# Partial Derivative of the Predicted Data and of the Total Error, with Respect to Parameters in the Autocovariance Matrix, in Gaussian Process Estimation (GPE) Problem

Bill Menke, August 21, 2020 with Example added September 15, 2020

This derivation relies on the exactly the same manipulations that are used in seismic adjoint methods (e.g. Menke, 2018), except that the linear operators involved are matrices, as contrasted to differential operators.

The $N$ training points $\mathbf{d}$ are assumed to be included in the $M$ model parameters $\mathbf{m}$, so that the predicted data $\mathbf{d}^{pre}$ can be recovered from the estimated model parameters $\mathbf{m}^{est}$ by $\mathbf{d}^{pre} = \mathbf{G}\,\mathbf{m}^{est}$, where $\mathbf{G}$ is an $N \times M$ matrix of zeros and ones.

The Gaussian Process Estimate (GPE) of the model parameters is:

$$\mathbf{m}^{est} = \mathbf{C}_{md}(p)\ \mathbf{u}(p) \quad \text{with} \quad \mathbf{u}(p) = \mathbf{A}^{-1}(p)\,\mathbf{d}^{obs} \quad \text{and} \quad \mathbf{A}(p) = \mathbf{C}_{dd}(p) + \sigma^2\mathbf{I}$$

Here, the $M \times N$ autocovariance matrix $\mathbf{C}_{md}(p)$ and the $N \times N$ symmetric autocovariance matrix $\mathbf{C}_{dd}(p)$ are functions of a parameter $p$. The matrix $\mathbf{A}(p)$ also is symmetric. The constant $\sigma^2$ represents a variance. The partial derivative of the predicted data with respect to this parameter is:

$$\frac{\partial \mathbf{d}^{pre}}{\partial p} = \mathbf{G}\,\frac{\partial \mathbf{m}^{est}}{\partial p} \quad \text{with} \quad \frac{\partial \mathbf{m}^{est}}{\partial p} = \frac{\partial \mathbf{C}_{md}}{\partial p}\,\mathbf{u} + \mathbf{C}_{md}\frac{\partial \mathbf{u}}{\partial p} \quad \text{and}$$

$$\frac{\partial \mathbf{u}}{\partial p} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial p}\mathbf{A}^{-1}\mathbf{d}^{obs} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial p}\,\mathbf{u} = -\mathbf{A}^{-1}\frac{\partial \mathbf{C}_{dd}}{\partial p}\,\mathbf{u}$$

Putting this together:

$$\frac{\partial \mathbf{d}^{pre}}{\partial p} = \mathbf{G}\frac{\partial \mathbf{C}_{md}}{\partial p}\,\mathbf{u} - \mathbf{G}\mathbf{C}_{md}\mathbf{A}^{-1}\frac{\partial \mathbf{C}_{dd}}{\partial p}\,\mathbf{u} \quad \text{with} \quad \mathbf{A}\mathbf{u} = \mathbf{d}^{obs}$$

Or:

$$\frac{\partial \mathbf{d}^{pre}}{\partial p} = \mathbf{G}\frac{\partial \mathbf{C}_{md}}{\partial p}\,\mathbf{u} - \mathbf{G}\mathbf{C}_{md}\mathbf{v} \quad \text{with} \quad \mathbf{A}\,\mathbf{v} = \frac{\partial \mathbf{C}_{dd}}{\partial p}\,\mathbf{u} \quad \text{and} \quad \mathbf{A}\mathbf{u} = \mathbf{d}^{obs}$$

Since $\mathbf{G}\mathbf{C}_{md} = \mathbf{C}_{dd}$ and $\mathbf{G}\,\partial\mathbf{C}_{md}/\partial p = \partial\mathbf{C}_{dd}/\partial p$, this result can also be written:

$$\frac{\partial \mathbf{d}^{pre}}{\partial p} = \frac{\partial \mathbf{C}_{dd}}{\partial p}\,\mathbf{u} - \mathbf{C}_{dd}\mathbf{v} \quad \text{with} \quad \mathbf{A}\,\mathbf{v} = \frac{\partial \mathbf{C}_{dd}}{\partial p}\,\mathbf{u} \quad \text{and} \quad \mathbf{A}\mathbf{u} = \mathbf{d}^{obs}$$

Thus, in order to calculate $\partial\mathbf{d}^{pre}/\partial p$, one must solve two instances of an $N \times N$ system, both with the same matrix, $\mathbf{A}$.

