

# Stochastic Interpolation as an Approach to Data Regularization

Daniel Howard<sup>1</sup> and Joseph Kolibal<sup>2</sup>

<sup>1</sup>QinetiQ Fellow, QinetiQ Group PLC, Malvern, UK

<sup>2</sup>Assoc. Prof. Math., The University of Southern Mississippi, USA

## Abstract

*This paper describes stochastic interpolation (SI), a framework for data regularization that can be particularly useful when data contains noise. Based on a probabilistic method the framework recovers well-behaved functional representations of input data. The SI framework allows for data approximation, data interpolation, and can also obtain a representation of the roughening of data. It provides a unique mechanism for linking statistical data analysis with conventional interpolation and approximation methods that are built on non-negative operators. SI splits the interpolation operator into two steps involving an initial step that corresponds to discrete de-convolution followed by a step that corresponds to discrete convolution of the data. At the heart of the process is a row stochastic matrix. It represents the approximation of the data by means of a probabilistic weighting of the data values which allows the direct inclusion of statistical models into the process of data regularization. The probabilistic approach that SI offers has potential connections to the solution of inverse problems of data regularization that are of interest to the EMRS DTC, such as problems of image formation, filtering of clutter, where an idea exists about the nature of the generating processes, the PSF of a camera, and concerning the noise in a system. Stochastic interpolation resulting from the SI framework can account for this knowledge in the construction of the convolution stochastic matrix, and this can be used to interpolate, approximate or de-convolve the data.*

## Introduction

A well-posed problem in the sense of Hadamard [1] is defined to be a problem for which (1) a solution exists; (2) the solution is unique, and (3) the solution depends continuously on the data (is stable under perturbations on data). The third condition is less obvious but can be explained with the problem of working out a previous temperature distribution from final temperature data, for instance, and the solution to this is highly sensitive to perturbations in the final data and therefore such a problem is ill-posed in the sense of Hadamard.

Inverse problems are often ill-posed in the sense of Hadamard. In terms of the EMRS DTC consider the direct problem of image for-

mation. This involves digitally capturing the scene of interest but a number of physical factors such as temperature effects, rain, blur, focus, different types of noise, dynamic range, yield an image or representation that may not contain sufficiently the detail that is of interest to a user. These physical factors can be modelled by means of mathematical operators and this is called the direct problem. The inverse problem, the counterpart to this direct problem, is to recover and obtain the original scene starting from the end result, from what has been captured. The solution to this inverse problem has to involve assumptions about the nature of the noise and physical details that were responsible for the loss of resolution and information

in the image formation. There are problems that are related to this inverse problem in the EMRS DTC, for example the problem of super-resolution from many image presentations of a scene. Starting from many images they attempt to reconstruct the scene and this seems to be breaking the diffraction limit of the imaging system or to be overcoming the limits of the digital imaging sensor. However, all of these inverse problems cannot escape the possibility of being ill-posed in the mathematical meaning of solution existence and solution uniqueness, i.e., depending on the degree and nature of the noise that affected the formation of the acquired images, the scene that we seek may not be a unique solution to this inverse problem or a solution cannot be assumed to exist, and various factors can make the inverse problem ill-posed (under-determined).

Numerical solutions of ill-posed problems will require that assumptions be made to overcome their ill-posed nature, whether this actually reflects reality or not. To do this it becomes necessary to introduce additional information about the solution such as an assumption of smoothness or a bound over some measure or norm. This activity is known in mathematics as data regularization. A popular example is the Tikhonov regularization, also known as ridge regression. The idea of regularization of an ill-conditioned problem that admits a large number of solutions is to give preference to a particular solution with desirable properties or one which is more probable in light of our knowledge about the context of the situation. An optimal regularization parameter is unknown and in practical problems determined by an ad-hoc method (for example using a Bayesian interpretation).

Data regularization has historically been accomplished with a range of well developed mathematical tools. Whether the interest is in approximation or interpolation, various techniques such as least squares, Fourier, wavelet and other transform methods can be applied.

Both approximation and interpolation are usually implemented as a linear or linearized problem using a variety of bases including polynomials, trigonometric polynomials, exponential functions and radial basis functions. Approximation and interpolation aim to recover a real valued function  $g(x)$  that approximates or interpolates the generating function  $f(x)$  from a presentation of  $f(x)$  at a set of values  $f = f_1, \dots, f_N$  taken at locations  $x = x_1, \dots, x_N$ .

Interpolation by means of radial basis functions is employed in reconstructing an object from scattered point cloud data, e.g., to model 3D surface data from a laser scan. An advantage of radial basis functions is that the resulting analytical function can be evaluated anywhere in 3D on and off the surface independent of the locations of the original data points. Another advantage is that gradients and higher derivatives can be determined analytically and that they are continuous and smooth.

Similarly, Stochastic Interpolation (SI) in much of the description that follows, can, in some cases, be seen to be similar to the approach used in the development of the radial basis function approach. SI achieves data regularization using non-negative operators in a manner that is related to radial basis functions (RBFs), thus it shares the rapid convergence and high accuracy associated with RBFs, however it derives from a totally different lineage, enforcing a constraint condition that requires that approximations be stochastic, and requiring a control volume defining the influence of each nodal point. Most significantly, it does not require a radial basis, although these are, from an analytical viewpoint, the most commonly utilized class of functions used in the development of SI. Intrinsically, data far away from an interpolated or approximated point is assumed to be less influential than data that is close to the point being interpolated, hence a dependence on the norm of this distance, i.e., a radial dependence, arises naturally in many applications.

