A Novel Bayesian Method for Fitting Parametric and Non-Parametric Models to Noisy Data

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

Daniel Keren
Department of Computer Science
Haifa University
Haifa 31905, Israel
dkeren@cs.haifa.ac.il

Abstract

We offer a simple paradigm for fitting models, parametric and non-parametric, to noisy data, which resolves some of the problems associated with classic MSE algorithms. This is done by considering each point on the model as a possible source for each data point.

The paradigm also allows to solve problems which are not defined in the classical MSE approach, such as fitting a segment (as opposed to a line). It is shown to be non-biased, and to achieve excellent results for general curves, even in the presence of strong discontinuities.

Results are shown for a number of fitting problems, including lines, circles, segments, and general curves, contaminated by Gaussian and uniform noise.

1 Introduction

It is common practice to fit parametric models (lines, circles, implicit polynomials etc.) to data points, by minimizing the sum of squared distances from the points to the model (the MSE, or Mean Square Error, approach). While the MSE algorithm may seem natural, it in fact implicitly assumes that each data point is the noised version of the point on the model which is closest to it. This assumption is clearly false, and results in strong bias, for instance when fitting circles to data contaminated with strong noise.

The MSE algorithm suffers from another drawback: it cannot differentiate between a "large" model and a "small" one. For instance, when fitting a line segment to bounded data, one would like to know not only the slope and location of the fitted line, but also its start and end points. The MSE criterion cannot differentiate between the "correct" segment and a segment which is too long, because both have the same MSE error with respect to the data.

We offer a simple paradigm for fitting parametric models, which solves these two problems. This is done by considering each point on the parametric model as a possible source for each data point. The model is also extended to non-parametric models, and gives good results even for curves with strong discontinuities.

We show results of the method for lines, segments, circles and general curves. We also show results using Gaussian and uniform noise models.

1.1 Previous Work

There are many papers written on using least square techniques in order to fit parameters to a noisy model, and on using different numerical techniques and linear approximations in order to do the computations. See for example [11, 18] and their references, and also [2], where an ordinary least squares estimate is shown to be consistent for a regression problem.

There are also many papers with different solutions and heuristics to fitting circles, ellipses and other parametric curves using different statistical or optimization techniques; see, for example [20, 14, 16, 3, 19]. There have been a few papers related to Bayesian techniques for specific cases of parametric or non parametric curve and surface fitting, [7, 9, 8, 1, 4, 5]. The idea of associating a "cloud of influence" with each data point is used to compute a better straight line fitting in [10, 12], by using a more general error criterion than the point-line distance. This paper differs from previous work mainly in that precisely the MAP estimate of the model is found, where usually the MAP estimate of the model together with the denoised data points is computed or approximated. Also, we extend the fitting to the general, non-parametric case.

1.2 Suggested Algorithm

Given a set of data points $D = \{p_i\}_{i=1}^{n}$, and a parametric model $M(d_1 ... \ d_m)$ defined by a set of parameters $\{d_j\}_{j=1}^{m}$, a very common fitting algorithm is to choose the instance of the model $M^0(d_1^0 ... \ d_m^0)$ such that the so-called *MSE (Mean Square Error) function*, defined by

$$\sum_{i=1}^{n} dist^2(M(d_1 ... \ d_m), p_i)$$

attains its minimum at $\{d_1^0 ... \ d_m^0\}$. $dist^2(M(d_1 ... \ d_m), p_i)$ is the squared distance between p_i and the model.

The "Bayesian justification" of minimizing the MSE function is as follows: one wishes to maximize the probability of a certain model instance, given the data. Using Bayes' formula, assuming a uniform distribution over the different model instances and independent data,

$$Pr(M|D) = \frac{Pr(D|M)Pr(M)}{Pr(D)} \propto$$
$$Pr(D|M) = \Pi_{i=1}^{n} Pr(p_i|M)$$

Assuming isodirectional Gaussian measurement noise with a variance of σ^2, it is common to approximate $Pr(p_i|M)$ by

$$\frac{const}{\sigma^n} \exp(-\frac{dist^2(p_i^M, p_i)}{2\sigma^2})$$

where p_i^M is the point on the model M closest to p_i. Multiplying over i and ignoring constants, it is easy to see that maximizing this approximate probability is equivalent to minimizing the MSE function.

