Short Papers

A Bayesian Method for Fitting Parametric and Nonparametric Models to Noisy Data

Michael Werman and Daniel Keren

Abstract—We present a simple paradigm for fitting models, parametric and nonparametric, to noisy data, which resolves some of the problems associated with classical MSE algorithms. This is done by considering each point on the model as a possible source for each data point. The paradigm can be used to solve problems which are ill-posed in the classical MSE approach, such as fitting a segment (as opposed to a line). It is shown to be nonbiased 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, elliptic arcs, segments, rectangles, and general curves, contaminated by Gaussian and uniform noise.

Index Terms—Bayesian fitting, parametric models, nonparametric models.

1 Introduction

It is common practice to fit 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, in fact it, 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 leads to bias, for instance, when fitting circles to data contaminated by 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 image data, one would often like to know not only the slope and location of the fitted segment, but also its end points. The MSE criterion does not 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 problems. This is done by considering each point on the parametric model as a possible source for each data point. The paradigm is also extended to nonparametric models and gives good results even for data with strong discontinuities.

We show results of the method for lines, segments, circles, elliptic arcs, rectangles, and general curves. Both Gaussian and uniform noise models are considered.

1.1 Previous Work

There are many papers which describe least-square techniques to fit parameters to noisy data and on using different numerical techniques and linear approximations needed for 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], [4], [19]. There have been a few papers related to Bayesian techniques for specific cases of parametric or nonparametric curve and surface fitting, [7], [9], [8], [1], [5], [6]. 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.

In [3], a very interesting approximate solution to the traveling salesman problem is offered, in which a (nonparametric) path is pulled towards the cities, controlled by a term which tries to keep it as short as possible. This work differs from ours in the Bayesian formulation and, in that, no treatment of parametric models is offered.

In general, 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, nonparametric case.

1.2 Suggested Algorithm

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

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

attains its minimum at $\{d_1^0 \dots d_m^0\}$. $dist^2(M(d_1 \dots d_m), p_i)$ 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 and assuming a uniform distribution over the different model instances and independent data,

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

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

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

where p_i^M is the point on the model M closest to p_i. Multiplying over i, taking logarithms 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 for some cases (notably, for instance, for large values of σ). The correct expression is

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

- M. Werman is with the Institute of Computer Science, Hebrew University at Jerusalem, Jerusalem 91904 Israel. E-mail: werman@cs.huji.ac.il.
- D. Keren is with the Department of Computer Science, University of Haifa, Haifa 31905, Israel. E-mail: dkeren@cs.haifa.ac.il.

Manuscript received 21 Sept. 1999; revised 10 May 2000; accepted 15 May 2000.
Recommended for acceptance by P. Meer.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 110634.

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.

We will not show examples of using different priors on parametric models other than the uniform prior, only because we want to concentrate on the effect of the integration over the choice of model. Also, experience has taught us that, if many data points are available, the effect of the model prior on the optimal fit is very small.

2 Fitting Parametric Models

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

2.1 Line

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

1. A priori all lines are equiprobable.
2. A priori all points on a line are equiprobable.
3. Noise is additive isodirectional Gaussian,

   $$N\left(0, \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}\right),$$

   the value of σ is irrelevant.
4. 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

$$Pr(L|D) = \frac{Pr(D|L)Pr(L)}{Pr(D)} \propto \prod_{i=1}^{n} Pr((x_i, y_i)|L)$$

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,

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

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 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 the 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

$$N\left(0, \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}\right),$$

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

and

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

While there is no closed form expression for this integral, it can be estimated quickly by expressing it as an infinite series which swiftly converges (the proof is left out due to lack of space). In general, the integral of a Gaussian over a circle can be expressed as

$$\int_{x^2+y^2=r^2} \exp\left(-\left[(x-a)^2 + (y-b)^2\right]\right)ds = \sum_{n \geq 0} \frac{[(a^2+b^2)r^2]^n}{r(n!)^2 \exp(r^2+a^2+b^2)}.$$

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. See Fig. 1 for some typical examples of the fitting method offered here compared to the MSE fit.

