Sensitivity Kernel for an Uninvertible Operator

Bill Menke, July 8, 2022

We start with the usual formulation of an inverse problem:

$$\text{minimize } E = (u-d, u-d) \text{ with the constraint } \mathcal{L}u - f = 0$$

and entertain the possibility that $\mathcal{L}$ is not invertble. The Lagrange equation is:

$$(u-d,u-d) + 2(\lambda, \mathcal{L}u - f) = (u-d,u-d) + 2(\lambda, \mathcal{L}u) - 2(\lambda, f)$$

where 2λ is the Lagrange multiplier. Fréchet differentiating with respect to λ and setting the result to zero returns the constraint differential, $\mathcal{L}u - f = 0$. In order to Fréchet differentiate with respect to u we must rewrite the equation as

$$(d-u,d-u) + 2(\mathcal{L}^\dagger\lambda, u) - 2(\lambda, f)$$

Fréchet differentiating yields

$$u = d - \mathcal{L}^\dagger\lambda$$

Here, we have used $\delta p(x)/\delta p(x') = \delta(x-x')$. Inserting this expression into the differential equation eliminates u, yielding

$$0 = \mathcal{L}u - f = \mathcal{L}[d - \mathcal{L}^\dagger\lambda] - f = \mathcal{L}d - \mathcal{L}\mathcal{L}^\dagger - f$$

$$\text{so} \quad \mathcal{L}\mathcal{L}^\dagger = \mathcal{L}d - f$$

$$\mathcal{L}\mathcal{L}^\dagger\lambda = \mathcal{L}d - f = \mathcal{L}(e+u) - f = \mathcal{L}e + (\mathcal{L}u - f) = \mathcal{L}e$$

$$\text{so} \quad \mathcal{L}\mathcal{L}^\dagger\lambda = \mathcal{L}e$$

This is a differential equation for the Lagrange multiplier, λ. However, it is differetnt than the one commonly encountered in the invertible case, which is $\mathcal{L}^\dagger\lambda = e$. After $\mathcal{L}\mathcal{L}^\dagger\lambda = \mathcal{L}e$ is solved, $u = d - \mathcal{L}^\dagger\lambda$ gives the solution. Note that when $\mathcal{L}^{-1}$ exists, $u = d - \mathcal{L}^\dagger[[\mathcal{L}\mathcal{L}^\dagger]^{-1}(\mathcal{L}d - f)] = \mathcal{L}^{-1}f$. That is, the minimization only adds information when $\mathcal{L}$ is uninvertible,

As an example, suppose that the problem is

$$\text{find the } \mathbf{u} \text{ that minimizes } (\mathbf{u}-\mathbf{d}, \mathbf{u}-\mathbf{d}) \text{ subject to the constraint } \nabla\cdot\mathbf{u} - f = 0$$

Note that the constraint does not fully specify $\mathbf{u}$, because $\nabla\times\mathbf{v}$, for any function, $\mathbf{v}$, can be added to $\mathbf{u}$ without violating it. Then, $\mathcal{L} = \nabla\,\cdot$, $\mathcal{L}^\dagger = -\nabla$, $\mathcal{L}\mathcal{L}^\dagger = -\nabla^2$ and λ solves Poisson's equation $\nabla^2\lambda = \nabla\cdot\mathbf{d} - \mathbf{f}$. Poisson's equation is invertible, and with a well-known Green function, so we can uniquely solve for λ. Once λ is determined, the solution is $\mathbf{u} = \mathbf{d} + \nabla\lambda$. Furthermore, $\nabla\times\mathbf{u} = \nabla\times\mathbf{d}$; that is, information on the curl of $\mathbf{u}$ arise from $\mathbf{d}$, alone, and not involve the constraint.

I haven't pursued the matter in any detail, but it seems to me that solving the adjoint equation when the error, e, is known at only are points might be problematical, because $\mathcal{L}e$ would seem to require derivatives of e.

