving, neuron is excitatory or inhibitory. The sign of the weight, on the other hand, depends on the excitatory or the inhibitory nature of the pro-
jecting or presynaptic neuron; the first (last) \( N/2 \) columns of \( M \) represent the projections of the excitatory (inhibitory) neurons and are positive (negative).

Since \( K \) is normal it can be diagonalized by a unitary transform: \( K = EWE^\dagger \), where \( W \) is as above, and \( E = (e_1, e_2, \ldots) \) is the matrix of the orthonormal eigenvectors \( e_b \) of \( K \), \( b = 1, \ldots, N/2 \) (with eigenvalues \( w_b \)). Then transforming to the basis

\[
\begin{pmatrix}
(e_1) & (0) & (e_2) & (0) & \cdots & (e_{N/2}) & (0) \\
(0) & (e_1) & (0) & (e_2) & \cdots & (0) & (e_{N/2})
\end{pmatrix}
\]

(where \( \mathbf{0} \) represents the \( N/2 \)-dimensional vector of 0’s) transforms the matrix to being block-diagonal with the \( 2 \times 2 \) matrices

\[
\frac{1}{2} \begin{pmatrix} w_b & -w_b \\ w_b & -w_b \end{pmatrix}
\]

along the diagonal, for \( b = 1, \ldots, N/2 \).

The \( b \)-th block becomes

\[
\begin{pmatrix} 0 & w_b \\ 0 & 0 \end{pmatrix}
\]

in its Schur basis

\[
\left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} e_b \\ e_b \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} e_b \\ -e_b \end{pmatrix} \right\}
\]

so the full matrix takes the form Eq. (5.147).

Thus, the \( b \)-th difference mode

\[
\frac{1}{\sqrt{2}} \begin{pmatrix} e_b \\ -e_b \end{pmatrix}
\]

feeds forward to the \( b \)-th sum mode

\[
\frac{1}{\sqrt{2}} \begin{pmatrix} e_b \\ e_b \end{pmatrix}
\]

with weight \( w_b \).\(^7\)

For a general diagonal \( W \) in Eq. (5.147) (or equivalently, for general normal \( K \) in Eq. 5.148), the eigenvalue density, \( \rho(z) \), of \( A = M + \sigma J \) is isotropic around the origin \( z = 0 \), and depends only on \( r = |z| \). The spectral support is a disk centered at the origin.

\(^7\)This feedforward structure leads to a specific form of nonnormal transient amplification, which Murphy and Miller (2009) named “balanced amplification”; small differences in the activity of excitatory and inhibitory modes feedforward to and cause possibly large transients in modes in which the excitatory and inhibitory activities are balanced (see Chapter 6).
Figure 5.5: The eigenvalue spectra of $A = M + \sigma J$ for a binary $J$ with $\sigma = 0.1$ and $M$ given by Eq. (5.148) with $K = 1$ (corresponding to $w_b = 1$ for all the diagonal $2 \times 2$ blocks in Eq. 5.147). The main panels show the eigenvalues for single realizations of $J$, with $N = 600$ (left) and $N = 60$ (right). The red circles mark the boundaries of the spectral support, Eq. (5.149). Since $A$ is real in this case, its eigenvalues are either exactly real, or come in complex conjugate pairs; the spectrum is symmetric under reflections about the real axis. However, such signatures of the reality of the matrix appear only as subleading corrections to the spectral density $\rho(z)$; they are finite size effects which vanish as $N \to \infty$. The insets show a comparison of the analytic formula Eq. (5.150) (black curve) and the empirical result, based on the eigenvalues of the realizations in the main panels, for the proportion, $n_<(r)$, of eigenvalues lying within a radius $r$ of the origin (red dots). The random fluctuations and the average bias of the empirical $n_<(r)$ are both already small for $N = 60$, and negligible for $N = 600$.

In cases in which all the weights $w_b$ are $O(1)$, the radius of this disk can be found directly from Eq. (5.4), which yields

$$r_0 = \sigma \left[ \frac{1}{2} + \sqrt{\frac{1}{4} + \frac{\langle |w_b|^2 \rangle_b}{2\sigma^2}} \right]^{1/2}.$$  \hspace{1cm} (5.149)

Here, $\langle |w_b|^2 \rangle_b$ is the average of the squared feedforward weights over all blocks of Eq. (5.147); equivalently, $\langle |w_b|^2 \rangle_b = 2 \text{tr}(M^T M) \equiv 2\mu^2$. As long as some $w_b$ are nonzero, $r_0$ is larger than the radius of the circular law, $\sigma$, with the difference an increasing function
Figure 5.6: The eigenvalue spectra of $A = M + \sigma J$ for the $M$ given by Eq. (5.151) in the balanced case, $\mathbf{v}^T \mathbf{u} = 0$. Here, $N = 800$, $\sigma = 1$ and $\mu = 12$ (see Eq. 5.152). The black dots are the superimposed eigenvalues of $A$ for 20 different realizations of complex Gaussian $J$. The small red circle enclosing the vast majority of the eigenvalues has radius $\sigma = 1$, corresponding to the standard circular law Eq. (5.153). A $\Theta(N)$ number of eigenvalues lie within this circle. A $\Theta(\sqrt{N})$ number lie just outside of this circle in a thin boundary layer which shrinks to zero as $N \to \infty$. Finally, a $\Theta(1)$ number of eigenvalues lie at macroscopic distances outside the unit circle. The dashed blue circle shows radius $r_0$ given by Eq. (5.149); outliers can even lie outside this boundary.

of $\langle |w_b|^2 \rangle_b$; thus the spreading of the spectrum of $M$ (originally concentrated at the origin) after the random perturbation by $\sigma J$, is larger the more nonnormal $M$ is.

In cases in which the feedforward weights of some of the $2 \times 2$ blocks of Eq. (5.147) grow without bound as $N \to \infty$, there is a corresponding singular value of $M_z \propto z - M$ for every such block which is nonzero for $z \neq 0$ but vanishes in the limit, scaling like $\sim \frac{|z|^2}{|w_b|}$ where $w_b$ is the unbounded weight of that block.\footnote{See Eq. (C.41) of Appendix C and its preceding paragraph. Also note that as stated after Eq. (5.2) we assume $\|M\|_F = \mu = \sqrt{\langle |w_b|^2 \rangle_b / 2}$ is $O(1)$, so that at most $o(N)$ number of weights can be unbounded, and...} In line with the general discussion after...
Eq. (5.11), in such cases the naive use of Eq. (5.4) may yield an area larger than the true support of \( \lim_{N \to \infty} \rho(z) \); the correct support must be found by using Eqs. (5.18)–(5.19), which in this case can yield a support radius strictly smaller than Eq. (5.149). We will see an example of this in the next subsection.

The results for \( \lim_{N \to \infty} \rho(z) \) can be calculated explicitly for the specific instance of \( M \) having the Schur form Eq. (5.147), if all the Schur weights have the same value which may be named \( w \). This example clearly belongs to the case of bounded \( w \)'s; hence, \( \lim_{N \to \infty} \rho(z) \) is here \( \Theta(1) \) within the entire disk \( r \leq r_0 \).

The eigenvalue density is found to be \( \rho(r) = \frac{1}{2\pi} \frac{\partial n_<(r)}{\partial r} \), where \( n_<(r) \) is the proportion of eigenvalues within a distance \( r \) from the origin and is given explicitly by

\[
n_<(r) = \frac{r^2}{\sigma^2} \left[ 1 - \frac{|w|^2}{\sigma^2 + \sqrt{\sigma^4 + |w|^4 + 4|w|^2 r^2}} \right]. \tag{5.150}
\]

Notice that \( n_<(r) \), as described by Eq. (5.150), reaches unity exactly for \( r = r_0 \) given by Eq. (5.149), while \( \rho(r) \) is \( \Theta(1) \) for any smaller \( r \). Figure 5.5 shows the close agreement of Eq. (5.150) with empirical results based on single binary realizations of \( J \), for \( N \) as low as 60.

**A minimal balanced network**

Another interesting network that complies with Dale’s law is that in which \( M \) simply captures the differences between the mean inhibitory and mean excitatory synaptic strengths and between the numbers of excitatory and inhibitory neurons, with no other structure as-
sumed, as was studied in Rajan and Abbott (2006). Thus, all excitatory projections have the same mean \( \mu_E/\sqrt{N} \), and all inhibitory ones have the mean \(-\mu_I/\sqrt{N}\).

Calling \( f \) the fraction of neurons that are excitatory, we can write \( M \) as

\[
M = uv^T
\]

(5.151)

where \( u = N^{-1/2}(1, \ldots, 1)^T \) is a unit vector, and the vector \( v \) has components \( v_i = \mu_E \) or \( v_i = -\mu_I \) for \( i \leq fN \) and \( i > fN \), respectively (for \( f = 1/2 \) and \( \mu_E = \mu_I \), Eq. (5.151) is a special case of Eq. (5.148)).

The single-rank matrix \( M \) has only one non-zero eigenvalue given by \( v \cdot u = \frac{1}{\sqrt{N}} \sum v_i \), with eigenvector \( u \). The case in which the excitatory and inhibitory weights are balanced on average, in the sense that \( \sum v_i = 0 \), is of particular interest; mathematically it is in a sense the least symmetric and most nonnormal case as \( v \cdot u = 0 \). In this case all eigenvalues of \( M \) are equal to zero.

Furthermore, since in this case \( u \) and \( v \) are orthogonal, we can readily read off the Schur decomposition of \( M \) from Eq. (5.151). The normalized Schur modes are given by \( u \), \( v/\|v\| \) and \( N - 2 \) other unit vectors spanning the subspace orthogonal to both \( u \) and \( v \). All feedforward Schur weights are zero, except for one very large weight, equal to \( \|v\| \propto \sqrt{N} \), which feeds from \( v/\|v\| \) to \( u \). Thus the Schur representation of \( M \) has the form Eq. (5.147) with \( w_1 = \|v\| \equiv \mu\sqrt{N} \) and \( w_{b \neq 1} = 0 \), where we defined

\[
\mu^2 \equiv \text{tr} (M^\dagger M) = \|v\|^2/N = f \mu_E^2 + (1 - f)\mu_I^2.
\]

(5.152)
Figure 5.7: The number of eigenvalues of $M + \sigma J$, for the $M$ given by Eq. (5.151), lying outside the circle of radius $\sigma$ vs. $N$ (red line). Here, $\sigma = 1$, $\mu = 12$ and $\mathbf{v} \cdot \mathbf{u} = 0$. The numbers (red points connected by solid red lines) are obtained by numerically calculating the eigenvalues and counting the outliers for 200 realizations of $J$, and taking the average of the counts over all realizations, for $N = 100, 200, 400, 800, 1600$ (error bars show standard error of mean). The black dashed line plots $\sqrt{N}$ for comparison with the theoretical result Eq. (5.154); the dashed blue line, which includes subleading corrections to $\sqrt{N}$, is obtained by numerically solving Eq. (C.46) and substituting the result in Eq. (C.47). These formulae will be given in Appendix C, together with their derivation from Eqs. (5.7)–(5.8).

This is a case of "balanced amplification": differences between excitatory and inhibitory activity, represented by $\mathbf{v}$, feed forward to balanced excitatory and inhibitory activity, represented by $\mathbf{u}$, with a very large weight. In the following we will present results only for this balanced case of Eq. (5.151), which as just noted is a special case of Eqs. (5.147).

The balanced network Eq. (5.151) with $\mathbf{u} \cdot \mathbf{v} = 0$ is a special case of the more general case of Eqs. (5.147, 5.148). However, unlike the example studied in Fig. 5.5, this example belongs to the case of unbounded $w_b$’s, where the limit density is only nonzero in a proper subset of the disk $r \leq r_0$.

Indeed, in this special case, all $w_b$ are zero in this case except for one very large, un-
bounded weight $w_1 = \mu \sqrt{N}$. The matrix $M_z \propto z - M$ has here, as we have mentioned, an $o(1)$ smallest singular value, approximately given by $\frac{|z|^2}{\mu \sqrt{N}}$.

Using Eqs. (5.18)–(5.19), one finds that the support of $\lim_{N \to \infty} \rho(z)$ is the disc with radius $\sigma$ (within the annulus $\sigma < |z| \leq r_0$ the eigenvalue density is $o(1)$), and solving Eqs. (5.7)–(5.8) for $|z| \leq \sigma$, we find that the spectral density is in fact identical to the circular law (the eigenvalue density for the $M = 0$ case), i.e.

$$\rho(r) = \begin{cases} \frac{1}{\pi \sigma^2} + o(1), & (r < \sigma) \\ o(1), & (r > \sigma). \end{cases}$$

(5.153)

It has been shown in Chafai (2010) that more generally, for any $M$ of rank $o(N)$ and bounded $\|M\|_F$, the eigenvalue density of $A = M + \sigma I$ is given by the circular law in the limit $N \to \infty$. For single rank $M$ (as in the present case) and a diagonal $R$, it was also shown in Wei (2012) that the eigenvalue density of $M + JR$ agrees with that of $JR$ as $N \to \infty$.

In the present example, it was observed in Rajan and Abbott (2006) that even though the majority of the eigenvalues are distributed according to the circular law, there also exist a number of “outlier” eigenvalues spread outside the circle $|z| = \sigma$, which unlike in the $M = 0$ case, may lie at a significant distance away from it (see Fig. 5.6). As we mentioned in Sec. 5.2.5, the non-crossing approximation cannot be trusted to correctly yield the $o(1)$ contributions to $\rho(z)$ by these outliers for $|z| > \sigma$. However, if we ignore this warning and use Eqs. (5.7)–(5.8), keeping track of finite-size, $o(1)$ contributions, we obtain results that
agree surprisingly well (although not completely) with simulations.

For the total number of outlier eigenvalues lying outside the circle \( |z| = \sigma \), we obtain

\[
N_\sigma(\sigma) \equiv N n_\sigma(\sigma) = \sqrt{N} + O(1)
\]  

(5.154)

where we defined \( n_\sigma(r) = 1 - n_\sigma(r) \) to be the proportion of eigenvalues lying outside the radius \( r \). See Fig. 5.7 for a comparison of Eq. (5.154) with simulations. The vast majority of the outlier eigenvalues counted in Eq. (5.154) lie in a narrow boundary layer immediately outside the circle \( |z| = \sigma \), the width of which shrinks with growing \( N \). In addition to these, however, there are a \( \Theta(1) \) number of eigenvalues lying at macroscopic, \( \Theta(1) \) distances outside the circle \( |z| = \sigma \).

We have used Eqs. (5.7)–(5.8) to calculate \( N_\sigma(r) \), the number of outlier eigenvalues lying outside radius \( r \) for \( r > \sigma \). Figure 5.8 shows a plot of \( N_\sigma(r) \) and compares it with the results of simulations for different \( N \). For roughly the inner half of the annulus \( \sigma < |z| < r_0 \), \( N_\sigma(r) \) agrees well with simulations. As \( r \) increases, however, it deviates significantly from the empirical averages. In particular, \( N_\sigma(r) \) calculated from Eqs. (5.7)–(5.8) vanishes at \( r_0 \) given by Eq. (5.149), while the empirical average of the number of outliers is nonzero well beyond \( r_0 \).

Finally, we note that the distribution of these eigenvalues is not self-averaging, and it can depend on the real vs. complex nature of the random matrix \( J \). In the real case, their distribution has been recently characterized as that of the inverse roots of a certain random power series with i.i.d. standard real Gaussian coefficients (Tao, 2013).
Figure 5.8: The number, $N_> (r)$, of outlier eigenvalues of $A = M + \sigma J$, for the $M$ given by Eq. (5.151), lying farther from the origin than $r$, as a function of $r$. Here, $\sigma = 1$, $\mu = 12$ and $\mathbf{v}^\top \mathbf{u} = 0$. The vertical line marks $|z| = r_0 \simeq 3.54$ where $r_0$ is given by Eq. (5.149). The colored connected points are $N_> (r)$ for realizations of $A$, based on 200 samples of $J$, each color for a different $N$, for $N = 100, 200, 400, 800, 1600$ and 3200 (error bars show standard error of sample mean). Note the lack of scaling of $N_> (r)$ with $N$.

A factorizable multipopulation network

The example Eq. (5.151) motivated by Dale’s law with neurons of either excitatory or inhibitory types can be generalized to a network of neurons belonging to one of $n$ different populations. These are meant to represent the various subtypes of excitatory or inhibitory that we will describe in Chapter 7. Not only the mean but also the variance of connection strengths may depend on the pre- and post-synaptic types. When this dependence is factorizable (in a way we will now describe) the connectivity matrix of such a network will be of the form Eq. (5.1) with non-trivial $L$ and $R$.

Let $\alpha(i) \in \{1, \ldots, n\}$ denote the type of neuron $i$, and let $f_\alpha$ denote the fraction of
neurons of type $\alpha$ (so $\sum_{\alpha=1}^{\#} f_{\alpha} = 1$). Assume $n$ and $f_{\alpha}$ are all $\Theta(1)$. Assume further that each synaptic weight is a product of a pre- and a post-synaptic factor, and that in each synapse these factors are chosen independently from the same distribution, except for a deterministic sign and overall scale that depend only on the type of the pre and post-synaptic neurons, respectively.

If $A_{ij}$ denotes the weight of the synaptic projection from neuron $j$ to neuron $i$, we have

$$A_{ij} = \frac{1}{\sqrt{N}}(l_{\alpha(i)}x_{ij})(r_{\alpha(j)}y_{ij})$$

(5.155)

where $x_{ij}$’s and $y_{ij}$ are positive random variables chosen i.i.d. from the distributions $P_x(x)$ and $P_y(y)$, respectively. Here, $l_{\alpha}$ and $r_{\alpha}$ determine the sign and the scale (apart from the overall $\frac{1}{\sqrt{N}}$) of the pre and post-synaptic factors of the neurons in cluster $\alpha$, respectively. Note that when all $l_{\alpha}$ are positive, $A_{ij}$ satisfies Dale’s law.

By absorbing appropriate constants into $l_{\alpha}$’s and $r_{\alpha}$’s, we can assume with no loss of generality that $\text{Var}[xy] = \langle x^2 \rangle \langle y^2 \rangle - \langle x \rangle^2 \langle y \rangle^2 = 1$. Then it is easy to see that $A$ can be cast in the form Eq. (5.1) with

$$L_{ij} = l_{\alpha(i)}\delta_{ij}$$

(5.156)

$$R_{ij} = r_{\alpha(j)}\delta_{ij}$$

(5.157)

$$J_{ij} = \frac{1}{\sqrt{N}}(x_{ij}y_{ij} - \xi)$$

(5.158)

$$M = sL uu^\top R$$

(5.159)
where \( \mathbf{u} \) is the unit vector \( \frac{1}{\sqrt{N}}(1, \ldots, 1)^T \),

\[
s \equiv \xi \sqrt{N},
\]

and \( \xi \equiv \langle x \rangle \langle y \rangle \) is dimensionless and \( \Theta(1) \) (note that \( J \), given by Eq. (5.158), indeed has i.i.d. elements with zero mean and variance \( N^{-1} \)).

Being single-rank, \( M \) has \( N - 1 \) zero eigenvalues; its only (potentially) non-null eigenvector is \( L\mathbf{u} \), with a generically large eigenvalue

\[
\lambda_M = s\mathbf{u}^T R L \mathbf{u} = s \frac{1}{N} \sum_{i=1}^{N} r_{\alpha(i)} l_{\alpha(i)} = \xi \sqrt{N} \langle \sigma_\alpha \rangle \tag{5.161}
\]

where we defined

\[
\begin{align*}
\sigma_\alpha & \equiv l_{\alpha r_\alpha}, \\
\langle X_\alpha \rangle & \equiv \sum_{\alpha=1}^{\nu} f_\alpha X_\alpha. 
\end{align*}
\]

As for the example Eq. (5.151), we will focus on the balanced case in which \( \lambda_M \propto \langle \sigma_\alpha \rangle \neq 0 \). From Eq. (5.159), \( M = \tilde{\mathbf{u}} \tilde{\mathbf{v}}^T \) with \( \tilde{\mathbf{u}} = L\mathbf{u} \) and \( \tilde{\mathbf{v}} = sR\mathbf{u} \). The balanced condition is equivalent to \( \tilde{\mathbf{u}} \cdot \tilde{\mathbf{v}} = 0 \) (see Eq. 5.161). Thus, similar to Eq. (5.151), the Schur representation of \( M \) has the form (5.147) with \( w_1 = \|\tilde{\mathbf{u}}\|\|\tilde{\mathbf{v}}\| \) and \( w_b = 0 \) for \( b > 1 \).

In Appendix C it will be proven that, just as for Eq. (5.151), for the ensemble Eqs. (5.156)–(5.159) the limit of the eigenvalue distribution, \( \lim_{N \to \infty} \rho(z) \), is also not affected by the nonzero mean matrix Eq. (5.159). Hence we can obtain \( \lim_{N \to \infty} \rho(z) \) for that
example by safely setting $M$ to zero, and using formulae Eqs. (5.12)–(5.15) with $L$ and $R$ given by Eqs. (5.156)–(5.157). Thus $\lim_{N \to \infty} \rho(z)$ is isotropic and its support is the disk with radius

$$r_0 = \|RL\|_F = \sqrt{\langle \sigma^2 \rangle_\alpha}. \quad (5.164)$$

As in the previous example, when the balance condition $\langle \sigma \rangle_\alpha = 0$ holds, use of the naive formula Eq. (5.4) with $M = \tilde{u}\tilde{v}^T$ would have yielded

$$\bar{r}_0 = r_0 \left[ \frac{1}{2} + \sqrt{\frac{1}{4} + \xi^2} \right]^{1/2}, \quad (5.165)$$

which is larger than the correct result Eq. (5.164). As discussed above, this result is not correct, but it indicates the existence of $O(1)$ number of outlier eigenvalues lying outside the boundary of $\lim_{N \to \infty} \rho(z)$ given by Eq. (5.164).

For $r < r_0$, the $N \to \infty$ limit of the proportion, $n_>(r)$, of eigenvalues lying farther than distance $r$ of the origin is given by $g^2(r)$ which is found by solving Eq. (5.14), or equivalently

$$\langle \frac{1}{g^2 + \sigma^{-2} r^2} \rangle_\alpha = 1. \quad (5.166)$$

The results Eqs. (5.16)–(5.17) also hold, wherein the normalized sums over $i$ can be replaced with appropriate averages $\langle \cdot \rangle_\alpha$.

**Two-population case**

In the special case of two neuronal types ($n = 2$), a closed solution can be obtained for $n_>(r)$ and $\rho(r)$.
We will identify the two types with excitatory and inhibitory neurons. Moreover, here and in Chapters 6-7 we will use E and I as indices (instead of \( \alpha = 1 \) and 2) when considering two populations. Assuming that \( l_\alpha = 1, \sigma_E \equiv \sigma_1 > 0 \) and \( \sigma_I \equiv \sigma_2 < 0 \), the ensemble Eqs. (5.156)–(5.159) describes a synaptic connectivity matrix in which all excitatory (inhibitory) connections are i.i.d. with mean \( \xi \sigma_E N^{-\frac{1}{2}} \) \((-\xi |\sigma_I| N^{-\frac{1}{2}})\) and variance \( \sigma_E^2 N^{-1} (\sigma_I^2 N^{-1}) \). In this case, Eq. (5.166) yields a quadratic equation. Differentiating the solution of that equation with respect to \( r^2 \) we obtain the explicit result

\[
\rho(r) = \frac{\sigma_E^{-2} + \sigma_I^{-2}}{2\pi} \left[ 1 - \frac{(\sigma_E^{-2} + \sigma_I^{-2})r^2 - 1}{2} + \frac{r^2 - 2r^2}{\sigma_E^{-2} + \sigma_I^{-2}} \right] \sqrt{\frac{(\sigma_E^{-2} + \sigma_I^{-2})r^2 - 1}{4} + \frac{r^2(r^2 - r^2)}{(\sigma_E^{-2} + \sigma_I^{-2})^2}}
\]

(5.167)

This result was first obtained by Rajan and Abbott (2006), in a less compact formulation.

Figure 5.9 shows two examples of spectra for single realizations of matrices of the form Eq. (5.155), with three neural types \( n = 3 \), where \( x_{ij} \) and \( y_{ij} \), and hence \( J_{ij} \), have log-normal distributions. The insets compare \( n > (r) \) based on the numerically calculated eigenvalues, with those found by solving Eq. (5.166).

In the right panel, the normally distributed log \( J_{ij} \) have a higher standard deviation, and hence the distribution of \( J_{ij} \) has a heavier tail. The right panel’s inset demonstrates that the convergence to the universal, \( N \to \infty \) limit can be considerably slow when the distribution of \( J_{ij} \) is heavy-tailed.
Figure 5.9: The eigenvalue spectra of $A = M + LJR$ with $M$, $L$ and $R$ given by Eqs. (5.156)–(5.159) with neurons belonging to one of three different types ($n = 3$). The main panels show the eigenvalues for two particular realizations of $J$. In both panels, $N = 2000$, $f_1 = 0.6$, $f_2 = f_3 = 0.2$, $\lambda_1 = 1$, $\sigma_1 = r_1 = 0.76$, $\sigma_2 = r_2 = -0.57$, $\sigma_3 = r_3 = -1.71$ (so $\langle \sigma_\alpha \rangle = 0$ and $r_\alpha^2 = \langle \sigma_\alpha^2 \rangle = 1$), and $J_{ij}$ had real entries with log-normal distribution; in the left (right) panel, the normally distributed $\log_{10} J_{ij}$ had standard deviation 0.5 (0.75). The solid red circles mark the boundaries of the spectral support as given by Eq. (5.164), and the dashed blue circles show the radii given by Eq. (5.165). The insets compare $n_\lambda(r)$ based on the numerically calculated eigenvalues shown in the main panels (connected red dots), with that found by solving Eq. (5.166) (black curve). The right panel’s inset shows the empirically calculated $n_\lambda(r)$ for a single realization with the same ensemble parameters, but with $N = 8000$ (green connected dots lying slightly above the red connected circles); the convergence to the universal limit at $N \to \infty$ is significantly slower in the right panel in which the distribution of $J_{ij}$ had a considerably heavier tail.

5.4.4 Linearizations of nonlinear networks

The linearization of nonlinear firing rate equations for a recurrent neural network of $N$ neurons around some stationary background was discussed in Chapter 2. We start again from Eq. (2.9), which we rewrite here as

$$T \frac{d\mathbf{x}(t)}{dt} = -\mathbf{x}(t) + W \Phi \mathbf{x}(t) + \delta \mathbf{I}'(t),$$

(5.168)
with $\Phi = \text{diag}(f'(v_\ast))$, where $v_\ast$ is the assumed steady state.

Now suppose that the original connectivity matrix can be written as $W = \langle W \rangle + \delta W$, with a quenched disorder part that is an i.i.d. random matrix: $\delta W = \sigma J$. Then multiplying Eq. (2.9) by $T^{-1}$ can convert Eq. (2.9) into the form Eq. (2.12) with $\gamma = 0$ and $A = M + LJR$ with

$$M = T^{-1}(-\mathbf{1} + \langle W \rangle \Phi) \quad (5.169)$$

$$L = T^{-1} \quad (5.170)$$

$$R = \sigma \Phi \quad (5.171)$$

and input

$$\mathbf{I}(t) = T^{-1} \delta \mathbf{I}^v(t). \quad (5.172)$$

It should be reminded, at this point, that the mathematical problem we have been discussing arises not only in neurobiology, but also from certain types of networks studied in other disciplines.

Notably, biochemical and ecological networks have often a connectivity that is stochastic about some underlying mean connectivity represented by a nonnormal matrix. This can happen because biological knowledge imparts a great deal of structure to models of both biochemical (Jeong et al., 2000; Barabasi and Oltvai, 2004; Zhu et al., 2007; Vidal et al., 2011) and ecological networks (May, 1972; Camacho et al., 2002; Valdovinos et al., 2010; Vermaat et al., 2009; Guimera et al., 2010), and matrices characterizing such interactions are typically nonnormal.
Generalized Lotka-Volterra (GLV) equations (Hofbauer and Sigmund, 1998) used in modeling the dynamics of food webs provide an example. Let \( \mathbf{n}(t) = (n_1(t), \ldots, n_N(t))^T \) denote the vector of population sizes of \( N \) species. The GLV equations take the form

\[
\frac{d\mathbf{n}}{dt} = n_i(r_i + \sum_j W_{ij} n_j)
\]

or

\[
\frac{d\mathbf{n}}{dt} = \text{diag}(\mathbf{r} + \mathbf{W}\mathbf{n})\mathbf{n}
\]  

(5.173)

where \( r_i > 0 \) are the species’ intrinsic growth rates and \( \mathbf{W} \) is the interaction matrix. Linearizing Eq. (5.173) around a fixed point, \( \mathbf{n}_* \), yields again a linear system of the form Eq. (2.12) with \( \gamma = \mathbf{I}(t) = 0 \). Starting with the same simple model \( \mathbf{W} = \langle \mathbf{W} \rangle + \sigma \mathbf{J} \), we find that \( \mathbf{A} \) can be written in the form Eq. (5.1) with

\[
\mathbf{R} = \sigma \mathbf{I}, \quad \mathbf{L} = \text{diag}(\mathbf{n}_*),
\]

(5.174)

\[
\mathbf{M} = \text{diag}(\mathbf{r} + \mathbf{W}\mathbf{n}_*) + \mathbf{L}\langle \mathbf{W} \rangle.
\]  

(5.175)

Note that if no species is extinct in the fixed point, i.e. if all \( n_i* > 0 \), then \( \mathbf{M} = \mathbf{L}\langle \mathbf{W} \rangle \).

The necessary and sufficient condition for the stability of a fixed point (without any change in the external input) is that all eigenvalues of the corresponding \( \mathbf{A} \) have negative real parts. The formula for the boundary of the eigenvalue distribution, Eq. (5.4), can be applied in these cases to map out the region in parameter space (parameters here mean the time constants or intrinsic growth rates in \( T \) or \( \mathbf{r} \), or the connectivity parameters determining the random ensemble for \( \mathbf{W} \), i.e. \( \sigma \) and the parameters of \( \langle \mathbf{W} \rangle \)) in which a particular fixed point is stable.

