Thoughts on Differential Data

Bill Menke, December 14, 2020

Summary:

(1) The use of differential data is well-justified when the original, undifferenced data have a covariance matrix in the form of a two-sided exponential function. It is not well-justified when the data have a covariance matrix in the form of a Gaussian function.

(2) When the original, undifferenced data have a covariance matrix in the form of a two-sided exponential function, and are regularly spaced in distance, the differential data are uncorrelated with a uniform variance.

(3) When the original, undifferenced data have a covariance matrix in the form of a two-sided exponential function, but are irregularly spaced in distance, the differential data are uncorrelated with a non-uniform variance that scales with the spacing of the original data.

(4) When the differential data are merely computed by taking differences of the original, undifferenced data, model parameters estimated using ordinary least-squares applied to the differential data are approximately equal to those estimated using weighed least squares applied to the original, undifferenced data (with the weights given by the inverse of the two-sided exponential covariance matrix). A different (better) solution only results when the differential data are directly estimated and their variance is smaller than is implied by differencing the original data.

Part 1. We first discuss the effect of transforming a data vector $\mathbf{d}$ into another vector $\mathbf{d}'$ by multiplying it by an invertible matrix $\mathbf{D}$:

$$\mathbf{d}' = \mathbf{D}\mathbf{d}$$

Let the $\mathbf{d}$ be a data vector of N Normally-distributed random variables with zero mean and symmetric covariance matrix $\mathbf{C}_d$. It is well-known that the transformation $\mathbf{d}' = \mathbf{D}\mathbf{d}$ with $\mathbf{D} = \mathbf{C}_d^{-1/2}$ (the symmetric square root of $\mathbf{C}_d^{-1}$) decorrelates the data and scales their variances to unity, so that the transformed data $\mathbf{d}'$ have covariance $\mathbf{C}_{d'} = \mathbf{I}$. This "full decorrelation" occurs for any invertible $\mathbf{D}$ that satisfies $\mathbf{C}_d^{-1} = \mathbf{D}^T\mathbf{D}$ (from whence it follows that $\mathbf{C}_d = \mathbf{D}^{-1}\mathbf{D}^{-1T}$), as can be verified using standard error propagation:

$$\text{if} \quad \mathbf{d}' = \mathbf{D}\mathbf{d} \quad \text{then} \quad \mathbf{C}_{d'} = \mathbf{D}\mathbf{C}_d\mathbf{D}^T = \mathbf{D}\mathbf{D}^{-1}\mathbf{D}^{-1T}\mathbf{D} = \mathbf{I}$$

Here, we have used the identities $[\mathbf{AB}]^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ and $[\mathbf{D}^{-1}]^T = [\mathbf{D}^T]^{-1}$. Note that **D** can be multiplied by any unary matrix **U** (obeying $\mathbf{U}^T\mathbf{U} = \mathbf{I}$) without changing $\mathbf{C}_d^{-1}$:

$$\mathbf{C}_d^{-1} = (\mathbf{UD})^T(\mathbf{UD}) = \mathbf{D}^T\mathbf{U}^T\mathbf{UD} = \mathbf{D}^T\mathbf{D}$$

Because many superficially different **D**s correspond to the same $\mathbf{C}_d^{-1}$, comparing $\mathbf{C}_d^{-1}$s is better than comparing **D**s.

For a problem in which the data are related to M model parameters $\mathbf{m}$ via $\mathbf{Gm} = \mathbf{d}$, the least squares solution $\mathbf{m}^{est}$ is the one that minimizes the error:

$$E(\mathbf{m}) = (\mathbf{d}^{obs} - \mathbf{Gm})^T\mathbf{C}_d^{-1}(\mathbf{d}^{obs} - \mathbf{Gm})$$

