# Bayesian Interpolation

**Daniel Keren**
Dept. of Mathematics and Computer Science
The University of Haifa
Haifa 31905, Israel
dkeren@mathcs1.haifa.ac.il

**Michael Werman** *
Institute of Computer Science
The Hebrew University
Jerusalem 91904, Israel
werman@cs.huji.ac.il

## Abstract

Regularization looks for an interpolating function which is close to the data and also "smooth". This function is obtained by minimizing an error functional which is the weighted sum of a "fidelity term" and a "smoothness term". However, using only one set of weights does not guarantee that this function will be the MAP estimate. One has to consider *all* possible weights in order to find the MAP function. Also, using only one combination of weights makes the algorithm very sensitive to the data.

The solution suggested here is through the Bayesian approach: A probability distribution over all weights is constructed and all weights are considered when reconstructing the function or computing the expectation of a linear functional on the function space.

## 1 Introduction and Previous Work

In computer vision, regularization [Tikhonov and Arsenin, 1977] is used to reconstruct objects from partial data [Terzopoulos, August 1986; Horn and Schunck, 1981; Bertero *et al.*, 1988]. The difficulty stems from the multitude of possible solutions, each satisfying the partial data; which one should be chosen? Regularization overcomes this difficulty by choosing an object which approximates the given data and is also "smooth". Formally, a *cost functional* $M(f)$ is defined for every object $f$ by $M(f) = D(f) + \lambda S(f)$, where $D(f)$ measures the distance of $f$ from the given data, $S(f)$ measures the smoothness of $f$, and $\lambda > 0$ is a parameter. The $f$ chosen is the one minimizing $M()$.

In the one-dimensional case, one minimizes $M(f) = \sum_{i=1}^{n} \frac{[f(x_i) - y_i]^2}{\sigma^2} + \lambda \int_0^1 f_{uu}^2 du$. In the two-dimensional case, the functional to minimize is $M(f) = \sum_{i=1}^{n} \frac{[f(x_i, y_i) - z_i]^2}{\sigma^2} + \lambda \int_0^1 \int_0^1 (f_{uu}^2 + 2f_{uv}^2 + f_{vv}^2) du dv$.

How does one choose $\lambda$ and $\sigma$? There are various methods for doing this. However, all regularization schemes we are familiar with choose one combination of weights and use them alone to interpolate the function.

There is a problem with this approach: It fails to find the MAP estimate for the interpolant $f$, as it uses only one set of weights $\lambda$ and $\sigma$ to construct $f$. But, what happens if the chosen function has a relatively small probability for a wide range of other weights? The MAP estimate should maximize $\int_w Pr(f/w) Pr(w) dw$, where $w$ varies over the set of all possible weights. The main contribution of this work is the computation of $Pr(w)$ (the computation of $Pr(f/w)$ is straightforward).

A popular method for determining the smoothing parameter $\lambda$ is Generalized Cross Validation (GCV) [Craven and Whaba, 1979]. The problem with GCV is that the choice of the value of $\lambda$ is sometimes very sensitive to the data. Since this value is crucial to the shape of the fitted curve or surface, it turns out that sometimes a small change in the data drastically changes the shape of the fitted function. Another problem is that although it can be proved that GCV has some nice asymptotic properties, the choice of the "optimal" values of $\lambda$ and $\sigma$ is heuristic in nature. One goal of this work is to suggest an improvement for the GCV algorithm. Bayesian approaches for choosing the weights are suggested in the pioneering work of Szeliski [Szeliski, 1989] and more recently in [MacKay, 1992]. Here we suggest a different approach, namely, computing the probability distribution by directly integrating over the (infinite-dimensional) space of all possible interpolants. Another novelty is in using all possible weights, not only the "optimal" ones.

## 2 Computing the Joint Probability for $\lambda, \sigma$

Suppose we have a data set $D$ and we want to describe or fit it with a member of some model $M$. The Bayes solution is to find $f$ which satisfies

$$\max_{f \in M} Pr(f/D) = \max_{f \in M} \frac{Pr(D/f) Pr(f/M)}{Pr(D)}$$

$$\propto \max_{f \in M} Pr(D/f) Pr(f/M)$$

*This research was sponsored by ARPA through the U.S. Office of Naval Research under grant N00014-93-1-1202, R&T Project Code 4424341—01.

Regularization can be formalized in this way because $Pr(D/f) = \frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}\exp(-\sum_{i=1}^{n}\frac{[f(x_i)-y_i]^2}{2\sigma^2})$ (assuming uncorrelated Gaussian noise of constant variance) and the prior distribution is $Pr(f) \propto \exp(-\lambda\int f_{uu}^2 du)$ [Szeliski, 1989; Geman and D.Geman, June 1984; Keren and Werman, October 1993; Szeliski, 1987]. Multiplying, we get that the $f$ chosen from $M$ should maximize $\exp(-M(f))$, or minimize $M(f)$.