The Implicit Function Theorem (I.F.T.) can be used to calculate $\partial\mathbf{u}/\partial p$. Let me fall back to the discrete problem:

$$(\mathbf{d}-\mathbf{u})^T(\mathbf{d}-\mathbf{u})+2(\mathbf{L}\mathbf{u}-\mathbf{f})^T\boldsymbol{\lambda} \quad \text{with} \quad \mathbf{L}(p) \quad \text{and fixed } p$$

$$\frac{\partial}{\partial u_i} \text{ gives } N \text{ equations}, h_i^A=0\text{: } \mathbf{d}-\mathbf{u}+\mathbf{L}^T\boldsymbol{\lambda}=\mathbf{0}$$

$$\frac{\partial}{\partial \lambda_i} \text{ gives } N \text{ equation }, h_i^B=0\text{: } \mathbf{L}\mathbf{u}-\mathbf{f}=\mathbf{0}$$

Here, the functions h_i^A and h_i^B are abbreviations for the left-hand-side of the Lagrange derivatives. The unknowns are $\mathbf{u}, \boldsymbol{\lambda}, p$, and the number of unknowns is $2N+1$. The number of equations is $2N+1$. Jacobian derivatives are:

$$\frac{\partial h_i^A}{\partial u_j}=-\delta_{ij} \quad \text{and} \quad \frac{\partial h_i^A}{\partial \lambda_j}=L_{ji} \quad \text{and} \quad \frac{\partial h_i^A}{\partial p}=\lambda_j\frac{\partial L_{ji}}{\partial p}$$

$$\frac{\partial h_i^B}{\partial u_j}=L_{ij} \quad \text{and} \quad \frac{\partial h_i^B}{\partial \lambda_j}=0 \quad \text{and} \quad \frac{\partial h_i^B}{\partial p}=\frac{\partial L_{ij}}{\partial p}u_j$$

We ill set up the I.F.T. with independent variable is x_1: p and dependent variables $y_1 \cdots y_{2N}$: $\mathbf{u}, \boldsymbol{\lambda}$. The two parts of the Jacobian are:

$$\mathbf{J_y}=\begin{bmatrix} \dfrac{\partial h_i^A}{\partial u_j} & \dfrac{\partial h_i^A}{\partial \lambda_j} \\ \dfrac{\partial h_i^B}{\partial u_j} & \dfrac{\partial h_i^B}{\partial u_j} \end{bmatrix}=\begin{bmatrix} -\mathbf{I} & \mathbf{L}^T \\ \mathbf{L} & \mathbf{0} \end{bmatrix} \quad \text{so} \quad \mathbf{J}_y^{-1}=\begin{bmatrix} \mathbf{0} & \mathbf{L}^T[\mathbf{L}\mathbf{L}^T]^{-1} \\ \mathbf{L}[\mathbf{L}^T\mathbf{L}]^{-1} & -[\mathbf{L}\mathbf{L}^T]^{-1} \end{bmatrix}$$

$$\mathbf{J_x}=\begin{bmatrix} \dfrac{\partial h_i^A}{\partial p} \\ \dfrac{\partial h_i^B}{\partial p} \end{bmatrix}=\begin{bmatrix} \boldsymbol{\lambda}^T\dfrac{\partial \mathbf{L}}{\partial p} \\ \dfrac{\partial \mathbf{L}}{\partial p}\mathbf{u} \end{bmatrix}$$

Finally, according to the implicit function theorem, for a $p_0, \mathbf{u}(p_0), \boldsymbol{\lambda}(p_0)$ for which the Lagrange derivatives are zero:

$$\begin{bmatrix} \dfrac{\partial \mathbf{u}}{\partial p} \\ \dfrac{\partial \boldsymbol{\lambda}}{\partial p} \end{bmatrix}_{p_0,\mathbf{u}(p_0),\boldsymbol{\lambda}(p_0)}=-\mathbf{J}_y^{-1}\mathbf{J_x}=-\begin{bmatrix} \mathbf{L}^T[\mathbf{L}\mathbf{L}^T]^{-1}\dfrac{\partial \mathbf{L}}{\partial p}\mathbf{u} \\ \mathbf{L}[\mathbf{L}^T\mathbf{L}]^{-1}\boldsymbol{\lambda}^T\dfrac{\partial \mathbf{L}}{\partial p}-[\mathbf{L}\mathbf{L}^T]^{-1}\dfrac{\partial \mathbf{L}}{\partial p}\mathbf{u} \end{bmatrix}_{p_0,\mathbf{u}(p_0),\boldsymbol{\lambda}(p_0)}$$