Eq. (5.4) may be also used in this way to determine the phase diagram of a clustered net-
Figure 5.10: The eigenvalues (black dots) of $A = M + JR$, with $M$ and $R$ given by Eqs. (5.177)–(5.178) with $g = 0.01$, $a = 1.02$ and $N = 2000$. This matrix governs the dynamics of small perturbations away from a non-trivial random fixed point in a clustered network of neurons (see Eq. 5.176), studied in Stern et al., 2012. The cyan dots on the real line are the eigenvalues of $M$, and the red curve is the boundary of support of the eigenvalue distribution, as calculated numerically from Eq. (5.4).

work of neurons, in which intra-cluster connectivity is large, but inter-cluster connectivity is random and weak (Stern et al., 2012). Because of the strong intra-cluster connectivity, each cluster behaves as a unit with a single self-coupling $a$. Letting the random inter-cluster couplings between $N$ clusters have zero mean and variance $g^2 / N$, one may start the analysis from the equation

$$\frac{d\mathbf{v}(t)}{dt} = -\mathbf{v}(t) + a \tanh(\mathbf{v}(t)) + gJ \tanh(\mathbf{v}(t))$$  

(5.176)

where $J$ is an i.i.d. random matrix as above. Here, $\mathbf{v}$ is a vector whose $\alpha$-th component is the mean voltage of cluster $\alpha$, while the nonlinear function $\tanh(\mathbf{v}(t))$ (with the hyperbolic
tangent acting component-wise) represents the vector of mean firing rates of the clusters.

Eq. (5.176) was the starting point for the analysis of Stern et al. (2012), who showed that there is a region of the phase plane \((a, g)\) where the self-connectivity, \(a\), is excitatory and sufficiently strong, in which the system eventually relaxes to non-zero random attractor fixed points \(v_s\); for smaller values of \(a\), the dynamics is chaotic (chaos in the \(a = 0\) case was established in Sompolinsky et al., 1988).

The form of these fixed points (the distribution of the elements of \(v_s\) as \(N \to \infty\) for a given \((a, g)\)) can be obtained using mean-field theory, and the linearization about \(v_s\) leads to an equation in the form of Eq. (2.12), with \(A = M + JR\), where \(M\) and \(R\) are the diagonal matrices

\[
M = \text{diag}(-1 + a \tanh'(v_s)) \quad (5.177)
\]

\[
R = \text{diag}(g \tanh'(v_s)). \quad (5.178)
\]

Given this form, it can be shown that the fixed point \(v_s\) is stable if \(z = 0\) is outside and to the right of the spectrum of the Jacobian matrix of the linearization, \(A\). The mean field solution for \(v_s\) determines the statistics of the elements of \(R^2M^{-2}\) for a given \((a, g)\). From these it can be determined if \(z = 0\) is outside the spectrum using formula Eq. (5.4) for the boundary of spectrum, which yields the requirement \(\text{tr} \left( \frac{R^2}{M^2} \right) < 1\).

The region of stability of the fixed points in the \((a, g)\) plane was mapped in Stern et al. (2012). Figure 5.10 shows a numerical example of the eigenvalue distribution for \(A\) for a given \((a, g)\) and the superimposed boundary calculated using Eq. (5.4).

Note a potential caveat in the applicability of the formulae in this Chapter to the lin-
earization analysis of systems like Eq. (2.9) and Eq. (5.173). We have derived the above formulae by assuming that $M$, $L$ and $R$ are independent of $J$. However, $M$ and $R$ as given by Eqs. (5.169) and (5.171) (or $M$ and $L$ in Eqs. (5.174)–(5.175)) depend on $J$ via their dependence on $v_s$ ($n_s$). Fortunately, this dependence is often too weak and indirect to render the present formulae inapplicable; an example is provided by the excellent agreement of the empirical spectrum and the red boundary given by formula Eq. (5.4) in Fig. 5.10, which also held for other parameter choices of the model of Stern et al. (2012).

Finally note that, although we assumed here the existence of a stable fixed point around which linearization is possible, the Jacobian of the dynamical system (expressed in terms of the randomly distributed eigenvalues) is a relevant quantity also for studying instances of chaotic dynamics of a network. These are known not only to exist in the phase space of model (2.2), but also to possibly provide optimal computational performance for brain function (Toyoizumi and Abbott, 2011).

### 5.5 Conclusions

This chapter has been devoted to the spectra of synaptic weight matrices that are stochastic about some underlying mean connectivity represented by a nonnormal matrix. The stochasticity, furthermore, may not be i.i.d. across elements of the matrix.

We have addressed the problem by studying large random $N \times N$ matrices of the form $A = M + LJR$, where $M$, $L$ and $R$ are arbitrary deterministic matrices and $J$ is a random matrix of zero-mean independent and identically distributed elements. $M$ can be nonnormal, and $L$ and $R$ allow correlations that have separable dependence on row and column
indices.

We proved the compactness of the eigenvalue density for large $N$, drew analytically the shape of the supports, and obtained explicit results for the density profiles. We have also provided a correct regularizing procedure for finding the support of the eigenvalue density for certain highly nonnormal instances of mean connectivity. These results are universal, i.e. only dependent on the first two moments of the distribution of disorder (although convergence to the large-$N$ limit is found to be slower with heavy-tailed distributions).

We have also considered the persistence as $N \to \infty$ of a finite number of randomly distributed outlying eigenvalues outside the support of the eigenvalue density of $A$. We have argued, and demonstrated numerically, that these arise in regions of the complex plane $\Omega$ where there are nonzero singular values of $L^{-1}(z\mathbf{1} - M)R^{-1}$ (for $z \in \Omega$) that vanish as $N \to \infty$.

When such singular values do not exist and $L$ and $R$ are equal to the identity, there is a correspondence in the normalized Frobenius norm (but not in the operator norm) between the support of the spectrum of $A$ for $J$ of norm $\sigma$ and the $\sigma$-pseudospectrum of $M$. We also showed that using random matrices to approximate pseudospectra may be inadequate because of the way the spectral density is distributed.

The results have been applied, finally, to a number of network types that are of interest to neurobiology.
Disordered neural dynamics

6.1 Dynamical functions of non-normality

We have seen in Chapter 5 how non-normality affects the spectral distribution of linear systems. Non-normality can also lead to important dynamical properties not seen with normal matrices.

The first such property is transient amplification, a simple but far-reaching phenomenon encountered in dynamical systems described by Eq. (2.12) with \( I(t) = 0 \) and governed by a nonnormal matrix \( A \).

In any stable system, the size of the response to an initial perturbation eventually decays to zero, with an asymptotic rate set by the system’s eigenvalues. In stable nonnormal systems, however, after an initial perturbation, the size of the network activity (as measured, e.g., by its norm squared \( \|x(t)\|^2 = x(t)^\top x(t) \)) can nonetheless exhibit a transient growth that may be large and long-lasting before it eventually decays to zero. Thus, a small perturbation from a fixed point of a stable system with nonnormal connectivity can lead to a large transient response over finite time. By contrast, in stable normal systems, \( \|x(t)\|^2 \) can only decrease with time.

The strength and even the time scale of transient amplification are set by properties
of the matrix $A$ beyond its eigenvalues; they depend on the degree of nonnormality of
the matrix, as measured, e.g., by the degree of non-orthogonality of its eigenvectors, or
alternatively by its hidden feedforward structure (see Eq. (5.143) for the latter’s definition).

Indeed, at the root of this there lies the fact that networks with a recurrent connectivity
pattern described by a nonnormal matrix can be regarded as having a hidden feedforward
connectivity structure between orthogonal activity patterns, each of which can also excite
or inhibit itself. In neural networks, as we have seen in Chapter 5, such hidden feedfor-
ward connectivity arises from the natural separation of excitatory and inhibitory neurons.
Transient amplification in neural network gives rise to an effect known as ”balanced am-
plification” (Murphy and Miller, 2009): small differences in the activity of excitatory and
inhibitory modes feedforward to and cause possibly large transients in modes in which the
excitatory and inhibitory activities are balanced.

This effect has been conjectured to play multiple functional roles in neural systems.
Murphy and Miller, from their simulations of cat visual cortex, have pointed out that tran-
sient amplification in cortical structures lends itself to amplifying steady-state responses
in a pattern-selective manner; thus, through balanced amplification certain patterns can be
amplified efficiently without any dynamical slowing.

Dixon et al. (2016) have shown that transient amplification plays an essential role in
determining baseline noise levels in stable intracellular networks; even simple systems are
capable of amplifying small fluctuations orders of magnitude beyond the levels predicted
by linear stability theory (see also Hennequin, 2012 and the references therein).

Goldman (2009) showed that networks containing long hidden feedforward chains can
generate long time scales and provide a substrate for working memory. It is known, more-
over, that systems with nonnormal connectivity can exhibit pseudo-resonance frequencies in their power-spectrum at which the system responds strongly to external inputs, even though the external frequency is not close to any of the system’s natural frequencies as determined by its eigenvalues (Trefethen and Embree, 2005). Such pseudo-resonances will be manifested in the frequency power spectrum of the response of the system to time dependent inputs.

Finally, transient amplification yields unexpected results also in ecological networks (Neubert and Caswell, 1997; Chen and Cohen, 2001; Tang and Allesina, 2014) and has been conjectured to play a key role in many biochemical systems (McCoy, 2013).

The treatment offered in the previous chapter was restricted to the spectrum of a stochastic weight matrix. For ensembles of matrices that are strongly nonnormal, it is clear that the eigenvalues do not suffice to specify the dynamics induced in a network. In this chapter, therefore, we will move on to focusing on the dynamics directly and will study quantities that provide insight into the time evolution of the system.

In Sec. 6.2, some general results will be stated for the magnitude of the response to a delta function pulse of input (which provides a measure of the time-course of potential transient amplification) and for the frequency power spectrum of the system’s response to external noise and to sinusoidal inputs. The average norm squared $\| x(t) \|^2$ and the power spectrum of response are examples of quantities that depend not only on the eigenvalues of $M + LJ R$, but also on its eigenvectors; the diagrammatic procedure for their calculation is described in Sec. 6.3. Sec. 6.4 will present the explicit results of analytical or numerical calculations based on these general formulae for some specific examples of $M$, $L$ and $R$. In Sec. 6.5, the further potential and limitations of the non-crossing approximation will
be discussed. In Sec. 6.6, we will consider disorder fluctuations of the covariance, which offer a possible window into the operational regime of networks. As much as possible, the treatment will be kept mathematically sparing to focus on the significance of results. Details of all calculations are given in Appendices C and D.

### 6.2 General results

#### 6.2.1 Preliminary remarks

In this section we will state the main general results of this chapter, which are valid for arbitrary \( M, L \) and \( R \). These include a formula for \( \langle \| x(t) \|_2^2 \rangle \), i.e. the ensemble average of the norm squared of the state vector, \( x(t) \), as it evolves under Eq. (2.12) with \( I(t) = 0 \), as well as a formula for the ensemble average of the power spectrum of the response of the network to external noisy and to sinusoidal inputs.

The results of this section are valid (and in the case of the power spectrum, meaningful) when the system Eq. (2.12) is asymptotically stable. As mentioned after Eq. (2.12), this means that \( M, L, R \) and \( \gamma \) must be chosen such that for any typical realization of \( J \), all eigenvalues of \(-\gamma I + M + LJR\) have negative real part.

In particular, the entire support of the eigenvalue density of \( M + LJR \), as determined by Eq. (5.4), must fall to the left of the vertical line of \( z \)'s with real part \( \gamma \). This necessary condition, however, may not be sufficient either at finite \( N \) or in cases where an \( O(1) \) number of eigenvalues remain outside this region of support even as \( N \to \infty \).
6.2.2 Result for the average squared norm

First, we consider the time evolution of the squared norm, \( \| x(t) \|^2 \), of the response of the system to an impulse input, \( I(t) = x_0 \delta(t) \), at \( t = 0 \), before which we assume the system was in its stable fixed point \( x = 0 \). This is of course equivalent, for \( t > 0 \), to the squared norm of the activity as it evolves according to Eq. (2.12) with \( I(t) = 0 \), starting from the initial condition \( x(0) = x_0 \).

In fact, a formula will be provided for the ensemble average of the more general quadratic function, \( x(t)^T B x(t) \), where \( B \) is any \( N \times N \) symmetric matrix; the norm squared corresponds to \( B = 1 \). The result for general \( B, M, L \) and \( R \) is given as a double inverse Fourier transform

\[
\langle x(t)^T B x(t) \rangle_J = \int \int \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} e^{it(\omega_1 - \omega_2)} \text{Tr}[B C_x(\omega_1, \omega_2; x_0 x_0^T)] , \quad (6.1)
\]

in terms of the \( N \times N \) Fourier-domain covariance matrix \( C_x(\omega_1, \omega_2; x_0 x_0^T) \equiv \langle \tilde{x}(\omega_1) \tilde{x}(\omega_2)^T \rangle_J \) (where \( \tilde{x}(\omega) \) is the Fourier transform of \( x(t) \)). The expression for the latter is given by

\[
C_x(\omega_1, \omega_2; C^I) = C_0^x(\omega_1, \omega_2; C^I) + \Delta C_x(\omega_1, \omega_2; C^I) \quad (6.2)
\]

where

\[
C_0^x(\omega_1, \omega_2; C^I) \equiv \frac{1}{\gamma + i\omega_1 - M^I} C^I \frac{1}{\gamma - i\omega_2 - M^I} , \quad (6.3)
\]

yields the result obtained by ignoring the randomness in the connectivity (i.e. by setting
\[ A = M \] and

\[
\Delta C^x(\omega_1, \omega_2; C^d) \equiv \frac{1}{\gamma + i\omega_1 - M} LL^\dagger \frac{1}{\gamma - i\omega_2 - M^\dagger} \text{tr}\left(R^\dagger R \frac{1}{\gamma + i\omega_1 - M} C^d \frac{1}{\gamma - i\omega_2 - M^\dagger}\right) \tag{6.4}
\]

is the contribution of the random part of connectivity \( LJR \).

For later use, these expressions have been provided with a general third argument in \( C^x(\cdot, \cdot, \cdot) \); for use in Eq. (6.1), \( C^d \) must be substituted with \( x_0x_0^\dagger \). In the special case of \( \langle \| x(t) \|^2 \rangle_J \) corresponding to \( B = 1 \), and i.i.d. disorder \( (L = 1, R = \sigma 1) \), the contributions from Eqs. (6.3)–(6.4) can be more compactly combined into

\[
\langle \| x(t) \|^2 \rangle_J = \int \int d\omega_1 \frac{d\omega_2}{2\pi} e^{i(t\omega_1 - \omega_2)} \frac{x_0^\dagger \frac{1}{\gamma - i\omega_2 - M^\dagger} \frac{1}{\gamma + i\omega_1 - M} x_0}{1 - \sigma^2 \text{tr}\left(\frac{1}{\gamma + i\omega_1 - M} \frac{1}{\gamma - i\omega_2 - M^\dagger}\right)} \tag{6.5}
\]

where the numerator was written using \( \text{Tr}\left(\frac{1}{z_1 - M} x_0x_0^\dagger \frac{1}{z_2 - M^\dagger}\right) = x_0^\dagger \frac{1}{z_1 - M} \frac{1}{z_2 - M^\dagger} x_0 \).

### 6.2.3 Results for the power spectrum

**Response to a noisy input**

Another quantity of interest is the power spectrum of the response of the system to a noisy input, \( I(t) \), that is temporally white, with zero mean and covariance

\[
I_i(t_1)I_j(t_2) = \delta(t_1 - t_2)C^d_{ij}; \tag{6.6}
\]

where the bar indicates averaging over the input noise (or by ergodicity, over a long enough time). The input is then a Wiener process, and the activity \( x(t) \) becomes an Ornstein-
Uhlenbeck process.

The matrix power spectrum of the response is, by definition, the Fourier transform of the steady-state response covariance,

$$C_{ij}^\omega(\omega) \equiv \int d\tau e^{-i\omega\tau} \overline{x_i(t+\tau)x_j(t)}, \quad (6.7)$$

and the ensemble average of this covariance matrix can be written as

$$\langle C^\omega(\omega) \rangle_J = C_0^\omega(\omega) + \Delta C^\omega(\omega). \quad (6.8)$$

where we have defined

$$C_0^\omega(\omega) \equiv C_0^\omega(\omega; C^I) \quad (6.9)$$

as the power spectrum matrix obtained by ignoring the randomness in the connectivity (i.e. by setting $A = M$), and

$$\Delta C^\omega(\omega) \equiv \Delta C^\omega(\omega; C^I) \quad (6.10)$$

as the contribution of quenched randomness $LJR$ to the power spectrum.

**Response to a sinusoidal input**

A closely related quantity is the total power of the steady-state response of the system to a sinusoidal input $I(t) = I_0 \sqrt{2} \cos \omega t$ (the $\sqrt{2}$ serves to normalize the average power of $\sqrt{2} \cos \omega t$ to unity, so that the total power in the input is $\|I_0\|^2$). For such an input, the steady-state activity, which we may denote by $x_\omega(t)$, is also sinusoidal, with a possible
phase shift. By ”total power of the steady-state response” we mean here the time average of the squared norm of the activity, $\|x_\omega(t)\|^2$, where now the bar indicates temporal averaging (which we call total power because the squared norm sums the power in all components of $x_\omega(t)$).

As in Eqs. (6.1)–(6.4), a formula is presented for the ensemble average of the more general quantity $x^T_\omega B x_\omega$. We have

$$\langle x^T_\omega B x_\omega \rangle_J = \text{Tr}(B \langle C^\omega(\omega) \rangle_J), \quad (6.11)$$

where $\langle C^\omega(\omega) \rangle_J$ is given by Eqs. (6.8)–(6.10) with $C^I$ replaced by $I_0 I_0^T$.

For the special case of $B = 1$, corresponding to the total power of the response at frequency $\omega$, using Eq. (6.3)–(6.4) with $\omega_1 = \omega_2 = \omega$, this formula can be simplified into

$$\langle \|x_\omega\|^2 \rangle_J = \left\| \frac{1}{z - M} I_0 \right\|^2 + \left\| \frac{1}{z - M} L \right\|^2_\text{F} \left\| \frac{1}{z - M} I_0 \right\|^2_\text{F}, \quad (6.12)$$

where $z = \gamma + i\omega$, $\| \cdot \|$ denotes the vector norm, and $\| \cdot \|_\text{F}$ denotes the Frobenius norm defined in Eq. (5.2).

Finally, for the case that the random part of the matrix is i.i.d., i.e. $L = \sigma I$ and $R = 1$, one can further simplify Eq. (6.12) into

$$\langle \|x_\omega\|^2 \rangle_J = \frac{\| (\gamma + i\omega - M)^{-1} I_0 \|^2_\text{F}}{1 - \sigma^2 \| (\gamma + i\omega - M)^{-1} \|^2_\text{F}}. \quad (6.13)$$

The stability of the $x = 0$ fixed point guarantees the positivity of the expressions Eqs. (6.12)–(6.13) for the power spectrum. This is true because, as noted above, stabil-
ity requires that the support of the eigenvalue density of $A$ is entirely to the left of the vertical line $\text{Re}(z) = \gamma$. By the result Eq. (5.6) for that support, this can only be true if the denominators of the last terms in Eq. (6.12)–(6.13) are positive, which guarantees the positivity of the full expressions.

### 6.2.4 The amplifying effects of disorder

Note that the first term in Eq. (6.12) and the numerator in Eq. (6.13) represent the power spectrum in the absence of randomness, i.e. if $A$, in Eq. (2.12) is replaced with $M$. Thus, formulae (6.12)–(6.13) show that the correct average power spectrum is always strictly larger than the naive power spectrum obtained by assuming that random effects will “average out”. Furthermore, due to the denominators of the last terms in Eqs. (6.12)–(6.13), the power spectrum will be larger for frequencies where the support of the eigenvalue density, Eq. (5.6), is closer to the vertical line with $\text{Re}(z) = \gamma$.

Similar, but less precise statements can also be made about the strength of transient amplification using formulae (6.1)–(6.5) for the squared norm of the impulse response. One measure of the strength of transient amplification up to time $T$ is $\int_0^T \|x(t)\|^2 dt$. Integrating formulae Eq. (6.1) (with $B = 1$) or Eq. (6.5) over $t$, one obtains formulae for $\int_0^T \|x(t)\|^2 dt$ that are the same as Eqs. (6.1)–(6.5), except for the factor $e^{it(\omega_1 - \omega_2)}$ in the integrands of Eqs. (6.1) and (6.5) being replaced by $\frac{i[1-e^{iT(\omega_1 - \omega_2 + i\epsilon)}]}{\omega_1 - \omega_2 + i\epsilon}$ (with $\epsilon \to 0^+$). Due to the denominator in this factor (for $T$ sufficiently large the numerator is constant), the main contribution to the integrals over $\omega_1$ and $\omega_2$ should typically arise for $\omega_1 \approx \omega_2$.

On the other hand, note that for $\omega_1 = \omega_2$ the denominators in Eqs. (6.4)–(6.5) reduce
to those in Eqs. (6.12)–(6.13), with the connection to the support of the spectral density noted above. Thus this dominant contribution to $\int_0^T \| x(t) \|^2 dt$ must be larger, the closer the support of the eigenvalue density, Eq. (5.6), is to the vertical line with $\text{Re}(z) = \gamma$. This also suggests that, as in the case of the power spectrum, the strength of transient amplification would typically be underestimated if randomness of connectivity is ignored and only its ensemble average $M$ is taken into account in solving Eq. (2.12).

### 6.2.5 Conditions for applicability

Numerical simulations indicate that the quantities $\| x(t) \|^2$ and $\| x_0 \|^2$ are self-averaging in the large $N$ limit. For large $N$, $\| x(t) \|^2$ or $\| x_0 \|^2$ for any typical random realization of $J$ will be very close to their ensemble averages, given by Eq. (6.5) and Eq. (6.12) respectively, with the random deviations from these averages approaching zero as $N$ goes to infinity (see Fig. 6.4-6.6 below). This conclusion is also corroborated by rough estimations based on Feynman diagrams of the variance of fluctuations of these quantities for different realizations of $J$.

On the other hand, it should be noted that the general formulae presented above are valid only for cases where the initial condition, $x_0$, or the input structure, $I_0$ or $C^I$, are chosen independently of the particular realization of the random matrix $J$ (e.g., cases where $x_0$ is itself random but independent of $J$, or when $x_0$ is chosen based on properties of $M$, $L$ or $R$). In particular, these results do not apply to cases in which the initial condition or the input is tailored or optimized for the particular realization of the quenched randomness, $J$, in which case the true result could be significantly different from those given by the formulae in this
Many of the remarks we made about the results of Chapter 5 apply also to the formulas of the previous sections. This concerns in particular the linearization of nonlinear systems (Sec. 5.4.4). If the fixed point around which one linearizes is stable, the results on $\|x(t)\|^2$ and $\|x_\omega\|^2$ can be thought of as characterizing directly the temporal evolution and the spectral properties of the linear response of the nonlinear system Eq. (2.8) (Eq. 5.173) in the given fixed point to perturbations.

On the other hand, a problematic aspect of the results of this chapter concerns their dependence on details of the distribution of synaptic weights.

For the dynamical quantities we are discussing, there is so far no rigorous result equivalent to the universality theorems for the spectrum established in Tao et al. (2010). Empirically, from limited simulations, universal behavior has been found to hold also for the present quantities (see Figg. 6.4-6.6). However, it is quite possible that universality for these quantities might require the existence of higher moments beyond the second. That was found to be the case for the universality of other properties of random matrices; see Tao and Vu (2011).

In order to demonstrate in practice the universality of the results, the numerical examples that will be shown in this Chapter have mostly been computed with non-Gaussian and/or real $J$'s.
6.3 Derivation of the general results

6.3.1 Formulation in terms of diffusive kernel

In this section, the definition of the observables of interest will be written in a form that is amenable to diagrammatic treatment.

Statistics of the network’s free motion

We will first consider the system’s response to an impulse input, \( I(t) = x_0 \delta(t) \), at \( t = 0 \), before which we assume the system was at rest in its fixed point \( x = 0 \).

As mentioned above, it will be assumed that \( x = 0 \) is a stable fixed point, i.e. all eigenvalues of \(-\gamma I + A\) have negative real parts, or equivalently, all eigenvalues of \( A \) have real parts less than \( \gamma \). More precisely, we assume that as \( N \to \infty \), this will be the case almost surely, i.e. for any typical realization of \( J \); in particular, the vertical line of \( z \)'s with real part \( \gamma \) must be to the right of the support of \( \rho(z) \), the average eigenvalue density for \( A \) as found by solving Eq. (5.4). This means that \( x(t) \) decays exponentially as \( t \to \infty \), and therefore its Fourier transform, \( \tilde{x}(\omega) \equiv \int_{-\infty}^{\infty} e^{-i\omega t} x(t) dt = \int_{0}^{\infty} e^{-i\omega t} x(t) dt \), is well-defined.

Fourier transformation of Eq. (2.12) with \( I(t) = x_0 \delta(t) \) yields \( i\omega \tilde{x}(\omega) = (-\gamma + A)\tilde{x}(\omega) + x_0 \). Solving algebraically for \( \tilde{x}(\omega) \), one obtains \( \tilde{x}(\omega) = (\gamma + i\omega - A)^{-1}x_0 \), or using Eqs. (5.23)–(5.24), \( \tilde{x}(\omega) = R^{-1}G(\gamma + i\omega; J)L^{-1}x_0 \). The inverse Fourier transform, \( x(t) = \int_{-\infty}^{\infty} e^{i\omega t} \tilde{x}(\omega) \frac{d\omega}{2\pi} \), then yields

\[
x(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{i\omega t} R^{-1}G(\gamma + i\omega; J)L^{-1}x_0. \tag{6.14}
\]
The present goal is to study the statistics of $x(t)$ (i.e., its moments) under the distribution Eq. (5.22). Equation (6.14) allows us to reduce this task to the calculation of various moments of $G(z; J)$ and its adjoint, and these can be found using the diagrammatic technique.

Note that, in general, these moments involve not only the statistics of the eigenvalues, but also that of the eigenvectors of $A = M + LJR$; this can immediately be seen from the spectral representation

$$R^{-1}G(z; J)L^{-1} = (z - A)^{-1} = V(z - \Lambda)^{-1}V^{-1}$$ (6.15)

where $\Lambda$ is a diagonal matrix of the eigenvalues of $A$, and $V$ is the matrix whose columns are the eigenvectors of $A$.

**Average norm squared**

We will start by looking at the simplest interesting statistic involving the eigenvectors: the average square norm of the state vector, namely, $\langle \|x(t)\|^2 \rangle$. As discussed in Sec. 6.2, its study is also motivated by the fact that transient amplification due to nonnormality of $A$ manifests itself in the transient growth of $\|x(t)\|^2 = x(t)^\dagger x(t)$. With a slight generalization, a formula will be derived for the average of a general quadratic function $x(t)^\dagger Bx(t)$ where $B$ is any symmetric matrix; the norm squared corresponds to $B = 1$.

Using $x(t)^\dagger = x(t)^\dagger$ ($x(t)$ is real), the identity $x^\dagger B x = \text{Tr}(Bxx\dagger)$, and Eq. (6.14), one finds

$$x(t)^\dagger B x(t) = \int \int \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} e^{i(\omega_1 - \omega_2)t} \text{Tr} \left[ B_R G(\gamma + i\omega_1; J) C_z G^4(\gamma + i\omega_2; J) \right]$$ (6.16)
where \( C_L \equiv L^{-1}x_0x_0^TL^{-1} \) and \( B_R \equiv R^{-1}BR^{-1} \).

Using Eq. (5.31) and \( G^\dagger(z; J) = -\lim_{\eta \to i0^\pm} G^{12}(\eta, z; J) \), and the \( 2 \times 2 \) matrices \( \pi^r \) defined in Eq. (5.39), one can rewrite the trace in Eq. (6.16) as
\[
\text{Tr}(\pi^2 \otimes B_R G(0, z_1; J) \pi^1 \otimes C_L G(0, z_2; J)),
\]
with \( z_i = \gamma + i\omega_i \), where now the trace is performed over \( 2N \times 2N \) matrices. Averaging over \( J \), one then obtains
\[
\langle x(t)^TBx(t) \rangle_J = \int \int \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} e^{i(t\omega_1 - \omega_2)} F(\gamma + i\omega_1, \gamma + i\omega_2; B, x_0x_0^T) \quad (6.17)
\]
where, for general matrix arguments \( B \) and \( C \), we have introduced the diffusive kernel
\[
F(z_1, z_2; B, C) \equiv \langle \text{Tr}(B G(0, z_1; J)C G(0, z_2; J)) \rangle_J. \quad (6.18)
\]

with
\[
B \equiv \pi^2 \otimes B_R, \quad B_R \equiv R^{-1}BR^{-1}, \quad (6.19)
\]
\[
C \equiv \pi^1 \otimes C_L, \quad C_L \equiv L^{-1}CL^{-\dagger}. \quad (6.20)
\]

Before proceeding with the calculation of \( F(z_1, z_2; B, C) \), it is convenient to reformulate also the other quantities presented in Sec. 6.2 so that they are expressed in terms of \( F(\gamma + i\omega, \gamma + i\omega; B, C) \), with appropriate \( B \)’s and \( C \)’s.