If few models are possible, the first step is to compute the probability of each model. In our case, the models are indexed by two continuous parameters, $\lambda$ and $\sigma$. Call the model that assumes $\lambda$ as a smoothing parameter and $\sigma$ as the measurement noise $M_{\lambda,\sigma}$. In this model, $Pr(f) \propto \exp(-\lambda\int f_{uu}^2 du)$. Given a data set $D$, we compute $Pr(M_{\lambda,\sigma}/D)$. Using Bayes rule:

$$Pr(M_{\lambda,\sigma}/D) = \frac{Pr(D/M_{\lambda,\sigma})Pr(M_{\lambda,\sigma})}{Pr(D)} \tag{1}$$

$$\propto Pr(D/M_{\lambda,\sigma}) = \frac{\int_{M_{\lambda,\sigma}} Pr(D/f)Pr(f/M_{\lambda,\sigma})\mathcal{D}f}{\int_{M_{\lambda,\sigma}} Pr(f/M_{\lambda,\sigma})\mathcal{D}f}$$

(where we assume a "uniform" prior on the models).

Although the space $M_{\lambda,\sigma}$ is infinite dimensional, it is possible to reduce Equation 1 to a quotient of two integrals defined on a finite dimensional space [Keren and Werman, 1994], which equals

$$\frac{\lambda^{\frac{n}{2}}}{\pi^{\frac{n}{2}}}|A+2\lambda\sigma^2 I|^{-\frac{1}{2}}\exp(-\lambda Y(A+2\lambda\sigma^2 I)^{-1}Y^t) \tag{2}$$

where $Y$ is the data vector and $A$ is a matrix which can be easily computed from the sample points (it does not depend on the data).

## 3 Computing the Expectation of the Value at $x$

If $L$ is a functional on $f$, its expectation given $D$ is $E[L(f)/D] = \int L(f)Pr(f/D)\mathcal{D}f$.

If we want to compute the value of a function at a point $x$, then $L(f)$ is simply the evaluation at $x$, and the expectation can be computed, according to Fubini's theorem, by first evaluating it for each $\{\lambda,\sigma\}$ pair, and then integrating over all such pairs, weighing each one by its probability conditioned by the data $D$: $E[f(x)/D] = \int\int E[f(x)/D, M_{\lambda,\sigma}]Pr(M_{\lambda,\sigma}/D)d\lambda d\sigma$

which can be shown to equal [Keren and Werman, 1994]

$$\frac{\int \frac{1}{\sqrt{v}}|A+vI|^{-\frac{1}{2}}(H_{x_1}(x)..H_{x_n}(x))(A+vI)^{-1}Y^t[Y(A+vI)^{-1}Y^t]^{\frac{4-n}{2}}dv}{\int \frac{1}{\sqrt{v}}|A+vI|^{-\frac{1}{2}}[Y(A+vI)^{-1}Y^t]^{\frac{4-n}{2}}dv} \tag{3}$$

where $H_{x_i}$ are simple polynomial functions in $x$.

We have presented a closed-form expression for evaluating the expectation of the value of $f$ at a point, given partial data about $f$. No optimization and no selection of an "optimal" parameter are necessary.

## 4 Experimental Results

Next, we demonstrate how the proposed approach results in stable fits. The algorithm was extended to 2D data [Keren and Werman, 1994]. In Figure 1, we show the GCV reconstruction for data which consist of a sinusoidal pattern contaminated by Gaussian noise. The data in Figure 2 slightly differs from that of Figure 1 at only one point. The GCV returns radically different results for the two data sets, because this slight change caused GCV to choose a very different value of $\lambda$. Figure 3 shows the reconstruction for the data of Figures 1 and 2 using Equation 3. The fits are nearly identical.

Figure 4 shows the GCV reconstruction to data created by adding Gaussian noise to the function $x(1-x)y(1-y)$. In Figure 5, the GCV fit to a data set differing from that of Figure 4 by one point is given. Again, GCV chooses a very different value of $\lambda$ and returns a very different fit. Applying Equation 3 to the data of Figure 4 results in the surface presented in Figure 6. Applying Equation 3 to the data of Figure 5 results in a fit which is almost similar.

## 5 Conclusion and Future Work

A novel approach for looking at regularization is suggested, which considers all possible values for the noise and smoothing parameter, by assigning a probability to each combination of these two parameters. This probability is used to compute the expectation of a linear functional over the function space. The main advantages of this interpolation paradigm are that it is free of instability problems which haunt interpolation schemes that choose a single value for these parameters; also, the computation of the parameter's probabilities is performed in a mathematically rigorous manner. We have also demonstrated that the optimal parameters can be found by solving a simple one-dimensional minimization problem [Keren and Werman, 1994].

In the future, we plan to demonstrate how the most probable function is computed, and compare it to the pointwise interpolation presented in this work.

