

We discuss a method to filter observed data,  $u_{\text{obs}}$ , in a Bayesian framework, where  $u_{\text{obs}}$  can be spatial, temporal, or higher dimensional data. We are also not limited to cases where  $u_{\text{obs}}$  is observed over a regular grid. Our filtered estimate,  $u_{\text{post}}$ , incorporates  $u_{\text{obs}}$  and prior knowledge of the statistical properties of the underlying signal. In this sense, this work is closely related to Kriging and, more generally, Gaussian process regression (e.g. Rasmussen and Williams, 2006). We constrain  $u_{\text{post}}$  with the observation equation

$$u_{\text{post}} = u_{\text{obs}} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{C}_{\text{obs}}), \quad (1)$$

and the prior model

$$u_{\text{prior}} \sim \mathcal{N}(0, \mathbf{C}_{\text{prior}}), \quad (2)$$

where  $\epsilon$  and  $u_{\text{prior}}$  are considered to be Gaussian processes with zero mean and covariances  $\mathbf{C}_{\text{obs}}$  and  $\mathbf{C}_{\text{prior}}$  respectively. The solution for  $u_{\text{post}}$  minimizes the objective function

$$\|u_{\text{post}} - u_{\text{obs}}\|_{\mathbf{C}_{\text{obs}}}^2 + \|u_{\text{post}}\|_{\mathbf{C}_{\text{prior}}}^2 \quad (3)$$

and is itself a Gaussian process with a distribution described by

$$u_{\text{post}} \sim \mathcal{N}(\bar{u}_{\text{post}}, \mathbf{C}_{\text{post}}). \quad (4)$$

We use  $\bar{u}_{\text{post}}$  and  $\mathbf{C}_{\text{post}}$  to denote the mean and covariance of  $u_{\text{post}}$  respectively. Using Bayesian linear regression (e.g. Tarantola, 2005) these values are found to be

$$\begin{aligned} \bar{u}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \mathbf{C}_{\text{prior}}^{-1})^{-1} \mathbf{C}_{\text{obs}}^{-1} u_{\text{obs}} \\ \mathbf{C}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \mathbf{C}_{\text{prior}}^{-1})^{-1}. \end{aligned} \quad (5)$$

$\mathbf{C}_{\text{obs}}$  is presumably well known, while  $\mathbf{C}_{\text{prior}}$  needs to be chosen based on an understanding of the underlying signal which we are trying to estimate. With a judicious choice of  $\mathbf{C}_{\text{prior}}$ , eq. (5) can be made equivalent to several well established smoothing methods. In particular, we demonstrate how eq. (5) can be viewed as a low-pass filter with a well defined cutoff frequency. This is first demonstrated for filtering one-dimensional data, and the extension to higher dimensions follows naturally.

## One-dimensional smoothing

For one-dimensional data we consider a prior which can be stated implicitly as

$$\mathbf{D}_n u_{\text{prior}} = q, \quad q \sim \mathcal{N}(0, \lambda^2), \quad (6)$$

where  $\mathbf{D}_n$  is an  $n^{\text{th}}$  order differentiation matrix, and  $q$  is white noise with constant variance  $\lambda^2$ . If we momentarily ignore the fact that  $\mathbf{D}_n$  is not invertible then we can explicitly write our prior covariance as

$$\mathbf{C}_{\text{prior}} = \lambda^2 (\mathbf{D}_n^T \mathbf{D}_n)^{-1}, \quad (7)$$

and the filtered mean and covariance are described by

$$\begin{aligned} \bar{u}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \frac{1}{\lambda^2} \mathbf{D}_n^T \mathbf{D}_n)^{-1} \mathbf{C}_{\text{obs}}^{-1} u_{\text{obs}} \\ \mathbf{C}_{\text{post}} &= (\mathbf{C}_{\text{obs}}^{-1} + \frac{1}{\lambda^2} \mathbf{D}_n^T \mathbf{D}_n)^{-1}. \end{aligned} \quad (8)$$

This filtered solution is closely tied to several well established methods of smoothing. For example, Kimeldorf and Wahba (1970) pointed out that eq. (8) is equivalent to a smoothing spline. This can be seen by ignoring data uncertainty (i.e.  $\mathbf{C}_{\text{obs}} = \mathbf{I}$ ) and noting from eq. (3) that  $\bar{u}_{\text{prior}}$  minimizes

$$\|u_{\text{obs}} - \bar{u}_{\text{prior}}\|_2^2 + \|\mathbf{D}_n \bar{u}_{\text{prior}}\|_2^2, \quad (9)$$

which is a discretized form of the definition of a smoothing spline (e.g. De Boor, 1978). As another example, we can loosely interpret our prior model in eq. (6) as Brownian motion or integrated Brownian motion when  $n = 1$  and  $n = 2$ , respectively. From this perspective, it is possible to see the connection between eq. (8) and Kalman filters, where these stochastic models are commonly used to describe time-dependent signals (e.g. Segall and Mathews, 1997).