**Response to white noise**

Let us obtain the desired expression for the matrix power spectrum, Eq. (6.7), of the steady-state response to a temporally white noisy input \( I(t) \) with covariance Eq. (6.6). Using the
Fourier transform of Eq. (2.12), and following similar steps to those leading to Eq. (6.14),
we can write the steady-state solution for \( x(t) \) as in Eq. (6.14) with \( x_0 \) replaced by the Fourier
transform of the input, \( \tilde{I}(\omega) \). Using this and exploiting \( x_j(t_2) = \tilde{x}_j^*(t_2) \) we can write (after
averaging over the input noise)

\[
\overline{x_i(t_1)x_j(t_2)} = \int \int \frac{d\omega_1}{2\pi} \frac{d\omega_2}{2\pi} e^{i\omega_1 t_1 - i\omega_2 t_2} K_{ij}(\omega_1, \omega_2) \quad (6.21)
\]

where the Fourier-domain covariance matrix, \( K(\omega_1, \omega_2) \equiv \overline{\tilde{x}(\omega_1)\tilde{x}(\omega_2)^*} \), is given by

\[
K(\omega_1, \omega_2) \equiv R^{-1}G(\gamma + i\omega_1; J)L^{-1}C^I(\omega_1, \omega_2)L^{-1}G^\dagger(\gamma + i\omega_2; J)R^{-\dagger}. \quad (6.22)
\]

where the bars indicate averaging over the input noise distribution, and \( C^I(\omega_1, \omega_2) \equiv \overline{\tilde{I}(\omega_1)\tilde{I}(\omega_2)^*} \).

On the other hand, the Fourier transform of Eq. (6.6) yields

\[
C^I(\omega_1, \omega_2) \equiv \overline{\tilde{I}(\omega_1)\tilde{I}(\omega_2)^*} = 2\pi \delta(\omega_1 - \omega_2)C^I, \quad (6.23)
\]

where we also exploited \( \tilde{I}_j^*(\omega) = \tilde{I}_j(-\omega) \) for a real \( I(t) \). Substituting Eq. (6.23) into
Eqs. (6.21)–(6.22), we obtain

\[
\overline{x_i(t_1)x_j(t_2)} = \int \frac{d\omega}{2\pi} e^{i\omega(t_1-t_2)} C^x_{ij}(\omega), \quad (6.24)
\]

where

\[
C^x(\omega) = R^{-1}G(\gamma + i\omega; J)L^{-1}C^dL^{-1}G^\dagger(\gamma + i\omega; J)R^{-\dagger}. \quad (6.25)
\]
and noting that Eq. (6.24) expresses the covariance of the response as an inverse Fourier transform, we see that $C^x(\omega)$ is indeed the power spectrum of the response, as defined in Eq. (6.7).

Finally note that the element, $C_{ij}$, of any matrix can be expressed as $\text{Tr}(e_j e_i^T C)$, where $e_i$ are the unit basis vectors (i.e. vectors whose $\alpha$-th component is $\delta_{\alpha \alpha}$). Using this trick with Eq. (6.25), and following the steps leading from Eq. (6.16) to Eq. (6.17), we see that after ensemble averaging, $\langle C_{ij}^x(\omega) \rangle_J$ can be written in the form

$$\langle C_{ij}^x(\omega) \rangle_J = \mathcal{F}(\gamma + i\omega, \gamma + i\omega; e_j e_i^T, C^x)$$

(6.26)

where $\mathcal{F}$ was defined by Eqs. (6.18)–(6.20).

**Response to sinusoidal input**

Next, consider the system Eq. (2.12) being driven by a sinusoidal input $I(t) = I_0 \sqrt{2} \cos \omega t$ (the factor of $\sqrt{2}$ serves to normalize the time average of $(\sqrt{2} \cos \omega t)^2$ to one), and consider the steady state response, which will also oscillate at frequency $\omega$. Decomposing the input $I(t)$ and the steady-state response $x_\omega(t)$ into their positive and negative frequency components (proportional to $e^{i\omega t}$ and $e^{-i\omega t}$, respectively), from Eq. (2.12) we obtain

$$x_\omega(t) = \sqrt{2} R^{-1} \text{Re}[e^{i\omega t} G(\gamma + i\omega; J)] L^{-1} I_0.$$  

(6.27)

Thus the norm squared of the steady state response, $\|x(t)\|^2 = x(t)^\dagger x(t)$, will have a zero frequency component, plus components oscillating at $\pm 2\omega$. Averaging over time kills
the latter, leaving the zero-frequency component intact, yielding

\[ \overline{\mathbf{x}_\omega(t)\mathbf{x}_\omega(t)} = \mathbf{I}_0 L^{-1} G^\dagger(z; J) R^{-1} R^{-1} G(z; J) L^{-1} \mathbf{I}_0 = \text{Tr} \left( R^{-1} R^{-1} G(z; J) \rho_t G^\dagger(z; J) \right) \]

(6.28)

where \( z = \gamma + i\omega \), the bar indicates temporal averaging, and we defined \( \rho_t \equiv L^{-1} \mathbf{I}_0 \mathbf{I}_0^\dagger L^{-\dagger} \).

Generalizing to \( \overline{\mathbf{x}_\omega(t) B \mathbf{x}_\omega(t)} \), averaging over the ensemble, and following the steps leading from Eq. (6.16) to Eq. (6.17), we arrive at

\[ \langle \overline{\mathbf{x}_\omega(t) B \mathbf{x}_\omega(t)} \rangle_J = \mathcal{F}(\gamma + i\omega; \gamma + i\omega; B, \mathbf{I}_0 \mathbf{I}_0^\dagger), \]

(6.29)

where \( \mathcal{F} \) is given by Eqs. (6.18)–(6.20). Comparing Eq. (6.29) with Eq. (6.26), we also obtain

\[ \langle \overline{\mathbf{x}_\omega(t) B \mathbf{x}_\omega(t)} \rangle_J = \text{Tr} \left( B \langle \mathbf{C}_\omega \rangle_J \right) \]

(6.30)

which is Eq. (6.11) of Sec. 6.2.3, it being understood that \( \mathbf{C}^\dagger \) in Eq. (6.26) is replaced by \( \mathbf{I}_0 \mathbf{I}_0^\dagger \) as in Eq. (6.29).

### 6.3.2 Summing up the diffuson

Now that we have expressed all the quantities of interest in terms of the diffusive kernel \( \mathcal{F} \) as defined in Eq. (6.18), the remaining task boils down to performing the average over \( J \) in Eq. (6.18) so as to obtain a closed formula for \( \mathcal{F} \) with general arguments \( B \) and \( C \).

To this end, we will now proceed to calculate the more general object

\[ F_{\mu_1 \nu_2; \mu_2 \nu_1} (1; 2) \equiv \langle \mathbf{G}_{\mu_1 \nu_1} (1; J) \mathbf{G}_{\mu_2 \nu_2} (2; J) \rangle_J, \]

(6.31)
Figure 6.1: Contributions to Eq. (6.31) in the non-crossing approximation. The first line shows Eq. (6.31) written using the expansion Eq. (5.41). The diagram shows the contribution of the $m$-th and $n$-th terms in the expansion for two Green’s functions, respectively. Thus the top (bottom) solid line contains $m$ ($n$) factors of $J$, shown by dashed lines. In the large $N$ limit, averaging each summand over $J$ boils down to summing all non-crossing pairings (NCP) of the dashed lines. The second row shows a specific non-crossing pairing for the diagram shown in the first line. Finally, summing over all $m$ and $n$ and all NCP’s, is equivalent to replacing all solid lines (representing $G(\eta_i, z_i; J = 0)$) with thick solid lines representing the non-crossing average Green’s function, $\overline{G}(\eta_i, z_i)$ (calculated according to Eqs. 5.47–5.45), and summing over all NCP’s with every pairing connecting the straight lines on top and bottom (and not each to itself). This procedure yields the ladder diagrams, the sum over which is shown in the third line.

using the non-Hermitian diagrammatic technique. The abbreviated notation we used here is (1) $\equiv (\eta_1, z_1)$ and (2) $\equiv (\eta_2, z_2)$ for the function arguments, and $\mu_i = (\alpha_i, a_i)$ (similarly for $\nu_i$) for indices in the $2N$ dimensional space (as in Sec. 5.3, $\alpha, \beta, \ldots$, and $a, b, \ldots$ denote indices in the 2 and $N$ dimensional spaces, respectively).

Once $F_{\mu_1\nu_2;\mu_2\nu_1}(1; 2)$ has been calculated, one can obtain $F(z_1, z_2; B, C)$, with the appropriate $B$ and $C$, via

$$F(z_1, z_2; B, C) = B_{\nu_2\mu_1} F_{\mu_1\nu_2;\mu_2\nu_1}(0, z_1; 0, z_2) C_{\nu_1\mu_2}, \quad (6.32)$$
where all indices are summed over, and $B$ and $C$ were defined in Eqs. (6.19)–(6.20).

As in Chapter 5, one may start by using the expansion Eq. (5.41) for the two Green’s functions in Eq. (6.31). This is shown diagrammatically in the first line of Fig. 6.1, for the contribution of $m$-th and $n$-th terms in the expansion of the first and the second Green’s function, respectively. As before, for large $N$, averaging over $J$ entails summing the contribution of all non-crossing pairings. This is indicated in the second line of Fig. 6.1.

Finally, the third line of Fig. 6.1 shows that summing over all $m$’s, all $n$’s and all non-crossing pairings, is equivalent to replacing all solid lines with thick solid lines representing the average Green’s function in the non-crossing approximation, $\mathcal{G}(\eta_i, z_i)$ (defined diagrammatically in the third line of Fig. 5.1, and given by Eq. (5.54) as we found in the previous section), and summing over all non-crossing pairings with every pairing connecting the thick arrow lines on top and bottom (and not each to itself). This procedure yields a sum over all ladder diagrams with different number of rungs, as shown in the third line of Fig. 6.1.

As shown in the first row of Fig. 6.2, the sum of all ladder diagrams can be written as a sum

$$ F = F^0 + F^D, $$

where

$$ F^0_{\mu_1 \nu_2: \mu_2 \nu_1} (1; 2) \equiv \mathcal{G}_{\mu_1 \nu_1} (1) \mathcal{G}_{\mu_2 \nu_2} (2), $$

is the disconnected average of the two Green’s functions, and $F^D_{\mu_1 \nu_2: \mu_2 \nu_1} (1; 2)$ is the sum of ladder diagrams in which the two Green’s function are connected by at least one wavy line.
Figure 6.2: The first row is the diagrammatic representation of Eqs. (6.33)–(6.35). In the last term, $\rho$'s and $\lambda$'s are summed over. It shows the sum of all ladder diagram contributing to Eq. (6.31) (i.e. the last line of Fig. 6.1) in terms of the "Diffuson" $\mathcal{D}$, which is defined in the second row. The first term on the right side of the first row equation (the ladder with zero rungs) is the disconnected average Eq. (6.34); it corresponds to taking the average of each Green’s function in Eq. (6.31) separately and then multiplying. The last row shows an iterative form of the equation in the second row, which can be solved to give the expression Eqs. (6.37) and (6.40) for $\mathcal{D}$.

The latter can be written in the form

$$F_{\mu_1 \nu_2; \mu_2 \nu_1}^{(1; 2)} \equiv G_{\lambda_2 \nu_2} (2) G_{\mu_1 \rho_1} (1) \mathcal{D}_{\rho_1 \lambda_2; \rho_2 \lambda_1} (1; 2) G_{\lambda_1 \nu_1} (1) G_{\mu_2 \rho_2} (2),$$

(6.35)

where all repeated indices are summed over, and the propagator $\mathcal{D}$, is given by the sum of all diagrams in the second row of Fig. 6.2; this is the quantity known in the theory of many-body non-equilibrium as "Diffuson" (see e.g. Kamenev, 2011). Its role is similar to the role played by the self-energy for a single Green’s function (Sec. 5.3.1).

To compute the Diffuson, it will help to first rewrite Eq. (5.40) as

$$\langle \mathcal{J}^{a \beta} \mathcal{J}^{\gamma \delta} \rangle_f = \frac{1}{N} \sum_{r,s=1}^2 (\pi^{r}_{ad} \delta_{ad}) \sigma^1_{rs} (\pi^{a}_{\gamma \beta} \delta_{eb}),$$

(6.36)
where $\sigma^1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ is the first Pauli matrix. This helps us because in the expansion of Fig. 6.2, the two factors in Eq. (5.40) involving $\pi^r$ and $\pi^s$ decouple and get absorbed in adjacent loops, or contribute to form factors in the left or right ends of the ladder diagrams. This is demonstrated in Fig. 6.3 for the second term in the series expansion of the Diffuson as shown in the second line of Fig. 6.2.

Extending this similarly to all the terms in that expansion, we obtain

$$D_{\mu;\lambda,\nu}(1; 2) = D_{\mu;\lambda,\nu}^{\alpha \beta \gamma \beta}(1; 2) = \frac{1}{N} \sum_{r,s=1}^2 (\pi^r_{\alpha \delta} \pi^s_{\gamma \delta}) \ D_{rs}(1; 2) \ (\pi^s_{\gamma \delta} \delta_{eb})$$

(6.37)

where $\mu = (\alpha, a)$, $\nu = (\beta, b)$, $\lambda = (\gamma, c)$, $\rho = (\delta, d)$; here we have defined the $2 \times 2$ matrices

$$D(1; 2) \equiv \sigma^1 + \sigma^1 \Pi^1 \sigma^1 + \cdots = \sigma^1 \sum_{n=0}^{\infty} (\Pi^1 \sigma^1)^n,$$

(6.38)

and the “polarization matrix” for the Diffuson

$$\Pi_{rs}^1(1; 2) \equiv \text{tr}(\pi^r G(1) \pi^s G(2)) = \text{tr}(G^{rs}(1) G^{sr}(2)).$$

(6.39)

with the trace performed over the $2N$-dimensional space, where Eq. (5.39) was used to obtain the last equality. Summing the geometric series in Eq. (6.38) we obtain

$$D(1; 2) = \sigma^1 \left( I_{2 \times 2} - \Pi^1(1; 2) \sigma^1 \right)^{-1}.$$

(6.40)
Figure 6.3: The contribution to the matrix element $D^{\alpha_1\beta_1\gamma_1\beta_2}(1; 2)$ of the Diffuson from the second term in the series shown in the second row of Fig. 6.2, in more detail. The covariance of $J$ in the form Eq. (6.36) is used to write this expression in a more manageable form. The repeated indices, $r, t, u, s$, are summed over 1 and 2. The matrices inside the loop multiply each other in cyclic order, giving rise to the trace $\text{Tr}(G(2)\pi^t G(1)\pi^u)$. The whole diagram gives $\frac{1}{N} \sum_{rs} (\pi^r \otimes 1)_{\alpha_1d_1}^{\alpha_1d_1} [\sigma^1 \Pi^D \sigma^1]_{rs} (\pi^s \otimes 1)_{\beta_2d_2}^{\beta_2d_2}$ where the "polarization matrix" $\Pi^D_{\mu\nu}$ was defined in Eq. (6.39).

The $2 \times 2$ matrix inversion then yields

$$D(1; 2) = \frac{1}{(1 - \Pi^D_{12})(1 - \Pi^D_{21}) - \Pi^D_{11}\Pi^D_{22}} \begin{pmatrix} \Pi^D_{22} & 1 - \Pi^D_{12} \\ 1 - \Pi^D_{21} & \Pi^D_{11} \end{pmatrix} \tag{6.41}$$

where all $\Pi^D$’s have arguments $(1; 2) = (\eta_1, z_1; \eta_2, z_2)$ which were suppressed for clarity.

Going back to Eq. (6.32), we can also break up the kernel $F(z_1, z_2; B, C)$ into a disconnected part and a connected part mirroring the decomposition Eqs. (6.33)–(6.35):

$$F(z_1, z_2; B, C) = F^0(z_1, z_2; B, C) + \Delta F(z_1, z_2; B, C), \tag{6.42}$$

where $F^0(z_1, z_2; B, C)$ and $\Delta F(z_1, z_2; B, C)$ are defined as in Eq. (6.32), but with $F_{\mu_1\nu_2;\mu_2\nu_1}$ on the right side replaced by $F^0_{\mu_1\nu_2;\mu_2\nu_1}$ and $F^D_{\mu_1\nu_2;\mu_2\nu_1}$, respectively.

Using Eqs. (6.34)–(6.35) and (6.37), one obtains for the disconnected and connected
kernels:

\[
F^0(z_1, z_2; B, C) = \text{Tr}(BG(0, z_1)CG(0, z_2)) = \text{Tr}(B_RG^{21}(0, z_1)C_LG^{12}(0, z_2)), \quad (6.43)
\]

\[
\Delta F(z_1, z_2; B, C) = \frac{1}{N} \sum_{r,s} \text{Tr}(B_RG^{2r}(0, z_1)G^{r2}(0, z_2)) \times
\]
\[
\times D_{rs}(0, z_1; 0, z_2) \text{Tr}(G^{s1}(0, z_1)C_LG^{1s}(0, z_2)), \quad (6.44)
\]

where \(r\) and \(s\) are summed over \(\{1, 2\}\).

According to Eq. (6.17) we are interested in \(z_i = \gamma + i\omega_i \ (i = 1, 2)\) for arbitrary real \(\omega_i\). As mentioned before Eq. (6.14), these trace a vertical line in the complex plane that is entirely to the right of the support of the average eigenvalue density \(\rho(z)\) of \(A\), i.e. they are in the region where the valid solution of Eq. (5.53) is the trivial \(g(0, z) = 0\). In this case, we have Eq. (5.60), and for \(\eta \to i0^+\), from Eq. (5.30) (replacing \(A\) with \(M\), corresponding to \(J = 0\)) we have

\[
G(0, z_i) = - \begin{pmatrix}
0 & M^{-\dagger}_{z_1} \\
M^{-1}_{z_i} & 0
\end{pmatrix}.
\quad (6.45)
\]

Using this in Eqs. (6.43)–(6.44) we obtain

\[
F^0(z_1, z_2; B, C) = \text{Tr}(B_RM_{z_1}^{-1}C_LM_{z_2}^{-\dagger}), \quad (6.46)
\]

\[
\Delta F(z_1, z_2; B, C) = \text{tr}(B_RG^{21}(0, z_1)G^{12}(0, z_2)) \times
\]
\[
\times D_{12}(0, z_1; 0, z_2) \text{Tr}(G^{21}(0, z_1)C_LG^{12}(0, z_2)). \quad (6.47)
\]
Using the definitions Eq. (5.5) and Eqs. (6.19)–(6.20), one can simplify Eq. (6.46) to

$$F^0(z_1, z_2; B, C) = \text{Tr} \left( B \frac{1}{z_1 - M} C \frac{1}{z_2 - M^T} \right).$$

(6.48)

From Eqs. (6.39) and (6.45) we see that (for \( z_i \) of interest and for \( \eta_i \) going to zero) \( \Pi^{rr} = 0 \) and \( \Pi^{12} = \Pi^{21} \), and from Eq. (6.41) we obtain

$$D_{12}(0, z_1; 0, z_2) = \frac{1}{1 - \Pi_{21}^{D}(0, z_1; 0, z_2)} = \frac{1}{1 - \text{tr} (G^{21}(0, z_1)G^{12}(0, z_2))}.$$

(6.49)

Substituting this in Eq. (6.47) and using Eq. (6.45) once again, we finally obtain

$$\Delta F(z_1, z_2; B, C) = \frac{\text{tr} (B \frac{1}{z_1 - M} M^{-1} \frac{1}{z_2 - M^T}) \text{Tr} (M^{-1} C \frac{1}{z_2 - M^T})}{1 - \text{tr} (M^{-1} M^T)},$$

(6.50)

and after simplification using Eqs. (5.5) and (6.19)–(6.20),

$$\Delta F(z_1, z_2; B, C) = \frac{\text{tr} \left( B \frac{1}{z_1 - M} M^{-1} \frac{1}{z_2 - M^T} \right) \text{Tr} \left( \frac{1}{z_2 - M} LL^\dagger \frac{1}{z_2 - M^T} \right)}{1 - \text{tr} \left( R^\dagger R \frac{1}{z_1 - M} C \frac{1}{z_2 - M^T} \right)}.$$  

(6.51)

The general formulae of Sec. 6.2 readily follow. Equations (6.1)–(6.4), with \( C^I \) replaced by \( x_0 x_0^\dagger \), for the case of response to an impulse input follow from Eqs. (6.17), (6.42), (6.48) and (6.51), respectively. Equations (6.8)–(6.10) (with \( C^0 \) and \( \Delta C^x \) defined in Eqs. 6.3–6.4) for the power spectrum of the response to a temporally white noisy input, are similarly obtained from Eq. (6.26) by using Eqs. (6.42), (6.48) and Eq. (6.51), after setting \( B = e_y e_i^\dagger \), \( C = C^I \) and \( z_1 = z_2 = \gamma + i\omega \) (with the traces involving \( B = e_y e_i^\dagger \) turned into matrices in Eqs. (6.8)–(6.10), using \( \text{Tr} (e_y e_i^\dagger X) = X_{ij} \)). The result Eq. (6.11) for the steady state
response to a sinusoidal input was already derived in Eq. (6.30).

We see that according to Eqs. (6.17) and (6.42)

$$\langle x(t)^TBx(t) \rangle_J = \left[ x(t)^TBx(t) \right]_{J=0} + \Delta f_B(t),$$

(6.52)

where the two terms on the right hand side are obtained by replacing $\mathcal{F}(\cdot, \cdot; B)$ in Eq. (6.17) with Eq. (6.48) and Eq. (6.51), respectively. The integrals over $\omega_1$ and $\omega_2$ decouple for the first term yielding the expected result for $J = 0$,

$$\left[ x(t)^TBx(t) \right]_{J=0} = e^{-2\gamma t} \text{Tr} \left( Be^tM x_0 x_0^T e^{tM^T} \right) = x_0^T e^{t(-\gamma+M)^T} Be^{t(-\gamma+M)} x_0.$$  

(6.53)

Although it has been doable for the $J = 0$ contribution, performing in closed form the double Fourier transform of Eq. (6.17) needed for obtaining $\Delta f_B(t)$ is not possible for arbitrary $M$, $L$ and $R$. In the next section, this will be calculated analytically for some special examples of $M$, with $L$ and $R$ proportional to the identity matrix (i.e. for i.i.d. quenched randomness).

### 6.4 Applications to specific networks

This section will present the results of explicit calculations of the average squared norm of response to impulse, Eqs. (6.1) and (6.5), and the total power in response to sinusoidal input Eq. (6.13), for some specific examples of $M$, $L$ and $R$. All the derivations will be omitted; details of the calculations for these results can be found in Appendix C.

Once again, the feedforward structure of the Schur decomposition will be used to char-
acterize the examples. Due to the invariance of the trace, the norm, and the adjoint operation under unitary transforms, the general formulae for the average squared norm in time and frequency space, Eqs. (6.5) and (6.13), take the same form in any basis, so in particular we can work in the Schur basis of $M$.

Hence, $M$ can be replaced by its upper triangular Schur form $T$ (such that $M = UTU^\dagger$ with unitary $U$), provided $L$ and $R$ are also expressed in $M$’s Schur basis and $x_0$ or $I_0$ are replaced by $U^\dagger x_0$ or $U^\dagger I_0$, respectively. The unitary invariance of these formulae is in turn a consequence of both the invariance of $\|x(t)\|^2$ and the invariance of the statistical ensemble for $J$ (whence the invariance of $LJR$ when $L \propto R \propto 1$) under unitary transforms. Therefore, the feedforward structure of the Schur decomposition will be used to characterize the different examples to be considered.

In keeping with the overall approach advocated in the Introduction, the examples below are chosen so as to demonstrate, in their simplest possible settings, features of nonnormal matrices that are of interest to neurobiology.

### 6.4.1 Feedforward chain with alternating eigenvalues

We will begin by considering networks where each and every Schur mode is only connected to its lower adjacent mode, forming a long feedforward chain of length $N$. As we did in Chapter 5, for simplicity we take all feedforward weights in this chain to have the same
value $w$, so that $M_{nm} = w \delta_{n+1,m} + \lambda_n \delta_{nm}$, or

$$M = T = \begin{pmatrix} \lambda_1 & w & 0 & \cdots \\ 0 & \lambda_2 & w & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (6.54)$$

In particular, we consider here a mean synaptic weight matrix $M$ that has the form Eq. (6.54) with alternating imaginary eigenvalues, $\lambda_n = (-1)^{n+1}i$. The eigenvalues of this network were studied in Sec. 5.4.2. Figure 5.3 shows the ensemble average of the total power spectrum of response of the system Eq. (2.12) to sinusoidal stimuli ($\langle \|x_\omega\|^2 \rangle_J$) as
given by the general formula Eq. (6.13) (green curve).

As can be seen, the theoretical curve perfectly matches the empirical average (red curve). The latter was obtained by generating 100 realizations of $J$, calculating $\|x_0\|^2$ for each realization, which is given by the numerator of Eq. (6.13) with $M$ replaced by $M + \sigma J$, and then averaging the results over the 100 realizations.

The pink (light gray) shading shows the standard deviation of the power spectrum over these 100 realizations. This will shrink to zero as $N$ goes to infinity, so that for large $N$ the power spectrum of any single realization of $A = M + \sigma J$ will lie very close to the ensemble average.

The system (2.12) in the zero disorder case, $\sigma = 0$, has two highly degenerate resonant frequencies (imaginary parts of the eigenvalues of $M$), $\omega_0 = \pm 1$, leading to possible peaks in the power spectrum at these frequencies. The smaller the decay of these modes (in this case given by $\gamma$) is, i.e. the closer the eigenvalues of the combined matrix $-\gamma + M$ are to the imaginary axis, the sharper and stronger are the resonances.

Comparing the zero disorder power spectrum (blue curve) with that for $A = M + \sigma J$, we see that the disorder has led to strong but unequal amplification of the two resonances relative to the case without disorder. This is partly due to the disorder scattering some of the eigenvalues of $-\gamma + A$ much closer to the imaginary axis, creating larger resonances.

### 6.4.2 Feedforward chain with null eigenvalues

The second example we will consider concerns $M$ of the form (6.54) with all eigenvalues zero ($\lambda_n = 0$).
Figure 6.5: The norm squared of the response to impulse, $\|x(t)\|^2$, of the system Eq. (2.12), for $A = M + \sigma J$, with binary $J$, and $M$ given by Eq. (5.144) (with $\lambda_n = 0$) describing a $N$-long feedforward chain with uniform weights $w$. Here, $w = 1$, $\sigma = 0.5$, $\gamma = 1.005\sqrt{\sigma^2 + w^2} \simeq 1.124$, and $N = 700$. The dashed green curve shows the result Eq. (6.55) for the average squared impulse response, $\langle\|x(t)\|^2\rangle$, which lies on top of the solid red curve showing the empirical average of $\|x(t)\|^2$ over 100 realizations of binary $J$. The five thin dashed black curves show the result for five particular realizations of $J$, and the pink area shows the standard deviation among the 100 realizations. The standard deviation shrinks to zero as $N \to \infty$, and $\|x(t)\|^2$ for any realization lies close to its average for large $N$. For comparison the lowest, purple curve shows $\|x(t)\|^2$ obtained by ignoring the effect of quenched disorder, i.e. by setting $A = M$.

Both the magnitude of the response $x(t)$ to impulse Eq. (6.5) and the total power spectrum of steady-state response, Eq. (6.13), may be calculated analytically for this network in the case where the initial condition is (or the input is fed into) the last Schur mode, i.e. the beginning of the feedforward chain: $x_0 = (0, \cdots , 0, 1)^T$ (or $I_0 \propto (0, \cdots , 0, 1)^T$).

For the evolution of the average norm squared, with the initial condition $x_0 =$
(0, ⋯, 0, 1)ᵀ, one obtains

\[ \langle \| x(t) \|^2 \rangle_\beta = e^{-2\gamma t} I_0(2t \sqrt{|w|^2 + \sigma^2}), \quad (t \geq 0) \]  \hspace{1cm} (6.55)

where \( I_\nu(x) \) is the \( \nu \)-th modified Bessel function.

Figure 6.5 plots the function Eq. (6.55) and compares it with the result obtained by ignoring the disorder (corresponding to \( \sigma = 0 \)). The main difference between the two curves is the slower asymptotic decay of the \( \sigma \neq 0 \) result (green) compared with the zero-disorder case (purple). This is the result of the disorder spreading some of the eigenvalues of \(-\gamma + A\) closer to the imaginary axis, creating modes with smaller decay.

Importantly, in neither case do we see transient amplification. By contrast, in the \( \sigma = 0 \) and for small enough decay, i.e. for \( \gamma < |w| \), the system Eq. (2.12) exhibits very strong transient amplification. In this case, starting from the initial condition \( x_0 = (0, \cdots, 0, 1)ᵀ \), the solution for the \((N - n)\)-th Schur component is \( x_{N-n}(t) = \frac{(w t)^n}{n!} e^{-\gamma t} \) (for \( 0 \leq n \leq N - 1 \)), which is maximized at \( t = n/\gamma \) with a value \( \max |x_{N-n}| \sim (\frac{|w|}{\gamma})^n \) for \( n \gg 1 \). Thus up to time \( t \sim N/\gamma \) the norm of the activity grows exponentially; \( \|x(t)\|^2 \gtrsim (\frac{|w|}{\gamma})^{2\gamma t} \) for \( t \lesssim N/\gamma \). For larger times the activity reaches the end of the \( N \)-long feedforward chain and starts decaying to zero; asymptotically \( \|x(t)\|^2 \sim e^{-2\gamma t} \) for \( t \gg N/\gamma \).

As we have seen in Sec. 5.4.2, the spectrum of \( M \) is extremely sensitive to perturbations; even for very small but nonzero \( \sigma \), the spectrum of \(-\gamma I + A\) has eigenvalues with real part as large as \( |w| - \gamma \). Therefore, in the limit \( N \to \infty \), the system Eq. (2.12) is unstable for \( |w| > \gamma \), as soon as \( \sigma \neq 0 \).

Conversely, we see that in the presence of disorder – even infinitesimally small disorder
in the $N \to \infty$ limit – as long as the system is stable (which from Eq. (C.17) requires \( \gamma > \sqrt{|w|^2 + \sigma^2} \)), it exhibits no transient amplification for the initial condition along the last Schur mode.

Notice however that, as was mentioned after Eq. (6.13), Eq. (6.5) and hence Eq. (6.55) do not yield the correct answer when the direction of the impulse is optimized for the specific realization of the quenched disorder \( J \); such disorder-tuned initial conditions can yield significant transient amplification even for the stable \( \sigma \neq 0 \) system.