By multiplying the data equation $\mathbf{d} = \mathbf{Gm}$ by **D**, we can show that the transformed data satisfy $\mathbf{Dd} = \mathbf{GDm} = \mathbf{G}'\mathbf{m}$, with $\mathbf{G}' = \mathbf{GD}$. Furthermore, we can show that the transformed error $E'(\mathbf{m})$ is equal to $E(\mathbf{m})$:

$$E(\mathbf{m}) \equiv (\mathbf{d}^{obs} - \mathbf{Gm})^T\mathbf{C}_d^{-1}(\mathbf{d}^{obs} - \mathbf{Gm}) = (\mathbf{d}^{obs} - \mathbf{Gm})^T\mathbf{D}^T\mathbf{D}\,(\mathbf{d}^{obs} - \mathbf{Gm})$$

$$= (\mathbf{Dd}^{obs} - \mathbf{DGm})^T(\mathbf{Dd}^{obs} - \mathbf{DGm}) = (\mathbf{d}'^{obs} - \mathbf{G}'\mathbf{m})^T(\mathbf{d}'^{obs} - \mathbf{G}'\mathbf{m}) \equiv E'(\mathbf{m})$$

That is, the error $E'(\mathbf{m})$ associated with $\mathbf{d}'^{obs}$ is equal to the error $E(\mathbf{m})$ associated with $\mathbf{d}'^{obs}$. Consequently, the same least squares solution is achieved, irrespective of whether $\mathbf{d}'^{obs}$ or $\mathbf{d}^{obs}$ is used. That the "ordinary" least squares solution $\mathbf{m}(\mathbf{d}') = [\mathbf{G}'^{\mathrm{T}}\mathbf{G}']^{-1}\mathbf{G}'^{\mathrm{T}}\mathbf{d}'$ is exactly the same as the weighted least squares solution $\mathbf{m}(\mathbf{d}) = [\,\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{d}$ is shown as follows:

$$\mathbf{m}(\mathbf{d}) \equiv [\,\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{d} = [\,\mathbf{G}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{D}\,\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{Wd}$$

$$= [\,(\mathbf{DG})^{\mathrm{T}}(\mathbf{DG})]^{-1}(\mathbf{DG})^{\mathrm{T}}\mathbf{d}' = [\,\mathbf{G}'^{\mathrm{T}}\mathbf{G}']^{-1}\mathbf{G}'^{\mathrm{T}}\mathbf{d}' \equiv \mathbf{m}(\mathbf{d}')$$

Here, we have used the rule, $[\mathbf{AB}]^{\mathrm{T}} = \mathbf{B}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}$.

Part 2. We now consider the special case where the data **d** are sampled from a continuous function $d(x)$, such that $d_n = d(x_n)$ and $x_n = n\Delta x$. We first state some well-known properties of the matrices of first and second differences, $\mathbf{D}_1$ and $\mathbf{D}_2$, respectively:

$$\mathbf{D}_1 = \frac{1}{\Delta x}\begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

$$\mathbf{D}_1^T = \frac{1}{\Delta x}\begin{bmatrix} 1 & -1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & -1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \cdots & 0 & 1 \end{bmatrix}$$

$$\mathbf{D}_1^{-1} = \Delta x \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 1 & 1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}$$

$$\mathbf{D}_1^{-1T} = \Delta x \begin{bmatrix} 1 & 1 & 1 & \cdots & \cdots & 1 \\ 0 & 1 & 0 & \cdots & \cdots & 1 \\ 0 & 0 & 1 & 1 & \cdots & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

$$\mathbf{D}_2 = \mathbf{D}_1^T\mathbf{D}_1 = (\Delta x)^{-2}\begin{bmatrix} -2 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 1 & -2 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & 1 & -2 & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & -2 & 1 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & -1 \end{bmatrix}$$

$$\mathbf{D}_2^{-1} = -[\mathbf{D}_1^T\mathbf{D}_1]^{-1} = -\mathbf{D}_1^{-1}\mathbf{D}_1^{-1T} = -(\Delta x)^2\begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 2 & 2 & \cdots & 2 \\ 1 & 2 & 3 & 3 & \cdots & 3 \\ 1 & 2 & 3 & 4 & \cdots & 4 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 2 & 3 & 4 & \cdots & N \end{bmatrix}$$

