MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

and

CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

Working Paper No. 264   December, 1984

ASSOCIATIVE LEARNING OF STANDARD REGULARIZING OPERATORS IN EARLY VISION

Tomaso Poggio and Anya Hurlbert

Abstract

Standard regularization methods can be used to solve satisfactorily several problems in early vision, including edge detection, surface reconstruction, the computation of motion and the recovery of color. In this paper, we suggest (a) that the quadratic variational principles corresponding to standard regularization methods are equivalent to a linear regularizing operator acting on the data and (b) that this operator can be synthesized through associative learning. The synthesis of the regularizing operator involves the computation of the pseudoinverse of the data. The pseudoinverse can be computed by iterative methods, that can be implemented in analog networks. Possible implications for biological visual systems are also discussed.

1. Introduction

Regularization theory denotes several techniques that can be used to solve ill-posed problems in the sense of Hadamard (1923). Recently, it has been pointed out that most of the problems in early computational vision are ill-posed in that sense (Poggio and Torre, 1984). One subset of regularization theory is what we call *standard* regularization techniques. With this term, we define the techniques devised by Tikhonov (1963, see Tikhonov and Arsenin, 1977), and the related techniques of Nashed (1974, 1976), Bertero (1982) and others. Standard regularization techniques typically lead to variational principles of the quadratic type. Although they have inherent limitations, discussed by Poggio and Torre (1984), standard regularization techniques are powerful methods for approaching ill-posed problems.

Several problems in early vision can be solved satisfactorily by standard regularization techniques. Examples are the computation of the optical flow (Horn and Schunck, 1981), surface reconstruction (Grimson, 1981, 1982), the computation of motion along contours (Hildreth, 1984) surface interpolation (Terzopoulos, 1984a, 1984b), filtering for edge detection (Poggio, Voorhees, and Yuille, 1984), the computation of color (Hurlbert and Poggio, 1984), stereo matching (Horn 1985, Poggio and Yuille, in preparation), and shape from shading (Ikeuchi and Horn, 1981).[1]

In this paper we show that the regularizing operator that provides the solution from the data can be learned by a simple associative learning scheme (as applied to colour vi sion in Hurlbert and Poggio, 1984) . Thus the regularizing algorithms needed to solve the problems mentioned above do not need to be formulated *explicitly*, but can be *learned associatively from a series of examples*. Furthermore, several of the associative memory schemes that have been proposed in the past could be used to synthesize the regularizing operator. Associative memories could be implemented easily (at least in principle) by biological systems. The synthesis of standard regularizing operators through associative learning is therefore especially attractive from the point of view of biological vision systems.

2. Standard Regularization Methods

Consider the equation

$$y = Az \tag{1}$$

where A is a known operator. The direct problem is to determine y from z, the inverse problem is to obtain z when y ("the data") are given. Though the direct problem is usually well-posed, the inverse problem is often ill-posed.

Standard regularization theories for "solving" ill-posed problems have been developed by Tikhonov (1963; Tikhonov and Arsenin, 1977) and Nashed (1974, 1976), amongst others. The basic idea of regularization techniques is to restrict the space of acceptable solutions by minimizing an appropriate functional. The standard regularization of the ill-posed problem of finding z from the data y requires the choice of norms $\|\cdot\|$ (usually quadratic) and of a linear *stabilizing functional* $\|Pz\|$. The choice is dictated by mathematical considerations, and, most importantly, by a physical analysis of the generic constraints on the problem. Among the methods that can be applied (see Poggio and Torre, 1984 and Bertero, 1982), the main one is to find z that minimizes

---

[1]The last two examples are only approximatively of the Tikhonov type.

$$\|Az - y\|^2 + \lambda \|Pz\|^2, \tag{2}$$

where λ is a regularization parameter. The regularization parameter λ controls the compromise between the degree of regularization of the solution and its closeness to the data. In this paper we assume that λ is given (for example, from estimates of the noise).

3. Associative Learning of Standard Regularizing Operators

Minimization of the regularization principle equation (2) corresponds to a *regularizing operator*, i.e. a system, acting on the input data y and providing, as an output, the regularized solution z. We want to show that this regularizing operator can be synthesized by associative learning from a set of examples. Our argument consists of two simple claims. The first claim is that the regularizing operator corresponding to quadratic variational principles is linear. The second is that any linear mapping between two vector input spaces can be synthesized by an associative scheme based on the computation of the pseudoinverse of the data. We explain now in more detail our argument.

