Determining the Optimal Weights in Multiple Objective Function Optimization

Michael A. Gennert

Department of Computer Science
Worcester Polytechnic Institute
Worcester, MA 01609

Alan L. Yuille

Division of Applied Sciences
Harvard University
Cambridge, MA 02138

September 21, 1998

Abstract

An important problem in computer vision is the determination of weights for multiple objective function optimization. This problem arises naturally in many reconstruction problems, where one wishes to reconstruct a function belonging to a constrained class of signals based upon noisy observed data. A common approach is to combine the objective functions into a single total cost function. The problem then is to determine appropriate weights for the objective functions. In this paper we propose techniques for automatically determining the weights, and discuss their properties. The Min–Max Principle, which avoids the problems of extremely low or high weights, is introduced. Expressions are derived relating the optimal weights, objective function values, and total cost.

1 Introduction

An important problem in computer vision is the determination of weights for multiple objective function optimization. This problem arises naturally in many reconstruction problems, where one wishes to reconstruct a function belonging to a constrained class of signals based upon noisy observed data. There is usually a tradeoff between reconstructing a function that is true to the data, and one that is true to the constraints. Instances of this tradeoff can be found in shape from shading [3], optical flow [4], surface interpolation [2], edge detection [7], visible surface reconstruction [9], and brightness-based stereo matching [1]. The recently popularized regularization method for solving ill-posed problems [6, 8] always requires the tradeoff of conflicting requirements.

The basic framework defines a cost or error functional which reflects the "badness" of a proposed solution to the reconstruction problem. Mathematical techniques, such as the calculus of variations, are used to find the best solution to the reconstruction problem. The contribution of each constraint to the cost funtional is weighted, and the weights may be adjusted to achieve a desired tradeoff.

In some cases, a priori knowledge may be used to determine a "best" set of weights. Typically, one must know some property of the observed data, such as the signal-to-noise ratio, before proceeding. This is not always possible. It would be better to have a method for determining weights that did not depend on a priori knowledge. In this paper we propose techniques for automatically determining weights, and discuss properties of these techniques.

Let us assume that there are several objectives to be achieved, and let there be a non-negative objective function for each objective which measures distance from that objective. We would like to combine the objective functions into a single cost function. In the following sections we will examine ways to construct the desired single objective function.

2 The Min–Max Principle

First consider a simple problem. This problem will address the tradeoffs involved in a two-objective optimization problem, where a cost function is to be minimized over a single variable. Our goal is to shed some insight into the more complex general problem, where the cost function incorporates arbitrarily many objectives to be minimized over a multiple-variable field.

2.1 Linear Weight Constraint

Let $y_1(z)$ and $y_2(z)$ be non-negative functions of a single state variable z. It may be assumed, without loss of generality, that $y_1(z) = 0$ and $y_2(z) = 0$ for some (not necessarily the same) z. These are the two objective functions. Let the overall cost function be a linear combination of the objectives, defined by

$$e(\lambda, z) = \lambda y_1(z) + (1 - \lambda) y_2(z), \qquad 0 \le \lambda \le 1.$$

For a given value of λ, there will exist a value of z which minimizes $e(\lambda, z)$. Let that minimum value be

$$e^*(\lambda) = \min_z e(\lambda, z) = e(\lambda, z^*(\lambda)).$$

Given λ, one generally tries to find $z^*(\lambda)$, the value with the minimum total cost. We can plot $e^*(\lambda)$ as a function of λ, as shown in figure 1. Since there exists some z such that $y_1(z) = 0$, and the y_i's are non-negative, $e^*(1) = \min y_1(z) = 0$. Likewise, $e^*(0) = 0$.