$$\mathbf{Z} = \mathbf{D}_1^{-1T}\mathbf{D}_1^{-1} = -(\Delta x)^2\begin{bmatrix} \mathrm{N} & \mathrm{N}-1 & \mathrm{N}-2 & \cdots & 2 & 1 \\ \mathrm{N}-1 & \mathrm{N}-1 & \mathrm{N}-2 & \cdots & 2 & 1 \\ \mathrm{N}-2 & \mathrm{N}-2 & \mathrm{N}-2 & \cdots & 2 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 2 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{bmatrix}$$

Note that the first row of $\mathbf{d}' = \mathbf{D}_1\mathbf{d}$ is not a difference, but rather the boundary condition that $d_1' = d_1$. It is needed to insure that $\mathbf{D}_1^{-1}$ exists. The matrix is $\mathbf{D}_1$ is not anti-symmetric, because while transposition flip the $0 \cdots 0, 1, -1, 0 \cdots 0$ pattern of rows to $0 \cdots 0, -1, 1, 0$, it shifts the pattern by one column. Nevertheless, $\mathbf{D}_1^T\mathbf{d} \approx -\mathbf{D}_1\mathbf{d}$, at least when d_n slowly varies with index n. Consequently, the approximation $\mathbf{D}_1^T \approx -\mathbf{D}_1$ is a useful one. The inverse of $\mathbf{D}_1$ is the integration (cumulative sum) matrix. The quantity $\mathbf{D}_1^{-1T}$ is the "acausal" integration matrix. The second difference operator is constructed from the first via $\mathbf{D}_2 = -\mathbf{D}_1^T\mathbf{D}_1$. Finally, we define $\mathbf{Z} \equiv \mathbf{D}_1^{-1T}\mathbf{D}_1^{-1}$.

We now examine the consequence of the choice $\mathbf{D} = \gamma^{-1}\mathbf{D}_1$; that is, the data are transformed by the first-difference matrix. Then, $\mathbf{C}_d^{-1} = \gamma^{-2}\mathbf{D}_1^T\mathbf{D}_1$ and $\mathbf{C}_d = \gamma^2[\mathbf{D}_1^T\mathbf{D}_1]^{-1} = \gamma^2\mathbf{D}_1^{-1}\mathbf{D}_1^{-1T} = -\gamma^2\mathbf{D}_2^{-1}$.

Unfortunately, $\mathbf{C}_d = -\gamma^2\mathbf{D}_2^{-1}$ is not a useful covariance matrix, for two reasons. First, correlation decreases linearly along the anti-diagonal at the fixed rate of $\gamma^2(\Delta x)^2$. Thus, this covariance matrix cannot model a process with a scale length. Furthermore, depending upon the magnitude of the slope and the size of the matrix, elements of $\mathbf{C}_d$ that are distal from the main diagonal may become negative, implying anticorrelation of data values. Second, the correlation increases along the main diagonal, implying that $\mathbf{d}$ is not stationary. This increase can be eliminated by a change to the boundary condition. When the first row of $\mathbf{D}_1$ is redefined as $1,0 \cdots 0,1$, the main diagonal of $\mathbf{C}_d$ is constant.

In the next section, we will discuss a choice of $\mathbf{C}_d$ that both explicitly contains a scale parameter s and for which $\mathbf{D}$ is approximately proportional to $\mathbf{D}_1$.

Part 3. We now construct a covariance matrix $\mathbf{C}_d$ that gives rise to a transformation involving $\mathbf{D}_1$. We assume that $d(x)$ is stationary, so that $\mathbf{C}_d$ is a symmetric Toeplitz matrix, with elements $[\mathbf{C}_d]_{nm}$ that depends only upon $|n-m|$.