The framework for SI is also different in that

it achieves interpolation by means of a split operator, e.g., instead of solving for the inverse of a matrix directly to obtain the interpolant, the process is split into what amounts to a deconvolution step followed by a convolution step, and this offers a lot of flexibility in data regularization, particularly for noisy data. In SI the stochastic constraint is important because it assures that constant functions get mapped to constant functions. For example, if the input data are of the same value, then the interpolated values at points between them are guaranteed to be that same value. This desirable behaviour carries over for slowly varying functions. More significantly, this constraint provides for flexibility since any non-negative function can be used to construct a stochastically normalized approximation method, including ones generated from the data itself, allowing for the convenient implementation of non-linear methods. This provides a powerful tool for generating non-linear interpolation models that are remarkably well behaved.

The next section explains the SI framework but owing to space limitations more details are available in our precursor report to the EMRS DTC, however an outline is necessary in order to have an understanding of the significance of the approach to data regularization. The key point is that the development leads to a *linear scheme* which can easily be non-linearized to account for rapidly varying, rough data, as was done in regard to image interpolation on non-uniform grids using limiters [2]. Limiters offered a robust solution to the Gibbs-like phenomenon. The non-linearity introduced in this study amounted to a modified estimate of the probability density function based on the data location, thereby altering the radial dependence of the data to one which was not necessarily radially dependent.

### Development Summary

The development of SI is most easily accomplished by starting with the Bernstein poly-

nomials since historically this contains the essence of the concept of probabilistic weighting of data to achieve an approximating polynomial. A modification of this approach leads to the construction of approximating functions rather than polynomials, and these were termed Bernstein functions in honour of this connection with the Bernstein polynomials.

The extension of this to an interpolation scheme is done through an argument that emphasizes the natural relationship of the method to inverse problems, asking the question, what data can be approximated to yield the current data? Solving for that directly is shown to correspond to a discrete de-convolution step. This step constructs a pre-image of the input data. This is followed by a re-convolution process that achieves interpolation. On exploring this construction, a generalized approach to approximation and interpolation emerges in which almost any cumulative probability distribution function could be used to construct a stochastic matrix.

### Development

Consider the Bernstein polynomial of order  $n$  approximating the function  $f(x)$ , sampled at the  $n + 1$  equally spaced data points  $(x_k, f(x_k))$ ,  $k = 0, 1, \dots, n$ ,

$$B_n(f, x) = \sum_{k=0}^n \binom{n}{k} f(x_k) x^k (1-x)^{n-k}. \quad (1)$$

where  $x \in [0, 1]$  and  $\binom{n}{k}$  denotes the combinatorial of  $n$  items taken  $k$  at a time, and where  $x_i = x_{i-1} + h$ ,  $h = 1/n$  with  $x_0 = 0$  and  $x_n = 1$ . By mapping  $[a, b]$ , i.e., the domain of  $f$ , to  $[0, 1]$ ,  $f$  may be considered to be sampled on the unit interval. The Bernstein polynomials are uniformly approximating, and have desirable properties that have been extensively catalogued.

The important observation to make, however, is that these polynomials are constructed by weighting each of the function's values, i.e., the data points  $f(x_k)$  by the binomial probability

distribution function. In essence, the polynomial  $B_n$  results from the re-assignment of the ‘mass’ contained at each node  $(x_k, f(x_k))$  to the interval  $[0, 1]$ , based on a probabilistic model.

The disadvantage with the Bernstein polynomials is that the data is required to be uniformly sampled, but even more than this technical limitation, the convergence of  $B_n$  to  $f$  is extremely slow. It means that the approximations to a known function  $f$  which are constructed using  $B_n$  depend on the number of sample points, and because of the slow convergence and the uniform approximation property, the representation of  $f$  by  $B_n$ , except when  $n$  is very large, acts as a low pass filter, smoothing out any useful high frequency information contained in  $f$ .

The intent was thus to replace the Bernstein polynomials by developing a continuum extension of these. The motivation was to allow approximation on any grid, and to eliminate the high frequency filtering which is correlated with the number of points  $n$  that is intrinsic to the construction of  $B_n(f, x)$ .

Denote a probability distribution function with mean  $\mu$  and variance  $\sigma^2$  as  $p_{\mu, \sigma^2}$ . Each term in the kernel of  $B_n(f, x)$  derives from the binomial distribution

$$p_{\mu, \sigma^2}^B = p_{nx, nx(1-x)}^B = \binom{n}{k} x^k (1-x)^{n-k}, \quad (2)$$

expressing the probability that  $n$  Bernoulli trials yield exactly  $k$  successes whenever the probability of success on each trial is  $x$ . For  $n$  sufficiently large it is natural to consider the normal probability distribution function

$$p_{\mu, \sigma^2}^N = p_{nx, nx(1-x)}^N = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma^2}, \quad (3)$$

as an approximation to the binomial distribution. Thus, if  $n \gg 1$ ,

$$\binom{n}{k} x^k (1-x)^{n-k} \approx (2\pi nx(1-x))^{-1/2} e^{-(k-nx)^2/2nx(1-x)}. \quad (4)$$

