Adjoint formulation for data kernel and error gradient when the data is a linear functional of the field

Bill Menke, December 26, 2020

Summary: We consider data d_i (at x_i) related to a field u by the linear functional $d_i = (h_i, u)$. Here the field u satisfies $\mathcal{U}u = f$. We show that the adjoint field λ_i part of the kernel calculation is the only part that depends on h_i, and that it does so through the adjoint field equation $\mathcal{U}^\dagger \lambda_i = h_i$. The adjoint equation and adjoint field are the same, irrespective of whether the model parameters m_j parameterize the differential operator $\mathcal{U}$ or the source of the field f.

For differential data, $h_i = -\frac{d}{dx}\delta(x - x_i)$; that is, the adjoint source is a dipole.

Consequently, a kernel $\mathbf{G}$ corresponding to differential data $\mathbf{d}$ can be calculated via a change to the adjoint source. This method is different from post-facto calculation of $\mathbf{G}$, where one starts with a kernel for the original data, $\mathbf{d}' = \mathbf{G}'\mathbf{m}$ and then multiplies by a differential operator $\mathbf{D}$ to get $\mathbf{d} = \mathbf{D}\mathbf{d}' = (\mathbf{D}\mathbf{G}')\mathbf{m} = \mathbf{G}\mathbf{m}$. I think that the direct method is less affected by round-off error.

Furthermore, the adjoint field calculation is the only the part of the calculation of the gradient of the total error that depends upon h_i. The adjoint equation and adjoint field are the same (though not the same as for the kernel), irrespective of whether the model parameters m_j parameterize the differential operator $\mathcal{U}$ or the source of the field f.

(Part 1) Kernel when source depends on a model parameters m_j

Data equation $d_i = (h_i, u)$ and its derivative $\frac{\mathrm{d}d_i}{\mathrm{d}m_j} = \left(h_i, \frac{\mathrm{d}u}{\mathrm{d}m_j}\right)$

Kernel equation $\mathrm{d}d_i = \sum_j \frac{\mathrm{d}d_i}{\mathrm{d}m_j}\,\mathrm{d}m_j$ with kernel $G_{ij} \equiv \frac{\mathrm{d}d_i}{\mathrm{d}m_j}$

Field eqn: $\mathcal{U}\,u(m) = f(m)$ and its solution $u(m) = \mathcal{U}^{-1} f(m)$

Field derivative: $\frac{\mathrm{d}u}{\mathrm{d}m_j} = \mathcal{U}^{-1}\frac{\mathrm{d}f}{\mathrm{d}m_j}$

Derivation of kernel:

$$\frac{\mathrm{d}d_i}{\mathrm{d}m_j} = \left(h_i, \frac{\mathrm{d}u}{\mathrm{d}m_j}\right) = \left(h_i, \mathcal{U}^{-1}\frac{\mathrm{d}f}{\mathrm{d}m_j}\right) = \left(\mathcal{U}^{-1\dagger}h_i, \frac{\mathrm{d}f}{\mathrm{d}m_j}\right) = \left(\lambda_i, \frac{\mathrm{d}f}{\mathrm{d}m_j}\right) \quad \text{with} \quad \mathcal{U}^\dagger \lambda_i = h_i$$

so the kernel is $G_{ij} = \left(\lambda_i, \frac{\mathrm{d}f}{\mathrm{d}m_j}\right)$

suppose $f = \sum_k m_k \delta(x - \xi_k)$ so $\frac{\mathrm{d}f}{\mathrm{d}m_j} = \delta(x - \xi_j)$

then $G_{ij} = \left(\lambda_i, \delta(x - \xi_j)\ \right) = \lambda_i(\xi_j)$

Differential data: $d_i = (du/dx)|_{x_i} = \left(-\frac{d}{dx}\delta(x - x_i), u\right)$ so $h_i = -\frac{d}{dx}\delta(x - x_i)$

and adjoint equation is $\mathcal{U}^{\dagger}\lambda_i = -\frac{d}{dx}\delta(x - x_i)$

In the case where $d_i = W(x)(du/dx)|_{x_i}$, one could set $h_i = -W(x)\frac{d}{dx}\delta(x - x_i) - \frac{dW}{dx}\delta(x - x_i)$ since then $d_i = \left[\frac{d}{dx}(Wu) - \frac{dW}{dx}u\right]\Big|_{x_i} = \left[W\frac{du}{dx}\right]\Big|_{x_i}$