We now consider the matrix equation:

$$(\mathbf{D}_1^{\mathrm{T}}\mathbf{D}_1 + s^2\mathbf{I})\,\mathbf{Q} = 2s\,\mathbf{I}$$

where s is a parameter and $\mathbf{Q}$ is unknown. Later in this section, we will show that:

$$Q_{nm} \approx \exp\{-s\Delta x|n-m|\}$$

(except possibly along its edges). Thus, s is a scale parameter. If we then choose the covariance to be the two-sided exponential function $\mathbf{C}_d = \gamma^2\mathbf{Q}$, the matrix equation implies:

$$\mathbf{C}_d^{-1} = \mathbf{D}^{\mathrm{T}}\mathbf{D} = \gamma^{-2}\,\mathbf{Q}^{-1} = \left(\gamma\sqrt{2s}\right)^{-2}(\mathbf{D}_1^{\mathrm{T}}\mathbf{D}_1 + s^2\mathbf{I})$$

This last expression is equivalent to:

$$\mathbf{C}_d^{-1} = \mathbf{D}^{\mathrm{T}}\mathbf{D} = \left(\gamma\sqrt{2s}\right)^{-1}(\mathbf{D}_1 + s\mathbf{I})^T \quad \left(\gamma\sqrt{2s}\right)^{-1}(\mathbf{D}_1 + s\mathbf{I})$$

The equivalence be verified by multiplying out the expression and by applying $\mathbf{D}_1^{\mathrm{T}} \approx -\mathbf{D}^{(1)}$:

$$\mathbf{C}_d^{-1} = \left(\gamma\sqrt{2s}\right)^{-2}(\mathbf{D}_1^{\mathrm{T}}\mathbf{D}_1 + \mathbf{D}_1^{\mathrm{T}}s + \mathbf{D}_1 s + s^2\mathbf{I})$$

$$\approx \left(\gamma\sqrt{2s}\right)^{-2}(\mathbf{D}_1^{\mathrm{T}}\mathbf{D}_1 - \mathbf{D}_1 s + \mathbf{D}_1 s + s^2\mathbf{I}) = \left(\gamma\sqrt{2s}\right)^{-2}(\mathbf{D}_1^{\mathrm{T}}\mathbf{D}_1 + s^2\mathbf{I})$$

Consequently,

$$\mathbf{D} \approx \left(\gamma\sqrt{2s}\right)^{-1}(\mathbf{D}_1 + s\mathbf{I})$$

When the data are highly correlated, s is small and $\mathbf{D} \approx \left(\gamma\sqrt{2s}\right)^{-1}\mathbf{D}_1$ (which is the anticipated result). Then, $\mathbf{C}_d^{-1}$ is then proportional to $\mathbf{D}^{(2)}$, the matrix of second differences:

$$\mathbf{C}_d^{-1} = \mathbf{D}^{\mathrm{T}}\mathbf{D} = \left(\gamma\sqrt{2s}\right)^{-2}\mathbf{D}_1^{\mathrm{T}}\mathbf{D}_1 = -\left(\gamma\sqrt{2s}\right)^{-2}\mathbf{D}_2$$

The formula for Q_{nm} is derived as follows. In the limit where $N \to \infty$ and $\Delta x \to 0$, and for the unbounded interval $-\infty < x < +\infty$, the matrix equation becomes the differential equation:

$$\left(\left(\frac{d}{dx}\right)^{\dagger}\left(\frac{d}{dx}\right) + s^2\right) q(x - x_0) = \frac{1}{2s}\delta(x - x_0)$$

$$\left(-\frac{d^2}{dx^2} + s^2\right) q(x) = \delta(x)$$

(with boundary conditions, $q([x] \to \infty) = 0$. Here, $\dagger$ denotes adjoint and $\delta(.)$ is the Dirac impulse function. After setting $x_0 = 0$ and Fourier transforming spatial coordinate x to wavenumber k, we find:

