Singular value analysis of Joint Inversion

Jodi Mead
Department of Mathematics

James Ford
Clearwater Analytics

Thanks to: Diego Domenzain, John Bradford

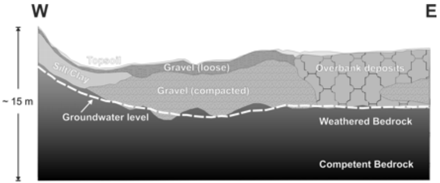
NSF DMS-1418714

BOISE STATE UNIVERSITY

Outline

- Subsurface earth imaging with electromagnetic waves
- Inverse methods and regularization
- Discrete joint inversion and the SVD
- Green's function solutions of differential equations
- Continuous inversion and the SVE
- Joint singular values and Galerkin approximation
- Example of joint inversion

Near subsurface imaging

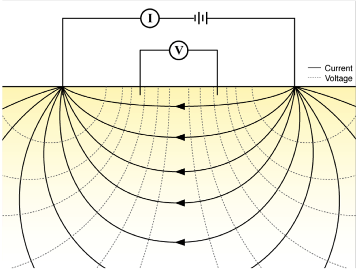


- Landfill investigation
- Mapping and monitoring of groundwater pollution
- Determination of depth to bedrock
- Locating sinkholes, cave systems, faults and mine shafts
- Landslide assessments
- Buried foundation mapping

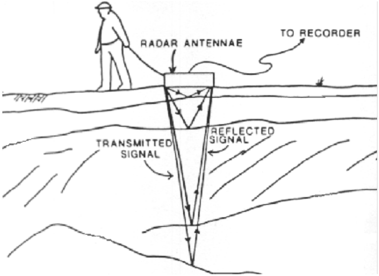
Boise Hydrogeophysical Research Site

Field laboratory about 10 miles from downtown Boise

Near subsurface imaging with electromagnetic waves



Electrical Resistivity Tomography



Ground Penetrating Radar

Ground Penetrating Radar (GPR)

$$\text{Damped Wave:} \quad \epsilon\frac{\partial^2 E}{\partial t^2} + \sigma\frac{\partial E}{\partial t} = \frac{1}{\mu}\nabla^2 E + f$$

- Radar signals f transmitted into the ground and energy that is reflected back to the surface is recorded.
- If there's a contrast in properties between adjacent material properties (permittivity ϵ, permeability μ and conductivity σ) a proportion of the electromagnetic pulse will be reflected back.
- Subsurface structures are imaged by measuring the amplitude and travel time of this reflected energy.

Electrical Resistivity Tomography (ERT)

$$\text{Diffusion:} \quad -\nabla \cdot \sigma \nabla \phi = \nabla \cdot J_s$$

- Current is passed through the ground via outer electrodes J_s and potential difference ϕ is measured between an inner pair of electrodes.
- Only responds to variability in electrical resistivity $\rho = 1/\sigma$ exhibited by earth materials.
- ERT data must be inverted to produce detailed electrical structures of the cross-sections below the survey lines.

Forward vs Inverse Modeling

	Forward	Inverse
$\epsilon \frac{\partial^2 E}{\partial t^2} + \sigma \frac{\partial E}{\partial t} = \frac{1}{\mu} \nabla^2 E + f$	Given ϵ, μ, σ Solve for E	Given E Solve for ϵ, μ, σ
$-\nabla \cdot 1/\rho \nabla \phi = \nabla \cdot J_s$	Given ρ solve for ϕ	Given ϕ solve for ρ

Nonlinear regression

Newton's method for $F(\mathbf{m}) = \mathbf{d}$

$$\begin{aligned} \mathbf{m}^{k+1} &= \mathbf{m}^k + \mathbf{J}^{-1}(\mathbf{m}^k)(F(\mathbf{m^k}) - \mathbf{d}) \\ &= \mathbf{m}^k + \triangle\mathbf{m} \end{aligned}$$

where $\mathbf{J}_{ij} = \frac{\partial F_i}{\partial m_j}$. We focus on the linear problem

$$\begin{aligned} \mathbf{J}(\mathbf{m}^k)\triangle\mathbf{m} &= F(\mathbf{m^k}) - \mathbf{d} \\ \mathbf{G}\mathbf{m} &= \mathbf{d} \end{aligned}$$

Full Disclosure....