Incidentally, from Eq. (6.55) we can also read the result for \( M = 0 \), by simply setting \( w = 0 \); this yields

\[
\langle \| x(t) \|^2 \rangle_J = e^{-2\gamma t} I_0(2\sigma t).
\]

Since in this case, all directions are equivalent, this is the answer for the (normalized) initial condition along any direction, again as long as the direction is chosen independently of the specific realization of \( J \).

Finally, the total power of response to a sinusoidal input with amplitude \( I_0 = (0, \cdots, 0, I_0)^T \) is given by

\[
\langle \| x_\omega \|^2 \rangle_J = \frac{\| I_0 \|^2}{\omega^2 + \gamma^2 - |w|^2 - \sigma^2}.
\]

The main effect of the disorder is to reduce the width of the resonance (the peak of \( \langle \| x_\omega \|^2 \rangle_J \) at \( \omega = 0 \)) and increase its height. This is partly a consequence of the scattering of the eigenvalues of \( -\gamma + A \) closer to the imaginary line by the disorder, creating modes with smaller decay.
6.4.3 Network with two-state feedforward components

In this section we will consider a matrix $M$ with a Schur form $T$ that is composed of $N/2$ disjoint feedforward chains, each connecting only two modes (thus assuming $N$ is even). Just as in Chapter 5, it will be instructive to focus on the case where all eigenvalues are zero, so that the mean connectivity in the Schur basis is

$$
T = W \otimes \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}
$$

where $W$ is the $N/2 \times N/2$ diagonal matrix of Schur weights $W = \text{diag}(w_1, w_2, \ldots, w_{N/2})$.

As noted in Chapter 5, the matrix $T$ in Eq. (5.147) arises as the Schur form of a mean matrix of the form

$$
M = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \otimes K
$$

where $K$ is a normal but otherwise arbitrary $N/2 \times N/2$ matrix. When $K$ has only positive entries, matrices of the form Eq. (6.59) satisfy Dale’s principle.

The magnitude of impulse response Eq. (6.5) and the power-spectrum of steady-state response Eq. (6.13) can be calculated for $A = M + \sigma J$, with $M$ given by Eqs. (6.58)–(6.59) with general $w_b$, when the (impulse or sinusoidal) input feeds into the second Schur mode in one of the $N/2$ chains/blocks of Eq. (6.58); the index for this block will be denoted as $a$.

For the average magnitude of impulse response, one finds:

$$
\langle \|x(t)\|^2 \rangle_t = \left[ \frac{1 + C_a}{2} I_0(2r_0 t) + \frac{1 - C_a}{2} J_0(2r_1 t) \right] e^{-2\gamma t}
$$

(6.60)
Figure 6.6: The squared norm of response to impulse, $\|x(t)\|^2$, of the system Eq. (2.12), for $A = M + \sigma J$, with log-normal $J$, and $M$ given by Eq. (5.147) describing $N/2$ doublet feedforward chains weights $w_b$. Here, $w_a = \sqrt{\langle|w_b|^2\rangle_b} = 3$, $\sigma = 0.4$, $\gamma = 1$, and $N = 1400$. The dashed green curve shows the result of Eqs. (6.60)–(6.61) for the average norm squared which, except for a small window around its peak, lies on top of the solid red curve showing the empirical average of $\|x(t)\|^2$ over 100 realizations of binary $J$. The five thin dashed black curves show the result for five particular realizations of $J$, and the pink area shows the standard deviation among the 100 realizations. The standard deviation shrinks to zero as $N \to \infty$ and $\|x(t)\|^2$ for any realization lies close to its average for large $N$. For comparison the lowest, purple curve shows $\|x(t)\|^2$ obtained by ignoring the effect of quenched disorder, i.e. by setting $A = M$.

where $r_1^2 \equiv r_0^2 - \sigma^2$, while $r_0$ is given by Eq. (5.149), $J_0(x)$ is the zero-th Bessel function, $I_0(x)$ is the corresponding modified Bessel function, and the parameter $C_a$ is defined as

$$C_a = \frac{1 + 2|w_a|^2/\sigma^2}{\sqrt{1 + 2\langle|w_b|^2\rangle_b/\sigma^2}}, \quad \text{(6.61)}$$

with $\langle|w_b|^2\rangle_b = 2 \text{tr}(M^TM)$ denoting the average squared feedforward weight among all the blocks of Eq. (6.58) (see derivation in Appendix C).
In Fig. 6.6, Eq. (6.60) is plotted and compared with the result obtained by completely ignoring the disorder, i.e. by setting \( \sigma = 0 \). In the latter case, the block \( a \) is decoupled from the rest of the network. Solving the \( 2 \times 2 \) linear system governed by the matrix

\[
\begin{pmatrix}
-\gamma & w_a \\
0 & -\gamma
\end{pmatrix}
\]

we find \( \|x(t)\|^2 = (1 + w_a^2 t^2) e^{-2\gamma t} \).

From the figure, it can be seen that the \( \sigma \neq 0 \) result (green) has a slower asymptotic decay compared with the zero-disorder case (purple); this is due to the disorder having spread some eigenvalues closer to the imaginary axis, creating modes with smaller decay, along with the fact that the coupling between the \( 2 \times 2 \) blocks induced by the disorder insures that these more slowly decaying modes will be activated.

Indeed, for large \( t \), \( \|x(t)\| \) decays like \( e^{-\gamma t} \) when \( \sigma = 0 \), while in the \( \sigma > 0 \) case, based on Eq. (5.149) it must decay like \( e^{-(\gamma - \rho_0) t} \), i.e. by a rate set by the largest real part of the spectrum shifted by \( -\gamma \). That is indeed what one can obtain from Eq. (6.60) by using the asymptotics of Bessel functions. In addition, both curves exhibit transient amplification where the magnitude of activity initially grows to a maximum, before it decays asymptotically to zero. The \( \sigma \neq 0 \) curve shows larger and longer transient amplification, which is most likely attributable both to the eigenvalues being closer to the \( \text{Re}(z) = \gamma \) line and to augmented nonnormal effects (e.g. larger effective feedforward weights, or longer chains).

It should be mentioned that, as in previous examples, if the input direction is optimized for the particular realization of \( J \), significantly larger transient amplification may be achieved.

Finally, the total power spectrum of response to a sinusoidal input, Eq. (6.12), is given
by the explicit formula

$$\langle \| x \omega \| ^2 \rangle _J = \frac{\omega ^2 + \gamma ^2 + |w_d|^2}{(\omega ^2 + \gamma ^2)^2 - \sigma ^2(\omega ^2 + \gamma ^2 + \mu ^2)} \| I_0 \|^2, \quad (6.62)$$

where $\mu ^2 \equiv \text{tr}(M^\dagger M) = \langle |w_d|^2 \rangle _b / 2$ and, as noted above, the direction of $I_0$ is that of the second Schur mode in block $a$.

### 6.5 Beyond the non-crossing approximation

In the foregoing sections, general formulae were derived for the magnitude of impulse response and frequency power spectrum in an $N$-dimensional linear dynamical system with a coupling given by partly random matrices. We have demonstrated the theory by tackling analytically some specific neural circuits and have qualitatively characterized the possible effect of disorder. Thus, the non-crossing diagrams have been able to provide information not only on the eigenvalues, but also on the transient dynamics of these systems.

This does not exhaust the possibilities of the technique. The same approximation may be used to calculate several other observables of interest, for matrix ensembles of the form $A = M + LJR$ and the prime candidate is the direct statistics of the eigenvectors, which for $M = 0$ were partly analyzed in Mehlig and Chalker (1999).

On the other hand, other quantities are not accessible to the non-crossing approximation. These include the random fluctuations of the eigenvalue density $\langle \delta \rho _j (z) \delta \rho _j (z + w) \rangle _J$ for arbitrary values of $z$ and $w$, which are of crucial interest for the study of eigenvalue repulsion. Another object that defies the non-crossing approximation is the statistics of
the "outlier" eigenvalues discussed after Eq. (5.11) and in the examples of Sec. 5.4.2 and Sec. 5.4.3, which is of importance to practical applications because it can drive the long-term dynamics. The calculation of such quantities may be possible, however, using the replica technique, e.g. in the form developed by Nishigai and Kamenev (2002).

Finally, there are important forms of disorder which are not covered by the general ensemble $A = M + LJR$ with i.i.d., and hence dense, $J$. Examples of relevance to neuroscientific applications include sparse $A$ (Rogers and Castillo, 2009; Slanina, 2011; Neri and Metz, 2012). Note that, e.g., binary matrices with probability of a nonzero weight, $p$, which is small but $\Theta(1)$ as $N \to \infty$ are covered by the above formulae; “sparse” disorder here means, for example, to the case $p = o(1)$.

In the ensemble considered in this article, and for real $J$, the covariance $\langle \delta A_{ij} \delta A_{i'j'} \rangle_j = (LL^T)_{ii'} (RT)^{jj'}$ is single rank. This is of course an idealization, and more general structures of correlation between the elements of $A$ may be considered. An important application would be the study of networks with local topologies, where for example the matrix $A$ has a banded structure (coming from translational invariance). Generalization to other forms of random disorder is, therefore, an important direction for future research.

In some very recent work, the spectrum of purely random matrices with a variance profile has been characterized (Kuczala et al., 2016; Aljadeff et al., 2016; Cook et al., 2017). The problem remains open for biologically plausible weight matrices having a structured mean connectivity (which serve also as a model for ecological and biochemical networks, as was discussed in Chapter 5).
6.6 Disorder fluctuations of the covariance

As was more than once in this chapter, stability of the network requires \( \max_{z \in \mathcal{S}} \Re[z] \geq \gamma \). For large \( N \), we can thus apply formula (5.4), to find that a dynamical instability at frequency \( \omega \) does occur if \( \sigma > \sigma(\omega) \), where

\[
\sigma_c(\omega) = \left[ \left\| (\gamma + i\omega - M)^{-1} \right\|_F \right]^{-1/2}
\]

(6.63)

Therefore, the natural parameter measuring the proximity of the disorder-driven instability is

\[
R(\omega) = \frac{\sigma}{\sigma_c(\omega)}
\]

This parameter is dimensionless and positive. When it is much smaller than unity, the linear description of the network is justified, as the network is stable with probability one in the thermodynamic limit. If \( R(\omega) \) close to unity, the system is close to a linear instability at the given frequency. The saturation point will be given by

\[
\delta_c = \inf_{\omega} \sigma_c(\omega)
\]

(6.64)

It is known (Sompolinsky at al., 1988) that a saturation mechanism in the input-output function may then turn the unstable regime into a chaotic one. A relevant question is whether this parameter can be estimated from sampling the actual activity of a neural circuit operating in a noisy environment. This is a motivation for studying the disorder fluctuations of the covariance, which has already been shown (Dahmen et al., 2016) to provide a
possible measurement of the parameter $R(0)$.

Right before Eq. (6.2), we included the disorder averaging in the definition of the covariance matrix $C^\times(\omega_1, \omega_2; x_0 x_0^T)$, hence we computed the mean value of covariance for motion in a random synaptic field. We are interested here in fluctuations around that mean value, hence we will define the covariance as dependent on the given configuration of disorder. Moreover, we will focus only on the same-frequency covariance matrix, which corresponds to setting $\omega_1 = \omega_2$. We define

$$c(\omega) = \overline{\mathbf{x}(\omega)\mathbf{x}^T(\omega)}$$

(6.65)

where the bar indicates noise averaging and $\mathbf{x}(\omega)$ is the Fourier transform of Ornstein-Uhlenbeck process

$$\frac{dx_i(t)}{dt} = -\gamma x_i(t) + \sum_{j=1}^{N} A_{ij} x_j(t) + I_i(t)$$

(6.66)

The noisy input $I(t)$ may be taken to be temporally white as in Eq. (6.6), with zero mean and an isotropic diffusion tensor $I_i(t_1)I_j(t_2) = D(t_1 - t_2)$; alternatively, we may describe the input as a Wiener process with distribution

$$\mu[I] \sim \left( \frac{dt}{2\pi D} \right)^{T/2} \exp \left( -\frac{1}{2D} \sum_{i=1}^{N} \int_{-T/2}^{T/2} I_i^2(t) dt \right).$$

(6.67)

For a fixed realization of the synaptic-weight matrix $A$, the covariance is given by Eq. (6.3), which we may rewrite here as

$$c(\omega) = (z I + A)^{-1} D(z^* I + A^T)^{-1};$$

(6.68)
with \( z = \gamma + i\omega \).

We would like to compute the first two moments in the distribution of this matrix; namely,

\[
C_{ij}(\omega) = \langle c_{ij} \rangle_J \quad \delta C_{ij}^2(\omega) = \langle (c_{ij}(\omega) - C_{ij}(\omega))^2 \rangle_J
\]

Just like the quantities considered so far, these too can be easily calculated by field-theoretical methods; for instance by direct diagrammatic expansion, or by introducing the ad-hoc generating function for the covariance:

\[
Z_{\omega}(K) = \left\langle \exp \left[ \frac{D}{2} K^T (z1 - A)^{-1} (z^* 1 - A^T)^{-1} K \right] \right\rangle_J
\]

which is expressed in term of a source field \( K \) and yields the desired momenta of covariance through the relationships

\[
\langle c_{ij}(\omega) \rangle_J = \left. \frac{\partial^2 Z_{\omega}}{\partial K_i \partial K_j} \right|_{K=0} \quad \forall i, j
\]

\[
\langle c_{ij}^2(\omega) \rangle_J = \frac{1}{2} \frac{\partial^4 Z}{\partial K_i^2 \partial K_j^2} - \frac{1}{2} \langle c_{ii} \rangle_J^2 \quad \text{if } i \neq j
\]

\[
\langle c_{ii}^2(\omega) \rangle_J = \frac{1}{3} \frac{\partial^4 Z}{\partial K_i^4}
\]

where we use the fact that \( c_{ii} \) is independent on the index.

The disorder averaging may be performed, most practically, by relying on a saddle-point expansion of the integrand in Eq. (6.70) equivalent the noncrossing approximation, hence valid \( N \to \infty \).
Details of the approximation are shown in Appendix D; to the lowest orders, we find

\[ Z(K) = \exp \left[ F_2(K) + F_4(K) + O(|K|^6) \right] \]  
(6.74)

where \( F_2 \) is quadratic and \( F_4 \) is quartic. In terms of the matrix \( B = (z 1 - M)^{-1} \), we can write

\[ F_2(K) = \frac{D}{2} \left( 1 - \frac{\sigma^2}{N \text{Tr} BB^T} \right)^{-1} K^T BB^T K + \frac{D \sigma^2}{2N} (K^T BB^T BB^T K) \]  
(6.75)

\[ F_4(K) = -\frac{D^2}{2s_2} \left( 1 - \frac{1}{1 - \sigma^2 s_2 / N} \right) (K^T BB^T K) (K^T BB^T BB^T K) - \frac{\sigma^2 D^2}{4N} \left( \frac{1}{N - \sigma^2 s_2} \right)^2 (K^T BB^T BB^T K) (K^T BB^T BB^T K) + \]

\[ + \frac{D^2 s_4}{4s_2^2} \left( 1 - \frac{1 + 2\sigma^2 s_2 / N}{(1 + \sigma^2 s_2 / N)^2} \right) (K^T BB^T K) (K^T BB^T K) \]

for \( s_2 = \text{Tr} (BB^T) \) and \( s_4 = \text{Tr} (BB^T BB^T) \).

Using Eqs. (6.71-6.73) in formulas (6.74-6.76) for a uniform mean connectivity \( M_{ij} = \mu \), one obtains

\[ \langle c_{ij}(\omega) \rangle_J = \frac{D}{1 - R^2(\omega)} \delta_{ij} + O \left( \frac{1}{N} \right), \]  
(6.77)

\[ \langle c_{ij}^2(\omega) \rangle_J - \langle c_{ij}(\omega) \rangle_J^2 = \frac{D}{N (1 - R^2(\omega))^2} \left[ \frac{1}{(1 - R^2(\omega))^2} + \frac{1}{1 - R^2(\omega)} \right] \delta_{ij} + O \left( \frac{1}{N} \right), \]  
(6.78)

It follows that the spectral parameter \( R(\omega) \) can be measured directly from the first two
moments in the distribution of covariances, through

$$R^2(\omega) \approx 1 - \left( 1 + N \frac{\delta C^2_{ij}(\omega)}{C^2_{ii}(\omega)} \right)^{-1/2}$$  (6.79)

Eq. (6.79) generalizes to finite frequencies the zero-frequency result obtained in Dahmen et al. (2016).

Another interesting problem to be tackled in future work is the calculation of disorder fluctuations for the two time covariance $\langle \langle c_i(t)c_j(t') \rangle \rangle$, which will require keeping into account the coupling between propagators at different frequencies. Moreover, the results in this section and those for the power spectrum of response await to be generalized to a dynamics driven by noisy inputs that are temporally colored, with a long-range correlation $I_i(t_1)I_j(t_2)$. These are promising directions for future work.\(^1\)

---

\(^1\)The work discussed in this chapter and in chapter 5 is the result of a collaboration between Yashar Ahmadian, F.F., and Kenneth Miller, and has first been presented in Ahmadian, Fumarola & Miller (2015)
Chapter 7

Response to targeted input perturbations

This chapter will be devoted to the problem of a neural network’s response to targeted perturbations of its external input. The aim is to derive some general formulas that may be useful for a rudimentary handling of phenomenological results. We begin, in Section 7.1, with a review of experimental methods used to apply such input perturbations. We then turn (Sec. 7.2) to the question of measuring, with currently available methods, whether excitatory subnetworks are independently stable. A basic answer is provided by the study of response to perturbations in a linear two-population network. The phenomenological importance of differentiation among cell types is then discussed (Sec. 7.3), especially as concerns inhibitory populations, and Sec. 7.4 is devoted to the calculation of input perturbation response in multipopulation networks.

7.1 Measuring the response to input perturbations

In 1979, Francis Crick foreshadowed some of today’s key developments in experimental neurobiology when he stated (Crick, 1979):

A method that would make it possible to inject one neuron with a substance that would then clearly stain all the neurons connected to it, and no others, would be invaluable. So would a method by which all neurons of just one type could be inactivated, leaving the others more or less unaltered.
As electrodes cannot be used to precisely target defined cells and drugs act much too slowly, Crick speculated that light might have the properties to serve as a control tool, but at the time neuroscientists knew of no clear strategy to make specific cells responsive to light. Forty years were to elapse before a reliable technique achieving just that was made possible through genetic targeting of light sensitive ion channels and pumps, a technique now known as optogenetics.

The response of cortical circuits to a perturbing stimulus had been traditionally measured with less effective methods (Douglas et al., 1989), such as extracellular electrical stimulation\(^1\) by current injection, and local perfusion of the brain with chemical agonists or antagonists of specific subtypes synaptic receptors (Bowery et al., 1984). Optogenetics, on the other hand, permits the targeted stimulation or suppression of chosen neuronal populations, i.e. of genetically defined cell classes, on a controllable time scale of milliseconds (Boyden et al., 2005; Han and Boyden, 2007; Zhang et al., 2007).

The key tools of this technique are optogenetic actuators, proteins that modify the activity of the cell in which they are expressed when that cell is exposed to light (Figure 7.1); this allows light to be used as the on-off switch. Actuators can be naturally occurring, or they can be chemically modified to become photosensitive. The most common use of optogenetics is for changing the membrane voltage potential of excitable cells. As was explained in Chapter 2, membrane depolarization leads to spiking, while membrane hyperpolarization leads to the inhibition of spiking; hence, changes to the membrane potential lead to suppression or stimulation of neurons. In general, actuators can be used to induce single

\(^1\)Which activates cells in a local vicinity. Intracellular stimulation, on the other hand, is not used because it would only affect the one cell involved and would presumably have little network effect.
Figure 7.1: A summary of optogenetic methods. Listed along the x-axis are some of the available optogenetic actuators. Color indicates the optimal frequency of light used for illumination, and $\tau_{\text{off}}$ is the speed of deactivation (fast opsins have a small $\tau_{\text{off}}$ and slow opsins have a large $\tau_{\text{off}}$), measured in msec. Excitatory and inhibitory opsins are available, as are opsins that can modulate intracellular signaling cascades. The list is adapted from Fenno et al. (2011) and Guru et al. (2015). The field is fast-moving and the list is quickly evolving.

or multiple action potentials (which can be organized into regular spike trains or which can be pseudo-random at a user-controlled rate), to modify the spontaneous firing rate of a neural population ("step-function actuators"), or to modify biochemical signaling pathways ("biochemical actuators"), with millisecond control over the timing of events.

The most powerful and widely used actuators are opsins — naturally occurring light-sensitive transmembrane proteins that are modified to improve various aspects of performance (Pastrana, 2011; Boyden, 2011; Guru et al., 2015).

The optogenetic toolbox is quickly expanding as a result of studies that aim to identify new light-sensitive proteins in different ecological niches or by re-engineering existing
variants. Notably, several of these tools can be used in combination to allow multimodal control of neuronal activity\textsuperscript{2}.

From the viewpoint of systems neuroscience, the possibility to use carefully targeted perturbations should help to observe and detect the computational mode of cortical circuits. However, due to the prevalence of recurrent interactions in the cortex, the outcome of such a perturbation may be unintuitive or difficult to predict. Theoretical modelling of perturbations is required to relate network architectures and operating regimes to the expected result of a particular perturbation, and to guide the choice of an appropriate experimental perturbation to optimal test hypotheses.

If the perturbations are sufficiently small, and the system will relax to a new steady state that is sufficiently close to the original one, linear response will apply. Considering the system at the basic level of linear rate models provides then a correct starting point for further insight. This is the approach that will be pursued in this chapter.

7.2 Modeling paradoxical responses

7.2.1 Are excitatory subnetworks independently stable?

One question to be thus addressed concerns the strength of recurrent excitation. Many mathematical models of mammalian cortex rely on crucial assumptions on the presence of strong recurrent excitation (Rutishauser and Douglas, 2009; Neftci et al., 2013; Muir and Cook, 2014).

\textsuperscript{2}The method has even been used, very recently, to control the firing of heart cells, thus extending the use of these tools to non-neuroscience applications (Wang et al., 2017).
However, strong recurrent excitation can only be a feature of a stable network if it is balanced by comparably strong local inhibition. Networks balanced by such a mechanism are known in the literature as "inhibition stabilized networks", or ISNs. Stated otherwise, ISN networks are circuits that, for a given feedforward input, have a stable fixed point that would be unstable (due to excitatory recurrence) if inhibitory firing rates were frozen, but which is stabilized by inhibitory feedback.

The alternative to "inhibition stabilized" network consists in the type of networks that relies on a weak excitatory population for stability. These networks would also be stable if the inhibitory feedback was frozen, but the computational mechanisms they support would be different, not relying on strong excitatory recurrence. In a simplified, firing-rate description, it may be said that such networks are stabilized by the first term on the RHS of Eq. (2.6), the "leak conductance" term.

It is an open question in which of these two balanced regimes the neocortex of mammals really operates. The question is of considerable importance because different regimes may support quite different computational mechanisms. Thus, several approaches have been devised to addressing the problem. Amongst them, we should mention the following:

(a) Firstly, an answer has been attempted on the basis of physiological and anatomical analysis. This generally leads to estimates that agree on recurrent excitation being very strong, especially so in the superficial layers of the cortex (Binzegger et al., 2004; Lefort et al., 2009).

(b) Another source of information comes from the experimental blocking of inhibition in cortex, where attempted, which has resulted in the observation of epileptiform activity (Avoli et al., 1995; Mann et al., 2009). This suggests that inhibitory feedback may be
actually necessary to stabilize the cortical network.

(c) An ISN regime may also be possible to detect “functionally”, that is, by the way cortical activity responds at the network level if the dynamics is experimentally perturbed. Long before the development of optogenetics, this approach to testing the ISN paradigm was examined in theoretical work by Tsodyks et al. (1997), which we will now discuss in some detail.

The prediction of Tsodyks et al. is that perturbations to inhibitory neurons of an ISN network will evoke a telltale response in the activity of the network, which they called "paradoxical response". Namely, if externally-induced activation of inhibitory cells (hereinafter, I cells) causes suppression in the firing rates of excitatory cells (hereinafter, E cells), the net inhibition received by the E cells will be paradoxically decreased. If, on the other hand, the suppression of I cells causes increase of E-firing, then the net inhibition received by E cells will paradoxically increase.

Such an unintuitive phenomenon produced by the perturbing of inhibition is a general prediction that comes from various nonlinear models (see also Rubin et al. 2015) and has thus emerged as a good candidate for a signature of the ISN regime.

In fact, there are two distinct ways in which paradoxical response can be defined: in terms of the I-current received, as we just did, or in terms of I-cell firing (with firing rate decreasing when input is increased, and vice versa). These two definitions are not equivalent. In an ISN, even if some I cells do not change their firing rate paradoxically while others do, the I-current received will be paradoxical. While Tsodyks et al.’s original result concerned I-cell firing, Rubin et al. (2015) and Litwin-Kumar et al. (2016) have showed that the paradoxical change in the I-input received is a more general result, also valid with mul-
tiple subtypes. Moreover, the received-current criterion distinguishes ISN regardless of the fraction of I cells perturbed. Whether the activity of perturbed cells increases or decreases depends, on the other hand, on how many are perturbed\(^3\).

In the laboratory, neurons are often being recorded and activated at the same time. Hence, a paradoxical inhibitory response of \textit{currents} is an easily accessible metric, that it is not difficult to discover empirically in cortical networks.

There are two ways to assay paradoxical responses:

- One way is intracellularly – i.e. by looking at the inhibition received by a cell. In an ISN circuit, there will be a paradoxical change in the inhibition received by E cells when a perturbation of I cells alters E cell firing rates. So far, results from experiments of this type are mixed. In cats, Ozeki et al. have thus demonstrated the relevance of the ISN regime to V1, by showing that measurements in cat V1 of inhibition received by E cells during surround suppression yielded paradoxical current response, at least when stimulated by a reasonably strong visual stimulus (Ozeki et al., 2009). For the mouse visual cortex, similar results are discussed in Adesnik (2017). Efforts have also been made to search for this kind of response in the auditory cortex, and there have been recent findings of ISNs using these methods in the A1 region of mouse (Kato et al., 2017). On the other hand, Atallah et al. (2012)

---

\(^3\)The received-current paradoxical effect can be worked out for a general network with one E type \((i = 1)\) and multiple I types \((i = 2, \ldots, N)\); calling \(r\) the firing-rate response and \(h\) the input perturbation, then if \(h_E = 0\), we have \((1 - w_{EE})r_E - \sum_{i \geq 2} w_{E(i)}^r (i) = 0\) or \((1 - w_{EE})r_E = \sum_{i \geq 2} w_{E(i)}^r (i)\). The right-hand side of the equation is the inhibition received by the E cell (as a negative number). We know that the ISN condition is \(w_{EE} > 1\); hence if \(r_E\) increases (or decreases), the net inhibition received has to paradoxically increase (or decrease). If there are multiple E subtypes, the same argument can be applied for any unstable eigenmode of the \(E \rightarrow E\) submatrix. Note that these arguments are independent of how many I cells are perturbed.
have failed to find evidence for a paradoxical inhibitory response in mouse, except in a small fraction of pyramidal cells. Of course, whether or not a given cortical circuit is an ISN may depend on conditions such as overall cortical firing rates, sampling, contextual variables, that may vary between experiments.

• The other way is extracellularly, i.e. through the extracellularly-measured spike response. There, an ISN can show a paradoxical change in firing rates. Here interpretation of results is clouded by two issues. First, only part of the I cell population may be perturbed by a given experimental intervention, and results may depend on what proportion is perturbed (Sadeh et al., 2017). Second, there are multiple I subtypes, and a perturbation of I cells may cause some types to increase firing rates and others to decrease, with the net result that (if the network is an ISN) the total inhibition the E cells receive will change paradoxically. However, in many conditions, particularly those increasingly used now to record many cells at once (Ca++ imaging; arrays of extracellular electrodes), only extracellular firing rates are experimentally accessible.

Therefore, both to assay whether or not a circuit is in the ISN regime, and more generally to infer circuit structure from results of perturbations, it is critical to know what can be inferred from changes in firing rates given incomplete perturbations of particular cell types and observations of responses in that and/or other cell types. These are the issues we will address here, beginning with the case of a single E and single I cell type, and then expanding

---

4 The intracellular measurement is more difficult, and must usually be performed one cell at a time; moreover, it is not accessible by the commonly used Ca++-imaging or array-electrode recording (not to mention single-cell electrode recording). Also, it is not widely accessible in primates so far; there appears to be a single publication where intracellular recording in primates was reported (Tan et al., 2014).
to the case of multiple cell types.

We will be focusing specifically on paradoxical response in firing rates (not in current received. Thus we will say that an inhibitory cell responds paradoxically if its firing rate goes down when the input is increased.

### 7.2.2 A uniform excitatory-inhibitory network

**The network**

We consider here a two-population network in which the mean synaptic weights can take only one of two values: $w_E$ if the presynaptic neuron is excitatory, and $w_I$ if the presynaptic neuron is inhibitory. Such a network is "uniform" in the sense that synaptic weights depend only on the presynaptic cell type.

In this case, the mean synaptic weight matrix $M$ has the form

$$M = \begin{pmatrix}
    w_E & \ldots & w_E & w_I & \ldots & w_I \\
    \vdots & \ddots & \vdots & \vdots & & \vdots \\
    \vdots & & \vdots & \vdots & & \vdots \\
    w_E & \ldots & w_E & w_I & \ldots & w_I
\end{pmatrix}, \quad (7.1)$$

where the first block occupies $N_E$ columns, the second block $N_I$ columns. In the special case where $N_E w_E = N_I w_I$ we recover, as a special case of network (7.1), the uniform balanced network of Eq. (5.151) that was introduced in Sec. 5.4.3.