(Part 2) Kernel when differential operator depends on parameters m_j

Data eqn: $d_i = (h_i, u)$ and its derivative $\frac{\mathrm{d}d_i}{\mathrm{d}m_j} = \left(h_i, \frac{\mathrm{d}u}{\mathrm{d}m_j}\right)$

Field eqn: $\mathcal{U}(m)\, u(m) = f$ so $u = \mathcal{U}^{-1} f$

Field derivative $\mathrm{d}u/\mathrm{d}m_j = \left(\mathrm{d}\mathcal{U}^{-1}/\mathrm{d}m_j\right) f = -\mathcal{U}^{-1}\left(\mathrm{d}\mathcal{U}/\mathrm{d}m_j\right) \mathcal{U}^{-1} f$

$$\frac{\mathrm{d}u}{\mathrm{d}m_j} = -\mathcal{U}^{-1}\left\{\left(\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}m_j}\right) u\right\}$$

Derivation of kernel:

$$\frac{\mathrm{d}d_i}{\mathrm{d}m_j} = \left(h_i, \frac{\mathrm{d}u}{\mathrm{d}m_j}\right) = -\left(h_i, \mathcal{U}^{-1}\left\{\left(\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}m_j}\right) u\right\}\right) =$$

$$= -\left(\mathcal{U}^{-1\dagger} h_i, \left\{\left(\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}m_j}\right) u\right\}\right) = -\left(\lambda_i, \left\{\left(\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}m_j}\right) u\right\}\right) \text{ with } \mathcal{U}^{\dagger}\lambda_i = h_i$$

suppose $\mathcal{U} = \sum_k m_k \delta(x - \xi_k)\frac{d}{dt^2}$ + (terms not depending upon m_k)

then $\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}m_j} = \delta\left(x - \xi_j\right)\frac{d}{dt^2}$ and $\left(\frac{\delta \mathcal{U}}{\delta m_j}\right) u = \ddot{u}\delta\left(x - \xi_j\right)$

$G_{ij} = \left(\lambda_i, \ddot{u}\delta\left(x - \xi_j\right)\right) = \lambda_i\left(\xi_j, t\right) \star \ddot{u}\left(\xi_j, t\right)$ (with $\star$ signifying time correlation)

Differential data: $d_i = (d/dx)u|_{x_i} = \left(-\frac{d}{dx}\delta(x - x_i), u\right)$ so $h_i = -\frac{d}{dx}\delta(x - x_i)$

and adjoint equation is $\mathcal{U}^{\dagger}\lambda_i = \frac{d}{dx}\delta(x - x_i)$ (same as in Part 1).

(Part 3) Equivalence in Part 1 to Green function integral

Let's write $\frac{\mathrm{d}u}{\mathrm{d}m_j} = \mathcal{U}^{-1}\frac{\mathrm{d}f}{\mathrm{d}m_j}$ as $v = \mathcal{U}^{-1}\varphi$ for short

$\mathcal{U}\, v(x) = \varphi(x)$

Standard Green function setup

$v(x_i) = \mathcal{U}_{x_i}^{-1}\varphi(x_i) = \int_x K(x_i, x)\varphi(x)\,\mathrm{d}x \equiv \left(K(x_i, x), \varphi(x)\right)_x$

With Green function satisfying $\mathcal{U}_{x_i} K(x_i, x) = \delta(x_i - x)$ (datum at x_i, source at x)

Adjoint method

$$v(x_i) = \left(\delta(x - x_i), v(x)\right)_x = \left(\delta(x - x_i), \mathcal{U}_x^{-1}\varphi(x)\right)_x = \left(\mathcal{U}_x^{-1\dagger}\delta(x - x_i), \varphi(x)\right)_x =$$

$$= \left(\lambda(x, x_i), \varphi(x)\right)_x \quad \text{with} \quad \mathcal{U}_x^{\dagger}\,\lambda(x, x_i) = \delta(x - x_i)$$

interchanging x with x_i in the differential equation $\mathcal{U}_{x_i}^{\dagger}\,\lambda(x_i, x) = \delta(x_i - x)$

Comparing inner products, we see that order of arguments of λ and K are reversed, so $\lambda = K^\dagger$
Comparing the differential equations, we also see that $\lambda = K^\dagger$ (see Part 4)
Thus, the result in Part 1 of $G_{ij} = (\lambda_i, \varphi)$ is just a Green function integral