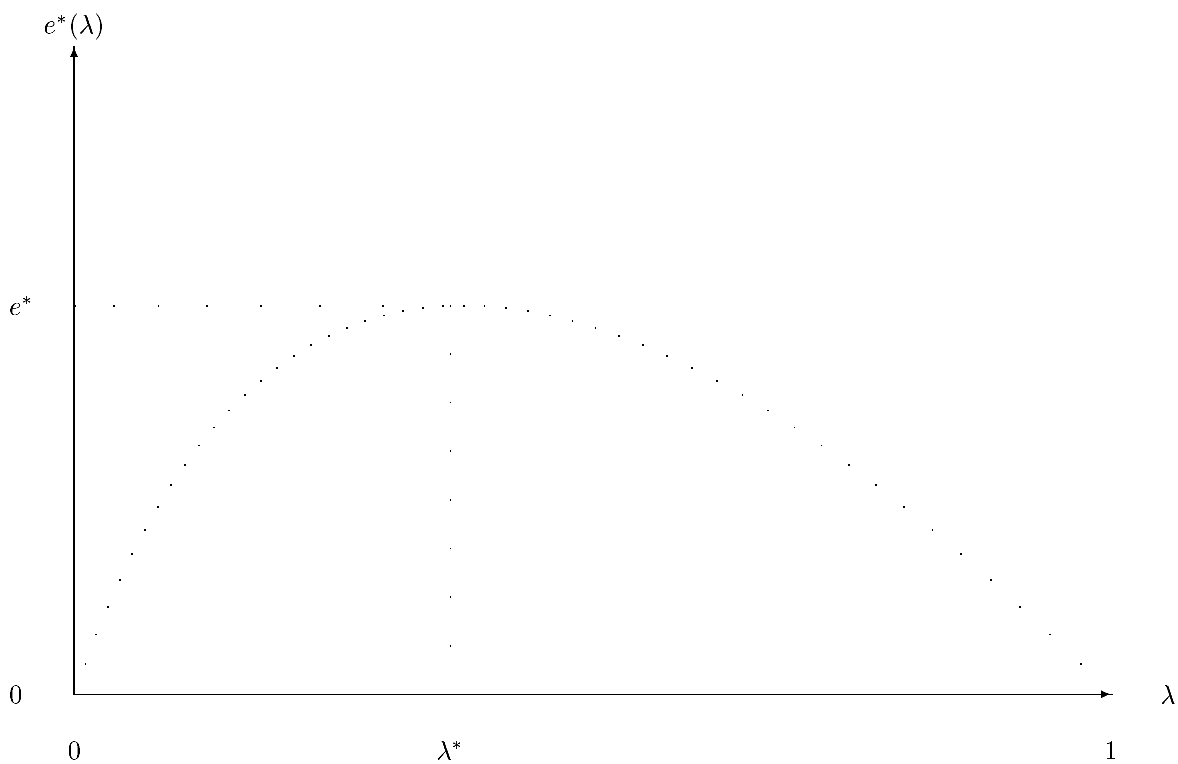

Figure 1: Total cost $e^*(\lambda)$ vs. λ. $e^*(\lambda)$ is convex with $e^*(0) = e^*(1) = 0$.

The total cost function must be convex, irrespective of the convexity of each objective function. Let $e_1^* = e^*(\lambda_1)$ and $e_2^* = e^*(\lambda_2)$. Then for any α between 0 and 1, $e^*(\alpha\lambda_1 + (1-\alpha)\lambda_2) \geq \alpha e_1^* + (1-\alpha)e_2^*$. The proof follows:

$$\begin{aligned}
e^*(\alpha\lambda_1 + (1-\alpha)\lambda_2) &= \min_z e(\alpha\lambda_1 + (1-\alpha)\lambda_2, z) \\
&= \min_z \big[\alpha(\lambda_1 y_1(z) + (1-\lambda_1)y_2(z)) \\
&\qquad + (1-\alpha)(\lambda_2 y_1(z) + (1-\lambda_2)y_2(z))\big] \\
&\geq \alpha \min_z \big[\lambda_1 y_1(z) + (1-\lambda_1)y_2(z)\big] \\
&\qquad + (1-\alpha)\min_z \big[\lambda_2 y_1(z) + (1-\lambda_2)y_2(z)\big] \\
&= \alpha e_1^* + (1-\alpha)e_2^*
\end{aligned}$$

Note that the proof does not depend on the convexity of either y_i.