2.2.2 *Comparison to Outlier Sensitivity vs. the Method of Moments*

A nonbiased estimator to the center of a circle under uniform noise is given by

$$C = \frac{1}{n}\sum_{i=1}^{n} p_i,$$

where p_i are the measured (noisy) points. A nonbiased estimator to the radius is then given by

$$\frac{1}{n}\sum_{i=1}^{n} ||p_i - C||^2 - 2\sigma^2.$$

We have compared the behavior of this estimator and the one proposed in this work under the presence of outliers. First, 100 points were selected on a circle of radius 10 and Gaussian noise with unit variance added to them. Then, 20 outliers were chosen uniformly in the square $[-A, A]$, where A ranged from 15 to 30 and the best fit circle was computed. This was repeated 1,000 times for

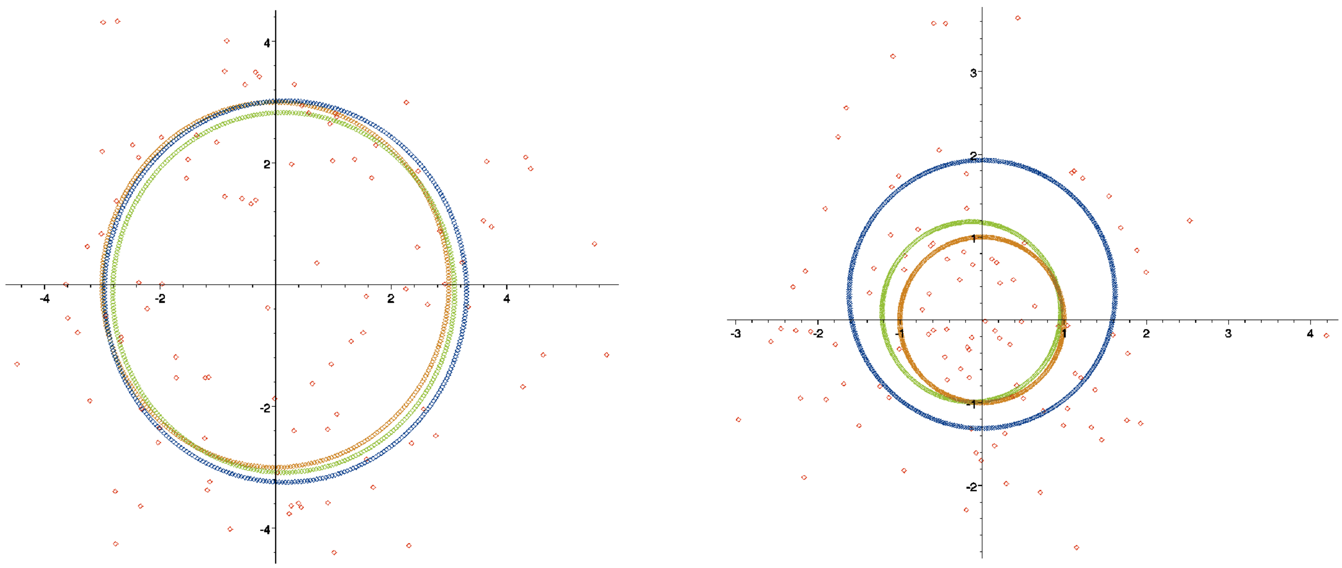


Fig. 1. Examples of MSE and suggested fit for circle. In both cases, the true circle is in gold, the noised data points are red (Gaussian noise with unit variance), the MSE fit is blue, and the suggested fit is green. These examples reflect the typical result that, when the noise is large with regard to the radius, the MSE fit is very biased, while the suggested fit is not. The improvement of the suggested method is much more apparent for the right circle (radius 1) than for the left circle (radius 3). The average optimal radius for the MSE method, computed over 100,000 noised circles (with unit radius and unit noise variance), was 1.49, indicating a very strong bias. The average radius for the method suggested in this work was 1.01.

each A. In Fig. 2, the average radius is plotted against A. Thus, for these type of outliers, the method suggested here is more robust than the method of moments. However, the method is not tailored to handle outliers and better results can probably be achieved by combining it with various methods from the realm of robust statistics.

2.3 Elliptic Arc

An elliptic arc is determined by seven parameters (five for the entire ellipse and two angles which define the arc). The algorithm for finding the MAP arc proceeds in the same way as for the circle the only difference being that the integration over the arc is more complicated. In Fig. 3, we present an example of an arc, the noisy data points sampled from it, and the reconstructed arc.