This work is motivated by inverting both damped wave and diffusion equation simultaneously (Joint Inversion). However, we have only obtained results for simplified equations:

$$\begin{aligned} u'' &= f \\ u'' + b^2 u &= f \end{aligned}$$

Inverse Problems

Consider solving problems of the form:

$$\mathbf{Gm} = \mathbf{d},$$

- $\mathbf{G} \in R^{m \times n}$ - mathematical model
- $\mathbf{d} \in R^{m}$ - observed data
- $\mathbf{m} \in R^{n}$- unknown model parameters

$\mathbf{G}$ cannot be resolved by the data $\mathbf{d}$ because it is ill-conditioned

$$\text{det}(\mathbf{G}^T\mathbf{G}) = \text{“large"}$$

and the solution $\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{d}$ is not possible.

Regularization

$$\mathbf{m}_{\mathbf{L}_p} = argmin_{\mathbf{m}} \left\{ \|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2^2 + \lambda ||\mathbf{L}_p(\mathbf{m} - \mathbf{m}_0)||_2^2 \right\}$$

$\mathbf{m}_0$ - initial estimate of $\mathbf{m}$

$\mathbf{L}_p$ - typically represents the first $(p = 1)$ or second derivative $(p = 2)$

λ - regularization parameter

This gives estimates

$$\mathbf{m}_{\mathbf{L}_p} = \mathbf{m}_0 + (\mathbf{G}^T\mathbf{G} + \lambda \mathbf{L}_p^T \mathbf{L}_p)^{-1} \mathbf{G}^T \mathbf{d}$$

Choice of λ

Methods:

L-curve, Generalized Cross Validation (GCV) and Morozov's Discrepancy Principle, UPRE, χ^2 method[1]...

- λ large $\rightarrow$ constraint: $\|\mathbf{L}_p(\mathbf{m}-\mathbf{m}_0)||_2^2 \approx 0$

$$\mathbf{m}_{\mathbf{L}_p} = argmin_{\mathbf{m}} \left\{ \|\mathbf{G}\mathbf{m}-\mathbf{d}\|_2^2 + \lambda||\mathbf{L}_p(\mathbf{m}-\mathbf{m}_0)||_2^2 \right\}$$

- λ small $\rightarrow$ problem may stay ill-conditioned

$$\mathbf{m}_{\mathbf{L}_p} = \mathbf{m}_0 + (\mathbf{G}^T\mathbf{G} + \lambda\mathbf{L}_p^T\mathbf{L}_p)^{-1}\mathbf{G}^T\mathbf{d}$$

[1] Mead et al, 2008, 2009, 2010, 2016

Choice of $\mathbf{L}_p$

$$\mathbf{m}_{\mathbf{L}_p} = argmin_{\mathbf{m}} \left\{ \|\mathbf{Gm} - \mathbf{d}\|_2^2 + \lambda ||\mathbf{L}_p(\mathbf{m} - \mathbf{m}_0)||_2^2 \right\}$$

$\mathbf{L}_0 = \mathbf{I}$ - requires good initial estimate $\mathbf{m}_0$, otherwise may not regularize the problem.

$\mathbf{L}_1$ - requires first derivative estimate, could be less information than $\mathbf{m}_0$ (just changes in $\mathbf{m}_0$).

$\mathbf{L}_2$ - requires second derivative estimate, leaves more degrees of freedom than first derivative so that data has more opportunity to inform changes in parameter estimates.

Joint Inversion as Regularization

$$\mathbf{m}_{12} = argmin_{\mathbf{m}} \left\{ \|\mathbf{G}_1\mathbf{m} - \mathbf{d}_1\|_2^2 + ||\mathbf{G}_2\mathbf{m} - \mathbf{d}_2||_2^2 \right\}$$

Objective function can be written

$$\left\| \begin{bmatrix} \mathbf{G}_1 \\ \mathbf{G}_2 \end{bmatrix} [\mathbf{m}] - \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix} \right\|_2^2 \equiv \|\mathbf{G}_{12}\mathbf{m} - \mathbf{d}_{12}\|_2^2$$