The generalized Bernstein function  $K_n(f, x)$  can now be developed by substituting (4) in place of (2). Since  $K_n(f, x)$  is evaluated not based on  $f$ , but on the piecewise constant approximation<sup>1</sup> to  $f$ , i.e.,  $\hat{f} = \sum f_k$ , in which  $\hat{f}$  is constant in the interval  $[y_k, y_{k+1}]$  containing each data point  $(x_k, y_k)$ . The recovery of  $f$  is done using

$$K_n(f, x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \hat{f}(z\sqrt{(2/n)x(1-x)} + x) e^{-z^2} dz,$$

with  $x \in [0, 1]$  and where  $\hat{f}$  is defined to have the values  $f(x_0)$  in  $(-\infty, y_1)$  and  $f(x_n)$  in  $(y_n, \infty)$ . The abuse of notation in denoting  $K_n(f, x)$  in place of  $K_n(\hat{f}; x)$  is obvious, and is consistent with the use in  $B_n(f, x)$ .

Since  $f_k$  is constant in each interval  $(y_{k-1}, y_k)$ ,  $-\infty = y_0, y_1, \dots, y_n = \infty$ , the stochastic approximant to the data is expressible as the sum

$$K_n(f, x) = \sum_{k=0}^n \frac{f(x_k)}{2} (g_1 + g_2) \quad (5)$$

with

$$g_1 = \operatorname{erf} \left( \frac{y_{k+1} - x}{\sqrt{(2/n)x(1-x)}} \right),$$

$$g_2 = \operatorname{erf} \left( \frac{x - y_k}{\sqrt{(2/n)x(1-x)}} \right).$$

From (5), it is evident that  $K(1; x) = 1$  for all  $x$  and for all  $n$ . Thus, this model has the satisfying property that  $K_n$  is exact for all constant functions, as is  $B_n$ , however the Bernstein functions<sup>2</sup>,  $K_n$ , are not exact for linear functions,  $x$ , as are the Bernstein polynomials.

The error function terms can immediately be recognized as the cumulative of the Gaussian

<sup>1</sup>This is the simplest model that can be applied. Higher order models can be developed by approximating  $f$  locally as a polynomial and then integrating.

<sup>2</sup>The name is in honour of S. Bernstein.

probability distribution function. As will become apparent, other probability distribution functions might be used by replacing the error function terms with the cumulative of the chosen function. This is a very useful feature of the method because a probability distribution function that is associated to a noise generating process can then be used via this framework to implement a principled data regularization for the problem of interest. Note that other functions are possible to be used instead. They may correspond to the PSF of a camera and the method can be applied in different stages to correspond to different steps of a process (e.g. image formation).

### Consequences of the construction

A connection between  $K_n$  in (5) and solutions to the heat or diffusion equation provided a method for achieving control of smoothing, and provided the necessary link to extending the process of approximation to interpolation in a natural manner.

Letting  $t = 1/n$  and defining  $u_k(x, 1/n)$  to be the  $k$ -th contribution to the sum  $K_n(f, x)$ , and making a minor change in the form of  $K_n$  in (5), gives

$$u_k(x, t) = \frac{f(x_k)}{2} (g_3 + g_4) \quad (6)$$

with

$$g_3 = \operatorname{erf} \left( \frac{y_{k+1} - x}{2\sqrt{x(1-x)/2\sqrt{t}}} \right),$$

$$g_4 = \operatorname{erf} \left( \frac{x - y_k}{2\sqrt{x(1-x)/2\sqrt{t}}} \right).$$

On setting  $\alpha = x(1-x)/2 \geq 0$  and then shifting  $x$  by  $\sigma_k = (y_{k+1} + y_k)/2$ , we have

$$u_k(x + \sigma_k, t) = \frac{f(x_k)}{2} \left[ \operatorname{erf} \left( \frac{L - x}{2\sqrt{\alpha t}} \right) + \operatorname{erf} \left( \frac{L + x}{2\sqrt{\alpha t}} \right) \right] \quad (7)$$

where  $L = (y_{k+1} - y_k)/2$ .

It is obvious that  $u_k(x + \sigma_k, t)$  is the solution to the partial differential equation

$$\frac{\partial u_k(z, t)}{\partial t} = \alpha \frac{\partial^2 u_k(z, t)}{\partial z^2} \text{ on } (-\infty, \infty) \text{ and}$$

$$u_k(z, 0) = \begin{cases} f(x_k) & x \in (-L, L) \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where the initial state is determined by the data,  $f_k$  shifted to the interval  $[-L, L]$ . The term  $\alpha = x(1-x)/2$  is the diffusion coefficient and there is a different differential equation for each value of  $x$  at which the Bernstein function,  $K_n(f, x)$  is evaluated. The construction of  $K_n(f, x)$  for each fixed  $x$  and  $n$  is then seen to be a discrete convolution of the solutions to the heat or diffusion equation with the function  $\hat{f}$ , i.e.,

$$K_n(f, x) = \sum_{k=0}^n f_k(x) u_k(x - \sigma_k).$$

With the diffusion coefficient given by  $\alpha = x(1-x)/2$ , i.e., the standard model<sup>3</sup> derived from the Bernstein polynomial, it is evident that  $u_k(0, 1/n) = f_k(0)$  and  $u_k(1, 1/n) = f_k(1)$ . Since  $\sum f_k(0) = f(0)$  and  $\sum f_k(1) = f(1)$ ,  $K_n(f, 0) = f(0)$  and  $K_n(f, 1) = f(1)$ , and so this Bernstein function shares the property of interpolating at the endpoints of the interval with the Bernstein polynomial.