However, this is only an approximation, which fails at some cases (notably, for instance, for large values of σ). The correct expression is

$$Pr(p_i|M) = \frac{const}{\sigma^n} \int_M \exp(-\frac{dist^2(p, p_i)}{2\sigma^2}) Pr(p|M)dp$$

where p is a point on M, or more generally Bayes rule;

$$Pr(M|p_i) = \frac{Prob(p_i|M)Prob(M)}{Prob(p_i)}$$

where

$$Pr(p_i|M) = \int_{p \in M} Pr(p_i|p)Pr(p|M)$$

where $Pr(p_i|p)$ is the noise model and $Pr(p|M)$ the a-priori distribution of points on M.

2 Fitting Parametric Models

We give some examples of applying the proposed method to fitting lines, segments and circles.

2.1 Line

We proceed to apply the fitting paradigm described in the introduction to the line, which by chance gives the classical MSE result, under the following assumptions.

a) *a priori* all lines are equiprobable

b) *a priori* all points on a line are equiprobable

c) noise is additive isodirectional Gaussian, $N(0, \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix})$, the value of σ is irrelevant.

d) points are independent samples from the line.

Given the data $D = \{(x_i, y_i)\}_{i=1}^{n}$, and denoting a line by L, we have $Pr(L|D) = \frac{Pr(D|L)Pr(L)}{Pr(D)} \propto \Pi_{i=1}^{n} Pr((x_i, y_i)|L)$ and

$$\begin{aligned} Pr((x_i, y_i)|L) &= \int_L Pr((x_i, y_i)|p)Pr(p|L)dp \\ &\propto \int_L \exp(-dist^2((x_i, y_i), p)Pr(p|L)dp \\ &= \int_{-\infty}^{\infty} \exp(-dist^2((x_i, y_i), L) + t^2))dt \\ &\propto \exp(-dist^2((x_i, y_i), L)) \end{aligned}$$

so that, $Prob(L|D) = \prod_{i=1}^{n} Prob(L|(x_i, y_i)) \propto \prod_{i=1}^{n} e^{-dist^2(P_i, L)}$.

Thus $-log(Prob(L|(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n))$ is equal up to an additive constant to:

$$\sum_{i=1}^{n} dist^2((x_i, y_i), L)$$

and the MAP estimate is the line L such that this is a minimum.

The same argument gives that the MAP estimate of a k-flat in R^m is the k-flat whose sum of squared distances from the data is smallest.

Thus, in this case, the paradigm suggested here agrees with the classical MSE paradigm. However, as we shall now show, this is not the case for other models.

2.2 Circle

We proceed to apply the fitting paradigm to the circle. Given the data $D = \{(x_i, y_i)\}_{i=1}^n$, and denoting the parameters of a circle C by (a, b) (for the center) and R (for the radius), we have, assuming noise is additive isodirectional Gaussian $N(0, \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix})$(the value of σ is irrelevant):

$Pr(a, b, R|D) \propto \Pi_{i=1}^n Pr((x_i, y_i)|a, b, R)$ and

$$Pr((x_i, y_i)|a, b, R) = \int_C Pr((x_i, y_i)|p)Pr(p|C)dp \propto$$

$$\int_0^{2\pi} \exp(-(\frac{[x_i - a - R\cos(\theta)]^2 + [y_i - b - R\sin(\theta)]^2)}{2\sigma^2})$$

$$(\frac{1}{2\pi R})(Rd\theta)$$

Where $\frac{1}{2\pi R}$ is $Pr(p|C)$, and $Rd\theta$ stands for the coordinate transformation in the integral (from C to the interval $[0, 2\pi]$). Hence, R cancels out. While there does not seem to be a closed form expression for this integral, it can be estimated quickly by expressing it as an infinite series which swiftly converges. The details are in Appendix I. The optimization over the circle's parameters a, b, R was performed using Powell's method [13]. See examples in Appendix II at the end (Fig. 1).