Goal: Improve condition number

$$\kappa(\mathbf{G}_{12}) < \kappa(\mathbf{G}_1), \kappa(\mathbf{G}_2)$$

where $\kappa(\mathbf{G}) = \frac{\sigma_{max}(\mathbf{G})}{\sigma_{min}(\mathbf{G})}$

Singular Value Decomposition (SVD)

$$\mathbf{G}_{12} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \rightarrow \mathbf{m}_{12} = \sum_{i=1}^{n} \frac{\mathbf{U}_{\cdot,i}^T \mathbf{d}_{12}}{\sigma_i} \mathbf{V}_{\cdot,i}$$

Truncated SVD (with decomposition on appropriate matrix)

$$\mathbf{m}_1 = \sum_{i=1}^{k_1} \frac{\mathbf{U}_{\cdot,i}^T \mathbf{d}_1}{\sigma_i} \mathbf{V}_{\cdot,i}, \;\; \mathbf{m}_2 = \sum_{i=1}^{k_2} \frac{\mathbf{U}_{\cdot,i}^T \mathbf{d}_2}{\sigma_i} \mathbf{V}_{\cdot,i}, \;\; \mathbf{m}_{12} = \sum_{i=1}^{l} \frac{\mathbf{U}_{\cdot,i}^T \mathbf{d}_{12}}{\sigma_i} \mathbf{V}_{\cdot,i}$$

Goal: Keep as many singular values as possible

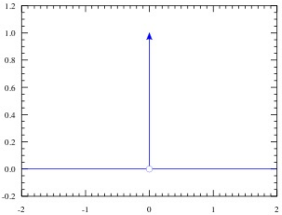
$$l >> k_1, k_2$$

Green's Functions

Let $\mathcal{L}_A = \mathcal{L}_A(t)$ be a linear differential operator. Then the corresponding Green's function $K_A(t,s)$ satisfies

$$\mathcal{L}_A K_A(t,s) = \delta(t-s), \tag{1}$$

where δ denotes the delta function, a generalized function:



Green's Function solutions of differential equations

Given the Green's function, we can find the solution to the inhomogeneous equation $\mathcal{L}_A u(t) = f(t)$:

$$u(t) = \int_\Omega K_A(t,s) f(s) ds.$$

Forward problem: Given f, find u; Inverse problem: Given u, find f

Conditioning of the inverse problem depends on the forward operator $A : H \to H_A$

$$Ah(t) \equiv \int_\Omega K_A(t,s) h(s) ds$$

Singular Value Expansion (SVE)

If A is compact

$$A = \sum_{k=1}^{\infty} \sigma_k \psi_k \otimes \phi_k$$

where $\{\phi_k\} \subset H$ and $\{\psi_k\} \subset H_A$ are orthonormal and $\{\sigma_k\}$ are the singular values of A. Thus

$$Ah = \sum_{k=1}^{\infty} \sigma_k \langle \phi_k, h \rangle_H \psi_k$$

or $A\phi_k = \sigma_k \psi_k$ for all k.

Singular Values

Adjoint $A^* : H_A \to H$ defined by $\langle Ah, h_A \rangle_{H_A} = \langle h, A^* h_A \rangle_H$

$$A^* = \sum_{k=1}^{\infty} \sigma_k \phi_k \otimes \psi_k$$

Singular values of A are $\sqrt{\text{eigenvalues}}$ of $A^* A : H \to H$:

$$A^* A \phi = \sigma^2 \phi$$

yields $\{(\sigma_k, \phi_k)\}_{k=1}^{\infty}$

Least Squares

$$\|Ah - f\|_{H_A}^2$$

has solution

$$h = A^\dagger f = \sum_{k=1}^{\infty} \frac{\langle \psi_k, f \rangle_{H_A}}{\sigma_k} \phi_k$$

Condition number: σ_1/σ_r; but as $r \to \infty$, $\sigma_r \to 0$.