If  $n$  is taken to be large, the resulting approximation given by  $K_n$  is indistinguishable from the approximations provided by  $B_n$ , as would be expected given that the binomial distribution converges to the normal distributions for large  $n$ . The functions,  $K_n$ , can also be shown to be uniformly approximating, however in contrast to the Bernstein polynomials, they are computable on any grid, and the quality of the approximation can be made independent of the number of sampled points. Indeed it can

<sup>3</sup>While of developmental interest, this model yields Bernstein functions that are polynomial-like in their behavior, and thus yields interpolants with similar pathologies. Thus  $\alpha$ , is usually taken to be a constant.

be shown that  $\lim_{\alpha \rightarrow 0} K_n(f, \alpha, x_k) = \hat{f}(x_k)$ , and thus taking the parameter  $\alpha$  small enough recovers the piecewise constant data representing  $f$  on the sample grid.

Thus far, a means for constructing an approximation to a set of data  $(x_k, f(x_k))$  sampled from the function  $f$  has been achieved, much the same as the Bernstein polynomial achieves this approximation. This approach can clearly be seen to be related to a radial basis function in which the interval centres define the radial distances to which a Gaussian is applied, however there is more to the construction than that since the last interval on either end is taken to infinity to correctly weight the contribution from the end points to assure that all of the probability is counted.

It is this probabilistic model that is of interest in this development since the intention is to continue within the framework of Bernstein's development of the Bernstein polynomials. Indeed, as will become evident, the process of approximation by  $K_n$  corresponds to a discrete convolution, hence by de-convolving the data, it is possible to conveniently construct a well-behaved interpolation operator. This will be the basis for constructing Stochastic Interpolation.

### Constructing Stochastic Interpolation based on the Bernstein functions

The description of Bernstein interpolation formally begins with the definition of the convolution process involving Bernstein functions, i.e.,

$$K_n(f, \alpha; s) = \sum_{k=0}^n \frac{y_k}{2} \left[ \operatorname{erf} \left( \frac{y_{k+1} - s}{2\sqrt{\alpha/n}} \right) - \operatorname{erf} \left( \frac{y_k - s}{2\sqrt{\alpha/n}} \right) \right] \quad (9)$$

where the smoothing inherent in  $K_n$  depends on  $\alpha$ , the diffusion coefficient, and where  $s$  corresponds to a point at which  $K_n$  is evaluated. The assumption is that the data  $\{(x_k, y_k)\}$  are obtained from a smooth function,  $u$ , which has been perturbed by random errors,  $\varepsilon$ , i.e.,

$y_k = f(x_k) = u(x_k) + \varepsilon_k$  with all data lying in  $[0, 1]$ . As before, the  $y_k$  are taken as midpoint cell boundaries<sup>4</sup>, i.e.,  $y_i = (x_{i+1} - x_i)/2$ , with  $y_0 = -\infty$  and  $y_{n+1} = \infty$ . In this description  $\alpha$  is no longer taken to be necessarily dependent on the functional form  $\sqrt{(2/n)x(1-x)}$ , and is interpreted more generally as a free parameter in the model.

Bernstein approximation, as described in (9), can be subsumed into a framework using matrix operations. Consider the matrix  $A_{mn} = (a_{ij})$  in which each entry in the matrix contains the term

$$a_{ij} = \operatorname{erf} \left( \frac{y_{j+1} - s_i}{2\sqrt{\alpha/n}} \right) - \operatorname{erf} \left( \frac{y_j - s_i}{2\sqrt{\alpha/n}} \right), \quad (10)$$

so that the  $i$ -th row of  $A_{mn}$  contains data pertaining to the  $s_i$ -th point,  $i = 0, \dots, m$ , at which  $K_n$  is evaluated, while each column,  $j = 0, \dots, n$ , contains the terms weighting this data by the window  $(y_j, y_{j+1})$  corresponding to the data point  $(x_j, y_j)$ . Expressed as a matrix operation, Bernstein approximation amounts to evaluating

$$K_n(\mathbf{f}, \alpha; \mathbf{s}) = A_{mn} \mathbf{f}, \quad (11)$$

where  $\mathbf{s} = (s_0, s_1, \dots, s_m)$ ,  $\mathbf{f} = (f_0, f_1, \dots, f_n)$ , and where  $f_k = f(s_k)$ . Note, the nonstandard notation on the size of  $A$ , i.e.,  $A_{mn}$  instead of  $A_{n+1, n+1}$  to avoid notational complexity.