In the following, Greek letters will be used to index the Schur-basis components, and
Latin letters to label individual neurons. Eq. (7.1) can be written as

\[ M = \lambda e_1 e_1^T + \eta e_1 e_2^T, \]  

(7.2)

where, as in Sec. 5.4.1, we have introduced the Schur modes

\[ e_1 = \frac{1}{\sqrt{N}} (1, 1, \ldots, 1)^T, \]  

(7.3)

\[ e_2 = \frac{1}{\sqrt{N_E N_I N}} \left( \frac{N_I, \ldots, N_I, -N_E, \ldots, -N_E}{N_E, N_I} \right)^T, \]  

(7.4)

which can be taken as the first two elements in an orthonormal basis to be built with the Gram-Schmidt method, while \( \lambda \) and \( \eta \) are the entries of \( M \) in its Schur form, namely

\[ \lambda = w_E N_E + w_I N_I \quad \eta = \sqrt{N_E N_I (w_E - w_I)}; \]  

(7.5)

since \( w_E > 0 \) and \( w_I < 0 \) we have \( \eta > 0 \), while the stability of the dynamical system requires \( \lambda < 1 \). The case of the balanced uniform network of Sec. (5.4.3) is recovered for \( \lambda = 0 \).

**The response**

Call \( I \) the feedforward input to which the perturbations are applied in experiments. As per Eq. (2.12), the steady state activity is then \( x_S = RI \), where \( R^{-1} = \gamma 1_N - M \); when the input is perturbed, and the system is given time to relax to a new steady state, the change in activity will be given by \( \Delta x = R \Delta I \).
For the synaptic weight matrix of Eq. (7.1), we have

$$R^{-1} = (\gamma - \lambda) e_1 e_1^T - \eta e_1 e_2^T + \gamma \sum_{\alpha \geq 2} e_\alpha e_\alpha^T. \quad (7.6)$$

with the inverse being

$$R = \frac{1}{\gamma - \lambda} e_1 e_1^T + \frac{\eta}{\gamma - \lambda} e_1 e_2^T + \frac{1}{\gamma} \sum_{\alpha \geq 2} e_\alpha e_\alpha^T. \quad (7.7)$$

Noticing that $\sum_{\alpha \geq 2} e_\alpha e_\alpha^T = 1_N - e_1 e_1^T$, we can write (7.7) as

$$R = \gamma 1_N + \left( \frac{1}{\gamma - \lambda} - \gamma \right) e_1 e_1^T + \frac{\eta}{\gamma - \lambda} e_1 e_2^T. \quad (7.8)$$

Having computed the inverse, using Eqs. (7.3-7.4) we rotate it back into the basis defined by individual neurons to obtain

$$R_{ij} = \gamma \delta_{ij} + \frac{1}{N} \left[ \frac{1 - \gamma^2 + \lambda}{\gamma - \lambda} + \frac{\eta}{\sqrt{N} (\gamma - \lambda)} (e_2)_{ij} \right], \quad (7.9)$$

that is,

$$R_{ij} = \frac{1}{\gamma - \lambda} \left[ (\gamma - \lambda) \delta_{ij} + \frac{1 - \gamma^2 + \lambda}{N} + \frac{\eta}{N \sqrt{N_E N_I}} (N_I \delta_{jE} - N_E \delta_{jI}) \right], \quad (7.10)$$

where $\delta_{jE}$ ($\delta_{jI}$) is equal to unity if $j$ is the index of an excitatory (inhibitory) neuron and zero otherwise.

Throughout this chapter, to obtain more compact formulas for the various response
functions, we will measure synaptic weights in units of the leak parameter $\gamma$. This amounts to fixing $\gamma = 1$; doing so in Eq. (7.10) and, moreover, plugging in Eqs. (7.5) for the Schur weights, Eq. (7.10) may be rewritten in the simple block-matrix form

$$R = \begin{pmatrix} 1_N + \frac{w_E}{D} & \frac{w_I}{D} \\ \frac{w_E}{D} & 1_N + \frac{w_I}{D} \end{pmatrix},$$

(7.11)

where $1_N$ is the identity in $N$ dimensions, and all the ostensibly scalar matrix elements should be understood as being constant across rows and columns, while $D = 1 - w_EN_E - w_IN_I$. Eq. (7.11) may also be written more compactly as

$$R_{ij} = \delta_{ij} + \frac{1}{D} \begin{pmatrix} w_E & w_I \\ w_E & w_I \end{pmatrix}_{T_IT_J},$$

(7.12)

where $T_i$ is the cell-type of the $i$-th neuron.

### 7.2.3 Basic network responses

As discussed in Sec. 7.2.1, a paradoxical response in firing rates is observed when a given neuron fires less if input is increased to a certain group of cells to which that neuron belong.

In current experimental setups, input perturbations always concern a multiplicity of neuron. The biologically relevant issue, therefore, is not the system’s response to a single cell being perturbed, but what happens when a substantial fraction of a population is perturbed. Nonetheless, as a first calculation we will look at the response to perturbations of a single cell, which is described by the sign of the diagonal elements in Eq. (7.11).
If perturbations are taken to concern a single neuron (which is never the case in real experiments), the signature of paradoxical response is simply a negative entry on the diagonal of the response matrix Eq. (7.11). A negative diagonal entry means then a response opposite to the direction of a perturbation in its input.

We being by considering (7.13) in the trivial case \( N_I = N_E = 1 \). We have then

\[
R = \frac{1}{1 - w_E - w_I} \begin{pmatrix} 1 - w_I & w_I \\ w_E & 1 - w_E \end{pmatrix}
\] (7.13)

From Eq. (7.13), it is clear that such a behavior cannot occur for excitatory neurons. We have \( R_{EE} = \frac{1 - w_I}{1 - w_E - w_I} \); here the denominator is the determinant of \(-1 + M\), which must be positive (as the product of two negative eigenvalues) in order to ensure stability, while the numerator is also positive because by definition we have \( w_I < 0 \).

On the other hand the inhibitory response can be paradoxical if \( w_E > 1 \), because then \( R_{II} < 0 \). This condition quantifies the intuition we describe above about ISN made necessary by the strength of excitation. Moreover, notice that the condition is independent on the value of \(|w_I|\), although considering the stability constraints yields the overall condition

\[
1 < w_E < 1 + |w_I|
\] (7.14)

for ISN in an unrealistic two-neuron network\(^5\).

\(^5\)More in general, from the linearized equations of motion for a two-neuron network with \( W = \begin{pmatrix} w_{EE} & w_{EI} \\ w_{IE} & w_{II} \end{pmatrix} \), we have \((1 - W)r = h\), where \( r \) is the firing-rate response and \( h \) is the input perturbation. The response of the inhibitory neuron to a perturbation \( h = (0, h_I) \) (no input to the E cell) is \( r_I = \frac{1 - w_{EE}}{\det(1 - W)} h_I \), which is paradoxical if \( w_{EE} > 1 \).
Let us now consider networks of the type (7.1) of arbitrary size but with the excitatory and inhibitory populations equally large. We have then $D = 1 - N(w_E + w_I)/2$. Since this must again be positive, it appears that we cannot consider the limit of large network ($N \gg 1$) unless we simultaneously assume that synaptic weights scale as $1/N$.

For a large, equal-population IE network, the response of a single excitatory neuron to variations in its own input cannot be paradoxical. Indeed, for $T_i = E$ we find $R_{ii} = \frac{1-(N/2-1)w_E-Nw_I/2}{1-Nw_E/2-Nw_I/2} > 0$. Here, the numerator is again positive, and the denominator is equal to the denominator plus a positive quantity, hence also positive.

On the other hand, if the input to an inhibitory neuron is changed, the response of that neuron may or may not respond paradoxically depending now on the relative magnitude of $w_E$ and $|w_I|$. Indeed, the condition $R_{ii} < 0$ for $T_i = I$ can be written as $D + w_I < 0$. If $w_E > |w_I|$, the condition becomes

$$N > N_c = \frac{2(w_I + 1)}{w_E + w_I},$$

which means that the response will always be paradoxical, as long as the network is sufficiently large. However, if instead $w_E < |w_I|$, the paradoxical effects disappears for sufficiently large populations (i.e. for $N > N_c$).

### 7.2.4 Perturbing the input to part of a population

It should be pointed out that the paradoxical response in firing rate predicted by Tsodyks et al. (1997) is expected to emerge when the *entire* inhibitory population of an ISN is perturbed simultaneously. In typical experiments, however, only a fraction of the inhibitory popula-
tion is usually affected by the perturbation. This raises the crucial question of whether the paradoxical effect will still be observed, possibly on a smaller scale, if only some sector of the input to inhibitory neurons is perturbed.

More specifically, the question is whether a the firing rate of a single inhibitory neuron will respond negatively to positive perturbations concerning the feedforward input to a fraction $\xi$ of that inhibitory population. The relevance of this to experiments is discussed extensively in Sadeh et al. (2017); to sum it up, when the firing-rate paradoxical response is not observed in experiments, the naive conclusion that the system is not ISN may be wrong, if the response also depends on the scale $\xi$ of perturbations.

It is possible to argue that paradoxical responses in a sufficiently large network should always require a minimal perturbation range. This follows from the fact that a neuron will perceive input variations both directly (if it lies within the scope of the perturbations) and indirectly (through recurrent interactions). The former contribution is necessarily positive and only the latter can be negative. On the other hand, the larger the system, the less it will "feel" single-neuron perturbations. Thus, the recurrent contribution can be made relevant (enough to outweigh the direct contribution and invert the sign), only if the perturbation addresses a sufficient portion of the network.

For the case of equal-size populations introduced in Sec. 7.2.3, the response of an I cell to a perturbation to a fraction $\xi$ of the I population is, from Eq. (7.10),

$$R_{II} = [1]_p + \frac{\xi w_I N}{2D},$$  \hspace{1cm} (7.16)

where the brackets $[...]_p$ mean that the term should only be included if the test neuron is
among those whose input has been perturbed.

Notice that setting $\xi = 1$ in Eq. (7.16), and imposing the requirement for paradoxical response, yields $\frac{2/N-w_E}{2/N-w_E-w_I} < 0$. Since the denominator is $D > 0$, it follows that $w_E > 2/N$, which is precisely the criterion for the ISN regime. More generally, the criterion for paradoxical response is $w_E > 2/N + (1 - \xi)|w_I|$. So the requirement on $w_E$ exceeds that of the ISN criterion.

Moreover, from Eq. (7.16), it’s clear that for neurons not concerned by the perturbation the response is always negative, whereas for neurons concerned by it the response is paradoxical only if $\xi > \xi_c$, where

$$\xi_c = \frac{2D}{|w_I|N} = \frac{2/N - w_E - w_I}{|w_I|}$$

(7.17)

where we have used the fact that stability requires $D > 0$.

For large $N$, this threshold tends to the limit

$$\lim_{N \to \infty} \xi_c = 1 - \frac{w_E}{|w_I|}.$$  

(7.18)

Hence, a paradoxical response may not be observed in observed where the fraction of the population to which input is perturbed is not large enough, as we anticipated discursively in the previous section. In its simplicity, formula (7.18) offers a rule of thumb for scenarios

---

6This is seen by requiring instability of the excitatory subnetwork, i.e. positivity of one of the eigenvalues of $M - 1$.

7We will reserve the word “paradoxical” for negative response of the perturbed cell, as a negative response of a cell to activating other I cells is not paradoxical.

8Positivity of the determinant also implies, once again, we cannot take the limit $N \to \infty$ unless $w_E; w_I$ also scale with $N$, to maintain $D > 0$. 

264
of the type that has been described numerically in Sadeh et al. (2017).

If the populations have different sizes, the response becomes

$$R_{II} = [1]_p + \frac{\xi w_I N_I}{D},$$  \hspace{1cm} (7.19)

with now $D = 1 - w_E N_e - w_I N_I > 0$.

The response is paradoxical if $\xi > \xi_c$, where

$$\xi_c = \frac{1 - w_E N_E + |w_I| N_I}{|w_I| N_I}. \hspace{1cm} (7.20)$$

And for $N_\alpha \gg 1/|w_\alpha|$ we can write $\xi_c \sim 1 - \frac{N_E}{N_I} \frac{w_E}{|w_I|}$.

Finally, consider the case where the test neuron were to be chosen at random among the neurons of the inhibitory population. In that case, the term $[1]_p$ would be equal to unity with probability $\xi$ and zero otherwise, so its expectation value would be $\xi$. This gives

$$R_{II} = \xi \left( 1 + w_I \frac{N_I - N_E}{2} \right) \hspace{1cm} (7.21)$$

which is negative if and only if $N_I - N_E > 2/|w_I|$, whatever the value of $\xi$. The sign of the response, in this latter case, is independent on $\xi$.  

265
7.3 Networks with multiple cell types

7.3.1 Differentiation of inhibitory cell-types

The classical description of cortical circuits as bipartite, i.e. consisting essentially of an inhibitory and an excitatory population, has been called into question by a number of recent studies on the proper classifications of neurons.

In particular, experimentalists have been studying specific cell types in V1 (Velez-Fort Margrie, 2014; Harris and Shepherd, 2015; Kim et al., 2015; Jiang et al., 2015; Markram et al., 2015; Tasic et al., 2016). In V1, three main subtypes of I cells are believed to comprise over 80% of inhibitory cells (Pfeffer et al., 2013): parvalbumin-(PV), somatostatin- (SOM), or vasoactive-intestinal-peptide-expressing (VIP) interneurons (e.g. Lee et al., 2010; Rudy et al., 2011; Pfeffer et al., 2013; Tremblay et al., 2016).

These 3 subtypes have striking differences in their visual responses (Pecka et al., 2014; Dipoppa et al., 2017), circuit effects (Wilson et al., 2012; Lee et al., 2012; Lee et al., 2014; El-Boustani et al., 2014; Seybold et al., 2015; Phillips and Hasenstaub, 2016), and modulation by locomotion (Polack et al., 2013; Fu et al., 2014; Dipoppa et al., 2017) and context (Adesnik et al., 2012).

Much interest has been focused on a putative disinhibitory circuit in which top-down or brain-stem inputs (e.g., induced by locomotion) can target VIP neurons, which inhibit SOM neurons and thus disinhibit E neurons, increasing the overall circuit gain (Polack et al., 2013; Lee et al., 2013; Pi et al., 2013; Fu et al., 2014; Zhang et al., 2014; Fu et al., 2015; Ayzenshtat et al., 2016; Jackson et al., 2016); and on a contextual modulation circuit in which SOM cells primarily integrate contextual influences (Adesnik et al., 2012).
Additional inhibitory subtypes have also been revealed that could modulate circuit function in specific ways (Olsen et al., 2012; Jiang et al., 2013; Bortone et al., 2014).

This new, complex phenomenological scenario is yet far from being matched by proper modeling efforts. The standard single-compartment inhibitory-excitatory ansatz is still a standard simplification to obtain many predictions in neuroscience, especially where the dynamics is nonlinear (e.g., Kadmon and Sompolinsky, 2015). In the rest of this chapter, we will take some elementary steps in the direction of including properties of multiple inhibitory cell-types, computing the response to elementary perturbations for linear networks with an arbitrary number of populations.

### 7.3.2 A model for the mean synaptic weight matrix

We will consider a model of neural network including $N$ neurons structured in $n$ populations $P_\alpha$; here, as in Sec. 5.4.3, we use Greek indices to label different populations. We will call $N_\alpha$ the size of population $P_\alpha$; $M_{ij}$ the mean synaptic weight from neuron $k$ to neuron $j$; and $\alpha_k$ the index of the population to which the $k$-th neuron belongs.

The mean synaptic weight matrix $M$ for the full neural network is an $N \times N$ matrix
\((N = \sum_\alpha N_\alpha)\) written as

\[
M = \begin{pmatrix}
w_{11} & \ldots & w_{11} & w_{12} & \ldots & w_{12} & \ldots & \ldots & w_{1n} & \ldots & w_{1n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_{11} & \ldots & w_{11} & w_{12} & \ldots & w_{12} & \ldots & \ldots & w_{1n} & \ldots & w_{1n} \\
w_{21} & \ldots & w_{21} & w_{22} & \ldots & w_{22} & \ldots & \ldots & w_{2n} & \ldots & w_{2n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_{21} & \ldots & w_{21} & w_{22} & \ldots & w_{22} & \ldots & \ldots & w_{2n} & \ldots & w_{2n} \\
w_{n1} & \ldots & w_{n1} & w_{n2} & \ldots & w_{n2} & \ldots & \ldots & w_{nn} & \ldots & w_{nn} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
w_{n1} & \ldots & w_{n1} & w_{n2} & \ldots & w_{n2} & \ldots & \ldots & w_{nn} & \ldots & w_{nn}
\end{pmatrix}, \quad (7.22)
\]

whose entries are

\[
M_{jk} = \sum_{\alpha\beta} \delta_{\alpha j}\delta_{\beta k}w_{\alpha\beta} = w_{\alpha j}\alpha k, \quad (7.23)
\]

where the \(n \times n\) matrix \(\{W_n\}_{\alpha\beta} \equiv w_{\alpha\beta}\) is what the synaptic weight matrix of the system would be if every population had a single neuron \((N_\alpha = 1)\).

This model is far more general than the mean connectivity model of Sec. 5.4.3, Eq. (5.155), but still simple enough so that some general conclusions can be easily extracted. The rest of the chapter will be focus on some general properties of a network with mean connectivity (7.22).
7.3.3 Eigenvalues of the mean connectivity

We will rotate Eq. (7.22) into a basis whose first \(n\) vectors are

\[
\mathbf{e}^{(1)} = \frac{1}{\sqrt{N_1}} \begin{pmatrix} 1, \ldots, 1, 0, \ldots, 0 \end{pmatrix} \frac{1}{N_1} \quad \frac{1}{N-N_1}
\]

\[
\mathbf{e}^{(\alpha)} = \frac{1}{\sqrt{N_\alpha}} \begin{pmatrix} 0, \ldots, 0, 1, \ldots, 1, 0, \ldots, 0 \end{pmatrix} \frac{1}{\sum_{j=1}^{\alpha-1} N_j} \quad \frac{1}{N_\alpha} \quad \frac{1}{\sum_{j=\alpha+1}^{N} N_j}
\]

\[
\mathbf{e}^{(n)} = \frac{1}{\sqrt{N_n}} \begin{pmatrix} 0, \ldots, 0, 1, \ldots, 1 \end{pmatrix} \frac{1}{N-N_n} \quad \frac{1}{N_n}
\]

The subsequent \(N-n\) basis vectors can be constructed by the Gram-Schmidt method, and their explicit form is not essential.

In the new basis, the mean connectivity can be written as

\[
M = \begin{pmatrix}
    w_{11} N_1 & w_{12} N_2 & \ldots & w_{1n} N_n & 0 & \ldots & \ldots & 0 \\
    w_{21} N_1 & w_{22} N_2 & \ldots & w_{2n} N_n & 0 & \ldots & \ldots & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    w_{n1} N_1 & w_{n2} N_2 & \ldots & w_{nn} N_n & 0 & \ldots & \ldots & 0 \\
    0 & \ldots & \ldots & 0 & 0 & \ldots & \ldots & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
    0 & \ldots & \ldots & 0 & 0 & \ldots & \ldots & 0
\end{pmatrix}
\]  \quad (7.24)

Thus, the eigenvectors of \(M\) belong to two classes. There are \(N-n\) modes corresponding to the null eigenvalue, each of them orthogonal to all the \(\mathbf{e}^{(\alpha)}\) (for \(\alpha = 1, \ldots, n\)). These
correspond to modes in which the firing of neurons within each population is incoherent. 

There are then \( n \) eigenvectors that are linear combinations of the \( e^{(\alpha)} \) (for \( \alpha = 1, \ldots, n \)), and corresponds therefore to modes where the neurons within each populations are either firing all together or not firing at all\(^9\). Let us call the corresponding \( n \) eigenvalues \( \Lambda_1, \Lambda_2, \ldots, \Lambda_n \), ordered by the magnitude of their real parts from the largest to the smallest. These can be found from the characteristic equation for the matrix \( M_n = W_n \times L \), which is the \( n \times n \) upper left portion of \( M \) in Eq. (7.24), as defined in terms of \( L = \text{diag} \left( N_1, N_2, \ldots, N_n \right) \)\(^{10}\).

The corresponding eigenvalues can be related to the eigenvalues \( \lambda_\alpha \) of \( W_n \) by noticing that

\[
\sum \ln |\lambda_\alpha| = \ln \left| \prod \lambda_\alpha \right| = \ln |\det L \det W_n| = \sum \ln |N_\alpha \lambda_\alpha|.
\]

We may derive further results by exploiting the inequality

\[
\text{Tr} (B) \text{Eig}_{\text{min}} \left( \frac{A + A^T}{2} \right) \leq \text{Tr} (AB) \leq \text{Tr} (B) \text{Eig}_{\text{max}} \left( \frac{A + A^T}{2} \right) \tag{7.25}
\]

(see Mori, 1988). Setting \( A = \text{diag} \left( N_1, N_2, \ldots, N_n \right) \), \( B = W_n \), we deduce

\[
\sum \lambda_\alpha \geq N_{\text{min}} \text{Tr} W_n. \tag{7.26}
\]

\(^9\)Some of these, of course, may also correspond to a zero eigenvalue \( \lambda_m = 0 \), with \( 1 \leq m \leq n \).

\(^{10}\)The eigenvalues of \( W_n L \) are, in fact, the same as those of the matrix \( L^p W_n L^{1-p} \) for any \( p \), where \( L^p \equiv \text{diag} \left( N_1^p, N_2^p, \ldots, N_n^p \right) \).
7.3.4 Eigenvectors of the mean connectivity

Let us finally consider the eigenvectors of the mean connectivity (7.22). The eigenvectors corresponding to coherent modes can be obtained in the following way. Call $v^\alpha$ the $\alpha$-th eigenvector of the reduced weight matrix $W_n$, and call $v^\gamma_\alpha$ its $\gamma$-th component (where both $\alpha$ and $\gamma$ range between 1 and $n$).

The first $n$ eigenvectors of $M$ are the vectors

$$V^{(\alpha)} = \sum_\beta v^{(\alpha)} e^{(\beta)},$$

(7.27)

where the vectors $e^{(\beta)}$ form the coherent basis defined above and

$$w_{\alpha 1} v^{(\gamma)}_1 + \sum_{\beta > 1} w_{\alpha \beta} N_\beta v^{(\gamma)}_\beta = \lambda_\gamma v^{(\gamma)}_\alpha$$

(7.28)

In the case where one of the populations (say $P_1$) is much more numerous than the others, it can be seen from Eq. (7.28) that firing of that population is suppressed in the subdominant modes. Indeed, in the limit $N_1 \to \infty$, we must have either $\lambda_\gamma = O(N_1)$ or $v^{(\gamma)}_1 = O(1/N_1)$. The first choice yields a single solution, i.e. the principal mode; the second choice yields $N - 1$ modes in which the firing of the leading population is suppressed.

For instance, consider two populations with $N_1 \gg N_2$. We have then

$$\Lambda_+ = N_1 w_{11} + O(N_2) \quad \Lambda_- = \left( w_{22} - \frac{w_{21} w_{12}}{w_{11}} \right) N_2,$$

(7.29)
and the eigenvectors, non-normalized, are

\[
V_+ \propto \begin{pmatrix} w_{11} + O(N_2/N_1) \\ w_{21} & \end{pmatrix} \quad V_- = \begin{pmatrix} O(N_2/N_1) \\ w_{11} & \end{pmatrix}, \quad (7.30)
\]

### 7.4 Response of multipopulation networks

#### 7.4.1 Response function

We will now derive some useful formulas regarding the response to input perturbations of the multipopulation network introduced in Sec. (7.3).

Defining the matrix \( \tilde{W} = \sqrt{L}W_n\sqrt{L} \), whose determinant is \( \det \tilde{W} = (\prod_{\alpha=1}^n N_\alpha) \det W_n \), the mean synaptic weight matrix \( M \) for the full neural network of Eq. (7.22) can now be expressed as \( M = V^T \tilde{W} V \), where \( V \) is a matrix with \( n \) rows and \( N \) columns, defined as

\[
V = \begin{pmatrix}
\frac{1}{\sqrt{N_1}} & \cdots & \frac{1}{\sqrt{N_1}} & \cdot & \cdots & \cdot & \frac{1}{\sqrt{N_n}} & \cdots & \frac{1}{\sqrt{N_n}} \\
\frac{1}{\sqrt{N_1}} & \cdots & \frac{1}{\sqrt{N_2}} & \cdot & \cdots & \cdot & \frac{1}{\sqrt{N_1}} & \cdots & \frac{1}{\sqrt{N_2}} \\
\frac{1}{\sqrt{N_1}} & \cdots & \frac{1}{\sqrt{N_2}} & \cdot & \cdots & \cdot & \frac{1}{\sqrt{N_n}} & \cdots & \frac{1}{\sqrt{N_n}} \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\
\end{pmatrix}
\]

(7.31)
where all the omitted entries are null. We can write

\[ V_{\alpha k} = \frac{1}{\sqrt{N_\alpha}} \Theta \left( \sum_{l=0}^{\alpha-1} N_\alpha < k \leq \sum_{l=0}^{\alpha} N_l \right) = \frac{1}{\sqrt{N_\alpha}} \delta_{\alpha \alpha k}, \]  

(7.32)

where \( \Theta \) is the truth function (equal to 0 or 1) and we notice that \( V V^T = I_n \), the identity matrix in \( n \) dimensions.

We now apply Woodbury’s identity, which yields

\[ R \equiv (1_N - M)^{-1} = (1_N - V^T \bar{W} V)^{-1} = 1_N - V^T \left(1_n - \bar{W}^{-1}\right)^{-1} V, \]

(7.33)

hence the individual entries of the \( R \) matrix are

\[ R_{ij} = \delta_{ij} - \frac{1}{\sqrt{N_{\alpha_i} N_{\alpha_j}}} \left[ \left(1_n - \bar{W}^{-1}\right)^{-1} \right]_{\alpha_i \alpha_j}, \]

(7.34)

Alternatively, we can write Eq. (7.34) as

\[ R_{ij} = \delta_{ij} - \left[ L^{-1/2} (1_n - L^{-1/2} W_1^{-1} L^{-1/2})^{-1} L^{-1/2}\right]_{\alpha_i \alpha_j}, \]

(7.35)

where \( L^{-1/2} = \text{diag} \left( \frac{1}{\sqrt{N_1}}, \frac{1}{\sqrt{N_2}}, \ldots, \frac{1}{\sqrt{N_n}} \right) \).

### 7.4.2 Two-population networks

We begin from the example of a non-uniform two-population network, with the response being measured on a neuron that is in principle different from the neuron to which the input has been perturbed. Here and in the following, we rescale the synaptic weights by the total
number of neurons in the network, defining the matrix elements through

\[ [W_n]_{\alpha\beta} \equiv \frac{n}{N} w_{\alpha\beta}, \quad (7.36) \]

where \( n \) is again the number of populations and the \( w_{\alpha\beta} \) are \( O(1) \). In the limit where \( N_\alpha \to 1 \) for all \( \alpha \), the matrix elements of \( M \) tend to those of \( W_n \).

Applying (7.34) for \( n = 2 \) gives

\[
R_{ij} = \delta_{ij} - \frac{1}{\sqrt{\kappa_{\alpha_i} \kappa_{\alpha_j}}} \left[ 1_n - L^{-1/2} W_n^{-1} L^{-1/2} \right]^{-1}, \quad (7.37)
\]

where we have redefined the \( L \)-matrix as \( L_{\alpha\beta} = \delta_{\alpha\beta} \kappa_\alpha \), with \( \kappa_\alpha = N_\alpha / N \). Working out the full content of Eq. (7.37) we finally arrive at

\[
R_{ij} = \delta_{ij} - \frac{\sqrt{\kappa_E \kappa_I} K}{N \sqrt{\kappa_{\alpha_i} \kappa_{\alpha_j}}} \left( \begin{array}{cc}
\kappa_E \kappa_I K - w_{EE} \kappa_E & w_{EI} \sqrt{\kappa_E \kappa_I} \\
w_{IE} \sqrt{\kappa_E \kappa_I} & \kappa_E \kappa_I K - w_{II} \kappa_I
\end{array} \right)_{\alpha_i \alpha_j}, \quad (7.38)
\]

where \( K = \det W_n = w_{EE} w_{II} - w_{IE} w_{EI} \).

Here the long denominator is equal, by construction, to the determinant of the full synaptic weight matrix. \( K \), on the other hand, is the determinant of the corresponding two-neuron network\(^{11}\).

\(^{11}\)The corresponding two-neuron network may of course be unstable per se. In this case, the network is only stabilized by the large size of the inhibitory population.
7.4.3 Formulas for different test neurons

Suppose we perturb experimentally a fraction $\gamma$ of randomly chosen neurons among those belonging to a given population $\mathcal{P}_\alpha$. Depending on the structure of the experiments, different formulas will need to be applied, and we list them here for reference.

If the test neuron (the neuron on which we measure variations in activity) is randomly chosen within the population whose input has been perturbed, then the first term on the right hand side of Eq. (7.35) (the Kronecker delta), summed over the $j$ index, has an expectation value exactly equal to $\gamma$, and from Eq. (7.34) we obtain

$$R_{\text{rand}}^\alpha(\gamma) = \gamma \left( 1_n - \tilde{W}^{-1} \right)_\alpha^\alpha. \quad (7.39)$$

If, on the other hand, the measurement is performed upon one of the neurons whose inputs are being perturbed, we have

$$R_{\text{pert}}^\alpha(\gamma) = -\gamma \left( 1_n - \tilde{W}^{-1} \right)_\alpha^\alpha. \quad (7.40)$$

If the test neuron belongs to population $\alpha$ but is among those whose input is not being perturbed, the response will be given by:

$$R_{\alpha\text{.n.p.}}^\alpha(\gamma) = 1 - \gamma \left( 1_n - \tilde{W}^{-1} \right)_\alpha^\alpha, \quad (7.41)$$

whose sign will change depending on the range of the perturbation.

If the response is measured on a neuron in population $\mathcal{P}_\beta$, with input perturbations
applied to the full population $\mathcal{P}_\alpha$ with $\alpha \neq \beta$, we have

$$R_{\beta\alpha} = -\sqrt{\frac{N_\alpha}{N_\beta}} \left[ (1_n - \bar{\tilde{W}}^{-1})^{-1} \right]_{\beta\alpha}, \quad (7.42)$$

Finally, if we are only perturbing the input to a fraction $\gamma$ of the neurons in $\mathcal{P}_\alpha$, then the RHS in Eq. (7.42) ought to be multiplied by an additional factor $\gamma$.