2.4 Line Segment

Another model that can be computed with the paradigm suggested here is the best fit segment. This cannot be done with the classical MSE methods, as they cannot distinguish between different length segments which have the same MSE error (see Fig. 4).

The Bayesian paradigm offered here naturally solves the problem of fits which are "too big." It is more rigorous than heuristics which penalize the area or volume of the fit (as in, for

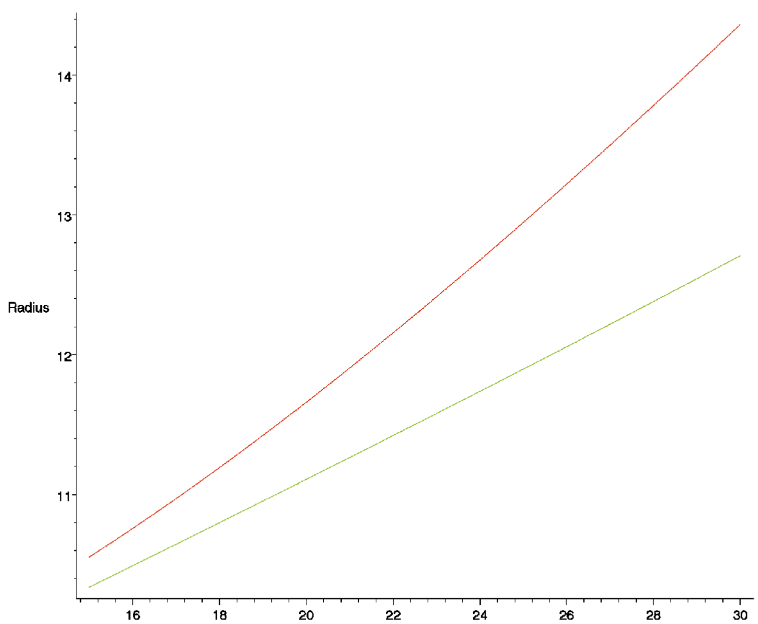


Fig. 2. Average radius for the method in this work (green) and the method of moments (red) as a function of outlier size (outliers were randomly selected in the square $[-A, A]$). The correct radius is 10.

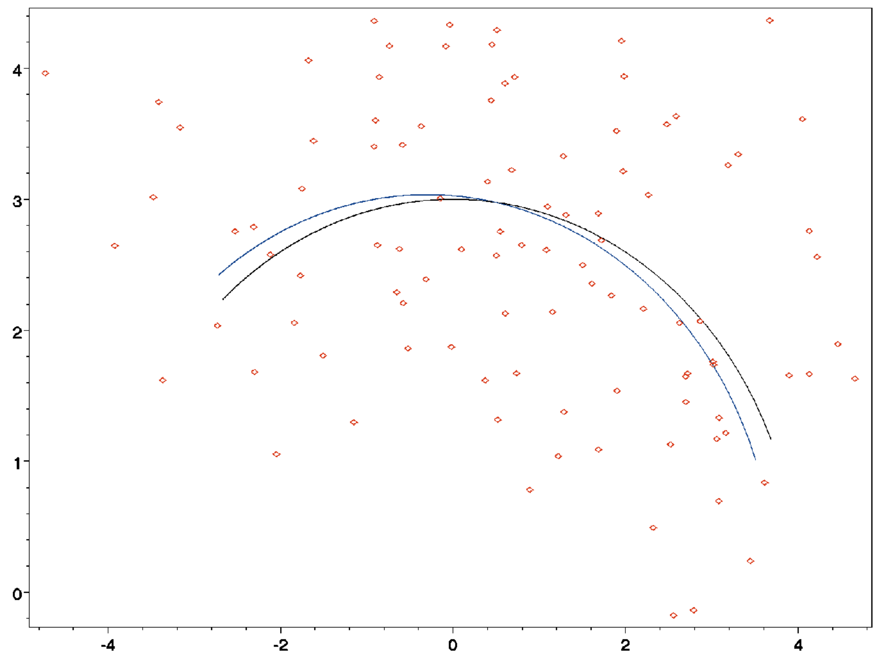


Fig. 3. Elliptic arc (dark), noised data points (red), and elliptic arc fitted using the suggested method (blue).

example, [6], [15]). Continuing as for the circle, the probability $Pr(p_i|S)$ of a point p_i given a segment S is proportional to

$$\int_S \exp\left(-\frac{dist^2(p_i,p)}{2\sigma^2}\right)dp. \quad (1)$$

This integral can be easily expressed using the error function (*erf*). As before, multiplying over the data points gives the overall probability. See Fig. 5 for an example of segment fitting.

