Fast Jacobian and Transpose of Jacobian Operation for EIT (and other inverse problems)

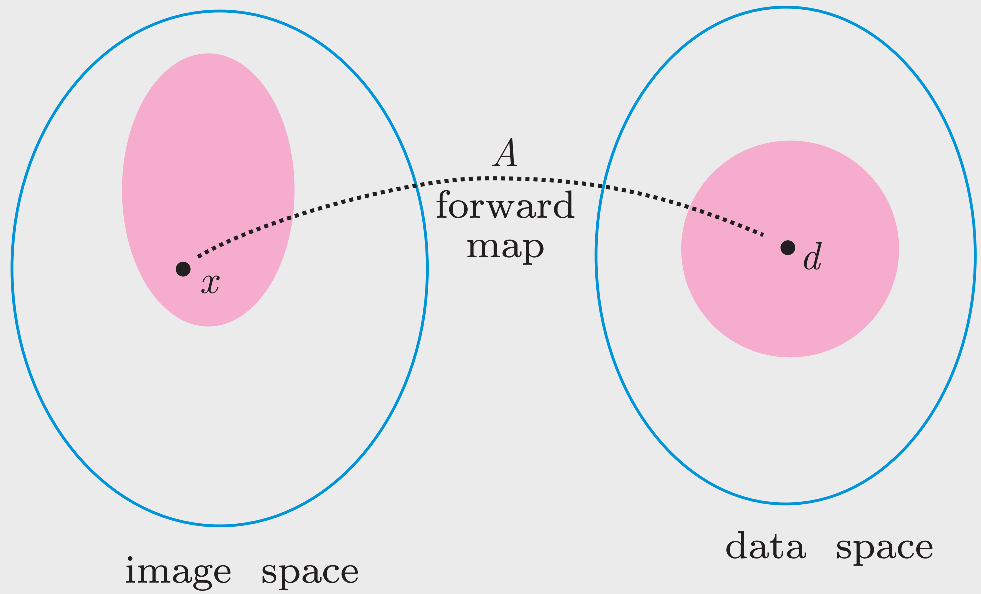
Colin Fox http://www.math.auckland.ac.nz/∼fox

Outline

- Why worry about the Jacobian and its transpose?
- Simplest example – symmetric matrix equation
- The real deal – FEM discretization of complete electrode model