(a) Variational principles of the form of equation (2) provide a regularized solution as a linear transformation of the data. The reason for this is that the Euler-Lagrange equations corresponding to equation (2) are linear partial differential equations and therefore define the solution z as a linear functional of the data y (depending on the boundary conditions).[2] For simplicity, we will restrict ourselves in this paper to the discrete case of equation (1), in which z and y are vectors and A and the Tikhonov stabilizer P are matrices. Equation (2) becomes then

$$\|Az - y\|^2 + \lambda \|Pz\|^2, \tag{3}$$

where $\|\cdot\|$ is a norm (for example l_2). We assume that P^{-1} exists. The minimum of this functional will occur at its unique stationary point z (see also Tikhonov and Arsenin, 1976). Setting to zero the gradient of the functional of equation (3) gives the minimum vector z as the solution of the system of linear equations

$$(A^T A + \lambda P^T P)z = A^T y \tag{4}$$

It follows that the solution z can be written as

$$z = Sy \tag{5}$$

and is therefore a linear transformation on the data vector y. It is important to notice that the form of the linear operator depends on the lattice in which the data lie, unless the operator is space- invariant, e.g., a convolution operator. In Hildreth's formulation of the computation of motion, for instance, the linear operator is parametrized by the specific contour or contour segment.

(b) The second observation is based on the following argument. Imagine that a set of input vectors y are available together with the corresponding correct solutions z. Arrange these vectors in two matrices Y and Z. The problem of synthesizing the regularizing operator S

---

[2] More about this can be found in Poggio and Koch (1984) and in Poggio, Voorhees and Yuille (1984).

2

that provides the regularized solution z for each vector y is the equivalent to " solving" the following equation

$$Z = SY \tag{6}$$

and finding the matrix S. A general solution to this problem is provided by

$$S = ZY^+ \tag{7}$$

where Y^+ is the pseudoinverse of Y. If no exact solution exists, the pseudoinverse of equation (7) is optimal in terms of least-squares, and is the most robust against errors.

Definitions and properties of the pseudoinverse can be found in several books (see for example Albert, 1982, and Campbell and Meyer, 1979). An extension to Hilbert spaces is given by Groetsch (1977). The pseudoinverse can be computed in an adaptive way by updating it when new data become available, a feature of particular interest for practical applications (and for biological systems).

Equation (7) shows that the standard regularizing operator S (parametrized by the lattice of data points) can be synthesized without need of an explicit variational principle, if a sufficient set of correct input-output data pairs is available to the system.

4. Discussion

We have shown that the system that provides a standard regularization solution from a data vector can be synthesized by a simple associative learning technique based on the computation of the pseudoinverse of a set of data vectors. An attractive feature of this method is that the pseudoinverse can be computed by adaptive techniques which improve its form as more data become available. Networks can be devised for implementing various approximations to these iterative schemes (see, for example, Kohonen, 1977). In particular, it has been proposed in the past that networks of neurons could synthesize close approximations to the pseudoinverse.

The associative learning scheme we have described takes its simplest form when the linear operator is space-invariant. Yet it is possible in more complicated cases with spatial dependence to synthesize the operator by piecing together operators correponding to sublattices of the data. In Hildreth's case, for example, operators corresponding to different segments of a contour may be computed if appropriate boundary conditions are available.

The scheme is furthermore restricted to the synthesis of linear mappings and therefore to quadratic variational principles. Nonlinear associative memories have also been proposed. In particular, one of us has extended the pseudoinverse technique to learning a nonlinear mapping of the polynomial type between two vector spaces. This simple result allows the synthesis of an arbitrary nonlinear map between two finite sets of vectors (see Poggio, 1982). Although these techniques may be of little practical use in general (because the matrix to be pseudoinverted can be very large), the approach shows that it is possible to synthesize by an associative learning scheme nonlinear regularizing operators associated with nonquadratic, nonstandard regularization principles. A closely related idea is to use a nonlinear *coding* of the input vectors *before* the (linear) associative memory (see Poggio, 1975). If the correct nonlinear coding is available, arbitrary nonlinear mappings can be learned effectively. This is connected with Kolmogorov's and Arnold's solution of Hilberth's 13th problem (see Poggio and Rosser, 1982 and Poggio, 1982).