2.5 Axis Aligned Rectangle

Axis aligned rectangles are useful descriptions often used in pattern recognition as they model Cartesian products of intervals (here, as opposed to the case of the circle, the model is the rectangle's interior).

Tenenbaum [17], considered the problem of learning *concepts* from small numbers of positive examples, which turned out to be an axis aligned rectangle fitting problem, which was solved using a similar Bayseian model but without noise.

We proceed to apply the fitting paradigm to the rectangle. Given the data $D = \{(x_i, y_i)\}_{i=1}^n$ and denoting the parameters of a rectangle R by (a, b, c, d) (for the lower left hand and upper right hand corners) we have, assuming noise is additive isodirectional Gaussian

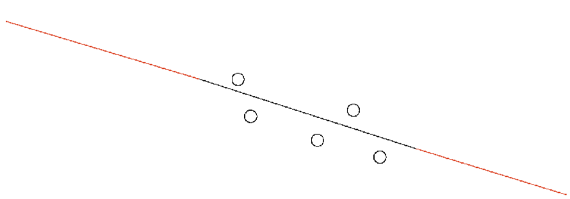


Fig. 4. The MSE paradigm for fitting a segment is inherently ill-posed because there is no increase in the error function for a segment that is too long. There is no penalty for the red portions of the segment above since they do not change the MSE function. The method suggested in this work does penalize such portions; since they are far from the data, the value of the integrand in (1) on them is small, but they increase the segment's length L, thus decreasing the integration element $dp = \frac{1}{L}$, the overall effect being a smaller value of the integral.

$$N\left(0, \begin{pmatrix} \sigma & 0 \\ 0 & \sigma \end{pmatrix}\right):$$

$$Pr(a,b,c,d|D) \propto \prod_{i=1}^n Pr((x_i,y_i)|a,b,c,d)$$

$$Pr((x_i,y_i)|a,b,c,d) = \int\int_R Pr((x_i,y_i)|p)Pr(p|R)dp \propto$$

$$\frac{1}{2(c-a)(d-b)}\left(erf\left(\frac{a-x_i}{\sqrt{2}\sigma}\right) - erf\left(\frac{c-x_i}{\sqrt{2}\sigma}\right)\right)$$
$$\left(erf\left(\frac{b-y_i}{\sqrt{2}\sigma}\right) - erf\left(\frac{d-y_i}{\sqrt{2}\sigma}\right)\right).$$

2.6 Two Axis Aligned Rectangles

We proceed to apply the fitting paradigm to two axis aligned rectangles. Given the data $D = \{(x_i, y_i)\}_{i=1}^n$ and denoting the two rectangles R_1, R_2 by $(a_1, b_1, c_1, d_1, a_2, b_2, c_2, d_2)$:

$$Pr(a_1,b_1,c_1,d_1,a_2,b_2,c_2,d_2|D) \propto \prod_{i=1}^n Pr((x_i,y_i)|a_1,b_1,c_1,d_1,a_2,b_2,c_2,d_2)$$

and $Pr((x_i,y_i)|a_1,b_1,c_1,d_1,a_2,b_2,c_2,d_2)$ is equal to

$$\frac{1}{2}(Pr((x_i,y_i)|a_1,b_1,c_1,d_1) + Pr((x_i,y_i)|a_2,b_2,c_2,d_2)).$$

This is just the sum of the probabilities of the two rectangles, each computed as in the one rectangle case. See Fig. 6 for an example.

2.7 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 $p \in p_i + (-S)$.

