Regularization: Problems and Promises

Terrance E. Boult*

CUCS Technical Report #321-88
Columbia University Department of Computer Science
New York City, New York, 10027. tboult@cs.columbia.edu

May 12, 1990

1 Introduction [1]

Regularization is becoming a popular framework for describing and solving many ill-posed problems of computer vision. Of course, a generalized framework is only useful if it provides additional insights or benefits unavailable without it. This paper discusses some of the benefits promised by the regularization framework. Additionally, as a mathematical paradigm for vision, regularization presents many difficulties for the vision researcher, and some of these difficulties are discussed in this paper. The paper then discusses the lack of development of most of the "promises of regularization" theory, and gives a brief look as some of the promises which have been realized.

In the context of smooth surface reconstruction, the paper addresses one of the most difficult problems with the use of regularization: the problem of determining an appropriate functional class, norm, and regularization stabilizing functional. In particular, results are discussed from an experiment which subjectively orders various functional classes and stabilizing functionals for a regularization-based formulation of the surface reconstruction problem. The conclusions drawn include the fact that there exist non-traditional formulations of this regularization problem which provide better results.

The paper concludes with a brief mention of two more general frameworks and their relationship to regularization.

[1] Supported in part by DARPA Grant #N00039-84-C-0165.

2 What a unifying mathematical formulation should do.

If a mathematical paradigm is to be useful as a unifying framework for a class of problems, it should have at least one of three basic properties:

1. provide new insights and/or better intuition into the structure or nature of the problems (this is often the case when the mathematical paradigm is borrowed from a well studied area),
2. provide new (mathematical) tools to simplify the analysis of the problems and aid in their solution,
3. predict new, experimentally verifiable, properties of the problems and/or variations of the problems.

If none of these properties is provided by a unifying mathematical formulation of a class of problems, it is, at best, an exercise in mathematics, and more likely, a useless attempt to import a mathematical framework into an area.

Using these as the basic criterion for the usefulness of a mathematical framework, we will examine the usefulness of regularization theory in computer vision. The specialization of the first two properties to vision problems is straight forward, and the last one requires either a relation to the physical world or psychological world to determine what one means be "experimentally" verifiable.

3 The problems with Regularization?

In recent years, a few vision researchers have put forth regularization as a framework for the solution of many low-level vision problems, e.g., [1] and [2]. In this section we will deal only with papers which have used "pure" regularization theory, e.g., [1] and the references therein, [3], and [4] Many other "regularization" papers actually have to do with either optimal approximation theory, e.g., [5], [6] or with extensions to regularization which render them as a class of Bayesian modeling, e.g., [7], [8], [9]. These extensions will be briefly mentioned in the final section of this paper.

Regularization is a term applied to three related techniques for turning an ill-posed mathematical problem into a well-posed one. A formal definition can be found in [10]. Let y be the data and let z be the "solution" such that $Sz = y$ where S is a linear operator. Because the problem is ill-posed z does not exist, is not unique, and/or does not depend continuously on the data y. Regularization techniques work by restricting the space of problems/solutions so as to insure a solution which is exists, is unique and does depend continuously on the data. To do this they require the choice of a *stabilizing function* P, a norm $\|\cdot\|_S$ on the space of stabilized solutions (i.e., $\|Pz\|$, this measure is often called the stabilizing functional), and a norm $\|\cdot\|_D$ on the space of data elements (for vision almost always the L^2 norm on a finite dimensional Euclidean space). And use these terms to define the restrictions, e.g., they may seek the z which minimizes $\|Sz - y\|_D + \lambda\|Pz\|_S$.

The choice of stabilizing function and norms is rarely unique and far from trivial. As pointed out in [10][p.59] "The choice of the stabilizing functional is often prompted by the nature of the problem. However, in a number of cases, more than one choice is possible". In that book over a dozen different methods are presented to deriving different "stabilizing functions", but *no* techniques are presented for determining which stabilizers are the "best".

Their are five main problems with regularization:

1. How smooth (regular) should the solution be, so as to be realistic? Also, should this smoothness be the same everywhere?

2. What should the tradeoff between error and smoothness be?

3. If the realistic solution requires non-convex stabilizers, can one still show that solving the regularized formulation is possible?

4. How does one incorporate other information into the solution process?

5. How does one derive meaningful error estimates?

All but the last problem really have to do with the question: How does one chose the function spaces, norms and stabilizing functions?

4 Failed Promises??

The promises of a unified framework for low-level vision problems are really the promises of any useful mathematical framework. So let us look at each of the three main promises of a mathematical formulation and comment on whether the promise was kept or failed when regularization theory is considered as a unifying framework for early vision problems.

4.1 Does it provide new insights and/or better intuition?

To date, there have been many "regularized" solutions to vision problems, e.g., [1] list eight problems. Unfortunately, few, if any, of these problems were actually solved using regularization. Instead, almost all were problems that had already been solved, using calculus of variations, smoothness assumptions, and/or approximation theory, and were reformulated, after the fact, into the terminology of regularization.

Further, most of the few new results which used regularization, e.g., [4], did not really use the mathematical properties of regularization, rather it was simply a way of stating a "smoothness" assumption in a mathematically formal way.

Thus, I would conclude that regularization has not provided better intuition or new insights, in fact, people use and understand regularization in terms of more traditional mathematics and heuristics used in computer vision.

4.2 Does it provide new mathematical tools?

I argue that this promise of regularization is also not met. Of the many problems reformulated into the regularization framework, few have been analyzed from the point of view of regularization, let alone with tools from regularization theory. This does not imply that no mathematical tools exists, they do. For example, [10] provides some mathematical tools for the determination of the regularization parameter (i.e., optimal tradeoff between error and smoothness) for some classes of problems. Instead, it is either the case that the mathematical overhead encountered when attempting to use these tools has been to high, or that the results of applying them to practical problems has been unsatisfactory.