References

Albert, A., *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York, 1972.

Bertero, M. "Problemi lineari non ben posti e metodi di regolarizzazione", in: *Problem non ben posti ed inversi*, Istituto di Analisi Globale, Firenze, 1982.

Campbell, S. L., and C. D. Meyer, Jr., *Generalized Inverses of Linear Transformations*, Pitman, London, 1979.

Grimson, W.E.L., *From Images to Surfaces: A computational study of the human early visual system*, MIT Press, Cambridge, Ma., 1981.

Grimson, W.E.L., "A computational theory of visual surface interpolation", *Phil. Trans. R. Soc. Lond., B*, **298**, 395–427, 1982.

Groetsch, C. W., *Generalized Inverses of Linear Operators*, Marcel Dekker, Inc., New York, 1977.

Hadamard, J., *Lectures on the Cauchy Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, 1923.

Hildreth, E.C., *The Measurement of Visual Motion*, MIT Press, Cambridge, Massachusetts, 1984.

Horn, B.K.P. "Robot Vision", M.I.T. Press, Cambridge, Mass, 1985.

Horn, B.K.P., and Schunck, B.G., "Determining optical flow", *Artificial Intelligence*, **17**, 185–203, 1981.

Hurlbert, A., and Poggio, T.. "A new color algorithm and learning it". AI Memo 716, 1984.

Ikeuchi, K., and Horn, B.K.P., "Numerical shape from shading and occluding boundaries", *Artificial Intelligence*, **17**, 141–184, 1981.

Kohonen, T. **Associative Memory**, Springer Verlag, Heidelberg, 1977.

Nashed, M. Z., "Approximate regularized solutions to improperly posed linear integral and operator equations in *Constructive and Computational Methods for Differential and Integral Equations* G. Auger, ed., Akademie Verlag; Berlin, 289–296, 1974.

Nashed, M. Z., ed. *Generalized Inverses and Applications*, Academic Press, New York, 1976.

Poggio, T. "On optimal nonlinear associative recall". *Biol. Cyber.*, 19, 201-209, 1975.

Poggio, T. "On optimal discrete estimation". In: **Proceedings of the First Symposium on Testing and Identification of Nonlinear Systems**, Mar. 17-20, California Institute of Technology, Pasadena, California, eds. G. D. McCann and P. Z. Marmarelis, 30-37, 1975.

Poggio, T. "Visual Algorithms". In: **Physical & Biological Processing of Images, eds.**, O. J. Braddick and A. C. Sleigh, Springer-Verlag, Berlin, pgs. 128-153, 1983.

Poggio, T. and Koch, C., "Analog networks: a new approach to neuronal computation", *Artificial Intelligence Lab. Memo*, **No. 783**, MIT, Cambridge, MA, 1984.

Poggio, T. and Rosser, B. "The computational problem of motor control." In: **Machine Intelligence** 10, Hayes, Mitchie & Pao, eds, 519-530, 1982.

Poggio, T. and Torre, V., "Ill-posed problems and regularization analysis in early vision" *Artificial Intelligence Lab. Memo*, **No. 773**, MIT, Cambridge, MA, April, 1984.

Poggio, T., Voorhees, H. and Yuille A., "Regularizing Edge Detection", *Artificial Intelligence Lab. Memo*, **No. 776**, MIT, Cambridge, 1984.

Terzopoulos, D., "Multiresolution computation of visible-surface representations," Ph.D. Thesis, Dept of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, January, 1984a.

Terzopoulos, D., "Multilevel reconstruction of visual surfaces: variational principles and finite element representations " *Artificial Intelligence Lab. Memo 671*, reprinted in *Multiresolution Image Processing and Analysis*, A. Rosenfeld (ed.), **237-310**, Springer-Verlag, Heidelberg, **1984b**.

Tikhonov, A. N. "Solution of incorrectly formulated problems and the regularization method", *Soviet Math. Dokl.*, **4**, 1035-1038, 1963.

Tikhonov, A.N. and Arsenin, V.Y., *Solutions of ill-posed problems*, Winston & Sons, Washington, D.C., 1977.