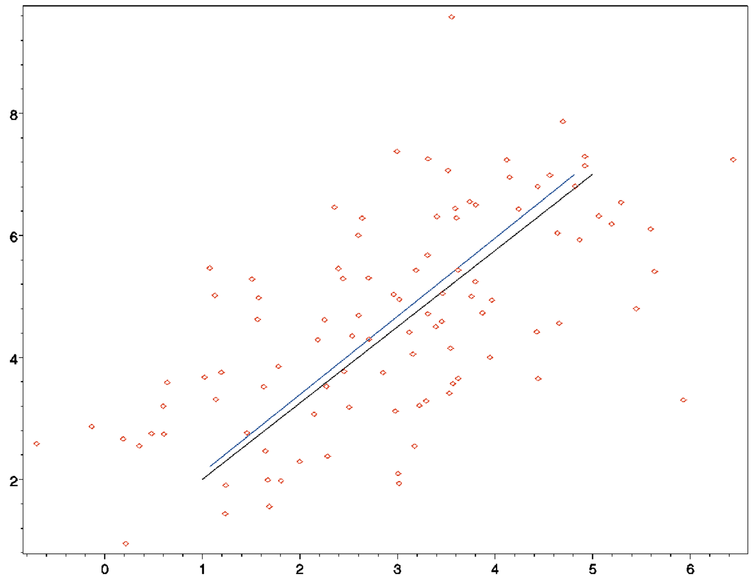
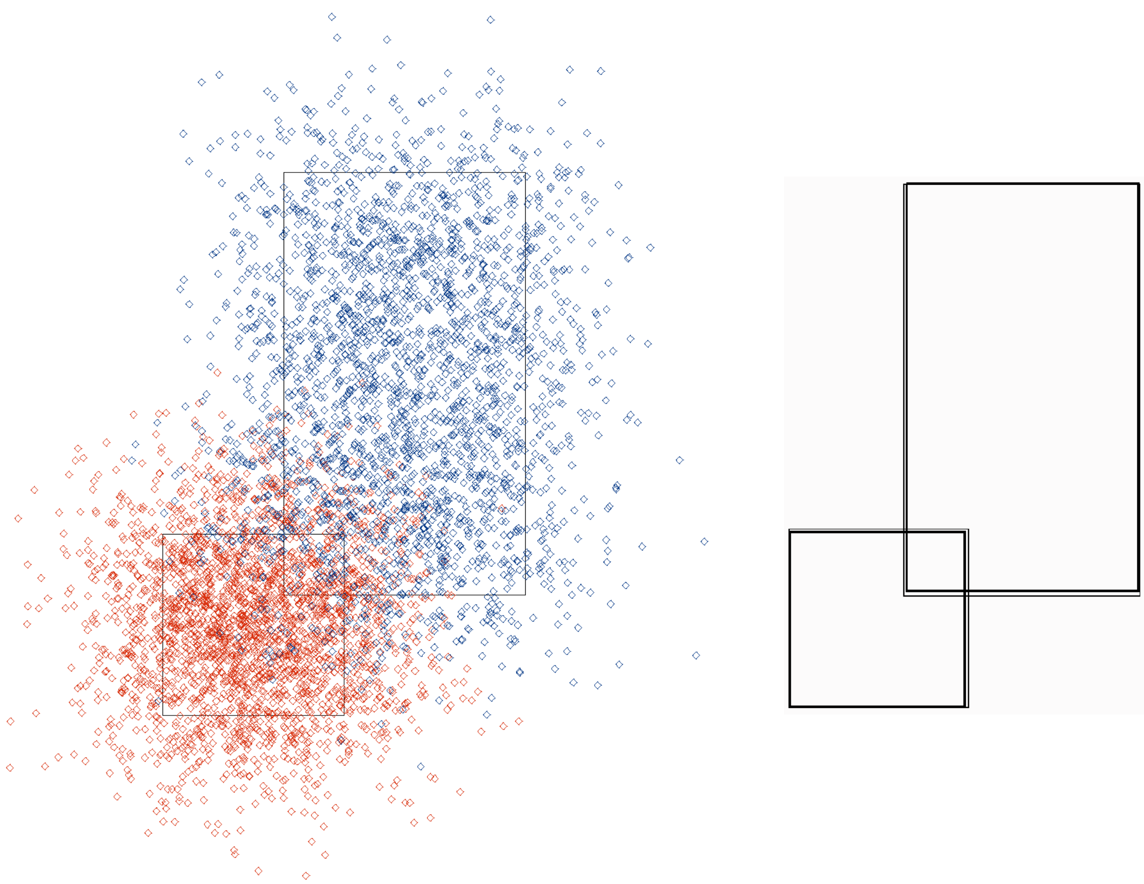


Fig. 5. Straight line segment (dark), noised segment points (red), and line segment fitted using the suggested method (blue). It is not a portion of the best MSE fit line.