The main difficulty in choosing λ is that if it is either too low or too high, one of the objective functions will be inadequately represented in the total cost, and the total cost will be too low. One way to ensure that the total cost will not be too low is to pick the maximum cost solution. That is, find λ^* such that $e^* = e^*(\lambda^*)$ is maximized. This may seem at first to be far from optimal, but if one recalls that e has already been minimized over z, it will be seen that the maximization over λ does make sense, while avoiding the problems of λ excessively low or high. This is the *Min–Max Principle.*

Note that if there exists a value of λ not equal to zero or one for which the total cost is zero, then the optimal cost found using the Min–Max Principle will also be zero. This follows from the convexity of $e^*(\lambda)$.

Let us see what the extremization of the total cost implies. If λ^* is the weight which maximizes $e^*(\lambda)$, then denote the value of the corresponding state variable by $z^* = z^*(\lambda^*)$. Since $e^*(\lambda)$ is maximized at λ^*, conclude that

$$\begin{aligned}
0 &= \left.\frac{d}{d\lambda}e^*(\lambda)\right|_{\lambda=\lambda^*} \\
&= \left.\frac{d}{d\lambda}e(\lambda, z^*(\lambda))\right|_{\lambda=\lambda^*} \\
&= \left.\frac{\partial}{\partial\lambda}e(\lambda, z^*)\right|_{\lambda=\lambda^*} + \left.\frac{\partial}{\partial z}e(\lambda^*, z)\frac{d}{d\lambda}z^*(\lambda)\right|_{\lambda=\lambda^*, z=z^*}
\end{aligned}$$

But the last term, $\frac{\partial}{\partial z}e(\lambda^*, z)$, equals zero, because z minimizes e. Therefore,

$$0 = \left.\frac{\partial}{\partial\lambda}e(\lambda, z^*)\right|_{\lambda=\lambda^*} = y_1(z^*) - y_2(z^*).$$

Each objective function assumes the same cost value, although the weights λ and $1-\lambda$ are not necessarily identical, and the weighted contribution of each objective function will not in general be identical.

2.2 Non-linear Weight Constraint

In this section we consider a variation on the previous solution technique in which the total cost function is not a linear combination of the weights. Instead, let the weights of the objectives, when squared, sum to a constant. The constant can be set to 1 without loss of generality. That is,

$$e(\lambda, z) = \lambda y_1(z) + \sqrt{1-\lambda^2}\,y_2(z), \qquad 0 \leq \lambda \leq 1.$$

As before, the optimal value of z given λ is $z^*(\lambda)$, where $e^*(\lambda) = \min_z e(\lambda, z) = e(\lambda, z^*(\lambda))$. A plot of $e^*(\lambda)$ would look much the same as in the linear combination case, figure 1. It can be shown that $e^*(\lambda)$ is convex.

The Min–Max Principle requires that λ^* be found which maximizes the total cost, avoiding the problems of λ excessively low or high. As before, $e^* = \max_\lambda e^*(\lambda) = e^*(\lambda^*)$. Using the chain rule for differentation we obtain:

$$0 = \frac{\partial}{\partial \lambda} e(\lambda, z^*)\bigg|_{\lambda=\lambda^*} = y_1(z^*) - \frac{\lambda}{\sqrt{1-\lambda^2}} y_2(z^*).$$

After rearrangement,

$$\frac{1}{\lambda} y_1(z^*) = \frac{1}{\sqrt{1-\lambda^2}} y_2(z^*).$$

Each objective function, when divided by its weight, is equal. The weights λ and $\sqrt{1-\lambda^2}$ will not in general be equal, therefore, the objective functions will not be equal. Also, the objective function with the greatest value will have the largest weight, so that its contribution to the total cost function is further increased. This contrasts with the linear weight case discussed above.

