Variance Estimates from Iterative Non-linear Least Squares
Bill Menke, April 18, 2019

Consider a 'toy" non-linear inverse problem with data $\mathbf{d}^{\text{obs}}$ and th model parameters $\mathbf{m}$, of the form:

$$d_1 = G_{11}m_1 + G_{12}m_2 + Q_{111}m_1^2 + (Q_{112} + Q_{121})m_1 m_2 + Q_{122}m_2^2$$
$$d_2 = G_{21}m_1 + G_{22}m_2 + Q_{211}m_1^2 + (Q_{212} + Q_{221})m_1 m_2 + Q_{222}m_2^2$$

(1)

with

$$G = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad Q_{ijk} = -\delta_{i1}\delta_{j1}\delta_{k1}$$

(2)

where δ_{ij} is the Kronecker Delta. I set $\mathbf{m}^{\text{true}} = [0.452, 0.690\ ]^{\text{T}}$ and $\mathbf{d}^{\text{obs}} = \mathbf{d}^{\text{true}}$. The true p.d.f. $p(\mathbf{m}|\mathbf{d}^{\text{obs}})$ is taken to be Normal in the error $\mathbf{e} = \mathbf{d}^{\text{obs}} - \mathbf{d}^{\text{pre}}(\mathbf{m})$ with covariance:

$$\mathbf{C}_d = \begin{bmatrix} (0.02)^2 & 0 \\ 0 & (0.1)^2 \end{bmatrix}$$

(3)

This p.d.f. is shon in Figure 1. Its maximum likelihood point is at $\mathbf{m}^{\text{true}}$. It's mean (calculated numerically) is not conincident with the maximum likelihood point, but rather is displaced to the right of it at $\mathbf{m}^{\text{mean}} = [0.485, 0.678\ ]^{\text{T}}$, because the p.d.f. is skewed. The covariance (canculated numerically) is:

$$\mathbf{C}_m = \begin{bmatrix} (0.1133)^2 & -0.0014 \\ -0.0014 & (0.0311)^2 \end{bmatrix}$$

(4)

The inverse problem is solved with interative linearized Generalized Least Squares (e.g. Menke, 2018, Equation 9.27 with no prior information). The gradient:

$$G_{ij}^{(n)} = \left.\frac{\partial d_i}{\partial m_j}\right|_{\mathbf{m}^{(n)}}$$

(5)

(where n indexes iteration number) is calculated by analytcally differentiating Equation (1). The solution rapidly converges on $\mathbf{m}^{\text{true}}$ (Figure 1) in about $N = 20$ iterations. The solution $\mathbf{m}^{(\text{N})}$ closely approximates the the maximum likeligood point $\mathbf{m}^{\text{true}}$. The covariance is estimated as:

$$\mathbf{C}_m^{(N)} = \left[\mathbf{G}^{(N)T}\mathbf{C}_d^{-1}\mathbf{G}^{(N)}\right]^{-\mathbf{1}} = \begin{bmatrix} (0.0924)^2 & 0.0004 \\ 0.0004 & (0.0204)^2 \end{bmatrix} \tag{6}$$

The values are somewhat smaller than the true covariance (but of the same order of magnitude). A Normal p.d.f. based on $\mathbf{m}^{(\mathrm{N})}$ and $\mathbf{C}_m^{(N)}$ approximates the true non-Normal p.d.f. reasonably well (Figure 2).