When  $m = n$  it is possible not only to convolve the vector  $\mathbf{f}$  as in (11), but to uniquely deconvolve it, i.e., to find the pre-image  $\mathbf{p}$  such that

$$A_{nn} \mathbf{p} = \mathbf{f}. \quad (12)$$

Each row of  $A_{mn}$  can be shown to be linearly independent – being generated by a normal probability distribution function with different mean for each row. Thus the inverse of  $A_{mn}$  always exists. This yields the pre-image data  $\{(x_i, p_i)\}$  corresponding to the data  $\{(x_i, y_i)\}$ . Constructing the Bernstein function approximation  $K_n(\mathbf{p}; x)$  to the pre-image  $\{(x_i, p_i)\}$  using

<sup>4</sup>They need not be assigned in this manner, and can be biased towards either end of the interval on a non-uniform grid.

the same diffusion model as was used to construct  $A_{nn}$  yields the identity operator,

$$A_{nn} \mathbf{p} = A_{nn}(A_{nn}^{-1} \mathbf{f}) = \mathbf{f}, \quad (13)$$

since  $A_{nn}A_{nn}^{-1} = I$ . The extension to constructing a workable interpolation method is now obvious.

To interpolate the data requires constructing the matrix  $A_{nn}$  that is used to deconvolve the data, and an extended matrix  $\tilde{A}_{mn}$  to convolve the pre-image. To construct  $A_{nn}$ , select each approximation point such that  $s_i = x_i$ . For the extended matrix  $\tilde{A}_{mn}$  each row is obtained by evaluating (10) at points  $s_i$ ,  $i = 1, \dots, m$  which lie at or between the values in the set  $\{x_i\}$ ,  $i = 0, \dots, n$ , i.e., choosing  $s_i$  such that  $x_l \leq s_i \leq x_n$ ,  $i = 0, \dots, m$ . Thus, the stochastic interpolation of the data using Bernstein functions is given by

$$\tilde{A}_{mn} \mathbf{p} = \tilde{A}_{mn} A_{nn}^{-1} \mathbf{f} \quad (14)$$

Clearly, if the  $i$ -th row of  $\tilde{A}_{mn}$  corresponds to a row generated by evaluating the Bernstein function at  $s_i = x_l$  for some  $l$ , then at this point the interpolation has exactly the value of  $y_l$ .

The matrices  $A_{nn}$  and  $A_{mn}$  are row stochastic, since each of the entries in each row represent a probability the sum of which over all columns in any row is exactly one by construction.

### Stochastic Interpolation

It is obvious that there is nothing special about the Gaussian which is used to generate the row space of the matrices  $A_{nn}$  and  $A_{mn}$  and that any probability distribution function can be used. For that matter, any non-negative function, properly normalized, can be used as the generator for the row space of  $A_{nn}$  and  $A_{mn}$  with the only requirement that  $A_{nn}$  be invertible. Even this condition can be relaxed, for example if least square solutions or other matrix pseudo-inverses can be applied, care is needed to ensure that this remains an interpolating process. From the viewpoint of data regularization, probability distribution functions that are generated from

the data values, or more properly, from analysis of the input data, represent an intriguing way to generate the required row spaces. By this means it is possible to generate an extensive range of non-linear interpolation schemes. These are schemes in which the interpolation of the data depends on attributes of the data, and yet are still constructed from linear matrix multiplications.

Stochastic Interpolation schemes all share the following attributes:

- The approximation process is a probabilistic weighting of the data values, similar to the process employed by Bernstein to construct the Bernstein polynomials. This means that constant data is always reconstructed exactly since constants are eigenvectors of any stochastic operator. Moreover, the inverse of a stochastic matrix is a sum-one matrix, i.e., one for which the row sums are equal to one, and thus stochastic interpolation also preserves constants exactly.
- The interpolation process can always be viewed as de-convolution process followed by re-convolution of a pre-image to yield the interpolant. Interpolation is split into the product of two linear operators, in effect factoring the operator. This means that it is possible to
  - Construct incomplete splittings in which de-convolution is emphasized, or in which convolution is emphasized, allowing for the smoothing or roughening of the data; and, to
  - Approximate the de-convolution process, i.e., construct an approximate inverse for  $A_{nn}$  thereby increasing the computational efficiency of the process.
- The resulting matrices for common probability distribution functions and most

smoothly varying positive functions used to generate the row space of the convolution and de-convolution matrices are usually well conditioned. For example, using the Bernstein functions, the condition numbers are on the order of a few hundred, even for very large matrices containing thousands of columns.

Let's examine the factorization of the linear interpolation process. The construction involving the product  $\tilde{A}_{mn}A_{mn}^{-1}$  can be accomplished by choosing values of  $\alpha$  to be different on each step, i.e.,  $\tilde{A}_{mn}(\alpha_a)A_{mn}^{-1}(\alpha_2)$ , where  $\alpha$  is related to variance of the underlying probability distribution function used to generate the stochastic approximant. If  $\alpha_1 \neq \alpha_2$ , then the row space of  $\tilde{A}_{mn}$  is different from that of  $A_{mn}$ , hence the product  $\tilde{A}_{mn}A_{mn}^{-1}$  will not be interpolating. Choosing  $\alpha_2 > \alpha_1$  will only result in an approximation scheme, but selecting  $\alpha_1 > \alpha_2$  leads to a variety of deconvolution schemes with peak sharpening, and high frequency detail enhancement. In particular, these properties can be adjusted to be local, thereby allowing interpolation of some of the data, smoothing of some of the data and roughening some of the data. This provides for tremendous flexibility. For example, because of the factorization of the operator, it is possible to replace  $A_{mn}^{-1}$  with an approximate inverse which has the property that high frequency data is smoothed, while low frequency data is interpolated.