One challenge that is common to each of these smoothing methods is selecting the hyperparameter  $\lambda$ . There are numerous ways that this could be accomplished. For example, one could use cross-validation, a trade-off

curve, or simply vary  $\lambda$  until  $\bar{u}_{\text{post}}$  looks appropriate when compared to  $u_{\text{obs}}$ . There is merit to using an entirely objective approach such as cross-validation, and this would be appropriate if there is no prior knowledge of the signal's characteristic wavelength. Otherwise, it may be better to chose  $\lambda$  such that eq. (8) damps out all the high frequency oscillations which are known to be noise. We elaborate on this point by demonstrating that eq. (8) can also be viewed as a low-pass filter with a cutoff frequency determined by  $\lambda$ .

We wish to transform  $\bar{u}_{\text{post}}$  in eq. (8) to the frequency domain and in order to do so we assume that  $u_{\text{obs}}$  has a constant sampling rate. We also require that  $\epsilon$  and  $u_{\text{prior}}$  are stationary stochastic processes (i.e. their statistical properties are invariant to time shifts). We require them to be stationary so that matrix multiplication by  $\mathbf{C}_{\text{obs}}$  and  $\mathbf{C}_{\text{prior}}$  can be viewed as convolution in the time domain and thus multiplication in the frequency domain. We then consider the simplifying case where  $\epsilon$  is white noise with constant variance  $\sigma^2$  (i.e.  $\mathbf{C}_{\text{obs}} = \sigma^2 \mathbf{I}$ ), and  $\mathbf{D}_n$  is the periodic spectral differentiation matrix (e.g. Trefethen, 2000). Under a discrete Fourier transform,  $\mathbf{D}_n$  has the properties

$$\mathcal{F}[\mathbf{D}_n f] = (2\pi i \omega)^n \hat{f} \quad (10)$$

and

$$\mathcal{F}[\mathbf{D}_n^T f] = (-2\pi i \omega)^n \hat{f}, \quad (11)$$

where  $\omega$  is the frequency domain variable,  $f$  is an arbitrary vector and  $\hat{f}$  is its discrete Fourier transform. The discrete Fourier transform of  $\bar{u}_{\text{post}}$  is then

$$\hat{u}_{\text{post}}(\omega) = \frac{\frac{1}{\sigma^2}}{\frac{1}{\sigma^2} + \frac{(2\pi\omega)^{2n}}{\lambda^2}} \hat{u}_{\text{obs}}(\omega). \quad (12)$$

We make the change of variables

$$\lambda^2 = (2\pi\omega_c)^{2N} \sigma^2 \quad (13)$$

which changes the hyperparameter from  $\lambda$  to  $\omega_c$ . The reason for this change of variables becomes apparent when we simplify eq. (12) to

$$\hat{u}_{\text{post}}(\omega) = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}} \hat{u}_{\text{obs}}(\omega). \quad (14)$$

We can recognize eq. (14) as an  $n$ 'th order low-pass Butterworth filter with cut-off frequency  $\omega_c$ . In the limit as  $N \rightarrow \infty$  eq. (14) becomes an ideal low-pass filter which removes all frequencies above  $\omega_c$  and leaves lower frequencies unaltered. Of course, an ideal low-pass filter is often undesirable because it will tend to produce ringing artifacts in the filtered solutions. When modeling  $u_{\text{prior}}$  as Brownian motion or integrated Brownian motion, where  $n = 1$  and  $n = 2$  respectively, the transfer function is tapered across  $\omega_c$ , which ameliorates ringing in the filtered solution.

By demonstrating that eq. (8) can be made equivalent to a low-pass filter, it may not be clear why we would ever use eq. (8) when it is far more efficient to filter in the frequency domain through the Fast Fourier Transform. In order to make use of the Fast Fourier Transform, the observations must be made at a constant sampling rate and the observation noise must be white with constant variance. In contrast, these conditions do not need to be met in order to evaluate eq. (8). The question is then whether eq. (8) still effectively acts as a low-pass filter when the idealized conditions are not met.