Decay rate q: $\sigma_k(A)$ decays like k^{-q}

Decay rate of singular values allow us to classify model conditioning

Decay rate example

Consider $A : H \to H_A$ with $H = H_A = L^2(0,1)$

$$Ah(t) = \int_0^t h(s)ds.$$

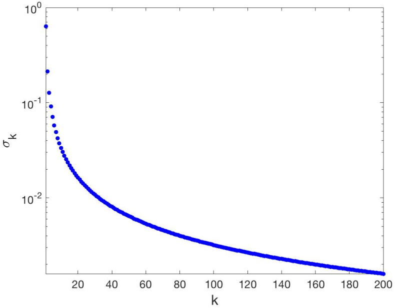
Solve $A^*A\phi = \sigma^2\phi$ for σ and ϕ. This problem is equivalent to

$$\lambda\phi'' + \phi = 0, \quad \text{with} \quad \phi(1) = \phi'(0) = 0.$$

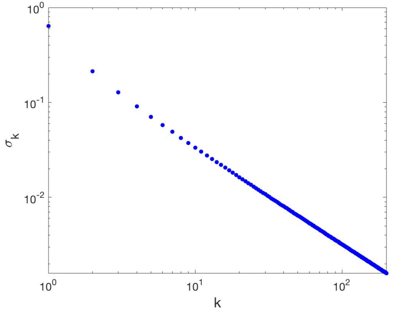
The singular functions and values are

$$\phi_k(t) = c_1 \cos\frac{2k-1}{2}\pi t \;\; \text{and} \;\; \sigma_k = \frac{2}{(2k-1)\,\pi},\; k \in \mathbb{N}.$$

Decay rate example



semi-log



log-log

Tikhonov Operator

$T_\lambda : H \to H_A \times H$ is defined by

$$T_\lambda h = (Ah, \sqrt{\lambda} h)$$

where $H_A \times H = \{(h_A, h) : h_A \in H_A, h \in H\}$ with inner product

$$\langle (h_{A,1}, h_1), (h_{A,2}, h_2) \rangle_{H_A \times H} = \langle h_{A,1}, h_{A,2} \rangle_{H_A} + \langle h_1, h_2 \rangle_H .$$

Tikhonov regularization

We minimize

$$\|T_\lambda h - (f,0)\|^2_{H_A \times H} = \|Ah - f\|^2_{H_A} + \lambda \|h\|^2_H .$$

Normal equations[2]

$$\begin{aligned} T^*Th &= T^*(f,0) \\ T^*(Ah, \sqrt{\lambda}h) &= T^*(f,0) \\ A^*Ah + \lambda h &= A^*f + \sqrt{\lambda} \cdot 0 \\ (A^*A + \lambda I)h &= A^*f. \end{aligned}$$

[2] Gockenbach, 2015

Pseudoinverse

Generalized inverse operator for the modified least squares problem is

$$A_\lambda^\dagger = (A^*A + \lambda I)^{-1}A^* = \sum_{k=1}^{\infty} \frac{\sigma_k}{\sigma_k^2 + \lambda} \phi_k \otimes \psi_k.$$

Notice that

$$\frac{\sigma_k}{\sigma_k^2 + \lambda} \to 0, \text{ as } k \to \infty$$

and λ restricts the solution space.

Joint Inversion

$$\|Ah - f\|_{H_A}^2 + \|Bh - g\|_{H_B}^2$$

with $A : H \to H_A$ and $B : H \to H_B$. Joint operator:

$$C : H \to H_A \oplus H_B = \{(h_A, h_B) : h_A \in H_A, h_B \in H_B\}$$

so that

$$Ch = (Ah \oplus B)\,,\ \ h \in H$$

Continuous analog to stacking matrices

Joint Operator Example

Define $H = L^2(0, 2\pi)$ and $H_A = H_B = \mathbb{R}$

$$Ah = \int_0^{2\pi} h(y)\delta(y-5)dy \text{ and } Bh = \int_0^{2\pi} h(y)\delta(y-7)dy.$$

Then $C : H \to H_A \oplus H_B$ is given by

$$Ch = (Ah, Bh) = \left(\int_0^{2\pi} h(y)\delta(y-5)dy, \int_0^{2\pi} h(y)\delta(y-7)dy\right),$$

a parametric curve in the space $H_A \oplus H_B = \mathbb{R}^2$.