2.2.1 Comparison to MSE Algorithm

For a circle, the MSE algorithm is well-known to be biased under noise (that is, it gives an estimate to the radius which, on the average, is larger than the true radius). We have empirically verified that the method suggested here is unbiased, by adding random noise and running the optimization process many times. The results always converged to the true radius.

2.3 Line Segment

Another model that can be computed with the paradigm suggested here is the best fit segment. As shown earlier, this cannot be done with the classical MSE methods, as they cannot distinguish between different length segments which have the same MSE error.

Continuing as for the circle, the probability $Pr(p_i|S)$ of a point p_i given a segment S is proportional to $\int_S \exp(-\frac{dist^2(p_i,p)}{2\sigma^2})dp$. This integral can be easily expressed using the error function (*erf*). As before, multiplying over the data points gives the overall probability. We have here, too, used the Powell method to optimize over the segment's four parameters; see a typical result in Appendix II (Fig. 2).

3 Line with uniform noise

Uniform noise with shape S (S can be a circle, square, etc.) is defined as follows:

$$Pr(p_i|p) = \begin{cases} \frac{1}{Area(S)} & \text{if } p_i \in p + S \\ 0 & \text{otherwise} \end{cases}$$

so that the probability $Pr(p_i|p)$ is positive iff $p \in p_i + (-S)$.

For example, let us fit a line to n noisy points, where the noise is uniform in unit size circles around the data; we have to omit the details for lack of space, however it is easy to see that the integral defining the probability for a data point p_i is proportional to the length of the line's intersection with the unit circle around p_i. Hence, finding the optimal line is equivalent to finding the line that pierces all the n circles centered at the data points, such that the product of its lengths of intersections with the circles is maximal.

It is interesting to note that, in the standard fitting paradigm (under uniform noise), the probability for *every* line which intersects the circles around the data points, is identical. The method described here therefore yields a "sharper" result (albeit not necessarily unique). See Appendix II for an example of fitting a line to points under uniform noise (Fig. 3).

4 Extending the Paradigm to Non-Parametric Models

The algorithms described and implemented in Sections 2,3 to parametric models, can be extended to general, non-parametric curves. Following the previous derivations, it is easy to see that the probability of a curve C, given sparse data $\{p_i\}_{i=1}^n$, is proportional to

$$\Pi_{i=1}^n \int_C Pr(p_i|p)Pr(p|C)dp$$

if the curve is represented by discrete points $\{c_j\}_{j=1}^m$, this probability may be approximated by the following expression:

$$\frac{1}{L^n(C)} \Pi_{i=1}^n (\sum_{j=1}^{m-1} \exp(-\frac{||c_j - p_i||^2}{2\sigma^2})||c_{j+1} - c_j||) \quad (1)$$

where $L(C)$ is the curve's length (as in the parametric models, we define $Pr(p|C)$ as $\frac{1}{L(C)}$). The factor $||c_{j+1} - c_j||$ stands for the length element of the curve.

In this work, we combined this term with a standard "smoothness term", such as $\int_C (C_{xx}^2 + 2C_{xy}^2 +$

$C_{yy}^2)dC$, to arrive at an optimal solution. Thus, the paradigm may be viewed as standard regularization, with the "data term" replaced by Eq. 1.

It is worthwhile to look at Eq. 1 and see how it leads to a curve which "sticks to the data". Parts of the curve which are far away from the data contribute little to the integrand, due to the presence of the

$$\exp(-\frac{\|c_j - p_i\|^2}{2\sigma^2})$$

term, which becomes smaller as we move away from the data. However, they result in a larger value of $L(C)$, which leads to a smaller value for the entire expression. This is amply demonstrated for data which consists of a noised version of a step function (Fig. 4 in Appendix II); note that the fitted curve does not suffer from the well-known "Gibbs phenomena", which yields spurious curve parts away from the data.

We have implemented the optimization of Eq. 1 via the Powell method [13]. The problem is non-trivial, as it involves many parameters (the x and y coordinates of the m points comprising the curve). However, convergence did not take more than a few minutes on a workstation.