3 General Case

In this section we consider the generalization to any number of objective functions and any dimensionality state space. Replace the scalar variable z by a vector field $\mathbf{z}(\mathbf{x})$ defined over domain $\mathbf{x}$. We use a vector field for $\mathbf{z}$ because there may be more than one quantity of interest. For example, in optical flow $\mathbf{z}$ would be the horizontal and vertical components of optical flow. In stereo $\mathbf{z}$ would be the horizontal and vertical disparities. If the image model has other parameters, they may be incorporated into $\mathbf{z}$.

The two objective functions are replaced by n objective functionals

$$\mathbf{y}(\mathbf{z}) = \int \mathbf{L}(\mathbf{z}(\mathbf{x}))\, d\mathbf{x},$$

where $\mathbf{L}$ is a set of possibly non-linear operators on the vector field $\mathbf{z}$. The weights are given by vector $\boldsymbol{\lambda}$, so that $e(\boldsymbol{\lambda}, \mathbf{z}) = \boldsymbol{\lambda}^T \mathbf{y}(\mathbf{z})$. The least cost solution for a given set of weights is $\mathbf{z}^*(\boldsymbol{\lambda})$, yielding a total cost of $e^*(\boldsymbol{\lambda}) = \min_{\mathbf{z}} e(\boldsymbol{\lambda}, \mathbf{z}) = e(\boldsymbol{\lambda}, \mathbf{z}^*(\boldsymbol{\lambda}))$. The cost functional is convex in $\boldsymbol{\lambda}$. The proof follows:

Let $e^*(\boldsymbol{\lambda}_1) = e_1^*$ and $e^*(\boldsymbol{\lambda}_2) = e_2^*$. $e^*(\boldsymbol{\lambda})$ is convex if, for any α between 0 and 1, $e^*(\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2) \geq \alpha e_1^* + (1-\alpha) e_2^*$.

$$\begin{aligned} e^*(\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2) &= \min_{\mathbf{z}} e(\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2, \mathbf{z}) \\ &= \min_{\mathbf{z}} (\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2)^T \mathbf{y}(\mathbf{z}) \\ &\geq \min_{\mathbf{z}} \alpha\boldsymbol{\lambda}_1^T \mathbf{y}(\mathbf{z}) + \min_{\mathbf{z}} (1-\alpha)\boldsymbol{\lambda}_2^T \mathbf{y}(\mathbf{z}) \\ &= \alpha e_1^* + (1-\alpha) e_2^* \end{aligned}$$

The Min–Max Principle requires that we find $\boldsymbol{\lambda}$ to maximize $e^*(\boldsymbol{\lambda})$; the maximum is obtained at $\boldsymbol{\lambda}^*$.

Note that if there exists a zero-cost solution with all positive weights, it will be found by the Min–Max proceduire. This follows from the convexity of $e^*(\boldsymbol{\lambda})$. Thus, in cases where there is an obvious optimal solution, the proposed method will discover it. This property is independent of any constraint on the weights, apart from the positivity assumption.

Usually, a constraint is needed for the weight vector $\boldsymbol{\lambda}$. Otherwise, as $\boldsymbol{\lambda}$ increases without limit, so does the cost $e^*(\boldsymbol{\lambda})$. Two methods of constraining $\boldsymbol{\lambda}$ have been considered earlier. When generalized, these correspond to

$$\mathbf{1}^T \boldsymbol{\lambda} = 1 \qquad \text{and} \qquad |\boldsymbol{\lambda}| = 1,$$

where $\mathbf{1}^T = (1, 1, \ldots)$ is a vector of 1's. In both cases we require that $\boldsymbol{\lambda} \geq 0$, that is, every component of $\boldsymbol{\lambda}$ must be non-negative. The first constraint forces the weights to lie on the hyperplane which is tangent to $\mathbf{1}$ (and which passes through point $\mathbf{1}/|\mathbf{1}|$). The second constraint forces the weights to lie on the unit sphere.