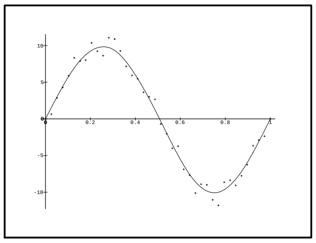

Figure 1

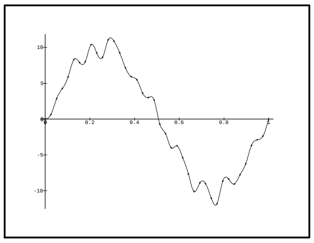

Figure 2

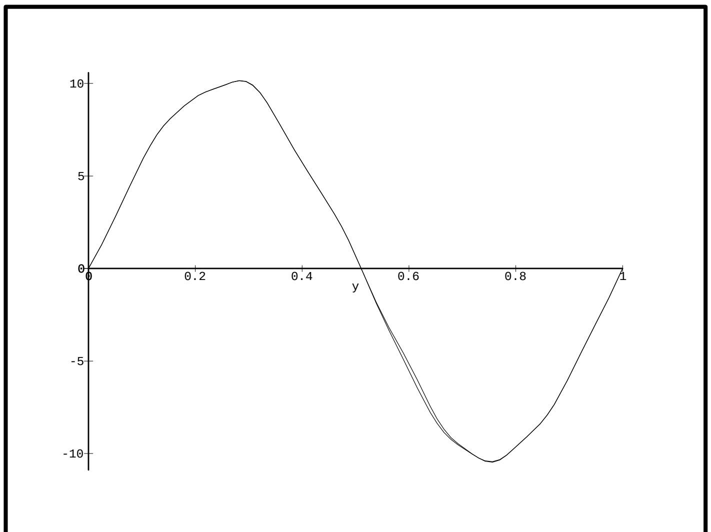


Figure 3

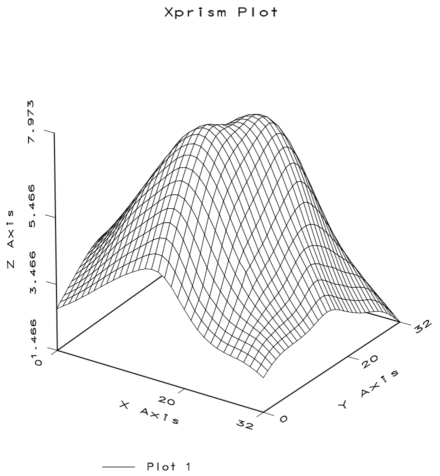


Figure 4

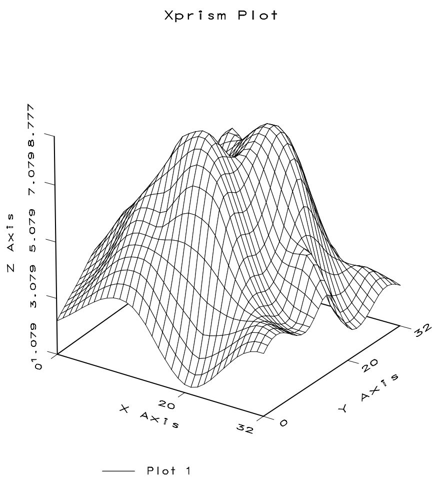


Figure 5

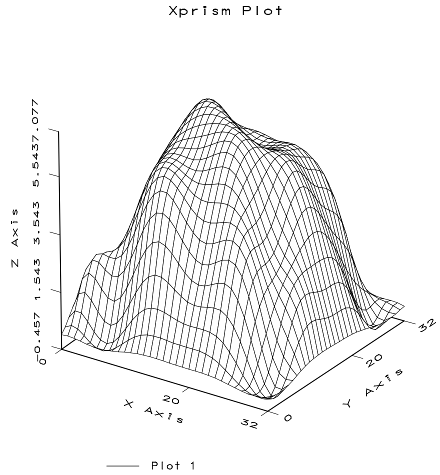


Figure 6

## References

[Bertero *et al.*, 1988] M. Bertero, T.A Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 8:869–889, 1988.

[Craven and Whaba, 1979] P. Craven and G. Whaba. Optimal smoothing of noisy data with spline functions. *Numerische Mathematik*, 31:377–403, 1979.

[Geman and D.Geman, June 1984] S. Geman and D.Geman. Stochastic relaxation, gibbs distribution, and the bayesian restorat ion of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6:721–741, June 1984.

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

[Keren and Werman, 1994] D. Keren and M. Werman. A bayesian framework for regularization. 1994. *submitted to IEEE Transaction on Pattern Analysis and Machine Intelligence.*

[Keren and Werman, October 1993] D. Keren and M. Werman. Probabilistic analysis of regularization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:982–995, October 1993.

[MacKay, 1992] David J.C. MacKay. *Bayesian Methods for Adaptive Models.* PhD thesis, California Institute of Technology, 1992.

[Szeliski, 1987] R. Szeliski. Regularization uses fractal priors. In *National Conference on Artificial Intelligence*, pages 749–754, 1987.

[Szeliski, 1989] R. Szeliski. *Bayesian Modeling of Uncertainty in Low-Level Vision.* Kluwer, 1989.

[Terzopoulos, August 1986] D. Terzopoulos. Regularization of visual problems involving discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8:413–424, August 1986.

[Tikhonov and Arsenin, 1977] A.N Tikhonov and V.Y Arsenin. *Solution of Ill-Posed Problems.* Winston and Sons, 1977.