5 Implicit Polynomials

In recent years, there is growing interest in fitting data with implicit models, mainly polynomials. This representation is very efficient both for recognition purposes [15], and for quickly determining whether a point is in the object or not [16, 5]. A problem which has still to be solved is that often spurious parts appear in the polynomial's zero-set, such as loops, components which are far from the data, folds, etc. Some heuristics have been suggested to overcome this problem [5, 17], and recently a more robust method was suggested which solves the problem in the case of a starshaped object [6]. However, there isn't yet any solution which is guaranteed to work in the general case.

The Bayesian fitting method described here has good potential to solve this problem, because it penalizes extraneous portions in the zero-set. We have applied it with success to some non-starshaped objects (lack of space prevents us from displaying the results here). There is, however, a computational problem: while, for the explicit models discussed in this work so far, there is a direct, simple relation between the parameters of the model and its geometric realization, that is not the case for the implicit model. Hence, in every step of the optimization process, the zero-set has to be computed, substantially increasing the running time.

6 Conclusion and Further Research

We have presented a fully Bayesian paradigm for fitting parametric and non-parametric models, which is natural, mathematically rigorous, and superior to the classical MSE method, albeit in a higher computational cost, mostly required in optimizing non-trivial cost functions for the fitted model. In the future, we hope to try and alleviate this problem, as well as to extend the paradigm to other models, such as splines.

References

[1] M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia. Sub-pixel Bayesian-estimation of albedo and height. In *International Journal of Computer Vision*, pages 289–300.

[2] C.L. Cheng and H. Schneeweiss. Polynomial regression with errors in the variables. *Journal of the Royal Statistical Society, B*, 60:189–199, 1998.

[3] A. W. Fitzbiggon, M. Pilu, and R. B. Fisher. Direct least squares fitting of ellipses. *Proceedings of International Conference on Pattern Recognition*, 1:253–257, 1996.

[4] S. Gull. Baysian data analysis: Straight line fitting. In *Maximum Entropy and Bayesian Methods*, pages 511–518.

[5] D. Keren, D. Cooper, and J. Subrahmonia. Describing complicated objects by implicit polynomials. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16:38–53, January 1994.

[6] D. Keren and C. Gotsman. Fitting curves and surfaces with constrained implicit polynomials. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(1):31–41, 1999.

[7] D. Keren and M. Werman. Variations on regularization. In *10'th International Conference on Pattern Recognition*, pages 93–98, Atlantic City, 1990.

[8] D. Keren and M. Werman. A full Bayesian approach to curve and surface reconstruction daniel keren and michael werman. *Journal of Mathematical Imaging and Vision*, 1999.

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

[10] N. Kiryati and A.M. Bruckstein. What's in a set of points? *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14(4):496–500, 1992.

[11] Y. Leedan and P. Meer. Estimation with bilinear contraints in computer vision. In *International Conference on Computer Vision*, Bombay, India, 1998.

[12] Y. Pnueli, N. Kiryati, and A.M. Bruckstein. Hough techniques for fast optimization of linear constant velocity motion in moving influence fields. *Pattern Recognition Letters*, 15:329–336, 1994.

[13] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling. *Numerical Recipes*. Cambridge University Press, 1986.

[14] P.D Sampson. Fitting conic sections to very scattered data: an iterative improvement of the bookstein algorithm. *Computer Vision, Graphics, and Image Processing*, 18:97–108, 1982.

[15] J. Subrahmonia, D. Cooper, and D. Keren. Practical reliable Bayesian recognition of 2D and 3D objects using implicit polynomials and algebraic invariants. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18:505–519, 1996.

[16] G. Taubin. Estimation of planar curves, surfaces and nonplanar space curves defined by implicit equations, with applications to edge and range image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13:1115—1138, November 1991.

[17] G. Taubin, F. Cukierman, S. Sullivan, J. Ponce, and D.J. Kriegman. Parameterized families of polynomials for bounded algebraic curve and surface fitting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16:287–303, 1994.

[18] G. Vosselman and R.M. Haralick. Performance analysis of line and circle fitting in digital images. In *Workshop on Performance Characteristics of Vision Algorithms*, Cambridge, 1996.

[19] M. Werman and Z. Geyzel. Fitting a second degree curve in the presence of error. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17:207–211, 1995.

[20] Z. Zhang. Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing*, 15:59–76, 1997.