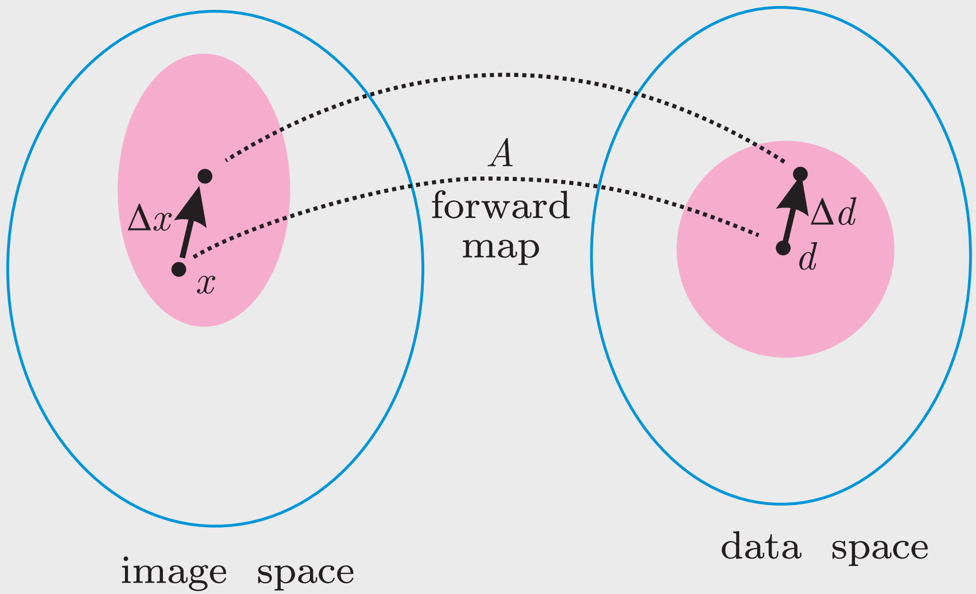
Jacobian and Jacobian Transpose

$d = Ax$: data d, image x, forward map$A$



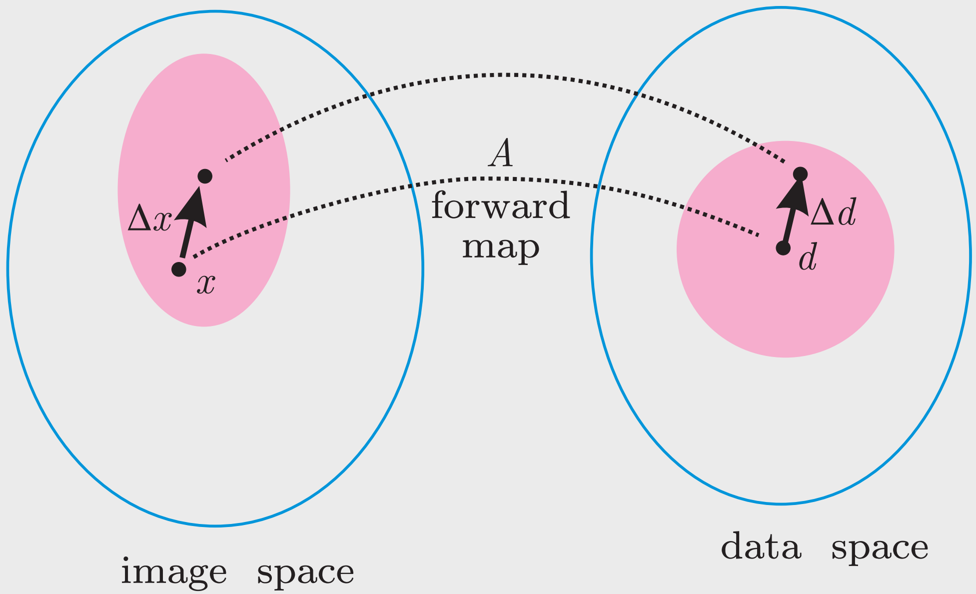
Jacobian and Jacobian Transpose

$d = Ax$: data d, image x, forward map$A$



Jacobian and Jacobian Transpose

$d = Ax$: data d, image x, forward map$A$



derivatives, gradients map as

$$\Delta d = J \Delta x \qquad \text{and} \qquad \nabla_x = J^{\mathsf{T}} \nabla_d$$

where the Jacobian is

$$J_{ij}(x) = \frac{\partial A_i}{\partial x_j}(x)$$

Least Squares

$$\hat{x} = \arg\min_{x} \|d_{\mathsf{m}} - A(x)\|_2^2$$

gradient-based optimization algorithms (quasi-Newton, conjugate gradients) use

$$\nabla_x \|d_{\mathsf{m}} - A(x)\|_2^2 = 2J^{\mathsf{T}}(d_{\mathsf{m}} - A(x))$$

Linearization

$$A(x + \Delta x) = A(x) + J\Delta x + O\left(\|x\|^2\right)$$

Gauss-Newton approximation

$$\nabla\nabla \|d_{\mathsf{m}} - A(x)\|_2^2 \approx 2J^{\mathsf{T}}J$$

Speeding-Up MCMC Sampling from $f(\cdot)$

1. At $x^{(t)}$ generate proposal y from $q(\cdot \mid x^{(t)})$
2. Let
$$g(x,y) = \min\left\{1, \frac{q(x \mid y)}{q(y \mid x)}\frac{f_x^*(y)}{f_x^*(x)}\right\}$$
W.p. $g(x^{(t)}, y)$, "promote" y. New proposal distribution is
$$q^*(y \mid x) = g(x,y)q(y \mid x) + (1 - r(x))\delta_x(y)$$
3. Let
$$\rho(x,y) = \min\left\{1, \frac{q^*(x \mid y)}{q^*(y \mid x)}\frac{f(y)}{f(x)}\right\}$$
W.p. $\rho(x^{(t)}, y)$ accept y setting $x^{(t+1)} = y$, otherwise $x^{(t+1)} = x^{(t)}$

J. Andrés Christen and Colin Fox, *MCMC using an Approximation*, Journal of Computational and Graphical Statistics, 2005/6

Speeding-Up MCMC Sampling from $f(\cdot)$

1. At $x^{(t)}$ generate proposal y from $q(\cdot \mid x^{(t)})$
2. Let

$$g(x,y) = \min\left\{1, \frac{q(x \mid y)}{q(y \mid x)} \frac{f_x^*(y)}{f_x^*(x)}\right\}$$

   W.p. $g(x^{(t)}, y)$, "promote" y. New proposal distribution is

$$q^*(y \mid x) = g(x,y)q(y \mid x) + (1 - r(x))\delta_x(y)$$

3. Let

$$\rho(x,y) = \min\left\{1, \frac{q^*(x \mid y)}{q^*(y \mid x)} \frac{f(y)}{f(x)}\right\}$$