$$(k^2 + s^2)\, q(k) = 1 \quad \text{and} \quad q(k) = \frac{2s}{(k^2 + s^2)}$$

Note that s represents a "corner" wavenumber, in the sense that when $k^2 \ll s^2$, $q(k)$ is constant, whereas when $s^2 \ll k^2$, $q(k) \propto k^{-2}$. It is well-known that the Fourier transform of $\exp\{-s|x|\}$ is $2s/(k^2 + s^2)$. Consequently, after invoking stationarity so that we can reintroduce x_0, we have:

$$q(x, x_0) = \gamma^2 \exp\{-s|x - x_0|\}$$

Part 4. We now contrast two approaches of solving for $\mathbf{m}$:

(A) The exact solution $\mathbf{m}(\mathbf{d}) = [\,\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{d}$ with $\mathbf{C}_d$ a two-sided exponential

(B) The approximate differential solution $\mathbf{m}(\mathbf{d}') = [\mathbf{G}'^{\mathrm{T}}\mathbf{G}']^{-1}\mathbf{G}'^{\mathrm{T}}\mathbf{d}'$ with $\mathbf{d}' = \mathbf{D}_1\mathbf{d}$.

We have deliberately omitted the factor of $\left(\gamma\sqrt{2s}\right)^{-1}$ from $\mathbf{d}'$, both because rarely is it used in practice, and because it cancels from the least-squares equation. Write $\mathbf{m}(\mathbf{d})$ in terms of $\mathbf{m}(\mathbf{d}')$:

$$\mathbf{m}(\mathbf{d}) = [\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\mathbf{G}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{d}$$

$$= [(\mathbf{D}_1^{-1}\mathbf{G}')^{\mathrm{T}}(\mathbf{D}_1^{\mathrm{T}}\mathbf{D}_1 + s^2\mathbf{I})\,(\mathbf{D}_1^{-1}\mathbf{G}')]^{-1}(\mathbf{D}_1^{-1}\mathbf{G}')^{\mathrm{T}}(\mathbf{D}_1^{\mathrm{T}}\mathbf{D}_1 + s^2\mathbf{I})\mathbf{D}_1^{-1}\mathbf{d}'$$

$$= [\mathbf{G}'^T\mathbf{G}' + s^2\,\mathbf{G}^T\mathbf{G}]^{-1}(\mathbf{G}'^T\mathbf{d}' + s^2\mathbf{G}^T\mathbf{D}_1^{-1}\mathbf{d}')$$

$$= [\mathbf{G}'^T\mathbf{G}' + s^2\,\mathbf{G}^T\mathbf{G}]^{-1}(\mathbf{G}'^T\mathbf{d}' + s^2\mathbf{G}'^T\mathbf{Z}\mathbf{d}')$$

$$= [\mathbf{G}'^T\mathbf{G}' + s^2\,\mathbf{G}^T\mathbf{G}]^{-1}\mathbf{G}'^T(\mathbf{I} - s^2\mathbf{Z})\mathbf{d}'$$

$$\approx ([\mathbf{G}'^T\mathbf{G}']^{-1} - s^2\,[\mathbf{G}'^T\mathbf{G}']^{-1}[\mathbf{G}^T\mathbf{G}][\mathbf{G}'^T\mathbf{G}']^{-1})\mathbf{G}'^T(\mathbf{I} - s^2\mathbf{Z})\mathbf{d}'$$

$$= [\mathbf{G}'^T\mathbf{G}']^{-1}\mathbf{G}'^T\mathbf{d}' - s^2\left[\mathbf{G}'^T\mathbf{G}'\right]^{-1}\left\{\,[\mathbf{G}^T\mathbf{G}]\left[\mathbf{G}'^T\mathbf{G}'\right]^{-1}\mathbf{G}'^T + \mathbf{G}'^T\mathbf{Z}\right\}\mathbf{d}'$$

$$= \mathbf{m}(\mathbf{d}') - s^2\left[\mathbf{G}'^T\mathbf{G}'\right]^{-1}\{\,[\mathbf{G}^T\mathbf{G}]\mathbf{m}(\mathbf{d}') + \mathbf{G}'^T[\mathbf{Z}\mathbf{d}']\}$$

$$= \left\{\mathbf{I} - s^2\left[\mathbf{G}'^T\mathbf{G}'\right]^{-1}[\mathbf{G}^T\mathbf{G}]\right\}\,\mathbf{m}(\mathbf{d}') - s^2\mathbf{m}(\mathbf{Z}\mathbf{d}')$$