7 Appendix I

We will now show how to evaluate the integral in Section 2.2. Without loss of generality, it can be expressed as

$$\int_{x^2+y^2=r^2} \exp(-[(x-a)^2+(y-b)^2])ds$$

substituting $x = r\cos(\theta)$, $y = r\sin(\theta)$ gives

$$r\exp(-(r^2+a^2+b^2)\int_0^{2\pi}\exp(2ar\cos(\theta)+2br\sin(\theta))d\theta$$

using the complex substitution

$$\zeta = \exp(i\theta), d\theta = \frac{d\zeta}{i\zeta}, \cos(\theta) = \frac{\zeta+\frac{1}{\zeta}}{2}, \sin(\theta) = \frac{\zeta-\frac{1}{\zeta}}{2i}$$

transforms this integral to

$$\begin{aligned}&\int_{\|\zeta\|=1}\exp(2ar\frac{\zeta+\frac{1}{\zeta}}{2}+2br\frac{\zeta-\frac{1}{\zeta}}{2i})\frac{d\zeta}{i\zeta} = \\ &\int_{\|\zeta\|=1}\exp(A\zeta+\frac{B}{\zeta})\frac{d\zeta}{i\zeta} \\ &(A = ar - ibr, B = ar + ibr)\end{aligned}$$

using the residue theorem, and since the free coefficient in the expansion of $\exp(A\zeta+\frac{B}{\zeta})$ is $\sum_{n\geq 0}\frac{(AB)^n}{(n!)^2}$, the final expression for the integral is

$$\frac{\sum_{n\geq 0}[(a^2+b^2)r^2]^n}{r(n!)^2\exp(r^2+a^2+b^2)}$$

which converges quickly due to the fast growing $(n!)^2$ factor in the denominator.

8 Appendix II: Some Results

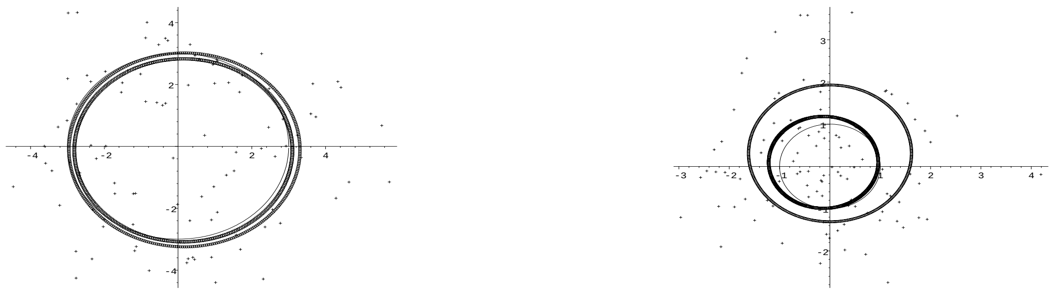

Figure 1: Examples of MSE fit and suggested fit for circle. In both cases, the true circle is a thin black line, the noised data points are designated by crosses (Gaussian noise with unit variance), the MSE fit is in small circles, and the suggested fit is in small squares. These examples reflect the typical result that, when the noise is large with regard to the radius, the MSE fit is very biased, while the suggested fit is not (see Section 2.2). The improvement of the suggested method is much more apparent for the right circle (radius 1) than for the left circle (radius 3).

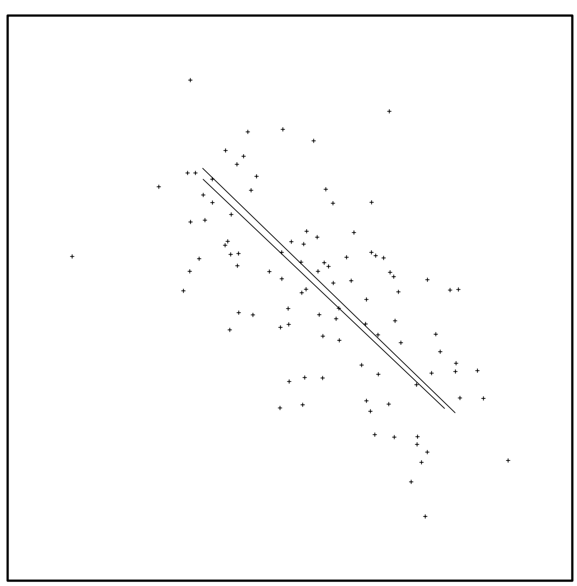

Figure 2: Straight line segment (upper segment), noised line points (little crosses), and line segment fitted using the suggested method (lower segment). It is not a portion of the best MSE fit line.