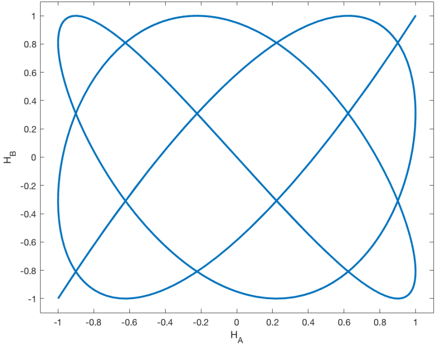
Joint Operator Example, con't

Consider $S = \{\cos kx : k \in \mathbb{R}, x \in [0, 2\pi]\} \subset H$, then

$$\begin{aligned} C(\cos kx) &= \left(\int_0^{2\pi} \cos(ky)\delta(y-5)dy, \int_0^{2\pi} \cos(ky)\delta(y-7)dy \right) \\ &= \left(\cos(k \cdot 5), \cos(k \cdot 7) \right). \end{aligned}$$

with image $[-1, 1] \times [-1, 1]$.

Joint Operator Example



Parametric curve defined by $C(\cos kx)$ for $k \in [-5, 5]$

Joint Singular Values

Adjoint $C^* : H_A \oplus H_B \to H$

$$C^*\left(h_A, h_B\right) = A^* h_A + B^* h_B.$$

so that

$$\begin{aligned} \sigma^2 \phi &= C^* C \phi \\ &= C^*\left(A\phi, B\phi\right) \\ &= A^* A \phi + B^* B \phi. \end{aligned} \tag{2}$$

Galerkin method for SVE

$A^{(n)}$ approximates A with orthonormal bases $\{q_i(s)\}_{i=1}^{n}$ and $\{p_j(t)\}_{j=1}^{n}$

$$\begin{aligned} a_{ij}^{(n)} &= \langle q_i, Ap_j \rangle \\ &= \langle q_i, \langle K_A, p_j \rangle \rangle \\ &= \int_{\Omega_s} \int_{\Omega_t} q_i(s) K_A(s,t) p_j(t) dt ds. \end{aligned} \tag{3}$$

so that

$$A^{(n)} = U^{(n)} \Sigma^{(n)} \left(V^{(n)} \right)^T$$

with $\Sigma^{(n)} = \text{diag}\left(\sigma_1^{(n)}, \sigma_2^{(n)}, \ldots \sigma_n^{(n)} \right)$

Convergence of Galerkin Method

Define

$$\left(\Delta_A^{(n)}\right)^2 = \|K_A\|^2 - \|A^{(n)}\|_F^2$$
$$= \sum_{i=1}^{\infty} \sigma_i^2 - \sum_{i=1}^{n} (\sigma_i^{(n)})^2.$$

Then the following hold for all i and n, independent of the convergence of $\Delta_A^{(n)}$ to 0: [3]

1. $\sigma_i^{(n)} \leq \sigma_i^{(n+1)} \leq \sigma_i$
2. $0 \leq \sigma_i - \sigma_i^{(n)} \leq \Delta_A^{(n)}$

[3] Hansen 1988, Renaut et al 2016

Convergence of Galerkin method for Joint SVE

Define

$$C^{(n)} = \begin{bmatrix} A^{(n)} \\ B^{(n)} \end{bmatrix} \text{ and } \left(\Delta_C^{(n)}\right)^2 = \sum_{i=1}^{\infty} \sigma_i(C)^2 - \sum_{i=1}^{n} \sigma_i(C^{(n)})^2.$$

Expanding

$$\begin{aligned} \left(\Delta_C^{(n)}\right)^2 &= \langle K_A, K_A \rangle + \langle K_B, K_B \rangle - \|A^{(n)}\|_F^2 - \|B^{(n)}\|_F^2 \\ &= \|K_A\|^2 - \|A^{(n)}\|_F^2 + \|K_B\|^2 - \|B^{(n)}\|_F^2 \\ &= \left(\Delta_A^{(n)}\right)^2 + \left(\Delta_B^{(n)}\right)^2. \end{aligned} \tag{4}$$

Thus if $\lim_{n\to\infty}(\Delta_A^{(n)})^2 = 0$ and $\lim_{n\to\infty}(\Delta_B^{(n)})^2 = 0$ the singular values of C are accurately approximated.