### 7.5 Conclusions

We have derived a number of formulas that can be useful for a basic understanding of the type of response to specific input perturbations such as are currently made possible by developments in optogenetics.

An analysis of a basic network structure led to a precise prediction concerning the detectability of a paradoxical response to input perturbations in inhibitory neurons. In particular, the theory suggests that this type of response, in a realistic experimental setup, may only arise for perturbations involving a sufficient number of neurons, even if the system is in an inhibition-stabilized regime. The threshold for paradoxical response is a quantity that, in principle, may be measured in current experimental setups, and it is useful to relate it to microscopic parameters of the circuitry.

We also modeled the possible multiplicity of cell-types, which is an increasingly relevant complication in theoretical studies, embodying specific properties of the given circuit under study. Allowing for mean connectivity weights that depend on the type of both the presynaptic and postsynaptic weights, we computed the response properties of the mean net-
work. These results may be regarded as preliminary steps in the direction of understanding the response behavior of multipopulation networks.
8.1 Review of the results

A number of problems in theoretical neuroscience amount to computing static and dynamic properties of the matrix of synaptic weights, also known as the connectivity of networks. This matrix lies at the core of the simplest working models of network interaction, hence it is crucial to find out how changes in its properties will affect the collective dynamics of neural circuits. In this thesis, the task has been tackled from a number of different angles, focusing on those aspects that would be relevant to open questions in cortical phenomenology.

We began by summarizing the simplifications underlying network models that involve synaptic weights (Chapter 2). While these simplifications are substantial, their consequences have been studied extensively in theoretical neuroscience. These simplified models can give substantial insight into the behavior of spiking networks, so long as their activity is not synchronized.

We then began to carry out the program for one of the best-documented functions of cortical circuits, namely, visual recognition of the orientations of contour elements. The selectivity of neuronal cells with respect to orientations depends on symmetry properties of the feedforward synaptic weights. We discussed how these symmetries emerge in a
Hebbian implementation of synaptic plasticity, and tackled a minimal model of orientation development, a single cell of the type proposed in Linsker (1986). The size of its dendritic arbor and the correlation width of visual input are its controlling parameters.

As we have shown, methods customarily applied to Shrödinger operators also lend themselves to the type of Hilbert-Schmidt operators involved in the theory. Exploiting this analogy, the problem was treated as one of zero-temperature phase transitions to derive information on the long term behavior of the system as a function of cortical and input parameters. We showed that, for sufficiently long-range input correlations, rotational symmetry is spontaneously broken, leading to the functional specialization of cells. We obtained a lower bound on the critical value of the correlation length, and computed the full set of feedforward weights (equivalent to the cell’s receptive field) in both the selective and nonselective phase.

Cortical cells are known to be arranged on the basis of their functional specialization. For instance, the input features that make a visual-cortex cell fire (color, direction of movement, orientation) depend continuously on the cell’s position within the cortex. In Chapter 4, we have performed a comprehensive analysis of how such functional organization (concerning orientation preferences) can be achieved across the cortex in the presence of homeostatic constraints. A mixture of perturbative, variational, and asymptotic techniques has been applied to various regions of the phase diagram to reveal a nontrivial structure of transitions between different symmetry classes, confirmed by numerical simulations, and to calculate the relevant observables.

In addition to its theoretical value, this study contributes to a possible resolution of the current debate on the conditions for the emergence of spatially organized orientation prefer-
ences. Hebbian models have predicted, so far, that this can only occur if the cortex receives input that correlates nonmonotonically at increasing distances (Miller, 1994; Wimbauer, 1998). These predictions have led experimentalists to search for traces of such a correlation profile (Ohshiro and Weliky, 2006) but their results contradict theoretical predictions.

The model solved in Chapter 4 takes into account the competition between developing arbors in a biological plausible fashion, and thus arrives at predicting basic features of the phenomenology in the presence of correlation functions that fall off monotonically, such as have been found in the experiments. This result shows that axonal competition may play a key role in developing a functional architecture of orientation preferences, by making a uniform arrangement of orientation preferences unstable.

The results of Chapters 3 and 4 illustrate how information about the long-term evolution of the synaptic matrix may be coaxed analytically from working models of plasticity. Static properties of the synaptic matrix, such as its spectral distribution, are also of interest because they drive the collective dynamics of neurons on shorter time scales.

However, computing these spectra is a nontrivial task for at least three reasons: (a) the size of typical neural populations makes the matrices involved correspondingly large; (b) plasticity makes synaptic weights dependent on sensory histories, hence disorder is ubiquitous, and may defy perturbative approaches; (c) the segregation of excitatory and inhibitory neuron, which makes these matrices extremely non-normal, disables random-matrix techniques from nuclear and mesoscopic physics, tailored on Hermitian operators.

In Chapter 5, we carried out a full characterization of the spectral properties of large synaptic weight matrices that possess a partly stochastic component. The starting point has been the universality theorem proved by Tao and Krishnapur (2010), which allows one to
pick any computationally convenient ensemble for calculations. By doing so, (a) we have rigorously proven that the support of the eigenvalues is compact in the thermodynamic limit; (b) we have been able to draw analytically its shape for an arbitrary mean connectivity; (c) we have computed in closed form the profile of the eigenvalue density within it.

Moreover, the correct regularizing procedure was provided for finding the support of the eigenvalue density for certain highly nonnormal instances of mean connectivity; the naive interpretation of the formulae fails in these cases, which are not yet discussed in the literature.

The potential applications are numerous and a number of them have been worked out in detail, especially where they concern network structures conventionally used in modeling. These have included: networks with a uniform Dale-compliant mean connectivity; exactly balanced inhibitory-excitatory networks; networks with a single-chain feedforward structure; networks with multiple, separable Schur components; networks with a factorizable subpopulation-dependent distribution of disorder; clustered networks with large intra-cluster connectivity.

All the results are valid for an arbitrary distribution of quenched disorder, and depend only on the distribution’s first two momenta. However, we have shown numerically that the convergence of observables to their large-N limit can be considerably slower if the distribution of disorder is heavy-tailed.

We have found that, in cases where the network is highly nonnormal, even an infinitesimal amount of disorder can be sufficient to destabilize it. We have also studied the phenomenon of outlying eigenvalues that survive even as the network size diverges, which is a subject of research in neuroscience and may be of importance to application (Rajan and
Abbott, 2006). A precise criterion has been introduced (related to the large-$N$ behavior of singular values) that delimits the regions of the complex plane where these outliers will arise.

The nontrivial link between the spectra of partially random matrices and the pseudospectra of their first moment has been then analyzed, yielding conclusions that apply to numerical applications. As a way to visualize a matrix’s pseudospectrum (Trefethen and Embree, 2005), dense random perturbations are sometimes added to a given matrix. Here, it was argued here that such a method can fail, and the failure was motivated and demonstrated with numerical examples. The connection with pseudospectra, moreover, allowed to discuss precisely how non-normality affects the sensitivity of spectra to quenched fluctuations.

Besides boosting sensitivity to disorder, non-normality in synaptic weight matrices gives rise to another crucial property of neural dynamics – large dynamical transients.

Transient amplification is supposed to be relevant to brain function in a variety of ways: (a) it multiplies the number of time scales involved, possibly enabling short-term memory (Goldman, 2009); (b) it can amplify steady-state responses in a pattern-selective manner (Murphy and Miller, 2009); (c) it plays an essential role in determining baseline noise levels in stable networks (Dixon et al., 2016).

In addition, systems with non-normal connectivity can also display pseudo-resonance frequencies in the power-spectrum when the external frequency is not close to any of the system’s natural frequencies as determined by its spectrum. It is clear that these properties cannot possibly be understood through the sole density of eigenvalues.

A full chapter of this thesis was consequently devoted to the calculation of dynamical
observables for systems with non-normal weight matrices (Chapter 6). Analytic expressions have been derived for the frequency power spectrum of steady-state response to white noise and to sinusoidal input, as well as for the magnitude of transients as measured by the average norm squared – quantities whose calculation had not been attempted analytically for stochastic networks.

The technique, which consists of a summation of ladder diagrams for the four-point propagators, seems not to have been applied before to the non-Hermitian case. The correct average power spectrum is found to be always strictly larger than the naive power spectrum obtained by assuming that random effects will average out. Moreover, the strength of transient amplification is such that it will be typically underestimated if the randomness of connectivity is ignored.

Applications have then been carried out in extensive detail for different types of networks (feedforward chains with null eigenvalues, feedforward chains with alternating eigenvalues, networks with two-state Schur components). It was found that disorder leads to a strong but unequal amplification of resonances. In cases where it does not destabilizes the network, it can also lead to both larger and longer transient amplification effects. Since transients have been hypothesized to serve important computational functions in cortical circuits, by enhancing them and strengthening them disorder may play a functional role in computation.

Recent technological advances in neurobiology, notably optogenetics, provide new tools to manipulate the activity of neurons in a specifically targeted fashion. Optogenetics is allowing experimentalists to test a number of theoretical predictions about the mechanisms inherent to circuit function. A notable example is the "paradoxical" effect predicted
by Tsodyks et al. (1997) for cortical circuits with strong excitation (hence stabilized by strong inhibition), where the activity of inhibitory neurons was predicted to increase if their feedforward input is reduced, and vice versa.

A more general form of the effect, involving a paradoxical variation in total inhibitory input to excitatory cells, has been studied experimentally through intracellular measurements. However, in many conditions, particularly those increasingly used to record many cells at once (Ca++ imaging; arrays of extracellular electrodes), only extracellular firing rates are experimentally accessible. To interpret such measurements, it will be crucial to understand how a realistic network responds to changes in the input, if these target only a given subpopulation or fractions thereof.

An obvious first step in understanding such responses, in the vicinity of a stimulus-driven fixed point, consists in linearizing the dynamics about possible fixed points. We have done so in a minimal model (Chapter 7), providing arguments for the existence of a threshold in the scale of perturbations for the emergence of paradoxical response; expressions have been derived linking this threshold to the network parameters. These findings, which confirm a suggestion by Sadeh et al. (2017), may serve as a guide for future optogenetic experiments.

Current developments in neuroscience are bringing to light unsuspected differences between various neuron types, especially those of the inhibitory variety, which appear to translate into different roles in network function (Pfeffer et al. 2013, Tremblay et al. 2016). Theoretical research is still far from meeting the challenge posed by these developments—that is, building theories with cell-type-specific inhibitory targeting that remain nonetheless simple enough to be treatable.
In the final sections of Chapter 7 we computed the response of networks with a multi-
population structure to targeted input perturbations. The general formulae derived therein
may be regarded as preliminary steps in this research direction.

8.2 Open problems

The overall approach followed throughout this thesis has been to seek minimal functioning
models with a manageable parameter space, as a way to single out relevant degrees of
freedom for each problem.

Such models may not account for biophysical details of single-cell behavior, but can
yield substantial insight into the mechanistic development of interaction, and offer a crucial
contribution to understanding the basis for functional behavior in the brain. Due to their
reduced parameter space, moreover, such models are mathematically tractable, and often
treatable through tools borrowed from theoretical physics.

Although this approach may come naturally to a physicist, it is not the sole approach
used in theoretical neuroscience. In studies on the connectivity matrix, for instance, one
crucial problem is how to infer its $N^2$ entries from large-scale measurements of spike trains.
This problem, requiring extensive inferential approaches of its own (Advani and Ganguli,
2016) is only touched tangentially by the type of theories we have discussed.

The methods we employed, moreover, allow for a number of developments that have
been barely grazed in this thesis. Further work will be necessary with the aims of (a) de-
veloping further the various models considered in this thesis, (b) treating other known models
with the techniques developed herein, (c) modeling additional parts of the phenomenology
in ways that will lend themselves to this type of treatment.

The variational study of Hebbian evolution, for instance, has been among the main tools used in Chapters 3 and 4, where its power was exemplified as a general mathematical possibility. This method had not been explored so far in the Hebbian literature. Although we only applied it to the development of orientation selectivity, it can certainly be applied to other existing problems in Hebbian development, such as those concerning the emergence of ocular dominance, direction selectivity, and phase selectivity – among others.

If the fastest-growing patterns emerging during development is treated as the ground state of a quantum-mechanical system, mapping the parameter space of the system becomes a standard exercise. The parameters of the model can be constrained by computing the conditions for the breaking of symmetries, or for the emergence of plausible forms of selectivity.

On a different note, little has been done in studying plausible nonlinear complications of Hebbian models beyond the mere aim of ensuring convergence. Yet, nonlinear variants of Linsker-type Hebbian dynamics may be key to creating bottom-up models of orientation development that are able to reproduce not only a salt-and-pepper organization such as was demonstrated here, but also the band-pass spectrum measured in the cortical architecture of several groups of animals – notably cats and monkeys. This would serve to bridge Hebbian theory to Landau-type models of universal behavior such as those of Kaschube et al. (2010).

The way we have represented disorder throughout the thesis can also be easily employed in other contexts. In Chapter 5, we applied a variant of the diagrammatic technique for propagators to matrices of the form \( A = M + LJR \) where all the stochasticity is in \( J \), and the method provided a wealth of information on the eigenvalue distribution \( \rho(z) \).
In Chapter 6, we followed an analogous procedure to obtain quantitative insight into the transient dynamics of the system. In Chapter 7, the same methods were employed when we averaged the steady-state response function of generic linear networks.

This technique may be further exploited to calculate other quantities of interest, in particular concerning the direct statistics of eigenvectors, which was not considered in this thesis. On the other hand, several biological quantities lie beyond the reach of the non-crossing approximation; thus, another direction of research will consist in developing the method a step further.

In this sense, a particularly important subject is the microscopic structure of spectral density fluctuations, \(\langle \delta \rho_j(z) \delta \rho_j(z + w) \rangle\) (for arbitrary values of the complex variables \(z\) and \(w\)) which is of crucial interest to the study of eigenvalue repulsion; another observable is the statistics of the outlier eigenvalues considered in Chapter 5. The calculation of such quantities may be possible using the replica technique, for example in the manner of Nishigaki and Kamenev (2002).

In the ensemble used to represent disorder throughout this thesis, the covariance of the synaptic weight matrix \((\langle \delta A_{ij} \delta A_{i'j'} \rangle) = (LL^T)_{ii'}(RR^T)_{jj'}\) is single-rank. There are important forms of disorder which are not covered by this general ensemble. Examples of relevance to neurobiology applications include sparse \(A\), or more general structure of correlations between the elements of \(A\), the latter being of importance in considering networks with local topologies.

The simplest case concerns matrix ensembles with independent elements but with nonuniform mean and variances, which however depend only on the difference \(i - j\) (hence statistically they are translation invariant) or specific cases in which the matrix \(A\) has a
banded structure. Recent work has characterized the spectrum of purely random matrices with a variance profile (Kuczala et al., 2016; Aljadeff et al., 2016; Cook et al., 2017), but the problem remains open for biologically plausible weight matrices, i.e. having a structured mean connectivity (which serve also as a model for ecological and biochemical networks, as was discussed in Chapter 5).

Optogenetic input perturbations require a more detailed level of modeling than has been attempted so far, both here and in the rest of the literature. An obvious first step in understanding the system’s response, in the vicinity of a stimulus-driven fixed point, consists in linearizing the dynamics. For spatially extended models, the analysis can be further carried on so as to determine the set of characteristic activity patterns that are most strongly amplified by the network and the input patterns that drive them. The formulas derived in Chapter 7 are only preliminary steps in this direction, and nonlinearities of the input-output function may prove important.

The different role seemingly played by various inhibitory cell types in cortical circuitry poses a huge problem for theorists. The current phenomenological scenario has become reminiscent of Fermi’s pre-QCD remark, ”If I could remember the names of all these particles I’d be a botanist” (Lederman, 1963); for theorists, the challenge will be to develop model with cell-type-specific inhibitory targeting, that will remain nonetheless simple enough to be insightful.

A first prototype may be the “disinhibitory model” of mouse visual cortex according to which brainstem input induced by running targets VIP neurons, which consequently inhibit SOM neurons, which are also inhibitory and thus disinhibit excitatory neurons, increasing the overall circuit gain (Lee et al., 2013; Pi et al., 2013). While this basic scenario is
supported by recent data, the functional advantage or evolutionary rationale behind it is far from clear.

More in general, the functioning of mouse primary visual cortex under the influence of locomotion and arousal may prove to be an ideal testing ground for the integrated understanding of cell specialization in complex networks. We have there a very well-documented system in which different cell types are more or less strongly activated by locomotion depending on the ongoing amount of visual stimulus (Adesnik et al. 2012; Polack et al., 2013; Fu et al., 2014; Velez-Fort et al, 2014); constructing and solving a simple but predictive model for this type of situation would be a major step toward the general goal of understanding the operation of circuits in the cortex.


Atallah B.V., Bruns W., Carandini M., Scanziani M., *Parvalbumin expressing interneu-


Biane P. and Lehner F., Computation of some examples of Brown’s spectral measure in


Brézin E., Itzykson C., Parisi G. and Zuber J.B., Planar diagrams. Communications in Mathematical Physics, 59 (1978), 35.


Cowan J.D. and Sharp D., Neural networks and artificial intelligence 117 (1988), 85.
Crair M.C., Gillespie D.C. and Stryker M.P. *The role of visual experience in the development of columns in cat visual cortex*. Science, **279** (1998), 566.

Crick F. H. C. *Thinking about the Brain*. Scientific American, **241** (1979), 3.


Fumarola F., Miller K.D., *Development of orientation selectivity without Mexican Hat input correlations*. Program No. 588.22. at Neuroscience 2017, SfN, Washington, DC.


Han X. and Boyden E.S., *Multiple-color optical activation, silencing and desynchronization of neural activity, with single-spike temporal resolution*. PLOs One 2 (2007), e299.


Hayes W.P. and Meyer R.L., *Optic synapse number but not density is constrained during regeneration onto surgically halved tectum in goldfish: HRP-EM evidence that optic fibers
compete for fixed numbers of postsynaptic sites on the tectum. Journal of Comparative Neurology, 274 (1989), 539.


Hubel D. H., Wiesel T. N., *Ordered arrangement of orientation columns in monkeys


Kuffler S.W., Discharge patterns and functional organization of mammalian retina. Journal of Neurophysiology, 16 (1953), 37.


Lefort S., Tomm C., Floyd Sarria J.C.C., and Petersen C.C.H. *The excitatory neuronal network of the c2 barrel column in mouse primary somatosensory cortex*. Neuron, **61** (2009), 301.


MacKay, D.J.C. and K.D. Miller *Analysis of Linsker’s simulations of Hebbian rules,*
Neural Computation 2 (1990), 169.


Miller K. D., *A model for the development of simple cell receptive fields and the ordered


Nishigaki S.M. and Kamenev A., Replica treatment of non-Hermitian disordered


Riva F. and Zanon D., *The saddle point method may be ineffective for the planar approximation.* Physics Letters, 110 (1982), 57


Tanaka S. and Miyashita M., *Constraint on the number of synaptic inputs to a visual cortical neuron controls receptive field formation*. Neural Comput. **9** (2009), 2554.


Appendix A: Asymptotics of the Linsker eigenfunctions

This Appendix is devoted to the asymptotic form of the unconstrained Linsker eigenfunctions defined by Eq. (3.13) (with the coefficients found by solving Eqs. 3.24) in the limit of short-range and long-range correlations.

1 A1: Eigenfunctions in the long-range limit

Equations (3.24) for the polynomial coefficients can be rewritten in the form

\[ a_n = \sum_{k=n+1}^{N} T_{nk} a_k \]  \hspace{1cm} \text{(A.1)}

and these are solved formally by

\[ \frac{a_n}{a_N} = T_{nN} + \sum_{k=n+1}^{N-1} T_{nk} T_{kN} + \sum_{k_1=n+1}^{N-2} \sum_{k_2=k_1+1}^{N-1} T_{nk_1} T_{k_1 k_2} T_{k_2 N} + \ldots \]  \hspace{1cm} \text{(A.2)}

or in a more compact form

\[ \frac{a_n}{a_N} = T_{nN} + \sum_{k=n+1}^{N-1} T_{nk} T_{kn} + \sum_{p=1}^{N-n-1} \sum_{\substack{i \in [0, N]^p \\text{s.t.} \\ n < i_1 < i_2 < \ldots < i_p < N}} T_{n i_1} \left( \prod_{k=1}^{p-1} T_{i_k i_{k+1}} \right) T_{i_p N} \]  \hspace{1cm} \text{(A.3)}
with the T-matrix given by

$$T_{nk} = \frac{k!}{n!} \frac{(k + m)!}{(k - n)!} \frac{(2\eta^2)^{k-n} \beta^{2N-n-k}}{1 - \beta^{2(N-n)}}$$  \hspace{1cm} (A.4)$$

In the case $\eta \gg \rho$, we have $\beta \gg 1$, and therefore we can approximate this with

$$T_{nk} = -\frac{k!(k + m)!}{n!(n + m)!(k - n)!} \left( \frac{\eta^2}{\beta} \right)^{k-n}$$  \hspace{1cm} (A.5)$$

Substituting this into (A.3), it can be seen the various exponents sum up to $N - n$, while the factorials cancel telescopically. Hence one may write

$$a_n = \frac{N!(N + m)!}{n!(n + m)!} \left( \frac{\eta^2}{\beta} \right)^{N-n} a_N c_n^{(N)}$$  \hspace{1cm} (A.6)$$

where

$$c_n^{(N)} = -\frac{1}{(N - n)!} + \sum_{k=n+1}^{N-1} \frac{1}{(N - k)!} \frac{1}{(k - n)!} + \sum_{p=2}^{N-n-1} (-1)^{p+1} \sum_{\vec{i} \in [0,N]^p \atop n < i_1 < i_2 < \ldots < i_p < N} \frac{1}{(N - i_p)!} \left( \prod_{k=1}^{p-1} \frac{1}{(i_{k+1} - i_k)!} \right) \frac{1}{(i_1 - n)!}$$  \hspace{1cm} (A.7)$$

In each term of this series, we can recognize a multiple of a Stirling number of the second kind

$$\sum_{\vec{i} \in [0,N]^p \atop n<i_1<i_2<\ldots<i_p<N} \frac{1}{(N - i_p)!} \left( \prod_{k=1}^{p-1} \frac{1}{(i_{k+1} - i_k)!} \right) \frac{1}{(i_1 - n)!} \left( \begin{array}{c} N - n \\ p + 1 \end{array} \right) = \frac{(p + 1)!}{(N - n)!} \left( \begin{array}{c} N - n \\ p + 1 \end{array} \right)$$
where $\{\frac{A}{B}\}$ is the Stirling number.

With the change of variable $q = p + 1$, and incorporating the first two terms in the sum over $q$, one obtains

$$c_n^{(n)} = \sum_{q=1}^{N-n} (-1)^q \frac{q!}{(N-n)!} \frac{N-n}{q} \left\{ \begin{array}{c} N-n \\ q \end{array} \right\}$$

(A.8)

If we choose to write the Stirling number in explicit form, we have

$$\sum_{\vec{i} \in [0,N]^p} \frac{1}{(N-i_p)!} \left( \prod_{k=1}^{p-1} \frac{1}{(i_{k+1} - i_k)!} \right) \frac{1}{(i_1 - n)!} = \frac{1}{(N-n)!} \sum_{k=1}^{p+1} (-1)^{p-k+1} \binom{p+1}{k} k^{N-n}$$

(A.9)

hence

$$c_n^{(N)} = \frac{1}{(N-n)!} \sum_{q=1}^{N-n} (-1)^q \sum_{k=1}^{q} \frac{q}{k} (-1)^{q-k} \binom{q}{k} k^{N-n}$$

(A.10)

We can then use $(-1)^q \times (-1)^{q-k} = (-1)^k$, and switch the order of the two sums to obtain

$$c_n^{(N)} = \frac{1}{(N-n)!} \sum_{k=1}^{N-n} (-1)^k k^{N-n} \sum_{q=k}^{N-n} \binom{q}{k}$$

(A.11)

and performing the sum over $q$, we find $c_n^{(N)} \equiv c_{N-n}$, where

$$c_j = \frac{1}{j!} \sum_{k=1}^{j} (-1)^k k^m \binom{m+1}{k+1}.$$  

(A.12)

not a Stirling number but with alternating signs in the summand. We have thus

$$a_n = \frac{N! (N+m)!}{n! (n+m)!} \left( \frac{\eta^2}{\beta} \right)^{N-n} a_N c_{N-n}$$

(A.13)
Notice now that, in this regime, $\beta \sim \frac{\nu^2}{\rho^2}$, and that the resulting factor $\rho^{-N}$ can be incorporated in the definition of $a_N$. This is sufficient to prove that the variable entering the polynomials is $\frac{x^2}{2\rho^2}$. Thus, the radial functions have the form

$$R_{N,m}(r) = \frac{r^m}{\rho^{m+1}} e^{-\frac{x^2}{2\rho^2}} P_N^{(m)}\left(\frac{r^2}{2\rho^2}\right)$$

where the functions $P_N^{(m)}$ are, for every $m$, a complete set of polynomials of degree $N$ defined by

$$P_N(x) = \sum_{n=0}^{N} \frac{N! (N+m)!}{n! (n+m)!} c_{N-n}(2x)^n$$

The factor $1/\rho^{m+1}$ arises from normalization, as the wave functions squared integrate to 1. By putting it there explicitly, we keep the coefficients of the polynomials entirely parameter-free\(^1\).

The polynomials in Eq. (A.15), taken as a set labeled by $N \geq 0$, are orthogonal in the same sense as the Laguerre polynomials, namely

$$\int_0^{\infty} x^\alpha e^{-x} P_n^{(\alpha)}(x) P_m^{(\alpha)}(x) dx \propto \delta_{mn} \quad \forall \alpha \geq 0$$

though of course that factor could be incorporated in the coefficients by adopting a different convention.

\(^1\)The 2 in the argument of the polynomials $\left(\frac{x^2}{2\rho^2}\right)$ can be eliminated by rescaling the coefficients by $2^n$, but that would mean defining the polynomials to have an orthogonality relation with a weight $e^{-x/2}$. 

313
.2 A2: Eigenfunctions in the short-range limit

The short-range limit of this simple, one-cell, unconstrained model is particularly interesting. We may compute the asymptotics of the set of eigenfunctions either in the polar-coordinates or Cartesian-coordinates representation, but the polar representation is more useful because it classifies eigenfunctions by their angular momentum, i.e. by whether they are orientation-selective.

Since $\beta \to 1$ in the limit $\eta/\rho \to 0$, it follows that the denominator in the definition of $T_{nk}$ tends to zero. Accordingly, every matrix elements of $T$ introduces in formula (A.3) a large factor of the same order.

The leading term, therefore, will be the one that contains the largest number of such large factors. This is evidently the term where all the matrix elements of $T$ are located right above the diagonal.

In the short-range limit, therefore, we can approximate eq. (A.3) with

$$a_n \sim T_{n,n+1} T_{n+1,n+2} \cdots T_{N-1,N} a_N \quad \text{(A.17)}$$

as only the matrix elements right above the diagonal matter. These are found to be

$$T_{n-1,n} = n (n + m) \frac{2\eta^2 \beta^{2N-2n-1}}{1 - \beta^{2(N-n)}} \quad \text{(A.18)}$$

We must now proceeding to solve Eq. (A.17). After some algebra, we find

$$a_n = \frac{1}{(N-n)!} \frac{N! (N+m)!}{n!} \frac{(-\rho \eta)^{N-n}}{(n+m)!} \quad \text{(A.19)}$$
and the radial eigenfunctions take the form

\[ R_{N,m}(r) = a_N e^{-\frac{r^2}{2\gamma^2}} r^m \sum_{n=0}^{N} \frac{N! (N + m)!}{n! (n + m)! (N - n)!} r^{2n} \] (A.20)

The factor \( N! \gamma^N \) can be pulled out of the sum. If we do so, in the remaining sum we can recognize another generalized Laguerre polynomial of degree \( m \). Absorbing the factors we pulled out into the definition of \( a_N \), we can write therefore

\[ R_{N,m}(r) = a_N r^m L_N^{(m)} \left( \frac{r^2}{\gamma^2} \right) e^{-\frac{r^2}{2\gamma^2}} \] (A.21)

Finally, the normalization constant is found from the identity

\[ \int_{0}^{\infty} x^m e^{-x} \left[ L_N^{(m)}(x) \right]^2 dx = \frac{(N + m)!}{N!} \] (A.22)

which gives

\[ a_N = \left( \pi \gamma^{2m+2} \frac{(N + m)!}{N!} \right)^{-1/2} \] (A.23)

and we conclude

\[ R_{N,m}(r) = \left( \frac{N!}{(N + m)!} \right)^{1/2} \frac{e^{-\frac{r^2}{2\gamma^2}}}{\sqrt{\pi} \gamma^{m+1}} r^m L_N^{(m)} \left( \frac{r^2}{\gamma^2} \right) \] (A.24)

where \( \gamma \sim \sqrt{\rho \eta} \) because \( \eta \ll \rho \).
Appendix B: Wavenumber of cortical modulation

In this appendix, we proceed to calculate the optimal wavenumber for cortical modulations of orientation maps in the model of Chapter 4. We will do so first in the regime where cortical interaction are short-ranged and LGN interaction long-ranged, then in the opposite case, and finally near the RT phase boundary.

3 B1: Optimal wavenumber in the region $\zeta \gg \rho \gtrsim \eta$

We consider the regime $\eta \lesssim \rho \ll \zeta$, where $\lesssim$ means that $\eta$ is either of the same order of magnitude as $\rho$ or smaller. This limit is clearly recovered by setting to zero the parameter $s = \frac{\rho^2}{\mu^2}$ in Eq. (4.93) while keeping $\alpha > 0$. Indeed, if $\rho/\mu \sim \rho/\zeta \rightarrow 0$, $\alpha$ can only stay finite as long as $\eta/\zeta$ is small to the same order as $\rho/\zeta$, which means that $\eta \lesssim \rho$.