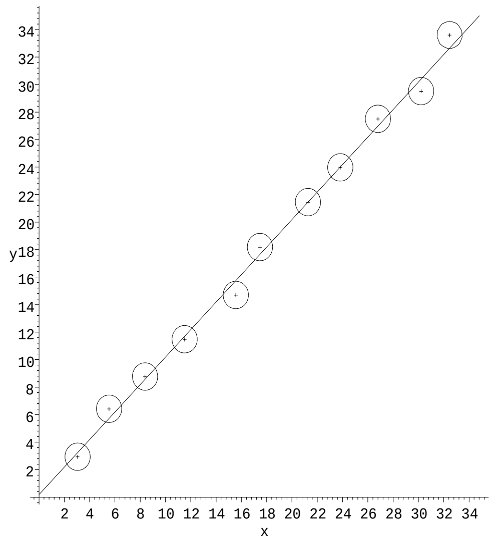

Figure 3: Fitting a line to points with uniform noise in the shape of a circle. The resulting line has a nice intuitive interpretation: it is the line which maximizes the product of lengths of its intersections with the circles around the data points.

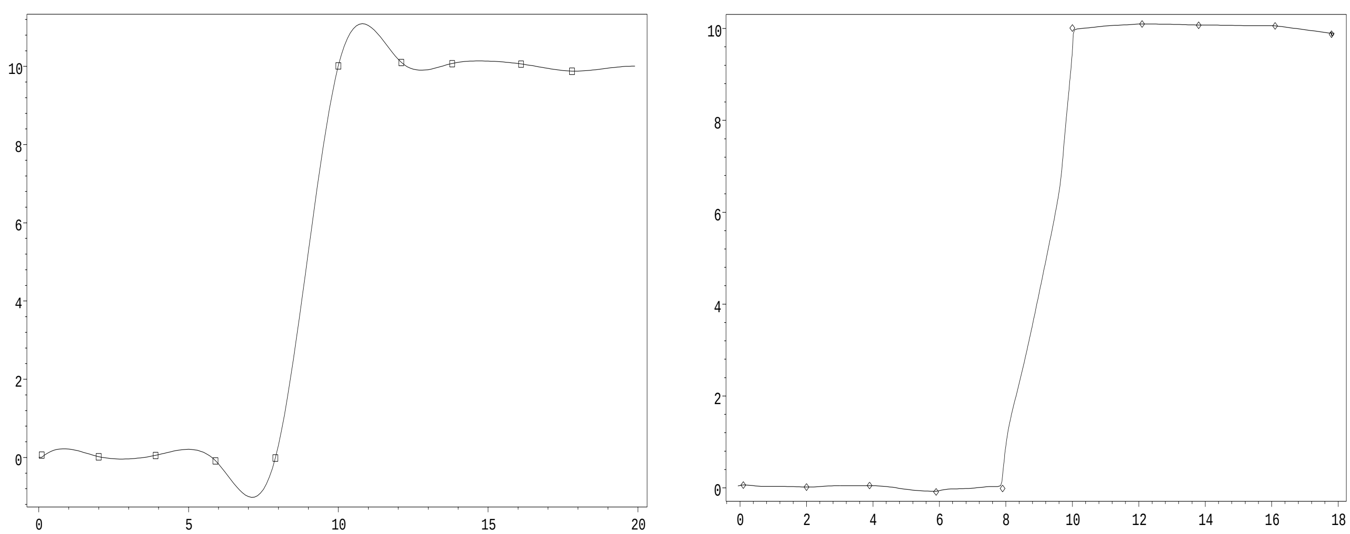

Figure 4: Regularized fit to a sampled step function, demonstrating the well-known Gibbs phenomena (left), and a fit to same data obtained using the novel Bayesian method suggested in this paper (right).