   W.p. $\rho(x^{(t)}, y)$ accept y setting $x^{(t+1)} = y$, otherwise $x^{(t+1)} = x^{(t)}$

e.g.

$$f_x^*(x + \Delta x | d) \propto \exp\left\{-\chi\left(d - \left(A(x) + J\Delta x\right)\right) - \rho\left(x\right)\right\}$$

J. Andrés Christen and Colin Fox, *MCMC using an Approximation*, Journal of Computational and Graphical Statistics, 2005/6

Simplest case: Matrix equation

Consider the inverse problem where simulation of measurements requires solving the matrix equation

$$Y_\sigma v = i$$

i is fixed, v is measured (data)

Y_σ is a symmetric nonsingular (positive definite) linear $N \times N$ matrix function of $\sigma \in \mathbb{R}^M$

Initially think of i as a single vector

Typically measurements are of $v_j : j \in E = 1, 2, \ldots, |E| \leq N$ for a set of fixed $i : i_j, j \in E$ are determined, with other components being zero.

Defines forward map

$$A : \sigma \mapsto \left(Y_\sigma^{-1}\right)_{E,E}$$

Inverse problem is to find σ from noisy measurement of $\left(Y_\sigma^{-1}\right)_{E,E}$

What is the Jacobian?

Change in v due to change in Y, σ

$$(Y_\sigma + dY_\sigma)(v + dv) = i \quad \Rightarrow \quad Y_\sigma dv = -dY_\sigma (v + dv)$$

To first order

$$\frac{dv}{d\sigma_j} = -Y_\sigma^{-1} \frac{dY_\sigma}{d\sigma_j} v$$

Chain rule gives a general change

$$\begin{aligned} dv = J d\sigma &= -\sum_j Y_\sigma^{-1} \frac{dY_\sigma}{d\sigma_j} v d\sigma_j \\ &= -Y_\sigma^{-1} \left(\sum_j \frac{dY_\sigma}{d\sigma_j} d\sigma_j \right) v \\ &= -Y_\sigma^{-1} Y_{d\sigma} v \end{aligned}$$

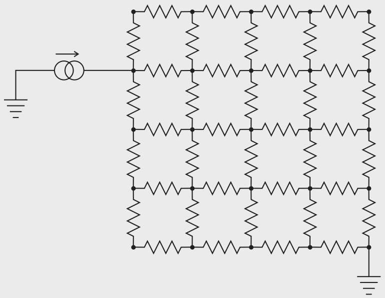
A minimum simulation of *all* measurements generates $G = \left(Y_\sigma^{-1}\right)_{:,E}$. Since $\left(Y_\sigma^{-1}\right)_{E,:} = G^\mathsf{T}$

$$J d\sigma = d\left(Y_\sigma^{-1}\right)_{E,E} = -G^\mathsf{T} Y_{d\sigma} G$$

Cheap calculation for sparse $Y_{d\sigma}$

For ‘local’ changes, $d\sigma$, when local stiffness matrix is small, $Y_{d\sigma}$ is sparse

e.g.



$$(Y_\sigma)_{lm} = \begin{cases} -\sigma_{lm} & l \neq m \\ \sum_{k=1}^{N} \sigma_{lk} & l = m \end{cases}$$

For single site change $\Delta\sigma_{lm}$

$$Y_{d\sigma} = \Delta\sigma_{lm} \begin{pmatrix} & \vdots & & \vdots & \\ \cdots & 1 & \cdots & -1 & \cdots \\ & \vdots & & \vdots & \\ \cdots & -1 & \cdots & 1 & \cdots \\ & \vdots & & \vdots & \end{pmatrix} = (e_l - e_m)(e_l - e_m)^{\mathsf{T}}$$

$$J d\sigma = -\Delta\sigma_{lm} \left(G_{l,E} - G_{m,E}\right)^{\mathsf{T}} \left(G_{l,E} - G_{m,E}\right)$$

Fox and Nicholls, *Sampling Conductivity Images via MCMC*, 1997

Fast Jacobian using low rank of Y_{σ_j}

When $Y_{d\sigma_j}$ is positive semi-definite with rank $p \approx 1$
(e.g. resistor network $p = 1$, FEM with triangulation for EIT $p = 2$)

$$Y_{e_j} = w_{j1} w_{j1}^{\mathsf{T}} + \cdots + w_{jp} w_{jp}^{\mathsf{T}}$$

$$Y_\sigma = \sum_j \sigma_j \sum_{l=1}^{p} w_{jl} w_{jl}^{\mathsf{T}}$$

Let $W_l = \begin{pmatrix} \vdots & \vdots & & \vdots \\ w_{1l} & w_{2l} & \cdots & w_{Nl} \\ \vdots & \vdots & & \vdots \end{pmatrix}$ for $l = 1, \ldots, p$

$$J d\sigma = -G^{\mathsf{T}} Y_{d\sigma} G = -\sum_{l=1}^{p} G^{\mathsf{T}} W_l \sigma W_l^{\mathsf{T}} G$$

Kolehmainen, Fox and Nicholls, *MCMC Inversion of Measured EIT Data*, 200?