For example, let us fit a line to n noisy points where the noise is uniform in unit size circles around the data; 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. See Fig. 7 for an example.



(a) (b)

Fig. 6. Example of fitting two axis aligned rectangles to a set of points. (a) Noisy data points with fitted rectangles. Red and blue points are noised interior points of the lower and upper rectangle, respectively. (b) Original rectangles superimposed with the fit.

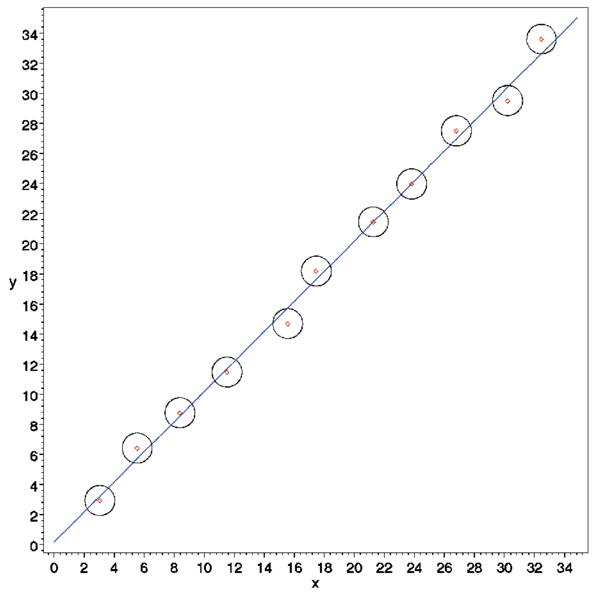


Fig. 7. Fitting a line to points with uniform noise in the shape of a circle. The resulting line has a nice intuitive interpretation. It is the line which maximizes the product of lengths of its intersections with the circles around the data points.

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

3 Extending the Paradigm to Nonparametric Models: Curve Data

The algorithms described and implemented in Section 2 to parametric models, can be extended to general, nonparametric 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

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

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

$$\frac{1}{L^n(C)} \prod_{i=1}^{n} \left(\sum_{j=1}^{m-1} \exp\left(-\frac{\|c_j - p_i\|^2}{2\sigma^2} \right) \|c_{j+1} - c_j\| \right), \tag{2}$$

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

$$\int_{C} (C_{xx}^2 + 2C_{xy}^2 + C_{yy}^2)dC,$$

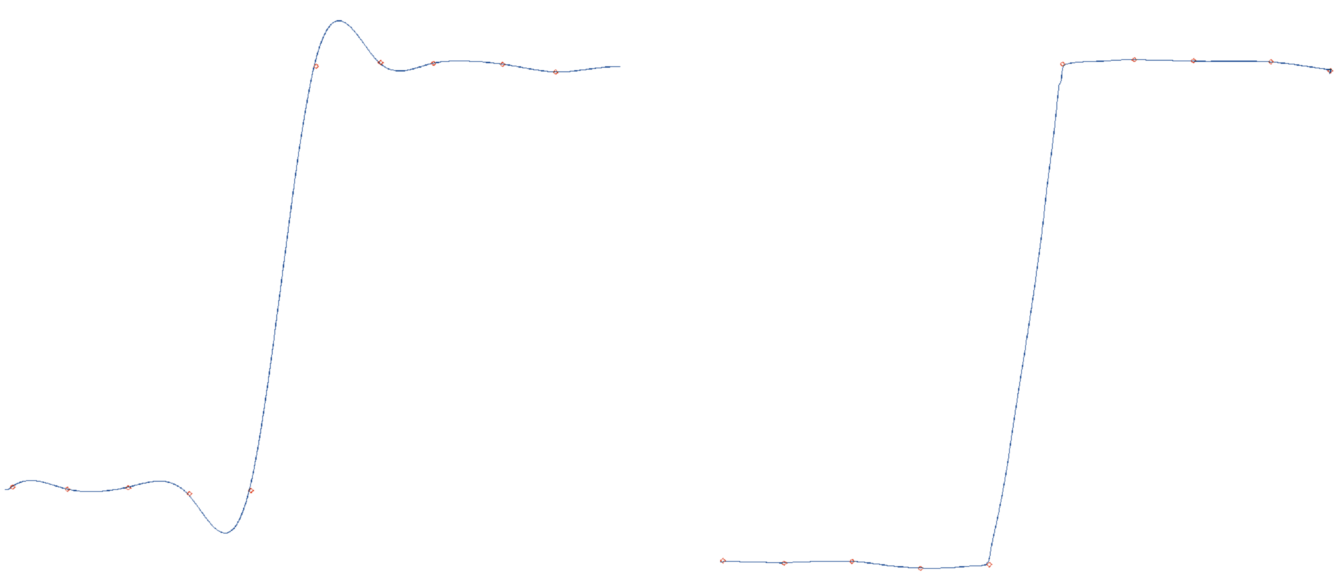
to arrive at an optimal solution. Thus, the paradigm may be viewed as standard regularization, with the "data term" replaced by (2).