This splitting also allows the less stable step associated with the inversion of  $A_{mn}$  to be treated separately, followed by the always stable step of approximation, i.e., multiplication by  $A_{mn}$ . For example, in a situation involving extremely rapidly varying differences in the grid size, the use of a mollifier such as the Bernstein function to construct  $A_{mn}^{-1}$  can yield excessive oscillations when interpolating in the vicinity of a very widely spaced grid that is next to a very tightly spaced grid. By limiting, or thresholding the value of the probability distribution func-

tion, i.e., by setting bounds on how rapidly  $u$  may vary, it is possible to eliminate or control these oscillations. This is accomplished by modifying the critical coefficients of  $a_{ij}$  of  $A_{mn}^{-1}$  by not allowing any to exceed a preset limit. The stochasticity of  $A_{mn}^{-1}$  is then assured by re-normalizing the rows of the matrix, i.e., we are now sampling from an altered probability distribution function.

Finally, this splitting opens possibilities for improving the efficiency of the inversion of  $A_{mn}$ . While all of the data points are usually not necessary in terms of numerically evaluating the interpolant, having a relatively large number of points included into the computational stencil is advantageous in regard to improving the accuracy of the scheme, and thus having efficient means for inverting  $A_{mn}$  are important.

An obvious improvement for any radially symmetric probability distribution function is achieved by limiting the range of integration described in (5) to extend not to  $-\infty$  and  $+\infty$ , but to a symmetric interval around  $x_0 = a$  and  $x_n = b$ . The resulting matrix can be shown to be Toeplitz, however it is not stochastic. The stochasticity is restored by noting that there is a diagonal matrix,  $D$ , such that  $B_{mn} = DA_{mn}$  is stochastic, hence the inverse for the stochastic convolution matrix  $B_{mn}$  is given by  $A_{mn}^{-1}D^{-1}$ . Since Toeplitz matrices can be inverted in  $\mathcal{O}(n^2)$  operations instead of  $\mathcal{O}(n^3)$  required for dense matrices, the savings can be significant when  $n$  is large. In practice, through the use of data blocking, only small matrices need to be inverted and hence the order of the scheme is not a significant consideration.

### The difference with RBFs

SI can be similar to RBFs in the case when the values of the entries of the matrices are also related to the distance between the input data points and, for instance, as arguments of a Gaussian function when that is used to generate the convolution matrix. Moreover, there is a difference, as SI requires that a control volume

be associated with each point in such a way as to produce a stochastic matrix that accounts for the geometric weighting of each data point.

SI achieves interpolation in a two step convolution and de-convolution process which is not practiced with RBFs and which gives SI flexibility at choosing between operational modes in which interpolation, approximation, or de-convolution is selected. While RBFs could be applied in a row-stochastic way, there is no natural measure of the geometric weighting of data as is intrinsic to SI. This weighting relative to the location of the data is constrained to have value one in SI, and consequently these values can be biased, e.g., upwinded by choosing the  $y_i$  in (9) so that they are not at the midpoints of the data values  $x_i$  and  $x_{i+1}$  but are closer to the data at one side than the other side of the interval.

Finally, the framework that is presented in this paper subsumes some of the development in RBFs. It is broader since the SI framework allows for the use of non-radially dependent terms.

### **Applications of stochastic interpolation**

We have developed the technique in different directions and have published this in the open literature [3] and reports to MOD. The publications cover our use of SI to design an alternative method of image compression, a methodology for zooming images [2], a method for closing a boundary in biomedical imaging [5], and a means to remove the baseline drift of instrumentation [4].

In principle this has the following potential advantage over rival methods when modeling the inverse problem counterpart of the direct problem of image formation. Knowledge about the nature of the noise that has resulted in the image data (whether it is white noise or K-means distributed, etc.) can be accounted for by using the appropriate probability distribution as the mollifier (the generator of the row space). Similarly, the PSF of a camera can inform the choice of mollifier. This can become a clean,

natural, easy to implement and straightforward way to achieve near-optimal data regularization for image data, i.e. to solve an inverse problem in super-resolution or in image restoration.

### **Advantages of stochastic interpolation**

Everything that a radial basis function may offer (analytical derivatives, simple processing of non-uniformly distributed input data, etc.) can be done using SI, and it offers much more. The stochastic constraint has allowed for the development of methods in which the data itself determines the shape of the probability density function.

The framework has also enabled us to construct blended probability distribution functions using the Gaussian and the Laplace function for example (Fig. 1 and Fig. 2). If  $u$  and  $v$  are two probability distribution functions, then  $u + cv$  is a probability distribution function after re-normalization, consequently it is possible to blend density functions so that combined schemes can be constructed. An excellent example involves the blending of the normal and Laplace distributions. While the Laplace is quite useful in representing pixel image data, it has the disadvantage of producing rapid drop off in response as the probability distribution function vanishes rapidly away from the point of interest. In contrast the normal distribution even with a small value of the variance, falls off too slowly.