4.3 Can the framework make interesting predictions?

Here we finally have a promise that is at least partially kept. Because regularization does not determine exactly the function spaces, norms and stabilizing functions, once a problem has been formulated in this framework one might ask about similar formulations with variations in the norm/class/stabilizer. Thus regularization theory can predict new variations of the problem. However, it provides no means for determining which of the various formulations is most appropriate.

As far a predicting experimentally verifiable properties, regularization does make such predictions, but they are identical to the predictions of the mathematical formulations originally used to solve the problem (e.g., the "barber-pole illusion" is predicted by the calculus of variations formulation of the optic-flow problem as well as by the regularization formulation.) Since regularization does not provide error estimates the predictions are qualitative, not quantitative.

5 Regularized surface reconstruction

This section briefly discusses the results on determination of the appropriate class, norm and stabilizers for the problem of surface reconstruction from sparse depth data. More details can be found in [11], [5].

As mentioned in the previous section, determining the exact mathematical form of the regularization of a problem is troublesome. It seems impossible to determine which of the uncountable number of potential formulations is "best" so we settled for a psychological (subjective) ranking of various formulations. However, for the results to be meaningful, we needed to be able to determine the relationship between the errors of solutions under different formulations. Regularization did not provide the necessary tools and so we turned to the results of information-based complexity. We were thus able to define "optimal error" algorithms which solved the regularization problems, and comparisons of the outputs of the algorithms was pursued.

Since the algorithms had optimal error properties, any errors in the reconstruction of the surfaces were inherent in the model. Thus by comparing the reconstructions generated by the algorithms we could draw conclusions about the underlying models. Again, while the models themselves were regularization type models, the experiment was only valid because we were able to show the algorithms used were optimal error algorithms.

For the problem of surface reconstruction, 9 different formulations were tested, and the conclusion was that 4-5 of these were "most appropriate". Only one of these, corresponding to thin-plate splines, had been previously used in vision research. The remaining classes all had smoothness "between" a membrane and the thin-plate spline, and some had "stabilizers" of considerably higher orders. Even more surprising was the conclusion that the order of the "stabilizer" was not as important as the effective smoothness of the functional class to which it was applied. We note the discrete regularization techniques in [3] or [9] can not deal with these intermediate smoothness classes.

6 Other related frameworks

As mentioned before, the above comments apply only to "traditional regularization theory". However, there are other frameworks in computer vision which include regularization as a special case. The two most important are optimal-approximation (generalized splines), and Bayesian modeling. We briefly mention how these extensions allow "regularization" to keep some of the promises of a unified framework. (In other words, why the extensions are themselves viable unified frameworks for early vision).

Optimal approximation theory (see also information-based complexity [12]) has advantages over regularization theory in that it also provides mathematical tools for deriving the error of an approximation. Optimal error algorithms derived this way can also allow one to use comparisons of the "output" of the algorithms as a means of comparing the models. Furthermore, this framework provides for extensions to average case analysis, see [13]. The main drawback of this approach is that it requires moderately heavy mathematics. Furthermore, like regularization theory, incorporation of certain types of a priori information is difficult.

Bayesian modeling has as a special case, stochastic solutions to regularization problems. However, it has the added advantage that it can handle non-convex stabilizers and can easily incorporate additional a priori information, see [7]. Furthermore, it can also be used to generate error models, see [9]. The disadvantage here is the class of underlying functional classes is greatly restricted if one uses the discritized versions of the functional equations.

References

[1] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, no. 6035, pp. 314–319, 1985.

[2] T. Poggio and V. Torre, "Ill-posed problems and regularization in early vision," Tech. Rep. lab memo 773, Massachusetts Institute Technology AI, 1984. Also appeared in the Proceedings of the DARPA Image Understanding Workshop.

[3] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, pp. 413–424, July 1986.

[4] G. Medioni and Y. Yasumoto, "Robust estimation of 3-d motion parameters from a sequence of image frames using regularization," in *Proceedings of the DARPA Image Understanding Workshop*, pp. 117–128, DARPA, 1985.

[5] T. Boult, "What is regular in regularization?," in *Proceedings of the IEEE Computer Society International Conference on Computer Vision*, pp. 457–462, IEEE, June 1987.

[6] D. Lee and T. Pavlidis, "One-dimensional regularization with discontinuities," in *Proceedings of the IEEE Computer Society International Conference on Computer Vision*, pp. 572–577, IEEE, June 1987.

[7] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, November 1984.

[8] A. Blake and A. Zisserman, "Invariant surface reconstruction using weak continuity constraints," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 62–68, IEEE, 1986.

[9] R. Szeliski, "Regularization uses fractial priors," in *Proceedings of AAAI 87*, American Association for Artificial Intelligence, Morgan Kaufman Publishers, Inc., 1987.

[10] A. Tikhonov and V. Arsenin, *Solution of Ill-Posed Problems*. Washington, D.D.: V.H. Winston & Sons, 1977. Distibuted by J.Wiley & Sons, NY.

[11] T. E. Boult, *Information Based Complexity in Non-Linear Equations and Computer Vision*. PhD thesis, Department of Computer Science, Columbia University, 1986.

[12] J. R. Kender, D. Lee, and T. Boult, "Information-based complexity applied to optimal recovery of the 2 $\frac{1}{2}$d sketch," in *Proceedings of the IEEE Computer Society Workshop on Computer Vision: Representation and Control*, pp. 157–167, Oct. 1985.

[13] J. Traub, G. Wasilkowski, and H. Woźniakowski, *Information-Based Complexity*. Academic Press, New York, 1988. To appear.