It is worthwhile to look at (2) and see how it leads to a curve which "sticks to the data." Portions of the curve which are far away from the data contribute little to the integrand, due to the presence of the

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

term, which becomes smaller as we move away from the data. However, these portions 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, note that the fitted curve does not suffer from the well-known "Gibbs phenomena," which yields spurious curve parts away from the data; see Fig. 8.

In order to force the resulting curve to be continuous, we parameterized it by the initial point, the distance between every two consecutive points (which is different for different curves, but fixed for every specific curve), and the angles between consecutive



(a) (b)

Fig. 8. Regularized fit to a sampled step function, demonstrating the well-known Gibbs phenomena (a) and (b) a fit to same data obtained using the method suggested in this paper.

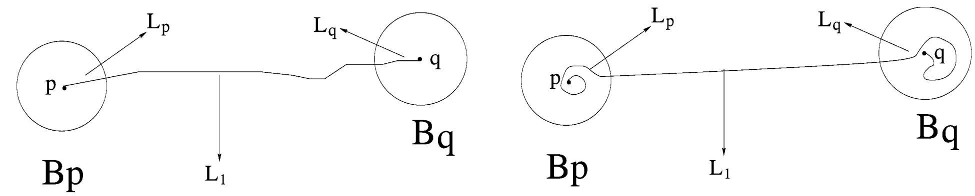


Fig. 9. Without regularization, the right curve has a higher probablility than the left one.

points. Keeping the distance between consecutive points fixed during every iteration helps to avoid "tears" in the restored curve.

3.1 Can We Forego Regularization?

A question that suggests itself is whether, in the case of fitting a general, nonparametric curve, we can drop the regularization term and use the purely Bayesian approach outlined in this paper. As will now be demonstrated, the answer is no—at least, if we expect to obtain "reasonable" results.

Let us look at a trivial case: only two data points (p and q) and measurement noise with an arbitrarily small standard deviation σ. Let us denote by B_p, B_q small circles around p, q, respectively. For a curve C connecting p and q, let us denote by L_p, L_q the lengths of $C \cap B_p$ and $C \cap B_q$, respectively, and, by L_1, the length of the portion of C which does not lie in $B_p \cup B_q$. The total length of the curve will be denoted by $L(C) = L_p + L_1 + L_q$.

Now, recall that the probability of the curve C, assuming a measurement noise σ, is proportional to

$$\frac{1}{L^2(C)} \int\limits_C \exp\left(-\frac{\|c-p\|^2}{2\sigma^2}\right) dc \int\limits_C \exp\left(-\frac{\|c-q\|^2}{2\sigma^2}\right) dc. \quad (3)$$

By making σ small enough relative to the radii of B_p, B_q, the integrand in each of the integrals can be made arbitrarily small on the portion of the curve with length L_1, hence, in the limit, the expression in (3) is bounded by

$$\frac{(L_p + L_q)^2}{(L_p + L_1 + L_q)^2}.$$

Therefore, it is clear that, in order for the probability to be large, L_1 should be as small as possible relative to L_p and L_q. Thus, a curve, such as the one on the left in Fig. 9, will have a smaller probability than the one on the right in Fig. 9.

