In the most general sense, we seek to estimate a smoothed signal, u_{smooth}, from observed data, u_{obs}, by incorporating prior knowledge of the signals statistical properties. In this paper, we describe the smoothed and observed signal as

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

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

We find u_{smooth} that simulataneously satisfies eq. (1) and eq. (2) in a least-squares sense by minimization the objection function

$$F(u_{\text{smooth}}) = ||u_{\text{smooth}} - u_{\text{obs}}||^2_{\mathbf{C}_{\text{obs}}} + ||u_{\text{smooth}}||^2_{\mathbf{C}_{\text{prior}}}. \tag{3}$$

The solution that minimizes eq. 3 is simply

$$u_{\text{smooth}} = (\mathbf{C}^{-1}_{\text{obs}} + \mathbf{C}^{-1}_{\text{prior}})^{-1} \mathbf{C}^{-1}_{\text{obs}} u_{\text{obs}}. \tag{4}$$

The challenge is in choosing an appropriate $\mathbf{C}_{\text{prior}}$. Below we discuss selection of $\mathbf{C}_{\text{prior}}$ is one-dimensional problems, which then naturally leads to an extension to selection of $\mathbf{C}_{\text{prior}}$ for higher dimensions.

One-dimensional smoothing

When the signal is only assumed to covary in time, u_{prior} is commonly treated as Brownian motion (eg. Me Segall, McQuire, Murray, etc.). We treat a u_{prior} as Brownian motion by assuming that its velocity is white noise with constant variance λ^2. That is to say

$$\mathbf{D}_1 u_{\text{prior}} = q, \quad q \sim \mathcal{N}(0, \lambda^2). \tag{5}$$

where $\mathbf{D}_N$ is a differentiation matrix which estimates an N'th order derivative. It is also common to treat u_{prior} as integrated Brownian motion, In such case, we would just replace $\mathbf{D}_1$ with $\mathbf{D}_2$. In either case, the appropriate choice of the $\mathbf{C}_{\text{prior}}$ is

$$\mathbf{C}_{\text{prior}} = \lambda^2 (\mathbf{D}_N^T \mathbf{D}_N)^{-1}. \tag{6}$$

It should be noted that $\mathbf{D}_N^T \mathbf{D}_N$ is generally not invertible and we write it as such for the sake for clarity. This is no problem because the inverse of $\mathbf{D}_N^T \mathbf{D}_N$ does not need to be taken when applying eq. 6 to eq. 4.

There is still a need to select λ, which described how rapidly we expect the smoothed signal to vary. There are numerous methods for selecting λ. For example, one could use maximum likelihood methods, a trade-off curve, or simply vary σ until the smoothed signal looks appropriate when compared to the observations. We note that smoothing is fundamentally a low-pass filter and so it is perhaps most useful to choose λ based on its cut-off frequency. We then consider the solution for u_{smooth} in the frequency domain.

For the purpose analytical tractability, we assume that that the observation noise is white and its covariance matrix is described by

$$\mathbf{C}_{\text{obs}} = \sigma^2 \mathbf{I}. \tag{7}$$

The solution for u_{smooth} in the frequency domain is

$$\hat{u}_{\text{smooth}}(\omega) = \frac{\left(\frac{1}{\sigma}\right)^2}{\left(\frac{1}{\sigma}\right)^2 + \left(\frac{(2\pi\omega)^N}{\lambda}\right)^2} \hat{u}_{\text{obs}}(\omega). \tag{8}$$

We make the change of variables

$$\lambda = (2\pi\omega_c)^N \sigma \tag{9}$$

where ω_c is now our free parameter. This simplifies eq. 8 to

$$\hat{u}_{\text{smooth}}(\omega) = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2N}} \hat{u}_{\text{obs}}(\omega) \tag{10}$$

and the solution in the time domain becomes

$$u_{\text{smooth}}(t) = \left(\mathbf{I} + \left(\frac{1}{2\pi\omega_c}\right)^{2N} \mathbf{D}_N^T \mathbf{D}_N\right)^{-1} u_{\text{obs}}(t). \tag{11}$$

We can recognize eq. 10 as a N'th order Butterworth filter, and ω_c is the cut-off frequency. Thus selecting the hyperparameter, λ, can be translated through eq. 9 into a more tangible question of identifying an upper bound on the frequency content of the underlying signal. In the limit as $N \to \infty$ eq. 10 becomes an ideal low-pass filter which removes all frequencies above ω_c and leaves lower frequencies unaltered. Of course, an ideal

low-pass filter is often undesirable in practice because it will tend to produce ringing artifacts in the smoothed solutions. When modeling u_{prior} as Brownian motion or integrated Brownian motion, where $N = 1$ and $N = 2$ respectively, the transfer function is tapered across ω_c, which ameliorates ringing in the smoothed solution.

It is worth pointing out that the solution for u_{smooth} in eq. 11 is identical to the solution that would be obtained through Kalman filtering followed by smoothing. Kalman filtering and smoothing are recursive algorithms that operate along the time dimension. It is thus difficult to envision how a Kalman filter can be extended to smoothing data in higher dimensions. In contrast, extending eq. 11 to higher dimensions comes simply and naturally as we describe in the next section.

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_{prior} to have non-zero covariance along multiple dimensions. In other words, we now treat u_{prior} as a random field rather than a stochastic process. The smoothed solution that we seek is still given by eq. 4 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_{prior} is white noise. That is to say

$$\mathbf{\Delta}_M u_{\text{prior}}(x_1, x_2) = q, \quad q \sim \mathcal{N}(0, \lambda^2) \tag{12}$$

where

$$\mathbf{\Delta}_M = \frac{\partial^{2M}}{\partial x_1^{2M}} + \frac{\partial^{2M}}{\partial x_2^{2M}}. \tag{13}$$

The corresponding covariance matrix is then

$$\mathbf{C}_{\text{prior}} = \lambda^2 \left(\mathbf{\Delta}_M^T \mathbf{\Delta}_M \right)^{-1}. \tag{14}$$

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

$$\lambda = (2\pi\omega_c)^{2M}\sigma \tag{15}$$

we obtain the solution

$$u_{\text{smooth}}(x_1, x_2) = \left(\mathbf{I} + \left(\frac{1}{2\pi\omega_c} \right)^{4M} \mathbf{\Delta}_M^T \mathbf{\Delta}_M \right)^{-1} u_{\text{obs}}(x_1, x_2), \tag{16}$$

which in the two-dimensional frequency domain is

$$\hat{u}_{\text{smooth}}(\omega_1, \omega_2) = \frac{1}{1 + \left(\frac{\omega_1^{2M} + \omega_2^{2M}}{\omega_c^{2M}} \right)^2} \hat{u}_{\text{obs}}(\omega_1, \omega_2). \tag{17}$$

The transfer function in eq. (17) can once again be recognized as a low-pass filter. Namely, in the limit as $M \to \infty$ all the frequency pairs, (ω_1, ω_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. (13) 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. (11 and 16) that eq. 4 can effectively act as a low pass filter with a judicious choice of $\mathbf{C}_{\text{prior}}$. One may then wonder why eq. (11 and 16) 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. (11 and 16) 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. (11 and 16) 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.