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

Furthermore, the adjoint field calculation is the only 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 \( \frac{d d_i}{d m_j} = (h_i, \frac{d u}{d m_j}) \)

Kernel equation \( d d_i = \sum_j \frac{d d_i}{d m_j} d m_j \) with kernel \( G_{ij} \equiv \frac{d d_i}{d m_j} \)

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

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

Derivation of kernel:
\[
\frac{d d_i}{d m_j} = \left( h_i, \frac{d u}{d m_j} \right) = \left( h_i, \mathcal{U}^{-1} \frac{d f}{d m_j} \right) = \left( U^\dagger h_i, \frac{d f}{d m_j} \right) = \left( \lambda_i, \frac{d f}{d m_j} \right)
\]

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

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

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 \( U^\dagger \lambda_i = -\frac{d}{dx} \delta(x - x_i) \)

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 \( d_i = \left[ \frac{d}{dx}(Wu) - \frac{dw}{dx} u \right]_{x_i} = \left[ W \frac{du}{dx} \right]_{x_i} \)

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

Data eqn: \( d_i = (h_i, u) \) and its derivative \( \frac{dd_i}{dm_j} = (h_i, \frac{du}{dm_j}) \)

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

Field derivative \( \frac{du}{dm_j} = (dU^{-1}/dm_j) f = -U^{-1}(dU/dm_j) U^{-1} f \)

\[
\frac{du}{dm_j} = -U^{-1} \left\{ \left( \frac{dU}{dm_j} \right) u \right\}
\]

Derivation of kernel:

\[
\frac{dd_i}{dm_j} = \left( h_i, \frac{du}{dm_j} \right) = \left( h_i, U^{-1} \left\{ \left( \frac{dU}{dm_j} \right) u \right\} \right) = \left( -U^{-1} h_i, \left\{ \left( \frac{du}{dm_j} \right) u \right\} \right) = \left( \lambda_i, \left\{ \left( \frac{du}{dm_j} \right) u \right\} \right) \text{ with } U^\dagger \lambda_i = h_i
\]

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

then \( \frac{du}{dm_j} = \delta(x - \xi_j) \frac{d}{dt} \) and \( \left( \frac{dU}{dm_j} \right) u = \ddot{u} \delta(x - \xi_j) \)

\( g_{ij} = \left( \lambda_i \ddot{u} \delta(x - \xi_j) \right) = \lambda_i (\xi_j, t) \star \ddot{u}(\xi_j, t) \) (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 \( U^\dagger \lambda_i = \frac{d}{dx} \delta(x - x_i) \) (same as in Part 1).

**Part 3** Equivalence in Part 1 to Green function integral

Let’s write \( \frac{du}{dm_j} = U^{-1} \frac{df}{dm_j} \) as \( v = U^{-1} \varphi \) for short

\( U \varphi(x) = \varphi(x) \)

Standard Green function setup

\( v(x_i) = U_i^{-1} \varphi(x_i) = \int_x K(x_i,x) \varphi(x) \ dx \equiv \left( K(x_i,x), \varphi(x) \right)_x \)

With Green function satisfying \( U_{x_i} K(x_i,x) = \delta(x_i - x) \) (datum at \( x_i \), source at \( x \))

Adjoint method

\( v(x_i) = (\delta(x - x_i), v(x))_x = (\delta(x - x_i), U_i^{-1} \varphi(x))_x = \left( U_i^{-1} \delta(x - x_i), \varphi(x) \right)_x = \left( \lambda(x,x_i), \varphi(x) \right)_x \)

with \( U_i^\dagger \lambda(x,x_i) = \delta(x - x_i) \)

interchanging \( x \) with \( x_i \) in the differential equation \( U_i^\dagger \lambda(x_i,x) = \delta(x_i - x) \)
Comparing inner products, we see that order of arguments of $\lambda$ 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 $\frac{dE}{dm_j} = 2 \left( e, \frac{de}{dm_j} \right)$

$e = d^{obs} - d$ so $\frac{de}{dm_j} = -\frac{dd}{dm_j}$

$-\frac{1}{2} \frac{dE}{dm_j} = \left( e, \frac{dd}{dm_j} \right) = \left( e, \left( h(x_i, x) \mathcal{U}^{-1} \frac{df}{dm_j} \right) \right)$

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

$-\frac{1}{2} \frac{dE}{dm_j} = \left( (h^\dagger(x, x_i), e(x_i)) \bigg| x \right, \mathcal{U}^{-1} \frac{df}{dm_j} \bigg| x = \left( \lambda, \frac{df}{dm_j} \bigg| x \right)$

with $\mathcal{U}^\dagger \lambda = \left( e, h^\dagger(x_i, x) \bigg| x \right)$

$\frac{dE}{dm_j} = -2 \left( \lambda, \frac{df}{dm_j} \bigg| x \right)$

(Part 6) Standard proof of rule

$\left( a(x_i), (B(x_i, x), c(x)) \bigg| x \right) = \left( (B^\dagger(x, x_i), a(x_i)) \bigg| x \right, c(x))$

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

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

$-\frac{1}{2} \frac{dE}{dm_j} = \left( e, \frac{dd}{dm_j} \bigg| x \right) = \left( e, \left( h(x_i, x) \mathcal{U}^{-1} \left\{ \left( \frac{dU}{dm_j} \bigg| \bigg| u \right) \right\} \bigg| x \bigg| x \right) = $
\[
\begin{aligned}
&= \left( u^{-1+} (e, h^+(x_i, x))_{x_i} \right)_{x_i} \left( \left\{ \frac{d \mathcal{U}}{d m_j} \right\}_x \right)_{x_i} = \\
&= \left( \lambda, \left\{ \frac{d \mathcal{U}}{d m_j} \right\}_x \right)_x \text{ with } U^+_{x} \lambda = (e, h^+(x_i, x))_{x_i}
\end{aligned}
\]