We answer this question with a demonstration. In this demonstration we generate 100 samples of synthetic data over the time interval  $0 < t < 1$  yr. The true signal in the synthetic data is a sine wave with a frequency of  $1 \text{ yr}^{-1}$  and 1 mm amplitude. We obscure the synthetic data with white noise that does not have a constant variance. We randomly pick a variance for each datum from a uniform random distribution ranging from 0.5 to  $2 \text{ mm}^2$ . For the period from 1 to 2 years, we assign a variance of  $25 \text{ mm}^2$  to the data, effectively making the data uninformative for the filtered solution. The synthetic data and its estimated power spectral density are plotted in Figure 1. Our goal is to use eq. (8) as a low-pass filter which damps out frequencies that are higher than our signal. Since our data does not have constant variance, we cannot choose a cutoff frequency from eq. (13). Instead we define a characteristic data variance as

$$\frac{1}{\sigma^2} = \frac{1}{P} \text{tr}(\mathbf{C}_{\text{obs}}^{-1}), \quad (15)$$

where  $P$  is the number of observations, and we relate  $\lambda$  to  $\omega_c$  by

$$\lambda^2 = (2\pi\omega_c)^{2n} \bar{\sigma}^2. \quad (16)$$

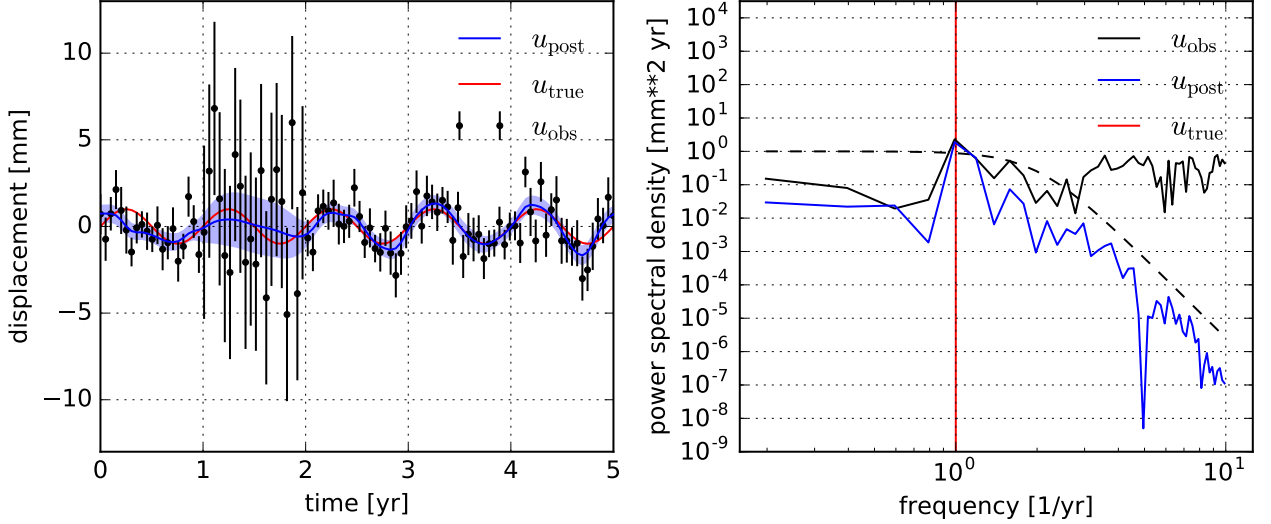


Figure 1: Left panel shows observed data (black scatter points), filtered data (blue line), and the true signal which we are trying to recover (red line). The lines on each scatter point and the light blue region show the one standard deviation uncertainty for the observations and filtered solution respectively. The right panel shows the estimated power spectral density for the observed, filtered, and true signal. The black dashed line is the squared transfer function in eq. (14) with  $\omega_c = 2 \text{ yr}^{-1}$ , which roughly indicates how eq. (8) scales the frequency content of  $u_{\text{obs}}$ . If the data variances were constant then the frequency content of  $u_{\text{obs}}$  and  $u_{\text{post}}$  would be exactly related by the black dashed line.

In this demonstration  $D_n$  is still the spectral differentiation matrix and we use  $n = 2$ . Our cutoff frequency is chosen to be  $\omega_c = 2 \text{ yr}^{-1}$ . The filtered solution and its frequency content are plotted Figure 1, where it is apparent that eq. (8) is indeed acting as a low-pass filter with the specified cut-off frequency even though the variances are not constant. It is worth noting that the frequency content of  $u_{\text{post}}$  appears to change depending on the variance of  $u_{\text{obs}}$ . For example, Over the interval  $1 < t < 2$ , where the variance of  $u_{\text{obs}}$  is higher,  $u_{\text{post}}$  lacks the higher frequency oscillations that can be seen in the remainder of the time series. This is desirable behavior because we do not want  $u_{\text{post}}$  contorting to fit dubious data. It is also worth pointing out that  $u_{\text{post}}$  asymptotically approaches a solution that is not significantly different from that shown in Figure 1 when the variance for  $u_{\text{obs}}$  increases to infinity over the interval  $1 < t < 2$ . This too is desirable behaviour because one could then surmise that an appropriate way to handle missing data with eq. (8) is to assign  $u_{\text{obs}}$  an infinite variance for the missing time period. With this insight, we can then extend the application of eq. (8) from data smoothing to data interpolation and extrapolation.