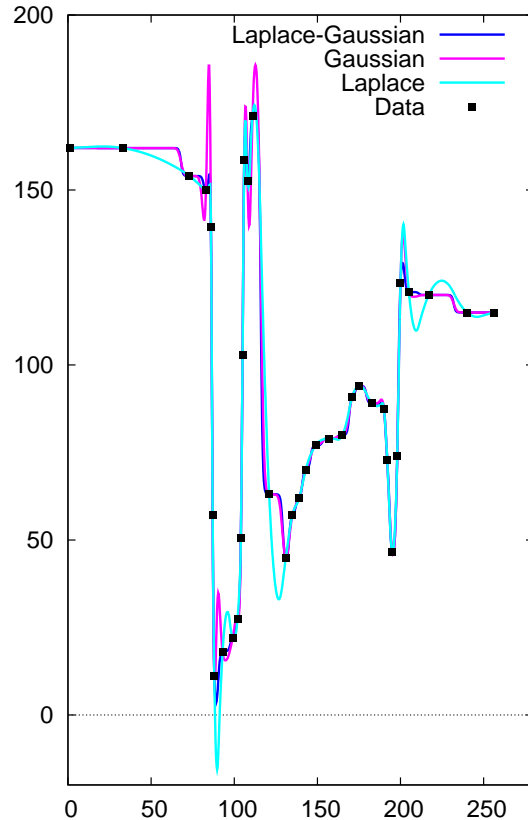
A combination of these can be blended to yield appropriate control of the shape of the interpolant for noisy image data. The facility with which renormalization yields new functions can be exploited to construct interpolation schemes in which the values of the probability distribution function can be limited to yield non-linear interpolation schemes. The SI approach also allows the construction of the approximate inverse to the convolution matrix so that it remains interpolating in regions of smoothly varying data since constants are exactly represented by stochastic operators, and

thus functions close to constants are also well represented.

The approach subsumes radial basis functions in that it is not necessary to have a radial basis function to do stochastic interpolation, and something that remains to be investigated are issues with non-linearity in which the probability distribution function itself is modified. We approached it one way, and found some difficulties, but there are other approaches, and in the context of Bayesian inference, it strikes us that the probability distribution functions could be generated by the data. If the basis for this generating function was again based on the distance between the points, then this would be a radial basis function that would be generated, although one that derived from a totally different approach, i.e., it would not necessarily be analytical. The disadvantage is that one would not have the derivative or other quantities if one used this basis.

Moreover, there are problems in the use of radial basis functions with ill-conditioning of the matrices, whereas with stochastic interpolation our examination of the condition number of the matrices using the Gaussian as the generator of the row space and even using other probability distribution functions that have the same shape, is that they produced matrices with condition numbers that were at most in the thousands for 1000 x 1000 matrices. Although there may exist many factors that contribute to this, we conjecture that the stochastic constraint contributes to this.

Finally, in regard to understanding the importance of splitting SI into a de-convolution followed by a convolution step, Fig. 3 shows the effects of partially de-convolving an image data set, selecting the smoothing on the first step, i.e.,  $\alpha_1$  to be larger than that on the second step, i.e.,  $\alpha_2$ . The result is a ‘snap-shot’ of the pre-image as it has moved toward the original data; the operator at this point is not interpolating, and so the resulting data is much rougher showing peaks that evolved into the data under

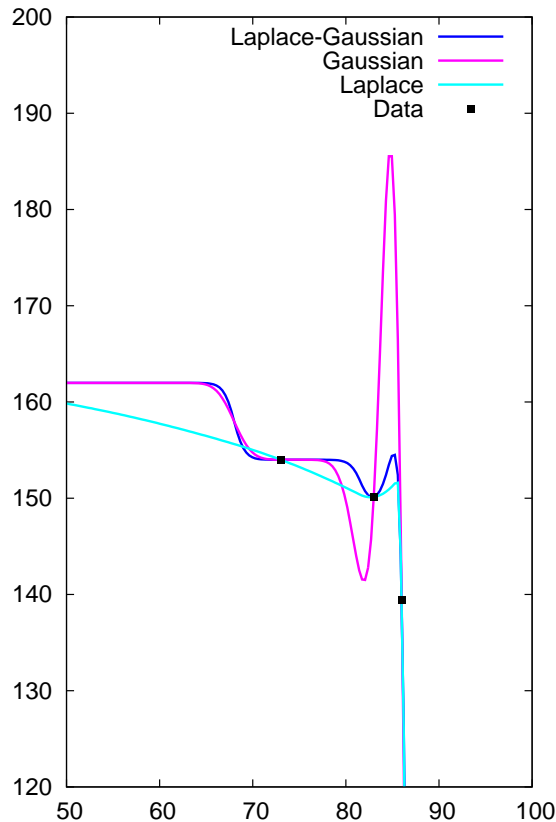


**Figure 1:** Discrete image pixel data on a highly non-uniform grid consisting of 35 data points interpolated to 1000 points using Gaussian, Laplace and blended Gaussian-Laplace probability distribution functions. Note that the Laplace undershoots at  $x = 80$  more than the Gaussian, however the blended distribution does not.

the stochastic smoothing model represented by  $A_{mn}$ . In this example, the arctan function was normalized to yield a pdf which was used to construct the de-convolution matrix. It is somewhat smoother than the Gaussian, having a behaviour somewhat different in the tail of the distribution.