Defining the total $L_2$ error as $E = \mathbf{e}^T\mathbf{e}$ with $\mathbf{e} = \mathbf{d}^{obs} - \mathbf{d}^{pre}$, we have:

$$\frac{\partial E}{\partial p} = 2\mathbf{e}^T\frac{\partial \mathbf{e}}{\partial p} = -2\mathbf{e}^T\frac{\partial \mathbf{d}^{pre}}{\partial p}$$

Or:

$$\frac{\partial E}{\partial p} = -2\mathbf{e}^T \frac{\partial \mathbf{d}^{pre}}{\partial p} = -2\mathbf{e}^T \mathbf{G} \frac{\partial \mathbf{C}_{md}}{\partial p}\, \mathbf{u} + 2\mathbf{e}^T \mathbf{G}\mathbf{C}_{md}\mathbf{v} =$$

$$= -2\mathbf{e}^T \frac{\partial \mathbf{C}_{dd}}{\partial p}\, \mathbf{u} + 2\mathbf{e}^T \mathbf{C}_{dd}\mathbf{v} \equiv T_1 + T_2$$

The partial derivative of the error is:

$$\frac{\partial E}{\partial p} = -2\mathbf{e}^T \frac{\partial \mathbf{C}_{dd}}{\partial p}\, \mathbf{u} + 2\mathbf{e}^T \mathbf{C}_{dd}\mathbf{v} \quad \text{with} \quad \mathbf{A}\,\mathbf{v} = \frac{\partial \mathbf{C}_{dd}}{\partial p}\, \mathbf{u} \quad \text{and} \quad \mathbf{A}\mathbf{u} = \mathbf{d}^{obs}$$

In order to calculate $\partial E/\partial p$, one must solve two instances of an $N \times N$ system, both with the same matrix $\mathbf{A}$. An alternate formulation can be achieved by writing $\partial E/\partial p = T_1 + T_2$ with:

$$T_1 = -2\{\mathbf{e}\}^T \mathbf{G} \frac{\partial \mathbf{C}_{md}}{\partial p}\, \mathbf{A}^{-1}\mathbf{d}^{obs} = -2\left\{\mathbf{A}^{-1\mathrm{T}} \frac{\partial \mathbf{C}_{md}^T}{\partial p} \mathbf{G}^T \mathbf{e}\right\}^T \mathbf{d}^{obs} =$$

$$= -2\, \mathbf{b}^T\, \mathbf{d}^{obs} \quad \text{with} \quad \mathbf{A}\,\mathbf{b} = \frac{\partial \mathbf{C}_{md}^T}{\partial p} \mathbf{G}^T \mathbf{e} = \frac{\partial \mathbf{C}_{dd}}{\partial p} \mathbf{e}$$

and with:

$$T_2 = 2\{\mathbf{e}\}^T \mathbf{G}\mathbf{C}_{md}\mathbf{v} = 2\{\mathbf{A}^{-1\mathrm{T}} \mathbf{C}_{md}^T \mathbf{G}^T \mathbf{e}\}^T \left\{\frac{\partial \mathbf{C}_{dd}}{\partial p}\, \mathbf{u}\right\} =$$

$$= 2\mathbf{c}^T \left\{\frac{\partial \mathbf{C}_{dd}}{\partial p}\, \mathbf{u}\right\} \quad \text{with} \quad \mathbf{A}\,\mathbf{c} = \{\mathbf{C}_{md}^T \mathbf{G}^T \mathbf{e}\} = \mathbf{C}_{dd}\mathbf{e}$$

So, the alternate formulation becomes:

$$\frac{\partial E}{\partial p} = -2\, \mathbf{b}^T\, \mathbf{d}^{obs} + 2\mathbf{c}^T \frac{\partial \mathbf{C}_{dd}}{\partial p}\, \mathbf{u} \quad \text{with} \quad \mathbf{A}\mathbf{u} = \mathbf{d}^{obs} \text{ and } \mathbf{A}\,\mathbf{b} = \frac{\partial \mathbf{C}_{dd}}{\partial p} \mathbf{e} \quad \text{and} \quad \mathbf{A}\,\mathbf{c} = \mathbf{C}_{dd}\mathbf{e}$$

A conceptual advantage of this formulation is that the linear systems explicitly involve either the observations $\mathbf{d}^{obs}$ or the error $\mathbf{e}$ as "source terms" (that is, on their right-hand sides). However, this implementation requires three $N \times N$ linear systems to be solved (all with the same matrix **A)**.