## Smoothing in higher dimensions

We expand our discussion from one-dimensional smoothing to smoothing in two-dimensional space. While the remainder of this paper discusses smoothing in two-dimensions, the extension to smoothing in higher dimensions will prove to be trivial. We now allow  $u_{\text{prior}}$  to have non-zero covariance along multiple dimensions. In other words, we now treat  $u_{\text{prior}}$  as a random field rather than a stochastic process. The smoothed solution that we seek is still given by eq. 5 and our task is once again to find an appropriate choice of  $\mathbf{C}_{\text{prior}}$ . Below we consider the case where the  $M$ 'th order Laplacian of  $u_{\text{prior}}$  is white noise. That is to say

$$\Delta_M u_{\text{prior}}(x_1, x_2) = q, \quad q \sim \mathcal{N}(0, \lambda^2) \quad (17)$$

where

$$\Delta_M = \frac{\partial^{2M}}{\partial x_1^{2M}} + \frac{\partial^{2M}}{\partial x_2^{2M}}. \quad (18)$$

The corresponding covariance matrix is then

$$\mathbf{C}_{\text{prior}} = \lambda^2 (\Delta_M^T \Delta_M)^{-1}. \quad (19)$$

We again assume that the observation noise is uncorrelated with constant variance  $\sigma^2$ . Using the change of variables

$$\lambda = (2\pi\omega_c)^{2M} \sigma \quad (20)$$

we obtain the solution

$$u_{\text{smooth}}(x_1, x_2) = \left( \mathbf{I} + \left( \frac{1}{2\pi\omega_c} \right)^{4M} \mathbf{\Delta}_M^T \mathbf{\Delta}_M \right)^{-1} u_{\text{obs}}(x_1, x_2), \quad (21)$$

which in the two-dimensional frequency domain is

$$\hat{u}_{\text{smooth}}(\omega_1, \omega_2) = \frac{1}{1 + \left( \frac{\omega_1^{2M} + \omega_2^{2M}}{\omega_c^{2M}} \right)^2} \hat{u}_{\text{obs}}(\omega_1, \omega_2). \quad (22)$$

The transfer function in eq. (22) can once again be recognized as a low-pass filter. Namely, in the limit as  $M \rightarrow \infty$  all the frequency pairs,  $(\omega_1, \omega_2)$ , where  $|\omega_1| > \omega_c$  or  $|\omega_2| > \omega_c$  are removed. That is, the transfer function becomes a two-dimensional box centered at  $(0, 0)$  with width  $2\omega_c$ . It is apparent from the transfer function why our prior model must be defined in terms of even order derivatives. If eq. (18) contained odd order derivatives, then all frequency pairs where  $\omega_1 = -\omega_2$  would be unattenuated, leaving undesirable artifacts in the smoothed solution.

We have demonstrated with eq. (?? and 21) that eq. 5 can effectively act as a low pass filter with a judicious choice of  $\mathbf{C}_{\text{prior}}$ . One may then wonder why eq. (?? and 21) would ever be used to smooth data when it is far more efficient to perform the equivalent operation through the Fast Fourier Transform domain. The power in eq. (?? and 21) resides in the fact that they makes no assumption about when or where the observations have been made. In order to filter data through the Fast Fourier Transform, the observations must be evenly spaced on a regular grid, which is often not the case for geophysical data. Indeed, the focus of this paper is to smooth data which has been observed at scattered locations. We can use eq. (?? and 21) regardless of where the observations were made provided that we are able to create the differentiation matrix. We use the recently developed Radial Basis Function Finite-Difference method to form the differentiation matrices and we describe the procedure in the follow section.

## References

- De Boor, C. (1978). *A practical guide to splines*, volume 27. Springer-Verlag, New York.
- Kimeldorf, G. S. and Wahba, G. (1970). A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. *The Annals of Mathematical Statistics*, 41(2):495–502.
- Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press.
- Segall, P. and Mathews, M. (1997). Time dependent inversion of geodetic data. *Journal of Geophysical Research*, 102(B10):22391–22409.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. SIAM.
- Trefethen, L. N. (2000). *Spectral Methods in Matlab*. SIAM.