3.1 Linear Weight Constraint

We now investigate the behavior of the solution to the first problem. The weights are constrained to lie on the hyperplane $\mathbf{1}^T\boldsymbol{\lambda} = 1$. Use the method of Lagrange multipliers, adjoining the total cost functional $e = \boldsymbol{\lambda}^T\mathbf{y}$ to the constraint to form

$$L = \boldsymbol{\lambda}^T\mathbf{y} + l(\mathbf{1}^T\boldsymbol{\lambda} - 1),$$

where l is the Lagrange multiplier. To extremize e, it must be true that

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = \mathbf{y}^T + l\mathbf{1}^T = \mathbf{0}^T$$

and

$$\frac{\partial L}{\partial l} = \mathbf{1}^T\boldsymbol{\lambda} - 1 = 0.$$

where $\mathbf{0}^T = (0, 0, \ldots)$. This shows that $\mathbf{y} = -l\mathbf{1} = \mathbf{1}e^*/n$, which is constant for all components of $\mathbf{y}$. Therefore, all objective functionals achieve the same value when the total cost functional is minimized.

3.2 Non-linear Weight Constraint

When the weights are constrained to lie on the unit sphere, the method of Lagrange multipliers gives

$$L = \boldsymbol{\lambda}^T\mathbf{y} + l(\boldsymbol{\lambda}^T\boldsymbol{\lambda} - 1).$$

Setting the partials to zero, we have

$$\frac{\partial L}{\partial \boldsymbol{\lambda}} = \mathbf{y}^T + 2l\boldsymbol{\lambda}^T = \mathbf{0}^T$$

and

$$\frac{\partial L}{\partial l} = \boldsymbol{\lambda}^T\boldsymbol{\lambda} - 1 = 0.$$

Therefore,

$$e^* = \boldsymbol{\lambda}^T\mathbf{y} = -2l$$

and

$$\mathbf{y} = e^*\boldsymbol{\lambda}.$$

The objective functionals are parallel to the weights, and every objective functional value, when divided by the corresponding weight, is equal.

4 Conclusions

In this paper we have proposed methods for determining the weights in multiple-objective optimization problems. The Min–Max Principle, which avoids the problems of extremely low or high weights, was introduced. Expressions were derived relating the optimal weights, objective functional values, and total cost.

The optimal weights are not described in closed form; they must be found by a search technique. The Fibonacci search technique [5] is a good choice because the total cost is a unimodal function. For each set of weights considered, the entire problem of determining the state $\mathbf{z}(\boldsymbol{\lambda})$ must be solved. This may be computationally intensive. However, if $\mathbf{z}$ is determined iteratively for each value of $\boldsymbol{\lambda}$, then the state $\mathbf{z}(\boldsymbol{\lambda}_k)$ may be used as the initial guess for the next set of weights $\boldsymbol{\lambda}_{k+1}$. This should greatly reduce the computation required.

Acknowledgements

The support of the Worcester Polytechnic Institute Research Development Council is gratefully acknowledged.

References

[1] M.A. Gennert, "Brightness-Based Stereo Matching," *These proceedings*.

[2] W.E.L. Grimson, "A Computational Theory of Visual Surface Interpolation," *Phil. Trans. Royal Soc. of London B*, Vol. 298, pp. 395–427, 1982.

[3] B.K.P. Horn, "Image Intensity Understanding," *Artificial Intelligence*, Vol. 8, No. 2, pp. 201–231, April 1977.

[4] B.K.P. Horn & B.G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, Vol. 17, Nos. 1–3, pp. 185–203, August 1981.

[5] D.A. Pierre, *Optimization Theory with Applications*, John Wiley & Sons, New York, 1969.

[6] T. Poggio & V. Torre, "Ill-Posed Problems and Regularization Analysis in Early Vision," MIT AI Laboratory Memo 773, April 1984.

[7] T. Poggio, H.L. Voorhees, & A. Yuille, "A Regularized Solution to Edge Detection," MIT AI Laboratory Memo 833, May 1985.

[8] D. Terzopoulos, "Regularization of Inverse Problems Involving Discontinuities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, No. 4, pp. 413–424, July 1986.

[9] D. Terzopoulos, "The Computation of Visible Surface Represenations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 10, No. 4, pp. 417–438, July 1988.