Example. We consider true model parameters:

$$m(x) = \gamma \cos(px)$$

where $p$ is a wavenumber. The true autocorrelation function and its derivative are:

$$C_{ij}(p) = \gamma^2 \cos\{p(x_i - x_j)\} \quad \text{and} \quad \frac{\partial C_{ij}}{\partial p} = -\gamma^2 (x_i - x_j) \cos\{p(x_i - x_j)\}$$

We consider a test scenario with $M = 101$ on the interval $0 \le x \le 100$, and with $\gamma = 1$ and $p^{true} = 0.15708$. The $N = 40$ data $d_i$ are randomly drawn from the $m$s and are perturbed by Normally-distributed random noise with zero mean and variance $\sigma^2 = (0.05)^2$. The GPE estimate is computed with an incorrect wavenumber $p = 0.95\, p^{true}$ (Figure 1).

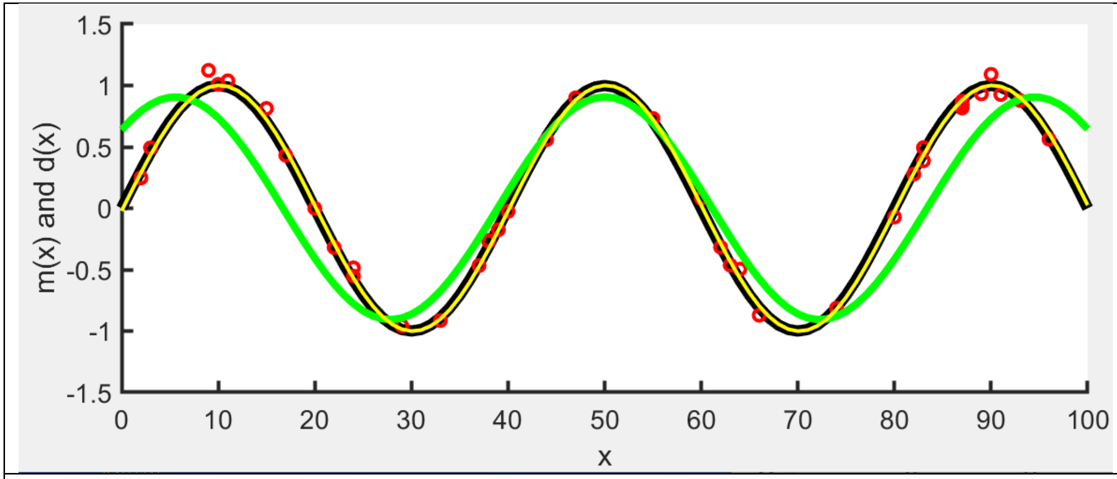

Fig. 1. True model parameters $\mathbf{m}$ (black curve) and data $\mathbf{d}$ (red circles). The GPE estimate with an incorrect wavenumber $p = 0.95\, p^{true}$ (green curve) fits the data poorly. Newton's method is used to iteratively improve $p$, leading to the improved fit (yellow curve).

Newton's method is used to iteratively improve $p$, using the derivative formula $\partial \mathbf{d}^{pre}/\partial p$. Its accuracy is verified against a finite-difference estimate (Figure 2).