In this limit one can write Eq. (4.94) as

$$f(x) \propto e^{-\alpha x/2} \left(1 - e^{-x}\right)$$  \hspace{1cm} (B.1)

and the maximum of the eigenfunction is at $x = \log (1 + 2/\alpha)$. This translates into

$$\omega_M = \omega(\rho \sim \eta \ll \zeta) = \frac{\mu^2}{\rho \zeta^2} \sqrt{\log \left(1 + \frac{2\zeta^2 \rho^2}{\mu^2 \eta^2}\right)} \sim \frac{1}{\rho} \sqrt{\log \left(1 + \frac{2\rho^2}{\eta^2}\right)}$$  \hspace{1cm} (B.2)
In the case $\rho \sim \eta$, the two addends in the argument of the logarithm are of the same order, and neither can be neglected. If $\eta \ll \rho$, we can further simplify to $\omega_M \sim \frac{\sqrt{\zeta}}{\rho} \sqrt{\log (\rho/\eta)}$.

We can now proceed to test the self-consistent hypothesis $\omega \ll \mu/\rho^2$ that was used in Sec. (4.5.1) to derive Eqs. (4.93-4.94).

Indeed, this hypothesis is automatically satisfied if $\eta \sim \rho$, because then the logarithm in Eq. (B.2) is of order one. If, however, we have $\rho \gg \eta$, the self-consistent condition translates into

$$\log \left( \frac{\rho}{\eta} \right) \ll \frac{\zeta^2}{\rho^2}. \quad (B.3)$$

Hence the region of validity of formula (B.2) is given by

$$\zeta \gg \rho \gg \eta \gg \rho \exp \left( -\frac{\zeta^2}{\rho^2} \right) \quad (B.4)$$

and the limiting value approached for small $\eta$ is found to be

$$\omega_0 \sim \sqrt{2} \zeta / \rho^2 \quad (B.5)$$

which may serve as our estimate of the saturation wavenumber found in the numerics.

Finally, notice that in this limit the eigenfunction (4.96) simplifies, as $q \ll 1$ and the $q$-proportional term in the paranthesis can be dropped. We thus have

$$\psi(r) \propto e^{-\frac{q^2}{4\beta^2} - \frac{q^2}{\mu^2} \omega x} \quad (B.6)$$

The key point of Eq. (B.2) is that the wavenumber becomes logarithm small if $\eta \gg \rho$.  

317
On the other hand, we know that it is identically zero for if $\eta > \sqrt{2}\zeta$; it follows that the condition to have discernible orientation maps is

$$\eta \lesssim \min (\zeta, \rho).$$ \hspace{1cm} (B.7)

For $\eta > \rho$, the orientation domains (columns) become gradually much larger than the arbors, hence undiscernible; for $\eta > \sqrt{2}\zeta$, the orientation map rapidly disappears, as confirmed by numerics.

### 4 B2: Optimal wavenumber far from the phase boundary

The derivative of the eigenvalue (4.93) is proportional to

$$D(x) = -\alpha(1-s)^2 + 2(1-s)(\alpha+1)e^{-x} + [4s - (1 + 4sx)(\alpha + 2)] e^{-2x} +$$

$$+ [(\alpha + 2)e^{-x} - \alpha(1+s)] \sqrt{(1-e^{-x}-s)^2 + 4sx e^{-2x}} \hspace{1cm} (B.8)$$

which we would like to maximize with respect to $x$.

We will thus solve $D(x) = 0$ asymptotically. If we assume self-consistently that $x \gg 1$, (which implies being far from the phase boundary), we find:

$$D(x) \sim \frac{1}{1-s}\left\{ -2\alpha(1-s)^2 + 2 \left[ 2 + 2\alpha - (4 + 3\alpha)s + (2 + \alpha)s^2 \right] e^{-x} \right\}. \hspace{1cm} (B.9)$$
The optimal wavenumber is given by $D(x) = 0$, that is,

$$e^{-x} = \frac{\alpha(1-s)^2}{2 + 2\alpha - (4+3\alpha)s + (2+\alpha)s^2}$$  \hspace{1cm} (B.10)

or, using $s \ll 1$,

$$x \sim \log \left( \frac{2 + \frac{\alpha}{1+\alpha}}{\alpha} \right) \sim \log \left( \frac{2}{\alpha} \right) \sim \log \left( \frac{1}{\alpha} \right) \sim \log \left( \frac{\rho}{\eta} \right)$$  \hspace{1cm} (B.11)

which is valid only for $\eta \ll \rho$ because self-consistence requires that $x \gg 1$; this regime overlaps with the one we considered previously, and the outcome confirms the above results. Further corrections may be obtained at will by Taylor-expanding in $e^{-x}$.

### .5 B3: Optimal wavenumber near the RT boundary

In order to look at the vicinity at the RT phase transition, we must assume $x \ll 1^2$.

We can then Taylor-expand in $x$, yielding:

$$D(x) = 2s (2 - \alpha s) + 4 (1 - 4s - \alpha s) x + (-12 - 2\alpha + 11\alpha s + 24s) x^2 + O(x^3)$$  \hspace{1cm} (B.12)

Or, using $s \ll 1$,

$$D(x)/2 \sim s (2 - \alpha s) + 2 (1 - \alpha s) x - (\alpha + 6) x^2$$  \hspace{1cm} (B.13)

$^2$Results will of course also be valid in other regions where the optimal wavenumber may be sufficiently small.
The solutions to $D(x) = 0$ are

$$x_\pm = 1 - \alpha s \pm \sqrt{(\alpha s - 1)^2 + s(\alpha + 6)(2 - \alpha s)} \over \alpha + 6 = 1 - \alpha s \pm \sqrt{1 + 12s - 6\alpha s^2} \over \alpha + 6$$

(B.14)

It is easy to see that $x_-$ is always negative, and we should keep in mind that, by definition, the variable $x$ must be positive. Let us look at Eq. (B.14) in the various regimes.

We first consider the regime $\eta < \zeta$, which means $\alpha s < 1$. In this regime, only $x_+$ is positive. The self-consistency is verified because the wavenumber (B.14) is indeed small in the parameter $1/\alpha$ and $\alpha s \lesssim 1$ implies $\alpha \sim 1/s \to \infty$. Using the smallness of $s \ll 1$, we may write

$$x_+ \sim {1 + \alpha s \pm \sqrt{1 - 6\alpha s^2} \over \alpha + 6}$$

(B.15)

and in the vicinity of $\eta \sim \zeta$ (just beneath the diagonal of the phase diagram) we have

$$x(\eta \lesssim \zeta) \sim (\alpha + 6)^{-1}$$

(B.16)

which, using $\alpha \gg 1$, yields $\omega \sim {\mu \over \zeta \eta} = \Omega$.

In the regime $\zeta < \eta < \sqrt{2}\zeta$, we have $1 < \alpha s < 2$, hence once again only the positive solution matters. In the vicinity of the phase boundary $\eta \sim \sqrt{2}\zeta$, we find

$$x \sim s {2 - \alpha s \over 2 \alpha s - 1}$$

(B.17)
which, translated in terms of wavenumber, yields:

\[ \omega_M^2 = \frac{\mu^2}{2\zeta^4} \frac{2\zeta^2 - \eta^2}{\eta^2 - \zeta^2} \]  \hspace{1cm} (B.18)

that is,

\[ \omega_M = \frac{\mu}{\sqrt{2\zeta^2}} \sqrt{\frac{2\zeta^2 - \eta^2}{\eta^2 - \zeta^2}} \]  \hspace{1cm} (B.19)

which describes the critical behavior of the orientation pattern. As the wavenumber goes to zero, the size of the orientation domains is diverging.

As for the region \( \eta > \sqrt{2\zeta} \), we have there \( \alpha > 2 \), and it is easy to see that both solutions \( x_\pm \) are always negative. Indeed, in that region the derivative of the eigenvalue with respect to wavenumber never vanishes, and the maximum is found in the cusp at \( \omega = 0 \).
Appendix C: Applications of the random-matrix results

This appendix will be devoted to the detailed calculations of the explicit expressions for the spectral density Eq. (5.7), the power spectrum Eq. (6.11), and the average squared norm Eqs. (6.1) and (6.5), for the specific examples of $M$, $L$ and $R$ presented in Sec. 5.4.

In the examples worked out in the sections C1 and C2, both $R$ and $L$ are proportional to the identity matrix; we will take $L = 1$ and $R = \sigma 1$. Furthermore, in such examples, calculations will be done by choosing the unit of time such that $\tau = 1$ (notice that given Eq. (2.12), the elements of $A$ and $M$ have dimensions of frequency); then at the end of the calculations, using the replacements $t \to t\sigma$, $z \to z/\sigma$, $\gamma \to \gamma/\sigma$, $M \to M/\sigma$, and $\rho \to \sigma^2 \rho$ (with the latter applying to both the eigenvalue density and the power spectral density) gives the result for general $\sigma$.

The eigenvalue density and the norm squared $\|x\|^2$ are invariant with respect to unitary transforms, and, for $L$ and $R$ proportional to the identity, so is the distribution of the random part of $A$, Eq. (5.22). Thus by effecting a unitary transform $M \to U^\dagger MU$, it may be assumed that $M$ is already in its Schur form Eq. (5.147) without loss of generality.
.6 C1: Single feedforward chain of length \( N \)

This section will be devoted to the example in Sec. 5.4.2, where \( M \) is

\[
M = T = \begin{pmatrix}
0 & w & 0 & \cdots \\
0 & 0 & w & \cdots \\
\vdots & \vdots & \vdots & \ddots
\end{pmatrix}
\]  

(C.1)

or \( M_{ij} = w \delta_{i+1,j} \).

Spectral support

Let us begin by calculating the shape of the spectral support. According to Eqs. (5.7)–(5.8), in order to calculate the spectral support, one should calculate first the inverse of

\[
M_z M_z^\dagger + g^2 = (z - M)(z - M)^\dagger + g^2
\]

(remember that we have set \( \sigma = 1 \)). To this aim, notice that

\[
K_{ij} \equiv \left[(z - M)(z - M)^\dagger\right]_{ij} = Q_{ij} - |w|^2 \delta_{iN} \delta_{jN}
\]  

(C.2)

where

\[
Q_{ij} \equiv (|z|^2 + |w|^2) \delta_{ij} - w \bar{z} \delta_{i+1,j} - \bar{w} z \delta_{i,j+1}.
\]  

(C.3)

As the difference \((z - M)(z - M)^\dagger - Q = -|w|^2 \mathbf{e}_N \mathbf{e}_N^\dagger\) is single rank, the Woodbury formula for matrix inversion can be applied, yielding

\[
\frac{1}{K + g^2} = \frac{1}{Q + g^2} + \frac{1}{Q + g^2} \frac{\mathbf{e}_N \mathbf{e}_N^\dagger}{Q + g^2} \frac{1}{Q + g^2} \left(\frac{1}{|w|^2 - \mathbf{e}_N^\dagger (Q + g^2)^{-1} \mathbf{e}_N}\right),
\]  

(C.4)

323
where \( \mathbf{e}_N^\top = (0, \ldots, 0, 1) \). The only conditions for the validity of Eq. (C.4) is that the factor in parenthesis is not singular, i.e. \( \mathbf{e}_N^\top (Q + g^2)^{-1} \mathbf{e}_N \neq |w|^{-2} \); the validity of this condition will be considered below.

Since \( Q \) is a Toeplitz matrix, and Hermitian, it can be diagonalized easily. Using standard methods, it can be seen that the eigenvalues and eigenvectors of \( Q \), satisfying \( Q \mathbf{v}_n = \lambda_n \mathbf{v}_n \), are given by

\[
\lambda_n = \left| z - |w| e^{i\phi_n} \right|^2, \quad \phi_n = \frac{\pi n}{N + 1} \quad \text{ (C.5)}
\]

\[
\mathbf{v}_n^j = \sqrt{\frac{2}{N + 1}} \left( \frac{\bar{w} z}{w \bar{z}} \right)^{j/2} \sin \phi_n j \quad \text{ (C.6)}
\]

for \( n = 1, \ldots, N \).

The eigenvectors are orthonormal \( \mathbf{v}_n^\dagger \mathbf{v}_m = \delta_{nm} \), which allows to use the following spectral representation

\[
\frac{1}{Q + g^2} = \sum_{n=1}^{N} \mathbf{v}_n \frac{1}{\lambda_n + g^2} \mathbf{v}_n^\dagger \quad \text{ (C.7)}
\]

Plugging Eqs. (C.5)–(C.7) into Eq. (C.4), and also noticing that Eqs. (C.6)–(C.7) allow to write

\[
\text{Tr} \left( \frac{1}{Q + g^2} \mathbf{e}_N \mathbf{e}_N^\dagger \frac{1}{Q + g^2} \right) = \left\| \frac{1}{Q + g^2} \mathbf{e}_N \right\|^2 = \frac{1}{N + 1} \sum_{n=1}^{N} \frac{2 \sin^2 \phi_n}{(\lambda_n + g^2)^2} = - \frac{\partial I_2(g, z)}{\partial g^2}, \quad \text{ (C.8)}
\]

one obtains

\[
\text{tr} \frac{1}{K + g^2} = I_1(g, z) + \frac{1}{N} \frac{- \partial I_2(g, z)}{\partial g^2} \quad \text{ (C.9)}
\]
where

\[ I_1(g, z) \equiv \text{tr} \frac{1}{Q + g^2} = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\lambda_n + g^2} \]  \hspace{1cm} (C.10)

\[ I_2(g, z) \equiv e^T \frac{1}{Q + g^2} e = \frac{1}{N + 1} \sum_{n=1}^{N} \frac{2 \sin^2 \phi_n}{\lambda_n + g^2}. \]  \hspace{1cm} (C.11)

In the \( N \to \infty \) limit, the sums in Eqs. (C.10)–(C.11) can be approximated by the integrals

\[ I_1(g, z) = \int_0^{2\pi} \frac{1}{|z| - |w|e^{i\phi}|^2 + g^2} \frac{d\phi}{2\pi}, \]  \hspace{1cm} (C.12)

\[ I_2(g, z) = \int_0^{2\pi} \frac{2 \sin^2 \phi}{|z| - |w|e^{i\phi}|^2 + g^2} \frac{d\phi}{2\pi}. \]  \hspace{1cm} (C.13)

Some elementary contour integration then yields

\[ I_1(g, z) = \left[ (|z|^2 + |w|^2 + g^2)^2 - 4|w|^2|z|^2 \right]^{-1/2}, \]  \hspace{1cm} (C.14)

\[ I_2(g, z) = \frac{|z|^2 + |w|^2 + g^2 - I_1(g, z)^{-1}}{2|w|^2|z|^2}. \]  \hspace{1cm} (C.15)

In particular, it can be seen that \( I_2(0, z) = \min(|w|^{-2}, |z|^{-2}) \), so that the condition for the validity of Eq. (C.4) would be violated for \(|z| < |w|\), if \( g \) turns out to be zero. However, note that \( I_2(g, z) \) is a decreasing function of \( g^2 \), so for finite \( g^2 > 0 \), the denominator in Eq. (C.9) is always positive (as is its numerator, for the same reason). Thus if one follows the correct procedure of Eq. (5.18)–(5.19), taking the \( N \to \infty \) limit before sending \( g^2 \) to zero, it is justifiable to use Eqs. (C.4) and (C.9). Furthermore, for \( g^2 > 0 \) the second term in Eq. (C.9) is \( O(N^{-1}) \), and should be neglected.
Solving Eq. (5.8) (with left hand side correctly interpreted as Eq. (5.18)), which now takes the form \( I_1(g, z) = 1 \), yields

\[
g(z)^2 = -|z|^2 - |w|^2 + \sqrt{4|w|^2|z|^2 - 1}. \tag{C.16}
\]

This is positive if and only if

\[
\sqrt{|w|^2 - 1} \leq |z| \leq \sqrt{|w|^2 + 1}, \tag{C.17}
\]

which after the proper rescaling yields Eq. (5.145) for general \( \sigma \), as was to be shown. Note that Eq. (C.17) is precisely the region given by Eq. (5.19), which in the present case reads \( I_1(0, z) \geq 1 \).

It is instructive to compare this result with what would be obtained by naively using Eq. (5.4), i.e. \( \text{tr} (K^{-1}) \geq 1 \), wherein \( g \) is set to zero before taking the \( N \to \infty \) limit; as will be shown, that only yields the second inequality in Eq. (C.17). To see this, first note that if \( |w| > |z| \), Eq. (C.9) may be used even for \( g^2 = 0 \) (since the denominator of the last term does not vanish), which yields \( \text{tr} [(M_z M_z^\dagger)^{-1}] = \text{tr} (K^{-1}) = I_1(0, z) + o(1) \), and by Eq. (5.4), the second inequality in Eq. (C.17).

For \( |z| < |w| \), however, one cannot set \( g = 0 \) in Eq. (C.9). In fact, when \( |z| < |w| \), the matrix \( z - M \) has an exponentially small singular value; to see this, note that the vector \( u \) with components \( u_i = (z_w)^{i-1} \) satisfies \( (z - M)u = w(z_w)^N e_N \), so that \( \| (z - M)u \| = |w||z_w|^N \), and since \( s_{\text{min}}(z - M) \leq \frac{\| (z - M)u \|}{\|u\|} \) and \( \|u\| \geq 1 \), it follows that \( s_{\text{min}}(z - M) \leq |w||z_w|^N \), which is \( O(e^{-cN}) \) for \( |z| < |w| \). For large enough \( N \), this singular value alone
suffices to make Eq. (5.10) (equivalent to Eq. (5.4)) hold for any \( |z| < |w| \), as \( \frac{1}{N} s_{\text{min}}(z)^{-2} \) diverges despite its \( \frac{1}{N} \) prefactor.

**Density profile**

Let us now calculate the eigenvalue density in the annulus Eq. (C.17). In order to use Eq. (5.7), let us first calculate

\[
\text{tr} \left( \frac{\bar{z} - M^\dagger}{\bar{K} + g(z)^2} \right) = \text{tr} \left( \frac{M^\dagger}{\bar{K} + g(z)^2} \right) \tag{C.18}
\]

where Eq. (5.8) was used to write the last expression.

To obtain \( \text{tr} \frac{M^\dagger}{\bar{K} + g(z)^2} \), once again Eq. (C.4) may be used. In the region Eq. (C.17), the contribution of the second term in Eq. (C.4) is again suppressed by \( 1/N \); from Eq. (C.1) and (C.7), one finds

\[
\text{tr} M^\dagger [Q + g^2]^{-1} = \bar{w} \frac{1}{N} \sum_{n=1}^{N} \left( \sum_{j=1}^{N-1} v_n^j \tilde{v}_n^{j+1} \right) \frac{1}{\lambda_n + g^2} \tag{C.19}
\]

A straightforward calculation using Eq. (C.6) (and the orthonormality of \( v_n \)) yields

\[
\sum_{j=1}^{N-1} v_n^j \tilde{v}_n^{j+1} = \left( \frac{w z}{w z} \right)^{1/2} \cos \phi_n. \]

Using this and approximating the sum over \( n \) with an integral leads to

\[
\text{tr} \left[ \frac{M^\dagger}{Q + g(z)^2} \right] \approx \frac{|w||z|}{z} \int_{0}^{2\pi} \frac{\cos \phi}{|z| - |w| e^{i\phi}|^2 + g(z)^2} \frac{d\phi}{2\pi},
\]

\[
= \frac{1}{2z} \left[ |z|^2 + |w|^2 + g(z)^2 I_1(g(z), z) - 1 \right]. \tag{C.20}
\]

Using Eq. (C.20) with \( I_1(g(z), z) = 1 \) (true in the region Eq. (C.17)), differentiating
Eq. (C.18) with respect to $\bar{z}$, and substituting in Eq. (5.7), one finds

$$\rho(z) = \frac{1}{\pi} \left[ 1 - \frac{|w|^2}{\sqrt{4|w|^2|z|^2 + 1}} \right], \quad (C.21)$$

for $z$ in the region Eq. (C.17). After the proper rescaling, this yields Eq. (5.146), as was to be shown.

**Average norm squared**

Let us now turn to the calculation of $\langle \|x(t)\|^2 \rangle_j$, using (6.5). To calculate the trace in the denominator of Eq. (6.5), first note that for Eq. (C.1) the expansion $(z-M)^{-1} = \sum_{n=0}^{N-1} \frac{M^n}{z^n}$ terminates and is exact, yielding

$$\left[ \frac{1}{z-M} \right]_{i,j} = \frac{1}{z} \left( \frac{w}{z} \right)^{j-i}, \quad (C.22)$$

for $j \geq i$, and zero otherwise.

Turning the sums in the trace into a sum over the nonzero diagonals of Eq. (C.22), this gives

$$\text{tr} \left( \frac{1}{\bar{z}_2 - M^t} \frac{1}{z_1 - M} \right) = \frac{1}{\bar{z}_2 z_1} \sum_{n=0}^{N-1} (1 - \frac{n}{N}) q^n \quad (C.23)$$

where $q \equiv |w|^2/(\bar{z}_2 z_1)$ and $z_i = \gamma + i\omega_i$. The condition of stability of Eq. (2.12) requires the entire spectrum of $-\gamma I + A = -\gamma I + M + J$ to be to the left of the imaginary axis.

By Eq. (C.17), this requires $\gamma > \sqrt{|w|^2 + 1} > |w|$. It follows that $|q| < 1$, and therefore the geometric series Eq. (C.23) converges as $N \to \infty$. Summing the series and retaining
terms of leading order as $N \to \infty$ gives

$$
\text{tr}\left( \frac{1}{z_2 - M^T z_1 - M} \right) = \frac{1}{z_2 z_1 - |w|^2}.
$$

(C.24)

Setting the initial condition $x_0$ in Eqs. (6.5) (or the input amplitude $I_0$ in Eq. (6.11)) to $e_N = (0, \cdots, 0, 1)^T$, and using Eq. (C.22), it appears that the numerator in Eq. (6.5) is also given by the RHS of Eq. (C.24). Using this and Eqs. (C.24) gives

$$
F(z_1, z_2) = \frac{1}{z_2 z_1 - |w|^2 - 1},
$$

(C.25)

for the integrand of Eq. (6.5) which was denoted by $F(z_1, z_2)$, with $z_i = \gamma + i \omega_i, (i = 1, 2)$.

By comparing the integrand of Eq. (6.5) with Eq. (6.13), it is seen that to obtain the total power spectrum for the input amplitude $I_0 = I_0(0, \cdots, 0, 1)^T$, one needs to multiply Eq. (C.25) by $I_0^2 = \|I_0\|^2$ and substitute $z_1 = z_2 = \gamma + i \omega$. With the proper rescaling, this yields Eq. (6.57) for general $\sigma$. The formula for $\langle \|x(t)\|^2 \rangle_I$ can be obtained by substituting Eq. (C.25) with $z_i = \gamma + i \omega_i$ for the integrand of Eq. (6.5). Changing the integration variables by $\omega_1 = \Omega + \omega/2$ and $\omega_2 = \Omega - \omega/2$ leads to

$$
\langle \|x(t)\|^2 \rangle_I = \int \frac{d\omega}{2\pi} e^{it\omega} \int \frac{d\Omega}{2\pi} \int \frac{d\Omega}{\Omega^2 + (\gamma + i\omega/2)^2 - |w|^2} \frac{1}{2\pi \sqrt{\gamma + i\omega/2}^2 - |w|^2 - 1},
$$

(C.26)

Finally, consulting a table of Laplace transforms (Abramowitz and Stegun, 1970) gives

$$
\langle \|x(t)\|^2 \rangle_I = e^{-2\gamma t} I_0(2t \sqrt{|w|^2 + 1}), \quad (t \geq 0)
$$

(C.27)
where \( I_0(x) \) is the 0-th order modified Bessel function. Implementing the rescalings \( t \to t\sigma, \gamma \to \gamma/\sigma \) and \( w \to w/\sigma \) finally yields Eq. (6.55), as was to be shown.

.7 C2: \( N/2 \) feedforward chains of length 2

This section of the Appendix will be devoted to the explicit calculations for the example of Sec. 5.4.3 where \( M \) is given by Eq. (5.147) (without loss of generality, \( M \) can be assumed to be in its Schur form), using formulae (5.7)–(5.8) for the spectral density and Eq. (6.5) for \( \langle \|x(t)\|^2 \rangle \).

Spectral Support

Let us begin by computing the shape of the spectral support. From Eq. (5.147), \( K \equiv M_z M_z^\dagger = (z - M)(z - M)^\dagger \) (setting \( L = R = 1 \) in Eq. (5.5); see the comments at the beginning of the Appendix) is a block-diagonal matrix with \( 2 \times 2 \) diagonal blocks, with the \( b \)-th block \((b = 1, \ldots, N/2)\) given by

\[
\begin{pmatrix}
    z & -w_b \\
    0 & z
\end{pmatrix}
\begin{pmatrix}
    \bar{z} & 0 \\
    -\bar{w}_b & \bar{z}
\end{pmatrix}
= \begin{pmatrix}
    |z|^2 + |w_b|^2 & -w_b \bar{z} \\
    -\bar{w}_b z & |z|^2
\end{pmatrix},
\]

(C.28)

where \( w_b \) is the corresponding Schur weight in Eq. (5.147).

Likewise, \((K + g^2)^{-1}\) whose trace appears in Eqs. (5.8) is given by a block-diagonal matrix with diagonal blocks

\[
\frac{1}{(|z|^2 + g^2)^2 + |w_b|^2 g^2}
\begin{pmatrix}
    |z|^2 + g^2 & w_b \bar{z} \\
    \bar{w}_b z & |z|^2 + g^2 + |w_b|^2
\end{pmatrix}
\]

Taking the
normalized trace yields

\[
\text{tr} \left( K + g^2 \right)^{-1} = \left\langle \frac{|z|^2 + g^2 + \frac{1}{2} |w_b|^2}{(|z|^2 + g^2)^2 + |w_b|^2 g^2} \right\rangle_b,
\]

where \( \langle \cdot \rangle_b \) means averaging over the \( N/2 \) blocks, i.e. \( \langle f(w_b) \rangle_b \equiv \frac{1}{N/2} \sum_{b=1}^{N/2} f(w_b) \).

Let us first calculate the support boundary of \( \rho(z) \). As discussed in Sec. 5.2.2, when (for \( |z| \neq 0 \)) all singular values of \( M_z = z - M \) are bounded from below as \( N \to \infty \), the support is correctly given by Eq. (5.4) (cases in which some \( s_i(z) \) are \( o(1) \) will be discussed further below). Setting \( g = 0 \) in Eq. (C.29), and substituting in Eq. (5.4) yields

\[
1 \leq \text{tr} K^{-1} = |z|^{-2} + \mu^2 |z|^{-4},
\]

where \( \mu^2 = \frac{1}{2} \langle |w_b|^2 \rangle_b = \text{tr} (M^\dagger M) \).

It follows that the support is the disk \( |z| \leq r_0 \), where

\[
r_0^2 = \frac{1}{2} + \sqrt{\frac{1}{4} + \mu^2}.
\]

The replacements \( \mu \to \mu/\sigma \) and \( r_0 \to r_0/\sigma \) then yield Eq. (5.149), as was to be shown.

**Density profile**

From Eqs. (5.8) and (C.29), within the support, \( g^2(z) \) is found by solving the equation

\[
\text{tr} \left( K + g^2 \right)^{-1} = \left\langle \frac{|z|^2 + g^2 + \frac{1}{2} |w_b|^2}{(|z|^2 + g^2)^2 + |w_b|^2 g^2} \right\rangle_b = 1,
\]

(C.32)
while for $|z| > r_0$ we have $g(z) = 0$. It is clear from Eq. (C.32) that $g^2(z)$ depends on $z$ and $\bar{z}$ only through $|z| \equiv r$. From Eq. (5.7), within its support the eigenvalue density is given by

$$\pi \rho(z) = \frac{\partial}{\partial \bar{z}} \text{tr}[M_\bar{z}^1(K + g^2)^{-1}] = 1 - \frac{\partial}{\partial \bar{z}} \text{tr}[M_\bar{z}^1(K + g^2)^{-1}] , \quad (C.33)$$

with the short-hand $g^2 = g^2(|z|)$ has been used (from the solution of Eq. (C.32)); to write the second equality in Eq. (C.33), Eq. (C.32) has been used.

From Eqs. (5.147) and (C.29), it can be seen that $M_\bar{z}^1(K + g^2)^{-1}$ has the same block-diagonal structure as Eq. (5.147), and a short calculation shows that $\text{tr}[M_\bar{z}^1(K + g^2)^{-1}] = \bar{z}I_3(|z|)$, where

$$I_3(r) \equiv \left\langle \frac{1}{2} |w_b|^2 \rightangle \left\langle \frac{r^2 + g(r)^2}{(r^2 + |w_b|^2 g(r)^2)^2} \rightangle , \quad (C.34)$$

$I_3(r)$ is manifestly positive (assuming some $w_b$ are nonzero), while when $g^2 > 0$, from Eq. (C.32) we have $I_3(r) \leq 1$, and thus

$$0 < I_3(r) \leq 1 , \quad (C.35)$$

Replacing this in Eq. (C.33), and using the fact that

$$\bar{z} \frac{\partial f(|z|)}{\partial \bar{z}} = 2r \left. \frac{\partial f(r)}{\partial r} \right|_{r = |z|} , \quad (C.36)$$

allows to write

$$\pi \rho(z) = \frac{1}{2r} \frac{\partial}{\partial r} \left[ r^2 - r^2 I_3(r) \right] , \quad (C.37)$$

for $r = |z| \leq r_0$, and $\pi \rho(z) = 0$ otherwise; the spectral density is rotationally symmetric
and depends only on $r = |z|$.