Fast Transpose of Jacobian

$J : \sigma \mapsto v$ where σ is $N \times 1$ and v is $|E| \times |E|$

Calculation of

$$J^{\mathsf{T}} : v \mapsto \sigma$$

is similar

$$\begin{aligned} J^{\mathsf{T}} v &= \sum_{ij} \frac{\partial v_{ij}}{\partial \sigma} v_{ij} \\ &= -\sum_{l=1}^{p} \left(G^{\mathsf{T}} W_l \right)^{\mathsf{T}} v \left(G^{\mathsf{T}} W_l \right) \end{aligned}$$

Complete Electrode Model for EIT

For fixed current patterns $\{I\}$

$$A : \sigma \mapsto \{U\}$$

Simulate A by solving the BVP

$$\nabla \cdot \sigma \nabla u = 0$$

$$\int_{e_l} \sigma \frac{\partial u}{\partial n} dS = I_l$$

$$\sigma \frac{\partial u}{\partial n}\bigg|_{\partial\Omega \setminus \bigcup_l e_l} = 0$$

$$\left(u + z_l \sigma \frac{\partial u}{\partial n} \right)\bigg|_{e_l} = U_l$$

Likelihood

$$L(\sigma | V) \propto \exp\left\{ -\frac{1}{2\epsilon^2} \| V - A(\sigma) \|_{\mathsf{F}}^2 \right\}$$

(Kuopio) FEM Discretization

$$u = \sum_{i=1}^{N_n} \alpha_i \varphi_i \quad U = \sum_{j=1}^{|E|-1} \beta_j n_j$$

n_j is j^{th} column of $\mathcal{D}$, the $|E|-1$ dim basis of current patterns. Weak form of BVP is

$$Mb = f$$

where

$$b = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad f = \begin{pmatrix} \mathbf{0} \\ \mathcal{D}^T I \end{pmatrix} \quad A = \begin{pmatrix} B & C \\ C^T & G \end{pmatrix}$$

and

$$B_{i,j} = \int_\Omega \sigma \nabla \varphi_i \cdot \nabla \varphi_j dr + \sum_{l=1}^{|E|} \frac{1}{z_l} \int_{e_l} \varphi_i \varphi_j dS \quad 1 \leq i,j \leq N_n$$

C and G due to electrode BC. Then $U = \mathcal{D}\beta$.

Assemble FEM matrix system, solve $|E|$ times.

Implementing 'Matrix' Scheme

Symmetrize calculation by picking a suitable set of $|E|$ currents that span space and solve for Green's functions, e.g.

$$K = \mathbf{1} - \frac{1}{|E|} \qquad G = M^{-1} f = \begin{pmatrix} \mathbf{0} \\ K \end{pmatrix}$$

Measurements patterns $M^\mathsf{T} = K M_1$, so $\left(A^{-1} M^\mathsf{T}\right)^\mathsf{T} = M_1^\mathsf{T} G$, and $f = K f_1$ so $b = G f_1$

$$J\sigma_l = -\left(A^{-1} I^\mathsf{T}\right)^\mathsf{T} \frac{\partial A}{\partial \sigma_l} b = -I_1^\mathsf{T} G^\mathsf{T} \frac{\partial A}{\partial \sigma_l} G f_1$$

turns out that in Kuopio FEM $K = I$, $M^\mathsf{T} = -K$

$$\begin{aligned} J\sigma &= \sum_{l=1}^{p} \left(G^\mathsf{T} W_l\right) \sigma \left(G^\mathsf{T} W_l\right)^\mathsf{T} \\ J^\mathsf{T} v &= \sum_{l=1}^{p} \left(G^\mathsf{T} W_l\right)^\mathsf{T} v \left(G^\mathsf{T} W_l\right) \end{aligned}$$

Summary

- Operating by Jacobian and transpose is ($\sim 10\times$) faster than forming Jacobian with matrix multiplication
- Useful in implementing Langevin diffusion, gradient ascent, linearization, etc
- Scheme works for EIT, narrow-band acoustic backscatter, etc