## Summary

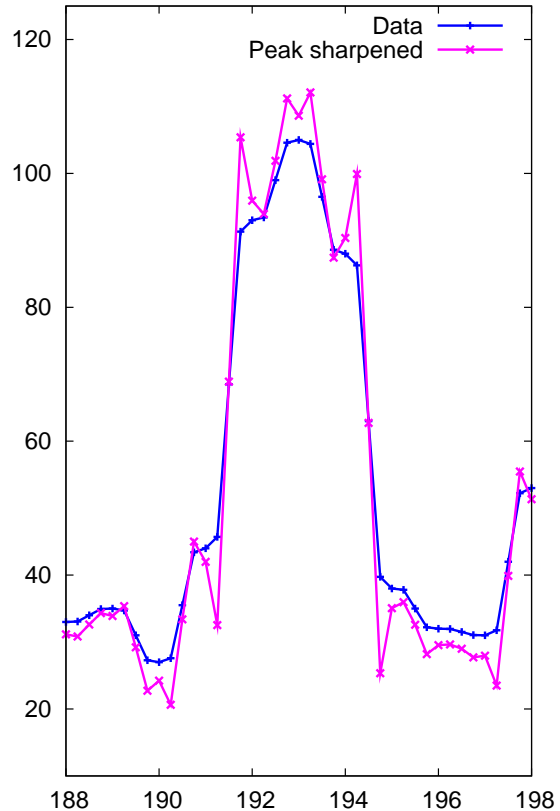
Stochastic Interpolation provides opportunities for function recovery from discrete data that derive from the probabilistic framework in which it is set, and provides the natural setting in which data approximation, interpolation, and



**Figure 2:** We zoom on Fig. 1 to show that Gaussian probability distribution function oscillates excessively in presence of step function, the Laplace probability distribution function is too tight, while the blended Gaussian-Laplace probability distribution function exhibits improved combination of behaviours.

de-convolution can be done for a wide variety of problems where statistical analysis is central to the estimation process.

Typically it has been applied in situations in which presumptions have been made about the statistical nature of the errors, e.g., using a normal distribution evaluated using the Bernstein function model, however even in these situations, innovative possibilities exist for working creatively with standard statistical models. For example, if  $u$  and  $v$  are two probability distribution functions, then  $u + cv$  is a probability distribution function after re-normalization, consequently it is possible to blend probability distri-



**Figure 3:** Discrete deconvolution of image data using a pdf based on the probability density function built from an arctan function. The data shown as peak sharpened is partially smoothed, i.e.,  $\alpha_1 = 0.3$ , and  $\alpha_2 = 0.1$ , and shows deliberately exaggerated peak enhancement to demonstrate the role of the pre-image in constructing rougher, i.e., more oscillatory approximations to the data.

bution functions so that combined schemes can be constructed. An excellent example related to image data involves the blending of the normal and Laplace distributions. While the Laplace is quite useful in representing pixel image data, it has the disadvantage of producing rapid drop off in response as the probability distribution function vanishes rapidly away from the point of interest. In contrast the normal distribution even with a small value of the variance, falls off too slowly. A combination of these can be blended to yield appropriate control of the shape of the interpolant for noisy image data.

### Future Work

Interesting and potentially productive applications of SI are in filtering images that suffer noise or poor spatial resolution, and in using the de-convolution step to yield a pre-image that can be used to register images instead of using pixel values. The pre-image, as typified in Fig 3, is rougher and more feature-rich than the data, and thus provides more structure for use in image alignment. Upon registration of the images, a larger image can be formed to contain the pixel values from multiple low-resolution images which are then approximately interpolated using different choices of generating functions of the stochastic matrix to obtain super-resolution (as alternative to Bayesian or POCS methods for temporal super-resolution). The approximate interpolation aspect of SI provides for a blending of approximation and interpolation in a manner such that higher frequency data is more nearly approximated, while lower frequency data is interpolated. Thus, SI might achieve data regularization to solve the inverse problem (restore the scene by removing atmospheric blurring etc., from the available imagery). This ambitious objective may see SI being applied as a series of stages.

Finally, probability distribution functions derived from discrete data may be used directly without the need for any analytical model. The discrete data may be interpolated using data derived estimates of the statistical model which then serves as the means for evaluating the weighting of  $|s - y_k|$  for any output point  $s$ , thereby generating the row space of the convolution and de-convolution matrices. While the assumption of a radial dependence of the functional form of the row space generator is the least complex, i.e., the form  $|s - y_k|$  seems to occur naturally in this setting, the framework for constructing stochastic interpolants is open to a broader class of generating functions, and it would be interesting to examine possibilities for row space generator functions based on analysis of discrete statistical data which result in

non-scalar correlations yielding more complex functional relationships.

### Acknowledgment

The work reported in this paper was funded by the Electro-Magnetic Remote Sensing (EMRS) Defence Technology Centre, established by the UK Ministry of Defence and run by the consortium of SELEX Galileo, Thales Defence, Roke Manor Research and Filtronic.

### References

- [1] Jacques Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52, 1902.
- [2] D Howard and J Kolibal. Image resize application of novel stochastic methods of function recovery. *Advances in Hybrid Information Technology: Revised Selected Papers*, Springer, 2007.
- [3] J Kolibal and D Howard. Implications of a novel family of stochastic methods for function recovery. *Proc IEEE IHH-MSP*, pages 495–498, 2006.
- [4] J Kolibal and D Howard. Maldi-tof baseline drift removal using stochastic bernstein approximation. *EURASIP Journal on Applied Signal Processing*, 2006.
- [5] J Kolibal and D Howard. Alternative parametric boundary reconstruction method for biomedical imaging. *Journal of Biomedicine and Biotechnology*, 2008.