$$\text{so}\ \left.\frac{\partial \mathbf{u}}{\partial p}\right|_{p_0,\mathbf{u}(p_0),\boldsymbol{\lambda}(p_0)} = -\,\mathbf{L}^T[\mathbf{L}\mathbf{L}^T]^{-1}\left.\frac{\partial \mathbf{L}}{\partial p}\mathbf{u}\right|_{p_0,\mathbf{u}(p_0),\boldsymbol{\lambda}(p_0)}$$

For an invertible matrix, $\mathbf{L}$, $\partial\mathbf{u}/\partial p = -\mathbf{L}^{-1}\frac{\partial \mathbf{L}}{\partial p}\mathbf{u}$, which is the result obtained by differentiating $\mathbf{u} = \mathbf{L}^{-1}\mathbf{f}$.:

$$\frac{\partial \mathbf{u}}{\partial p} = \frac{\partial}{\partial p}(\mathbf{L}^{-1}\mathbf{f}) = \frac{\partial \mathbf{L}^{-1}}{\partial p}\mathbf{f} = -\mathbf{L}^{-1}\frac{\partial \mathbf{L}}{\partial p}\mathbf{L}^{-1}\mathbf{f} = -\mathbf{L}^{-1}\frac{\partial \mathbf{L}}{\partial p}\mathbf{u}$$

Finally

$$\left.\frac{\partial \mathrm{E}}{\partial p}\right|_{p_0,\mathbf{u}(p_0),\boldsymbol{\lambda}(p_0)} = -\mathbf{e}^T\,\partial\mathbf{u}/\partial p = \mathbf{e}^T\mathbf{L}^{-1}\frac{\partial \mathbf{L}}{\partial p}\mathbf{u} = (\mathbf{L}^{T-1}\mathbf{e})^T\frac{\partial \mathbf{L}}{\partial p}\mathbf{u} =$$

$$\boldsymbol{\lambda}(p_0)^T\left.\frac{\partial \mathbf{L}}{\partial p}\right|_{p_0}\mathbf{u}(p_0) \quad \text{with} \quad \mathbf{L}(p_0)^T\boldsymbol{\lambda}(p_0) = \mathbf{e}(p_0)$$

So, it its established that, in the formulas, **u and λ** must solve the Lagrange conditions for the p under consideration. That is, they must minimize the error for this fixed value of p.

I was hoping to find a way to calculate $\partial\mathrm{E}/\partial p$ directly, without using the chain rule, but don't see a pathway to doing it.

As an ancillary matter, the derivation also shows that in cases in which **L, and therefore L**^{T} $\mathbf{L}^T$ is not invertible,

$$\left.\frac{\partial \mathrm{E}}{\partial p}\right|_{p_0,\mathbf{u}(p_0),\boldsymbol{\lambda}(p_0)} = -\mathbf{e}^T\,\partial\mathbf{u}/\partial p = \mathbf{e}^T\mathbf{L}^T[\mathbf{L}\mathbf{L}^T]^{-1}\frac{\partial \mathbf{L}}{\partial p}\mathbf{u}\frac{\partial \mathbf{L}}{\partial p}\mathbf{u} = ([\mathbf{L}\mathbf{L}^T]^{-1}\mathbf{L}\mathbf{e})^T\frac{\partial \mathbf{L}}{\partial p}\mathbf{u}$$

$$= \boldsymbol{\lambda}^T\frac{\partial \mathbf{L}}{\partial p}\mathbf{u} \quad \text{with} \quad [\mathbf{L}\mathbf{L}^T]^{-1}\boldsymbol{\lambda} = \mathbf{L}\mathbf{e}$$

That is, the adjoint equation is different than in the invertible case.