(Part 4) Standard proof for relationship between Green function of original and adjoint equations

$(u, \mathcal{U}v) = (\mathcal{U}^\dagger u, v)$ with $u = G^\dagger(x,z)$ and $v = G(x,y)$

and $\mathcal{U}^\dagger G^\dagger(x,z) = \delta(x-z)$ and $\mathcal{U}G(x,y) = \delta(x-y)$

$\left(G^\dagger(x,z), \delta(x-y)\right) = (\delta(x-z), G(x,y)) = 0$

$G^\dagger(y,z) = G(z,y)$

(Part 5) Gradient of total error, following Part 1.

$E = (e,e)$ so $\frac{\mathrm{d}E}{\mathrm{d}m_j} = 2\left(e, \frac{\mathrm{d}e}{\mathrm{d}m_j}\right)$

$e = d^{obs} - d$ so $\frac{\mathrm{d}e}{\mathrm{d}m_j} = -\frac{\mathrm{d}d}{\mathrm{d}m_j}$

$$-½\frac{\mathrm{d}E}{\mathrm{d}m_j} = \left(e, \frac{\mathrm{d}d}{\mathrm{d}m_j}\right)_{x_i} = \left(e, \left(h(x_i,x), \mathcal{U}_x^{-1}\frac{\mathrm{d}f}{\mathrm{d}m_j}\right)_x\right)_{x_i}$$

using rule $\left(a(x_i), \left(B(x_i,x), c(x)\right)_x\right)_{x_i} = \left(\left(a(x_i), B^\dagger(x,x_i)\right)_{x_i}, c(x)\right)_x$ (see Part 6)

$$-½\frac{\mathrm{d}E}{\mathrm{d}m_j} = \left(\left(h^\dagger(x,x_i), e(x_i)\right)_{x_i}, \mathcal{U}_x^{-1}\frac{\mathrm{d}f}{\mathrm{d}m_j}\right) =$$

$$= \left(\mathcal{U}_x^{-1\dagger}\left(e, h^\dagger(x_i,x)\right)_{x_i}, \frac{\mathrm{d}f}{\mathrm{d}m_j}\right)_x = \left(\lambda, \frac{\mathrm{d}f}{\mathrm{d}m_j}\right)_x \quad \text{with} \quad \mathcal{U}_x^\dagger\lambda = \left(e, h^\dagger(x_i,x)\right)_{x_i}$$

$$\frac{\mathrm{d}E}{\mathrm{d}m_j} = -2\left(\lambda, \frac{\mathrm{d}f}{\mathrm{d}m_j}\right)_x$$

(Part 6) Standard proof of rule

$$\left(a(x_i), \left(B(x_i,x), c(x)\right)_x\right)_{x_i} = \left(\left(B^\dagger(x,x_i), a(x_i)\right)_{x_i}, c(x)\right)_x$$

$$\int_{x_i} a(x_i)\left(\int_x B(x_i,x)c(x)dx\right)dx_i = \int_x\left(\int_{x_i} B(x_i,x)a(x_i)dx_i\right)c(x)\,dx =$$

$= \int_x\left(\int_{x_i} B^\dagger(x,x_i)\,a(x_i)dx_i\right)c(x)\,dx$ with $B^\dagger(x,x_i) = B(x_i,x)$

(Part 7) Gradient of total error, following Part 2.

$$-½\frac{\mathrm{d}E}{\mathrm{d}m_j} = \left(e, \frac{\mathrm{d}d}{\mathrm{d}m_j}\right)_{x_i} = \left(e, \left(h(x_i,x), \mathcal{U}^{-1}\left\{\left(\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}m_j}\right)u\right\}\right)_x\right)_{x_i} =$$

$$= \left(\mathcal{U}^{-1\dagger} \left(e, h^{\dagger}(x_i, x) \right)_{x_i}, \left\{ \left(\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}m_j} \right) u \right\} \right)_x =$$

$$= \left(\lambda, \left\{ \left(\frac{\mathrm{d}\mathcal{U}}{\mathrm{d}m_j} \right) u \right\} \right)_x \quad \text{with} \quad \mathcal{U}_x^{\dagger} \lambda = \left(e, h^{\dagger}(x_i, x) \right)_{x_i}$$