(We have checked these formulas numerically). Here, we define $\mathbf{m}(\mathbf{Z}\mathbf{d}') \equiv \left[\mathbf{G}'^T\mathbf{G}'\right]^{-1}\mathbf{G}'^T[\mathbf{Z}\mathbf{d}']$. We have used the first order approximation $[\mathbf{A} + \varepsilon\mathbf{B}]^{-1} \approx \mathbf{A}^{-1} - \varepsilon\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$.

As expected, the exact solution $\mathbf{m}(\mathbf{d})$ differs from the approximate solution $\mathbf{m}(\mathbf{d}')$ by very small "correction" terms of order s^2. Ironically, although this treatment began with a covariance function with a tunable scale parameter s, the approximate solution $\mathbf{m}(\mathbf{d}')$ is not a function of s. Scale only enters through the corrections terms.

Though small, the correction terms can have an important effect. Because the quantity

$$\mathbf{Z}\mathbf{d}' = \mathbf{D}_1^{-1T}\mathbf{D}_1^{-1}\mathbf{D}_1\mathbf{d} = \mathbf{D}_1^{-1T}\mathbf{d}$$

is the acausal integral of $\mathbf{d}$, it will tend to contain more long-wavelength features than does $\mathbf{d}'$. Depending upon the structure of $\mathbf{G}'$, these features may carry over into $\mathbf{m}(\mathbf{D}_2^{-1}\mathbf{d}')$. Because it omits this term, the approximate differential solution may be deficient in long-wavelength features, compared to the exact solution.

We conclude that the approximate differential solution $\mathbf{m}(\mathbf{d}')$ is very close to the exact solution $\mathbf{m}(\mathbf{d})$. Whether or not the former is more simply computed will depend upon the relative ease of computing $\mathbf{G}$ and $\mathbf{G}'$. If $\mathbf{G}'$ is computed by first computing $\mathbf{G}$ and then applying $\mathbf{G}' = \mathbf{D}\mathbf{G}$, no significant computational savings results. On the other hand, if $\mathbf{G}'$ can be computed directly, and with more ease than $\mathbf{G}$, then the approximate solution will be more efficient.

Part 5. We now demonstrate that the transformation $\mathbf{d}' = \mathbf{D}_1\mathbf{d}$ is not necessarily an appropriate choice to decorrelate spatially-correlated data. Our argument starts with the function:

$$q(x, x_0) = [\sin\{-s|x - x_0|\} + \cos\{-s|x - x_0|\}]\exp\{-s|x - x_0|\}$$

Like the two-sided exponential, this function has a scale parameter s, but differs in that it has no cusp at $x = x_0$. Its overall shape is reminiscent of a Gaussian function, except that its tails decay more slowly than a Gaussian's and have a small degree of "overshoot". It is known to satisfy the "plate flexure" equation:

$$\left[\left(\frac{d^2}{dx^2}\right)^{\dagger}\frac{d^2}{dx^2} + 4s^4\right]q(x - x_0) = 8s^3\,\delta(x - x_0)$$

Consequently, the covariance matrix $\mathbf{C}_d = \gamma^2\mathbf{Q}$ approximately corresponds to $\mathbf{d}' \propto \mathbf{D}_2\mathbf{d}$; that is, a transformation involving second differences.