It is easy to see that, in the limit, the optimal curve "wanders about" an unbounded amount of time as close as possible to p and q, while taking the shortest route between them. This is hardly surprising; as in the aforementioned case of a segment, the model tries to "stick" to the data. Thus, regularization has to be applied, else an "unreasonable" curve will result.

4 A Note on Convergence

The fitting paradigm offered here usually leads to the optimization of nonconvex functions. We have found that when an arbitrary starting point is used, the optimization sometimes converges to a local minimum. This problem was solved by initializing the optimization at the MSE fit.

We have used either the Powell or Nedler-Mead methods [13]; convergence was very fast for the parametric models. For the nonparametric fitting of general curves (Section 3), it took up to a minute on an XP1000 Digital workstation.

5 Conclusions and Further Research

We presented a Bayesian paradigm for fitting parametric and nonparametric models, which is natural, mathematically rigorous, and superior to the classical MSE method, albeit with a higher computational cost, mostly required in optimizing nontrivial 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.

Acknowledgments

The authors would like to thank Dr. Alexander Goldenshluger of the Department of Statistics in the University of Haifa, for some interesting discussions. This research was supported by the Israeli Ministry of Science.

References

[1] M. Berthod, H. Shekarforoush, M. Werman, and J. Zerubia, "Sub-Pixel Bayesian-Estimation of Albedo and Height," *Int'l J. Computer Vision,* vol. 19, no. 3, pp. 289-300, 1996.

[2] C.L. Cheng and H. Schneeweiss, "Polynomial Regression with Errors in the Variables," *J. Royal Statistical Soc. B,* vol. 60, pp. 189-199, 1998.

[3] R. Durbin, R. Szeliski, and A. Yuille, "An Analysis of the Elastic Net Approach to the Travelling Salesman Problem," *Neural Computation,* vol. 1, no. 3, pp. 348-358, 1989.

[4] A.W. Fitzbiggon, M. Pilu, and R.B. Fisher, "Direct Least-Squares Fitting of Ellipses," *Proc. Int'l Conf. Pattern Recognition,* vol. 1, pp. 253-257, 1996.

[5] S. Gull, "Baysian Data Analysis: Straight Line Fitting," *Maximum Entropy and Bayesian Methods,* pp. 511-518, 1989.

[6] D. Keren, D. Cooper, and J. Subrahmonia, "Describing Complicated Objects by Implicit Polynomials," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, pp. 38-53, Jan. 1994.

[7] D. Keren and M. Werman, "Variations on Regularization," *Proc. 10th Int'l Conf. Pattern Recognition,* pp. 93-98, 1990.

[8] D. Keren and M. Werman, "A Full Bayesian Approach to Curve and Surface Reconstruction," *J. Math. Imaging and Vision,* vol. 11, no. 1, pp. 27-45, 1999.

[9] D. Keren and M. Werman, "Probabilistic Analysis of Regularization," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 15, no. 10, pp. 982-995, Oct. 1993.

[10] N. Kiryati and A.M. Bruckstein, "What's in a Set of Points?" *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 4, pp. 496-500, Apr. 1992.

[11] Y. Leedan and P. Meer, "Estimation with Bilinear Contraints in Computer Vision," *Proc. Int'l Conf. Computer Vision,* pp. 733-738, 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,* vol. 15, pp. 329-336, 1994.

[13] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes.* Cambridge Univ. 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,* vol. 18, pp. 97-108, 1982.

[15] F. Solina and R. Bajcsy, "Recovery of Parametric Models from Range Images: The Case for Superquadrics with Global Deformations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 12, pp. 131-147, 1990.

[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. Pattern Analysis and Machine Intelligence,* vol. 13, pp. 1115-1138, Nov. 1991.

[17] J.B. Tenenbaum, "Bayesian Modeling of Human Concept Learning," *Advances in Neural Information Processing Systems,* 1998.

[18] G. Vosselman and R.M. Haralick, "Performance Analysis of Line and Circle Fitting in Digital Images," *Proc. Workshop Performance Characteristics of Vision Algorithms,* 1996.

[19] M. Werman and Z. Geyzel, "Fitting a Second Degree Curve in the Presence of Error," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 17, pp. 207-211, 1995.

[20] Z. Zhang, "Parameter Estimation Techniques: A Tutorial with Application to Conic Fitting," *Image and Vision Computing,* vol. 15, pp. 59-76, 1997.