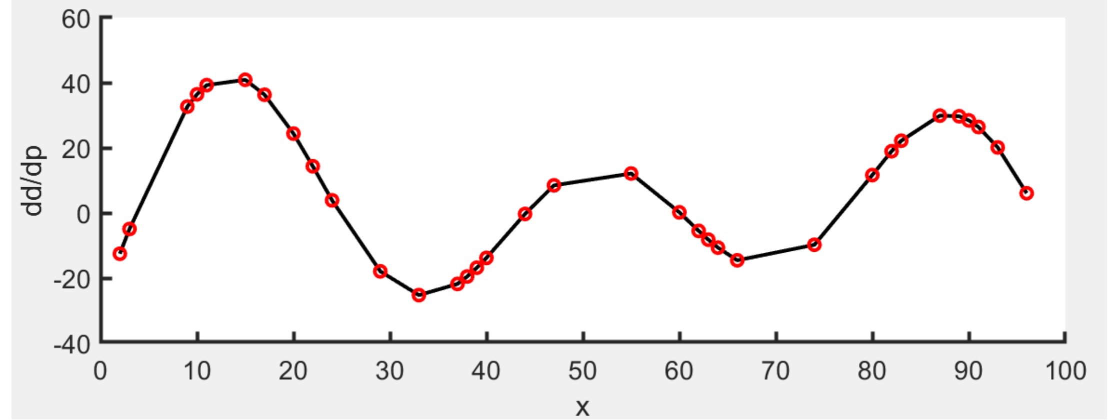

Fig. 2. The derivative $\partial \mathbf{d}^{pre}/\partial p$ (black curve), computed using the formula from this paper, compares favorably with the result of a finite difference calculation (red circles).

Newton's method converges rapidly with about three iterations (Figure 3).

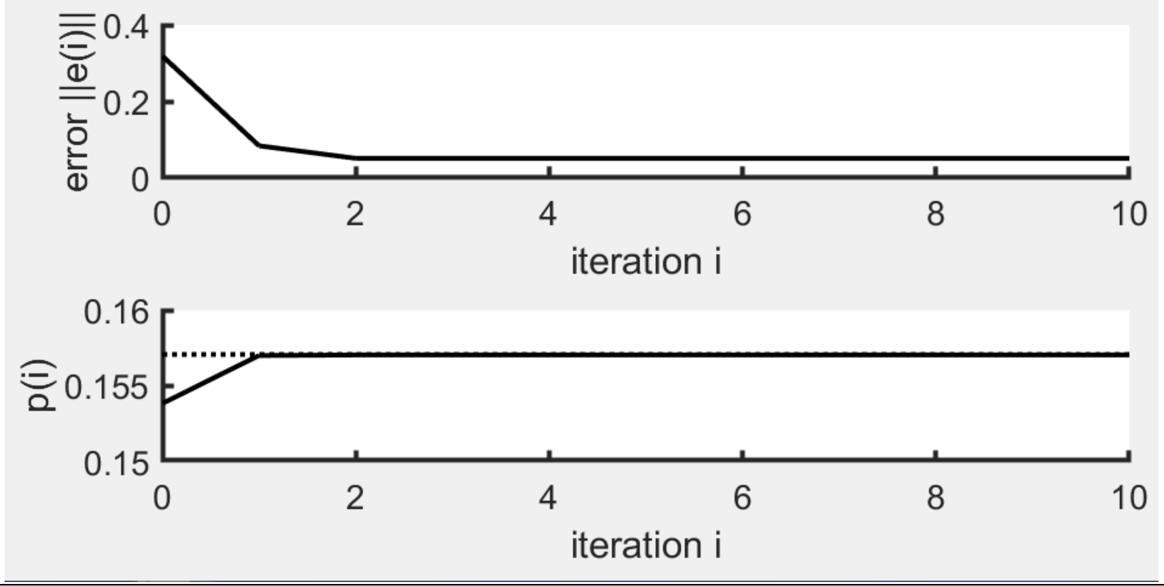

Fig. 3. Convergence of Newton's method. (A) The error $\|\mathbf{e}\| = (\mathbf{e}^T\mathbf{e}/N)$ where $\mathbf{e} = \mathbf{d}^{obs} - \mathbf{d}^{pre}$, as a function of iteration $i$. Note the error decreases to a constant level after about three iterations. (B) The estimated wavenumber $p$ (solid curve) as a function of iteration $i$, compared with the true wavenumber $p^{true}$ (dotted line).

Menke, W., Geophysical Data Analysis: Discrete Inverse Theory, Fourth Edition (Textbook), Elsevier, pp 350, 2018, ISBN: 9780128135556.