The transformation $\mathbf{D}$ corresponding to a Gaussian $\mathbf{C}_d$ can be stated formally but has limited usefulness. Let:

$$C_d(x) = \exp\{-½s^2x^2\}$$

Since the convolution of a Gaussian with itself is a Gaussian with twice the variance of the original, we can write:

$$q(x) = \exp\left\{-½\frac{x^2}{s^{-2}}\right\} = D^{-1}(x) * D^{-1}(x) \quad \text{with} \quad D^{-1}(x) = \sqrt{\pi}s\exp\left\{-½\frac{x^2}{½s^{-2}}\right\}$$

Here $D^{-1}(x)$ is the inverse operator to $D(x)$, in the sense that $D^{-1}(x) * D(x) = \delta(x)$. It follows that $q(x) = D^{-1\dagger}(x) * D^{-1}(x)$, since convolution with a symmetric function is self-adjoint. Now let us consider the equation $d(x) = D^{-1}(x) * d'(x)$. Taking the Fourier transform yields $d'(k) = D(k)\,d(k)$ where $D(k)$ is the reciprocal of $D^{-1}(k)$. The Fourier transform of a Gaussian of variance σ^2 is a Gaussian of variance σ^{-2}, so:

$$D^{-1}(k) = \sqrt{\pi}s\exp\left\{-½\frac{k^2}{2s^2}\right\} = \sqrt{\pi}s\exp\{-s^{-2}k^2\}$$

$$D(k) = \left(\sqrt{\pi}s\right)^{-1}\exp\left\{½\frac{k^2}{2s^2}\right\} = \left(\sqrt{\pi}s\right)^{-1}\exp\{s^{-2}k^2\}$$

The power series for an exponential is $\exp(x) = 1 + x + ½x^2 + (\frac{1}{6})x^3 + \cdots$, so:

$$d'(k) = D(k)d(k) = \left(\sqrt{\pi}s\right)^{-1}\left\{1 + s^{-2}k^2 + ½s^{-4}k^4 + (\frac{1}{6})s^{-6}k^6 + \cdots\right\}d(k)$$

Taking the inverse Fourier transform yields:

$$d'(x) = \left(\sqrt{\pi}s\right)^{-1}\left\{1 + s^{-2}\frac{d^2}{dx^2} + ½s^{-4}\frac{d^4}{dx^4} + (\frac{1}{6})s^{-6}\frac{d^6}{dx^6} + \cdots\right\}d(x)$$

Taking the discrete analog yields $\mathbf{d}' = \mathbf{Dd}$ with:

$$\mathbf{D} = \left(\sqrt{\pi}s\right)^{-1}\left\{\mathbf{I} + s^{-2}\mathbf{D}_2 + ½s^{-4}\mathbf{D}_4 + (\frac{1}{6})s^{-6}\mathbf{D}_6 + \cdots\right\}$$

Formally, the transformation **D** associated with a Gaussian covariance function is an infinite sum of even order-derivatives. The terms involve the length ratio $s^{-1}/\Delta x$ which, by inference, is greater than unity, implying that the sum diverges. While the solution indicates that higher order derivatives can appear in **D**, the result does not appear to have practical application here.

Part 5. Suppose that the data d_n are irregularly distributed at positions x_n, with $x_{n+1} > x_n$. It seems reasonable to consider the transformed data as $d'_1 = d_1$ and for $n > 1$, $d'_n = d_n - d_{n-1}$. This is just the transformation $\mathbf{D}_2$ with $\Delta x = 1$, which corresponds to covariance $\mathbf{C}_d = \mathbf{D}_2^{-1}$ (again, with $\Delta x = 1$). Thus, covariance between two data, d_n and d_m falls off linearly with the number of intervening data (as contrasted to the physical distance $\Delta x_{nm} = (x_n - x_m)$ between them.

This behavior might be acceptable if the data are randomly-distributed along the x-axis, with some mean spacing, Δx.