Reference

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

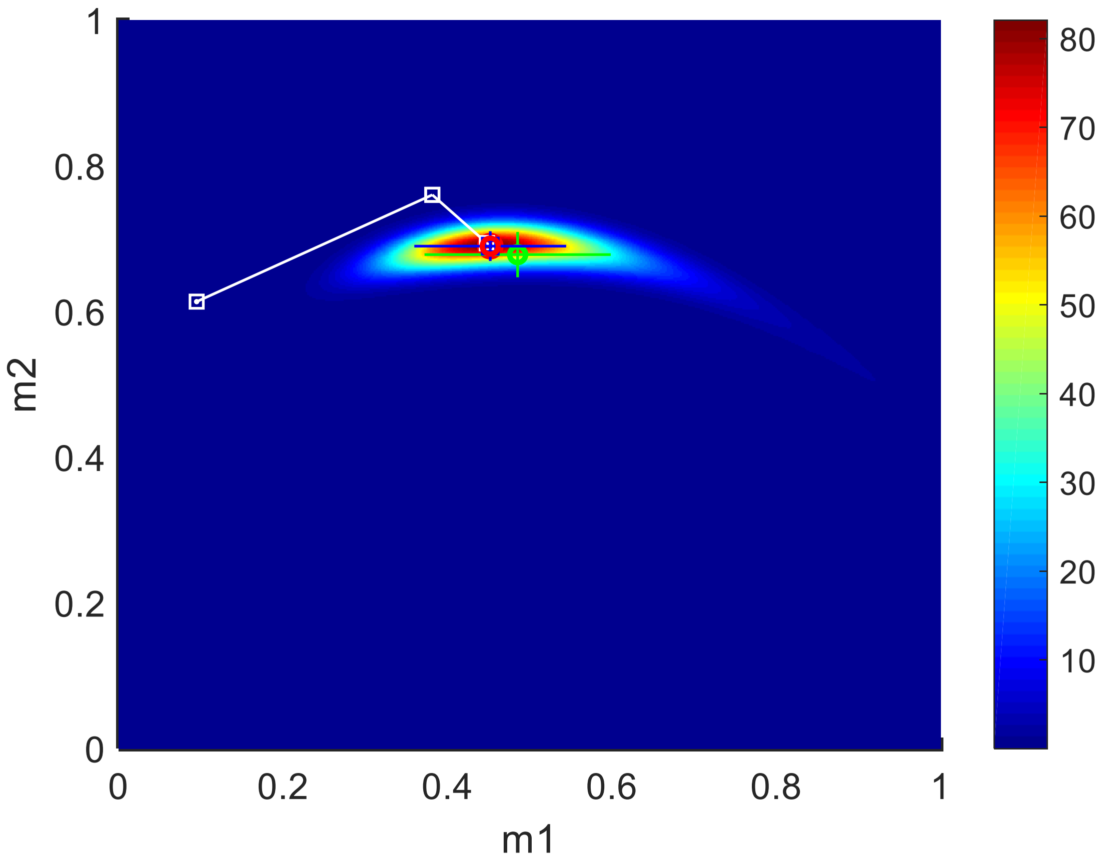

Figure 1. The non-Normal p.d.f. $p(\mathbf{m}|\mathbf{d}^{\mathrm{obs}})$ associated with the inverse problem (colors). The maximum likelihood point (red) and the true mean and standard deviation (green) are shown. The trajectory of a gradient-based iterative inversion (white) converges on the maximum likelihood point in a few iterations. The estimated maximum likelihood point (blue) and estimated standard deviation (blue) of the inversion are shown; they match the true values reasonably well

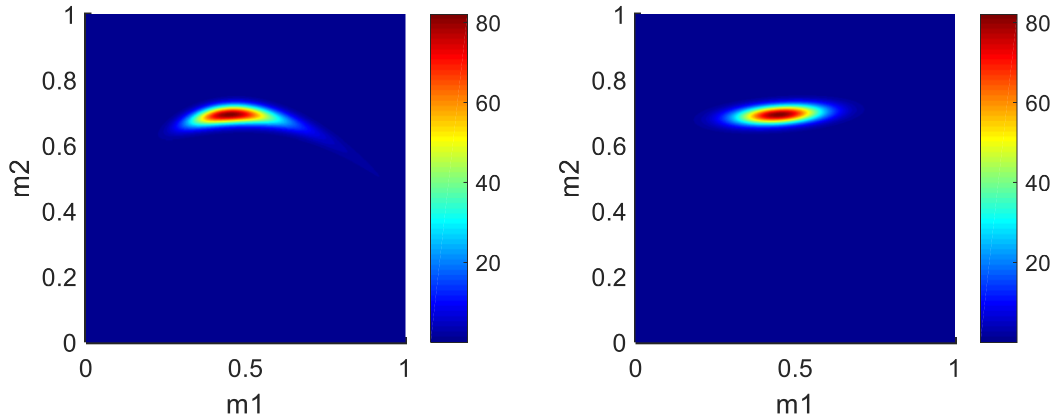

Figure 2. (A) The non-Normal p.d.f. $p\left(\mathbf{m}|\mathbf{d}^{\mathrm{obs}}\right)$ associated with the inverse problem (colors). (B) The Normal approximation for $p\left(\mathbf{m}|\mathbf{d}^{\mathrm{obs}}\right)$ computed from the maximum likelihood point and covariance matrix associated with the gradient-based iterative inversion.