The advantage of writing the density as a complete derivative is that it can be immediately integrated to yield $n_<(r)$, the proportion of eigenvalues with modulus smaller than some radius $r$. Indeed, $n_<(r) = 2\pi \int_0^r \rho(r') r' dr'$, which upon substitution of Eq. (C.37), yields

$$n_<(r) = r^2 (1 - I_3(r)) \quad (r \leq r_0).$$  \hfill (C.38)

Likewise, let $n_>(r) \equiv 1 - n_<(r)$ be the proportion of eigenvalues with modulus larger than $r$. From these definitions, it follows that

$$\rho(r) = \frac{1}{2\pi r} \frac{\partial n_<(r)}{\partial r} = - \frac{1}{2\pi r} \frac{\partial n_>(r)}{\partial r},$$  \hfill (C.39)

and from Eqs. (C.38) and $n_>(r) = 1 - n_<(r)$, after some manipulation exploiting Eq. (C.32), we obtain

$$n_>(r) = g(r)^2 (1 + I_3(r)).$$  \hfill (C.40)

We see that beyond the radius $r$ at which $g^2$ vanishes (which when all $w_b$’s are bounded is $r = r_0$), $n_>(r)$ and $\rho(r) = - \frac{1}{2\pi r} \frac{\partial n_>}{\partial r}$ vanish identically, while for smaller $r$ they are positive.

**Regularization procedure**

In cases in which some $w_b$ grow without bound as $N \to \infty$, some singular values $s_i(z)$ of $M_z = z - M$ are $o(1)$, and more care is needed.

First, to see this, note that by definition $s_i(z)^2$ are the eigenvalues of the block-diagonal
\( K = M_x M_x^\dagger \); thus they come in pairs composed of the eigenvalues of \( K \)'s \( 2 \times 2 \) blocks, given by Eq. (C.28).

Let us denote the pair of eigenvalues corresponding to block \( b \) by \( s_{b,\pm}(z)^2 \), with the plus and minus subscripts denoting the larger and smaller singular value, respectively. The sum \( s_{b,+}(z)^2 + s_{b,-}(z)^2 \) and the product \( s_{b,+}(z)^2 s_{b,-}(z)^2 \) are given by the trace and determinant of Eq. (C.28), i.e. by \( |w_b|^2 + 2 |z|^2 \) and \( |z|^4 \), respectively. It follows that for blocks where the feedforward weight \( w_b \) is \( O(1) \), both \( s_{b,\pm}(z) \) will be \( \Theta(1) \) for \( |z| \neq 0 \), while for blocks in which \( w_b \to \infty \) as \( N \to \infty \), we have

\[
\begin{align*}
    s_{b,+}^2(z) &= |w_b|^2 + O(1) \to \infty \\
    s_{b,-}^2(z) &\approx \frac{|z|^4}{|w_b|^2} = o(1).
\end{align*}
\]

Note that as stated after Eq. (5.2), the quantity \( \| M \|_F^2 = \mu^2 = \langle |w_b|^2 \rangle_b / 2 \) may be assumed to be \( O(1) \), so that at most \( o(N) \) number of weights can be unbounded, and each such \( w_b \) can at most be \( O(\sqrt{N}) \).

If all the \( w_b \) are \( O(1) \), and hence all singular values are \( \Theta(1) \) (for \( |z| \neq 0 \)), Eq. (C.31) yields the correct support radius as noted above, and for \( r \leq r_0 \), Eq. (C.32) yields a \( \Theta(1) \) solution for \( g(r)^2 \), which leads to a \( \Theta(1) \) solution for \( n_+(r) \) and \( \rho(r) \) via Eqs. (C.40)–(C.39).

In cases in which some \( w_b \) are unbounded, however, Eq. (C.31) (derived from Eq. (5.4)) may not yield the correct support boundary. Such cases are examples of the highly non-normal cases mentioned in the general discussion after Eq. (5.11), for which the support of \( \lim_{N \to \infty} \rho(z) \) must be found by using Eqs. (5.18)–(5.19). This is equivalent to solving Eq. (C.32) after the limit \( N \to \infty \) is taken (assuming \( g^2 > 0 \)), and then finding where the
solution for $g^2(|z|)$ vanishes, which yields the correct support radius. From Eq. (C.40) this is indeed the radius at which $\lim_{N \to \infty} n_>(r)$ and hence $\lim_{N \to \infty} \rho(r)$ vanish as well. This radius is in general smaller than $r_0$ as given by Eq. (C.31).

The density $\rho(z)$ will now be calculated for two specific examples of $M$, one from each group.

**A regular example: uniform feedforward weights**

The first example is that of equal and $O(1)$ feedforward weights in all blocks, which will be called $w$ (in terms of Eq. (5.148), this case corresponds to $K = w1$). Here the block averages in Eqs. (C.32) and (C.34) may be dropped, replacing $w_b$ with $w$.

Solving Eqs. (C.32) for $g^2(|z| = r)$ yields

$$g^2(r) = \frac{1}{2} - \frac{w^2}{2} - r^2 + \frac{1}{2} \sqrt{1 + w^4 + 4w^2r^2}. \tag{C.43}$$

Substituting this into Eq. (C.34) and Eq. (C.38) gives then

$$n_<(r) = r^2 - \frac{w^2r^2}{1 + \sqrt{1 + w^4 + 4w^2r^2}}. \tag{C.44}$$

The replacements $w \to w/\sigma$ and $r \to r/\sigma$ then finally Eq. (5.150) for general $\sigma$ (as was to be shown) and $\rho(r)$ can be calculated using Eq. (C.39).

**An irregular example: the minimal balanced networks**

The second case to be now considered is that of Eq. (5.151). In that case, only one of the blocks has a nonzero Schur weight given by $|w_1|^2 = \text{Tr}(M^\dagger M) = N\mu^2 = O(N)$, where
\( \mu = O(1) \) is given by Eq. (5.152).

Equation (C.32) now yields

\[
1 = \frac{1}{r^2 + g^2} + \frac{\mu^2}{(r^2 + g^2)^2 + N\mu^2g^2} \frac{r^2 - g^2}{r^2 + g^2},
\]

or

\[
\frac{r^2 + g^2 - 1}{r^2 - g^2} = \frac{\mu^2}{(r^2 + g^2)^2 + N\mu^2g^2}.
\]

The RHS of this last equation is \( I_3(r) \), as follows from Eq. (C.34); thus using Eq. (C.46) we can rewrite Eq. (C.40) as

\[
n_>(r) = g^2(r) \frac{2r^2 - 1}{r^2 - g^2(r)}.
\]

Let us now solve Eq. (C.46) to find \( g(r)^2 \). As noted above, and in accordance with the general prescription given after Eq. (5.11), for the purpose of obtaining \( \lim_{N \to \infty} \rho(z) \) one needs first to take the \( N \to \infty \) limit in Eq. (C.45), keeping \( g^2 > 0 \) fixed, and only then solve for \( g^2 \). Doing so makes the last term in Eq. (C.45) vanish, and we obtain \( g^2(r) = 1 - r^2 \).

This is positive for \( r \leq 1 \) and vanishes at \( r = 1 \), the correct support radius of \( \lim_{N \to \infty} \rho(z) \), which is strictly smaller than \( r_0 \) given by Eq. (C.31). From Eq. (C.47) we obtain \( n_>(r) = g^2(r) = 1 - r^2 \).

It then follows from Eq. (C.39) that the \( N \to \infty \) limit of the eigenvalue density is identical with the circular law (the result for the \( M = 0 \)), i.e. \( \lim_{N \to \infty} \rho(r) = \frac{1}{\pi} \) for \( r \leq 1 \) and zero otherwise. With the correct scaling, this yields Eq. (5.153).
Contrary to the general prescription given after Eq. (5.11), equations (C.32), (C.34) and (C.40) will now be solved for $r > 1$, without taking the limit $N \to \infty$ first. As will be see, the resulting solution for $g(r)^2$ (and therefore, by Eqs. (C.35) and (C.40), the solutions for $n_>(r)$ and $\rho(r)$) will be nonzero but $o(1)$ for $1 < r \leq r_0$. As discussed in Sec. 5.4.3, these finite-size corrections, which in general are not trustworthy, in the present case are in surprisingly good agreement with simulations for some range of $r$’s beyond $r_{\Theta(1)}$, but deviate from the true $n_>(r)$ for larger $r$ (see Fig. 5.8).

At finite $N$, it can indeed be checked that Eq. (C.32) has a positive solution for $g^2$ if and only if $r < r_0$, with $r_0$ given by Eq. (C.31). Simplifying Eq. (C.46) yields a cubic equation in $g^2$. However, it turns out that ignoring the cubic term in $g^2$ is harmless for large $N$; the quadratic approximation has the positive solution

$$g^2(r) = \sqrt{\left[\frac{1 - r^2}{2}\right]^2 + \frac{r^2(r^2 + \mu^2) - r^6}{\mu^2 N} + \frac{1 - r^2}{2}}, \quad \text{(C.48)}$$

and for all $r < r_0$, corrections to Eq. (C.48) when the cubic term is reinstated decay faster than the leading contribution from Eq. (C.48). Nevertheless, we numerically solved the full cubic equation (C.32) to obtain the black curve in Fig. 5.8, and the blue trace in Fig. 5.7).

First, analyzing Eq. (C.48) we see that $g^2(r)$ is indeed $\Theta(1)$ only for $r < 1$, where as we already found $g^2(r) = 1 - r^2 + o(1)$. Furthermore, for a fixed $r > 1$ (such that $r - 1$ does not vanish as $N \to \infty$), the solution for $g(r)$ is $O(N^{-1})$. Thus from Eq. (C.47) we see that $Nn_>(r)$, i.e. the total number of eigenvalues with modulus larger than $r$, for $1 < r < r_0$ (and $r - 1 = \Theta(1)$) is only $O(1)$; the solution for $Nn_>(r)$ is shown in Fig. 5.8.

Correspondingly, from Eqs. (C.47) and (C.39) we see that $\rho(r)$ is $o(1)$ in this region and
vanishes in the limit \( N \to \infty \), as was already found. Now let us calculate the total number of eigenvalues lying outside the circle \(|z| = 1\). This is given by \( N n_>(1) \). Eq. (C.48) gives \( g^2(1) = \frac{1}{\sqrt{N}} \), and substituting in Eq. (C.47) yields

\[
N_>(1) \equiv N n_>(1) = \sqrt{N} + O(1).
\]

(C.49)

With the proper rescaling this yields Eq. (5.154) for general \( \sigma \). Note that, according to Eq. (C.48), \( g(r) \) (and hence \( n_>(r) \)) remains \( \Theta(N^{-1/2}) \) (as opposed to \( O(N^{-1}) \)) in a thin boundary layer outside of width \( \Theta(N^{-1/2}) \) just outside of the circle \(|z| = 1\).

**Average norm squared**

Let us now work out the formula for \( \langle \|x(t)\|^2 \rangle \), Eqs. (6.5)–(6.1), when the initial condition \( x_0 \) is the second Schur-vector in block \( b = a \), which will be called \( e_{a2} \).

In the Schur representation, Eq. (5.147), this is written as \( e_{a2} = (0, 1)^T \) (writing only the components of \( e_{a2} \) in block \( a \)). To calculate the numerator in Eq. (6.5) one should first calculate \( (z - T_a)^{-1}e_{a2} \), where \( T_a = \begin{pmatrix} 0 & w_a \\ 0 & 0 \end{pmatrix} \) denotes the \( a \)-th diagonal \( 2 \times 2 \) block of Eq. (5.147). Since \( T_a^2 = 0 \), we have \( (z - T_a)^{-1} = z^{-1} + z^{-2}T_a \), which yields \( v_a(z) \equiv (z - T_a)^{-1}e_{a2} = (w_a z^{-2}, z^{-1})^T \).

We thus obtain

\[
\frac{1}{z_2} - M^\dagger \frac{1}{z_1} M x_0 = v_a(z_2)^T v_a(z_1) = \frac{1}{z_1 z_2} + \frac{|w_a|^2}{z_1^2 z_2^2}.
\]

(C.50)
On the other hand, we have

\[
\text{tr} \frac{1}{\bar{z}_2 - M^\dagger} \frac{1}{z_1 - M} = \left\langle \frac{1}{2} \text{Tr}_{2 \times 2} (\bar{z}_2 - T_b^\dagger)^{-1} (z_1 - T_b)^{-1} \right\rangle_b = \frac{1}{z_1 \bar{z}_2} + \frac{\langle |w_b|^2 \rangle_b / 2}{z_1^2 \bar{z}_2^2}. \tag{C.51}
\]

Substituting Eqs. (C.50)–(C.51) in Eq. (6.5) we obtain

\[
F(z_1, z_2) = \frac{z_1 \bar{z}_2 + |w_a|^2}{(z_1 \bar{z}_2)^2 - (z_1 \bar{z}_2 + \mu^2)}. \tag{C.52}
\]

where we used \( \mu^2 = \langle |w_b|^2 \rangle_b / 2 \), and we denoted the integrand of Eq. (6.5) by \( F(z_1, z_2) \) with \( z_i = \gamma + i \omega_i, \ (i = 1, 2) \).

By comparing the integrand of Eq. (6.5) with Eq. (6.13), we see that substituting \( z_1 = z_2 = \gamma + i \omega \) into Eq. (C.52) yields the total power spectrum, \( \left\langle \|x(t)\|^2 \right\rangle_J \). After the proper rescalings, this yields Eq. (6.62) for general \( \sigma \). To obtain \( \left\langle \|x(t)\|^2 \right\rangle_J \), on the other hand, we should substitute Eq. (C.52) into Eq. (6.1) with \( z_i = \gamma + i \omega_i \).

Let us use the change of variables \( \omega_1 = \Omega + \omega \) and \( \omega_2 = \Omega - \omega \). Then we have \( z_1 \bar{z}_2 = \Omega^2 + (\gamma + i \omega)^2 \), and from Eq. (6.1) we obtain

\[
\left\langle \|x(t)\|^2 \right\rangle_J = \int \frac{d\omega}{2\pi} e^{2it\omega} f_a(\gamma + i \omega) \tag{C.53}
\]

where we defined

\[
f_a(u) = 2 \int \frac{d\Omega}{2\pi} \frac{\Omega^2 + u^2 + |w_a|^2}{(\Omega^2 + u^2)^2 - (\Omega^2 + u^2 + \mu^2)}. \tag{C.54}
\]
Let us rewrite the integrand in Eq. (C.54) as

\[
\frac{\Omega^2 + u^2 + |w_a|^2}{(\Omega^2 + u^2 - r_0^2)(\Omega^2 + u^2 + r_1^2)} = \frac{\Omega^2 + u^2 + |w_a|^2}{r_0^2 + r_1^2} \left[ \frac{1}{\Omega^2 + u^2 - r_0^2} - \frac{1}{\Omega^2 + u^2 + r_1^2} \right],
\]

(C.55)

where \(r_0^2\) was defined in Eq. (C.31) and

\[
r_1^2 \equiv r_0^2 - 1 \geq 0.
\]

(C.56)

One can calculate the integral over \(\Omega\) in Eq. (C.54) by contour integration, closing the contour, say, in the upper half of complex plane. The poles of the first and the second terms on the second line of Eq. (C.55) are located at \(\Omega_{0,\pm} = \pm i\sqrt{u^2 - r_0^2}\) and \(\Omega_{1,\pm} = \pm i\sqrt{u^2 + r_1^2}\), respectively. For \(u = \gamma + i\omega (\gamma > 0)\) the roots falling in the upper half plane are \(\Omega_{0,+}\) and \(\Omega_{1,+}\), independently of \(\omega\).

From their residues we obtain

\[
f_a(u) = \frac{1}{r_0^2 + r_1^2} \left[ \frac{r_0^2 + |w_a|^2}{\sqrt{u^2 - r_0^2}} + \frac{r_1^2 - |w_a|^2}{\sqrt{u^2 + r_1^2}} \right].
\]

(C.57)

The integral of Eq. (C.57) in Eq. (C.53) is essentially the inverse Laplace transform of Eq. (C.57). Consulting a table of Laplace transforms (Abramowitz and Stegun, 1970) yields

\[
\langle \|x(t)\|^2 \rangle_t = e^{-2\gamma t} \left[ \frac{r_0^2 + |w_a|^2}{r_0^2 + r_1^2} J_0(2r_0 t) + \frac{r_1^2 - |w_a|^2}{r_0^2 + r_1^2} J_0(2r_1 t) \right].
\]

(C.58)

where \(J_0(x) (I_0(x))\) is the 0-th order modified Bessel function.

From Eqs. (C.31) and (C.56) it follows that \(r_0^2 + r_1^2 = \sqrt{1 + 4\mu^2}\), and using \(\mu^2 = \)
\langle |w_b|^2 \rangle_b / 2 \text{ once again, we obtain}

\langle \| x(t) \|^2 \rangle_J = \left[ \frac{1 + C_a}{2} I_0(2r_0 t) + \frac{1 - C_a}{2} J_0(2r_1 t) \right] e^{-2\gamma t} \quad (C.59)

where we defined

\[ C_a \equiv \frac{1 + 2|w_a|^2}{\sqrt{1 + 2\langle |w_b|^2 \rangle_b}}. \quad (C.60) \]

Effecting the proper rescalings we obtain the result for general \( \sigma \), Eqs. (6.60)–(6.61), as was to be proven.

.8 C3: Composite network with factorizable weights

This final section of the appendix will be devoted to the explicit calculations for a composite network network consisting of \( n \) neural types, as presented Sec. 5.4.3, with \( M, L \) and \( R \) given by Eqs. (5.156)–(5.159).

Eqs. (5.5) and (5.156)–(5.159) allow to write \( M_z = z(RL)^{-1} - suu^\top \), and

\[ M_z M_z^\dagger = |z|^2(RL)^{-2} - zsvu^\top - zsuv^\top + s^2uu^\top, \quad (C.61) \]

where \( v \equiv (RL)^{-1}u \). Applying the Woodbury matrix identity yields

\[ \frac{1}{g^2 + M_z M_z^\dagger} = Q - QU \frac{1}{D^{-1} + U^\top QU} U^\top Q \quad (C.62) \]
where the $N \times 2$ matrix $U = \begin{pmatrix} u & v \end{pmatrix}$ has been defined, and

$$Q \equiv \frac{1}{g^2 + |z|^2(RL)^{-2}} \quad \text{(C.63)}$$

$$D \equiv \begin{pmatrix} s^2 & -zs \\ -zs & 0 \end{pmatrix} \quad \text{(C.64)}$$

It can be seen that for $g > 0$, $\text{tr} \left( g^2 + M_z M_z^\dagger \right)^{-1} = \text{tr} Q$ up to $o(1)$ corrections. Indeed, from Eq. (C.62), the remainder $\Delta(g, z) \equiv \text{tr} \left( g^2 + M_z M_z^\dagger \right)^{-1} - \text{tr} Q$ is found to be

$$\Delta(g, z) = -\frac{1}{N} \text{Tr} \left[ \frac{U^\dagger Q^2 U}{D^{-1} + U^\dagger Q U} \right] \quad \text{(C.65)}$$

where the trace is now over $2 \times 2$ matrices.

Inverting $D$ yields

$$D^{-1} = -\begin{pmatrix} 0 & (zs)^{-1} \\ (zs)^{-1} & |z|^{-2} \end{pmatrix} \quad \text{(C.66)}$$

and for $p = 1, 2$ we obtain

$$U^\dagger Q^p U = \begin{pmatrix} u^\dagger Q^p u & u^\dagger Q^p v \\ u^\dagger Q^p v & v^\dagger Q^p v \end{pmatrix} = \begin{pmatrix} I_{p,0} & I_{p,1} \\ I_{p,1} & I_{p,2} \end{pmatrix}, \quad \text{(C.67)}$$

where

$$I_{p,k}(g, z) \equiv \frac{1}{N} \sum_{i=1}^{N} \frac{(l_{\alpha(i)} r_{\alpha(i)})^{-k}}{g^2 + |z|^2(l_{\alpha(i)} r_{\alpha(i)})^{-2}} = \left\langle \frac{\sigma_{\alpha}^{-k}}{(g^2 + \sigma_{\alpha}^{-2}|z|^2)^{p}} \right\rangle_{\alpha} \quad \text{(C.68)}$$
where the averaging notation has been defined by Eq. (5.163), and \( \alpha(i) \) is the label of the population to which neuron \( i \) belongs (the explicit \( g \) and \( z \) dependence of \( I_{p,k} \) will be dropped when convenient).

Note that all \( I_{p,k}(g, z) \) are \( O(1) \) and for even \( k \) they are positive. Inverting \( D^{-1} + U^\dagger QU \), we obtain

\[
\Delta(g, z) = \frac{1}{N - \det(D^{-1} + U^\dagger QU)} T(g, z)
\]

where

\[
T(g, z) \equiv \text{Tr} \left[ \begin{pmatrix} I_{2,0} & I_{2,1} \\ I_{2,1} & I_{2,2} \end{pmatrix} \begin{pmatrix} I_{1,2} - |z|^{-2} & (zs)^{-1} - I_{1,1} \\ (zs)^{-1} - I_{1,1} & I_{1,0} \end{pmatrix} \right]
\]

\[
= I_{2,2}I_{1,0} + I_{2,0} \left( I_{1,2} - \frac{1}{|z|^2} \right) - 2I_{2,1} \left( I_{1,1} - \frac{1}{s\text{Re} z} \right)
\]

and

\[
-\det(D^{-1} + U^\dagger QU) = I_{1,0} \left( \frac{1}{|z|^2} - I_{1,2} \right) + \left| I_{1,1} - \frac{1}{s\text{Re} z} \right|^2 = \frac{g^2}{|z|^2} I_{1,0}^2 + \left| I_{1,1} - \frac{1}{s\text{Re} z} \right|^2.
\]

We see that both \( T(g, z) \) and \( -\det(D^{-1} + U^\dagger QU) \) are \( O(1) \) (to obtain their limits as \( N \to \infty \), one may set \( s^{-1} = O(N^{-1/2}) \) equal to zero) and since \( -\det(D^{-1} + U^\dagger QU) \geq \frac{g^2}{|z|^2} I_{1,0}^2 \) and \( I_{1,0} > 0 \), it follows that for \( g > 0 \), the denominator in Eq. (C.69) is bounded away from zero, and hence \( \Delta(g, z) = O(N^{-1}) \) and can be safely ignored for \( g > 0 \).

We deduce that, indeed, \( \text{tr} \left( g^2 + M_z M_z^\dagger \right)^{-1} = \text{tr} Q + o(1) \). Eq. (C.63) moreover allows
to write $\text{tr} \, Q = I_{1,0}(g, z)$ and hence, from Eqs. (5.18),

$$K(g, z) = \lim_{N \to \infty} \text{tr} \, Q = \left\langle \frac{2}{g^2 + \sigma_\alpha^{-2}r^2} \right\rangle_\alpha$$

where $r \equiv |z|$.

Note that the approximation $\text{tr} \, (g^2 + M_zM_z^\dagger)^{-1} \approx \text{tr} \, Q$ is equivalent to using $M_zM_z^\dagger = |z|^2 (RL)^{-2}$ instead of the full expression Eq. (C.61) and hence to setting $M = 0$. Accordingly, the support of the eigenvalue distribution is given by Eq. (5.12), or equivalently by Eq. (5.164), and within this support, $g^2$ is depends only on $|z| = r$ and is found by solving Eq. (5.14), or equivalently Eq. (5.166). Similar considerations show that, in using Eq. (5.7) to obtain $\lim_{N \to \infty} \rho(z)$, one can set $M = 0$, yielding an isotropic eigenvalue density.

From Eqs. (5.13)–(5.15), the proportion, $n_\geq$, of eigenvalues lying a distance larger than $r$ is equal to $g^2(r)$, which is found by solving Eq. (5.166). The results Eqs. (5.16)–(5.17) also hold, wherein the normalized sums over $i$ can be replaced with appropriate averages $\langle \cdot \rangle_\alpha$.

Let us now go back to the expression for $\Delta(g, z)$, and consider the case $g = 0$. In this case

$$I_{n,k}(0, z) = |z|^{-2n} \langle \sigma_\alpha^{2n-k} \rangle_\alpha,$$

which allows to write

$$T(g, z) = |z|^{-6} \left( \langle \sigma_\alpha^2 \rangle_\alpha^2 - 2 \langle \sigma_\alpha^3 \rangle_\alpha \langle \sigma_\alpha \rangle_\alpha \right) + 2 \langle \sigma_\alpha^3 \rangle_\alpha \frac{s^{-1}}{|z|^4 \Re z}$$

(C.73)
and

\[ -\det(D^{-1} + U^\dagger Q U) = \left| \frac{z^{-2}}{\langle \sigma_\alpha \rangle - \frac{s^{-1}}{z}} \right|^2. \quad \text{(C.74)} \]

In the special case in which \( \langle \sigma_\alpha \rangle = 0 \) (corresponding to the special case of the example Eq. (5.151) with \( f \mu_E - (1 - f) \mu_I \propto u \cdot v = 0 \), which was considered above), the determinant will have a vanishing limit as \( N \to \infty \) (or \( s^{-1} \to 0 \)). This leads to a finite limit for \( \Delta(0, z) \) and we obtain

\[ \Delta(0, z) = \frac{s^2 \langle \sigma_\alpha^2 \rangle^2}{N |z|^4} = \frac{\zeta^2 \langle \sigma_\alpha^2 \rangle^2}{|z|^4}, \quad \langle \sigma_\alpha \rangle = 0. \quad \text{(C.75)} \]

Adding this to \( \text{tr} Q \) in the right side of Eq. (C.71), and using the naive formula Eq. (5.4) or \( \mathcal{K}(0, z) = 1 \) for the spectral boundary, we would have obtained the equation

\[ 1 = \frac{\langle \sigma_\alpha^2 \rangle}{r^2} + \frac{\zeta^2 \langle \sigma_\alpha^2 \rangle^2}{r^4}. \quad \text{(C.76)} \]

This in turn yields the radius Eq. (5.165) which is larger than the true boundary of the support of \( \lim_{N \to \infty} \rho(z) \) given by Eq. (5.164).
Appendix D: Moments of the same-frequency covariance

This Appendix will be devoted to deriving the results (6.77-6.78), stated at the end of Chapter 6.

In accordance with the procedure of generating functions, the starting point is to perform the disorder average in equation (6.70). This is done by writing \( Z_\omega(K) = \langle Z_\omega(K|A) \rangle_A \) as

\[
Z_\omega(K|A) = \int dX \ p_\omega(X|K) \exp(K^T X), \tag{D.1}
\]

with \( p_\omega(X) \) given by

\[
p_\omega(X) = (2\pi D)^{-N/2} \det(z \mathbf{1} - W) \exp\left(- \frac{1}{2D} \sum_{ijk} X_i(z \ast \delta_{ij} - W_{ji})(z \delta_{jk} - W_{jk})X_k\right), \tag{D.2}
\]

where the exponent is quadratic in \( W = A - i\omega \mathbf{1} \). It can be made linear through the simple Hubbard-Stratonovich transformation

\[
(2\pi D)^{-N/2} e^{-\frac{1}{2D} x^T (z\ast - W)^T(z\mathbf{1} - W)x} = \int \frac{d\tilde{X}}{(2\pi i)^N} e^{\frac{\partial}{\partial \tilde{x}^T} \tilde{x}^T \tilde{x} - \tilde{x}^T (z\mathbf{1} - W) \tilde{x}} \tag{D.3}
\]

where the variables \( \tilde{X}_j \) takes values along the purely imaginary axis between \(-i\infty\) and \(+i\infty\).
Let us define \( \int D\tilde{X} = \prod_j \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tilde{x}_j \), whereas \( \int DX = \prod_j \int_{-\infty}^{\infty} dX_j \), we can then write

\[
p_{\omega}(X) = \det (1 - W) \int D\tilde{X} \exp \left[ \frac{D}{2} \tilde{X}^T \tilde{X} - \tilde{X}^T (z1 - W) X \right] \quad (D.4)
\]

Plugging this into equation 6.70 yields

\[
Z_{\omega}(K|W) = |\det (z1 - W)| \int DX \int D\tilde{X} \exp \left[ \frac{D}{2} \tilde{X}^T \tilde{X} - \tilde{X}^T (z1 - W) X + K^T X \right]
\]

\[
Z_{\omega}(K) = \int DX \int D\tilde{X} \exp \left[ \frac{D}{2} \tilde{X}^T \tilde{X} + (K^T - \tilde{X}) X \right]
\]

\[
\times \int dW F(W) \delta (z1 - W) \exp (\tilde{X}^T W X)
\]

\[
\quad \quad \quad \quad (D.6)
\]

where the convergence of the W-integral is apparent because the determinant can only be polynomial in \( W_{ij} \), and the last exponent is purely imaginary.

We can write the distribution of \( W \) as

\[
F(W) = \left( \frac{\sqrt{N}}{\sqrt{2\pi} \sigma} \right)^{N^2} \exp \left( -\frac{N}{2\sigma^2} \sum_{ij} (W_{ij} - \mu)^2 \right)
\]

\[
\quad \quad \quad \quad (D.7)
\]

where \( \mu = M - i\omega 1 \).

Substituting (D.7) into Eq. (D.6), we find that

\[
Z(J) = \left( \frac{\sqrt{N}}{\sqrt{2\pi} \sigma} \right)^{N^2} \int DX \int D\tilde{X} \exp \left[ \frac{D}{2} \tilde{X}^T \tilde{X} + (J^T - \tilde{X}) X \right] I(X, \tilde{X})
\]

\[
\quad \quad \quad \quad (D.8)
\]
where

$$I \left( X, \tilde{X} \right) = \int dW \exp \left[ \log \det (1 - W) + \sum_{ij} \tilde{X}_i W_{ij} X_j - \frac{N}{2\sigma^2} \sum_{ij} (W_{ij} - \mu_{ij})^2 \right]$$  \hspace{1cm} (D.9)

Now,

$$\log \det (1 - W) = \text{Tr} \log (1 - \mu) + \text{Tr} \log \left[ 1 - (1 - \mu)^{-1} J \right]$$  \hspace{1cm} (D.10)

from which it is easy to check that for large matrices only the first term need be kept.

This allows to perform the integral over $J$ with a standard saddle point expansion of the exponent, which is equivalent to a non-crossing approximation and thus valid in the thermodynamic limit\(^3\), leading to Eqs. (6.74-6.76) of Chapter 6, from which Eqs. (6.77-6.78) ensue by the application of Eqs. (6.71-6.73).

\(^3\)See also Gurau (2010); some exceptions to this general equivalence were studied by Riva and Zanon (1981).