On the other hand, suppose that $x_n = b(n-1)^2$, where b is a constant, so that nearest-neighbor distances $\Delta x_{n+1,n} = 2n - 1$ grow linearly with n. The covariance matrix is:

$$[\mathbf{C}_d]_{nm} = \gamma^2 \exp\{-s|x_n - x_m|\}$$

While we are not able to derive analytically the corresponding transformation **D**, it easily can be computed. Numerical experiments indicate that rows of $\mathbf{C}_d^{-1}$ are very close to being proportional the rows of $-\mathbf{D}_2$, but that the overall magnitude of the rows varies approximately as $1/n$.

This last result can be explained by the following analysis. Suppose that the data **d** are evenly spaced in x and that their covariance matrix $\mathbf{C}_d$ is a two-sided exponential function. As shown previously, the transformation matrix **D** such that $\mathbf{C}_d^{-1} = \mathbf{D}^T\mathbf{D}$ completely decorrelates the vector of evenly spaced differences, $\mathbf{d}' = \mathbf{Dd}$, so that $\mathbf{C}_{d'} = \mathbf{I}$. Now suppose that we define a vector $\mathbf{d}'' = \mathbf{Sd}' = \mathbf{SD}_1\mathbf{d}$, of irregularly spaced differences between an ordered subset of **d**; for example, d_1, d_3 and d_6. The structure of the matrix **S** can be ascertained through the following example:

$d''_1 = d'_1 = d_1$

$d"_2 = d_3 - d_1 = (d_3 - d_2) + (d_2 - d_1) = (d_3 - d_2) + (d_2 - d_1) = d'_3 + d'_2$

$d"_3 = (d_6 - d_3) = (d_6 - d_5) + (d_5 - d_4) + (d_4 - d_3) = d'_6 + d'_5 + d'_4$

Or:

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

Here, the first row implements the boundary condition $d"_1 = d'_1 = d_1$ and subsequent rows implement the non-adjacent differences. The rows of $\mathbf{S}$ are mutually orthogonal, with the nth row consisting of, say, y_n instances of unity (with $y_1 = 1$). Thus, $\mathbf{SS}^T = \mathbf{Y}$, with $\mathbf{Y} \equiv \mathrm{diag}(\mathbf{y})$. Since $\mathbf{d}" = \mathbf{SDd}'$:

$$\mathbf{C}_{d"} = \mathbf{SDC}_d\mathbf{D}^T\mathbf{S}^T \approx \left(\gamma\sqrt{2s}\right)^{-2}\mathbf{SD_1C}_d\mathbf{D}_\mathbf{1}^T\mathbf{S}^T = \mathbf{SS}^T = \mathbf{Y}$$

(since $\mathbf{C}_{d'} = \mathbf{DC}_d\mathbf{D}^T = \mathbf{I}$). Thus, when the data have a two-sided exponential covariance matrix, irregularly spaced first-difference are uncorrelated but have unequal variances proportional to their spacing.

When the vector of irregularly spaced differences are defined as $\mathbf{d}^* = \mathbf{Y}^{-½}\mathbf{SD}_1\mathbf{d}$, then $\mathbf{d}^*$ is fully uncorrelated:

$$\mathbf{C}_{d"} = \left(\gamma\sqrt{2s}\right)^{-2}\mathbf{Y}^{-½}\mathbf{SD}_1\mathbf{C}_d\mathbf{D}_1^T\mathbf{S}^T\mathbf{Y}^{-½} = \mathbf{Y}^{-½}\mathbf{SS}^T\mathbf{Y}^{-½\mathrm{T}} = \mathbf{Y}^{-½}\mathbf{YY}^{-½\mathrm{T}} = \mathbf{I}$$

This result justifies the use of first-differences for irregularly spaced data, but it emphasizes that they will not in general have uniform variance. Instead, their variance increases with their physical separation. Irregular first-differences data $\mathbf{d}"$ need to be properly normalized as $\mathbf{d}^* = \mathbf{Y}^{-½}\mathbf{d}"$ for them to be compatible with the assumption that the original data are stationary.