Special case: Self-Adjoint Operator

Use singular functions in Galerkin method

$$
\begin{aligned}
a_{ij}^{(n)} &= \langle \phi_j, A\phi_i \rangle && (5)\\
&= \langle \phi_j, \sigma_i \phi_i \rangle && (6)\\
&= \begin{cases} \sigma_i & i = j \\ 0 & i \neq j \end{cases} && (7)
\end{aligned}
$$

then

$$A^{(n)} = \Sigma_A^{(n)} \quad \text{and} \quad B^{(n)} = \Sigma_B^{(n)}$$

Joint Singular Values for Self-Adjoint Operator

$C^{(n)} = \begin{bmatrix} A^{(n)} \\ B^{(n)} \end{bmatrix}$ gives

$$\left(C^{(n)}\right)^T \left(C^{(n)}\right) = \left(\Sigma_A^{(n)}\right)^2 + \left(\Sigma_B^{(n)}\right)^2$$

so that

$$\sigma_i\left(C^{(n)}\right) = \sqrt{\sigma_i\left(A^{(n)}\right)^2 + \sigma_i\left(B^{(n)}\right)^2}$$

Joint Singular Value Example

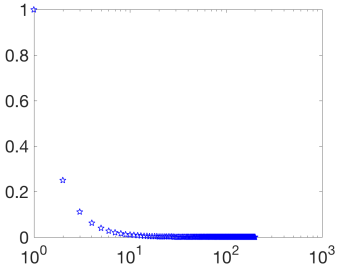
$$L_A u = -u'', \quad u(0) = u(\pi) = 0,$$
$$L_B u = u'' + b^2 u, \quad u(0) = u(\pi) = 0, \quad \text{and } b \notin \mathbb{Z}$$

Green's functions:

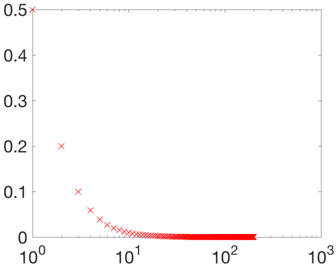
$$K_A = \begin{cases} \frac{1}{\pi}(\pi - x)\, y, & 0 \le y \le x \le \pi, \\ \frac{1}{\pi}(\pi - y)\, x, & 0 \le x \le y \le \pi. \end{cases}$$

$$K_B = \begin{cases} -\frac{\sin(by)\sin[b(\pi - x)]}{b\sin(b\pi)}, & 0 \le y \le x \le \pi, \\ -\frac{\sin(bx)\sin[b(\pi - y)]}{b\sin(b\pi)}, & 0 \le x \le y \le \pi. \end{cases}$$

Example - Individual Singular Values

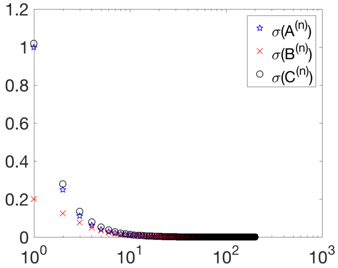


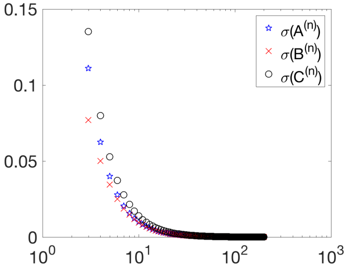
$\sigma_k(A^{(200)}) = \frac{1}{k^2}$



$\sigma_k(B^{(200)}) = \frac{1}{k^2+\pi^2}$

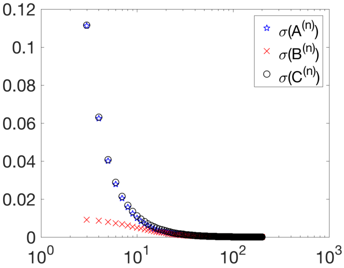
Example - Joint Singular Values



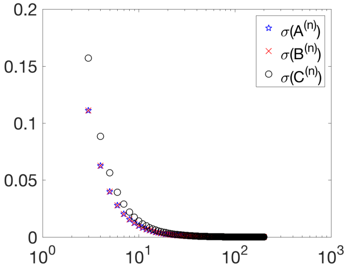


$$\sigma_k(C^{(200)}) = \sqrt{\left(\frac{1}{k^2}\right)^2 + \left(\frac{1}{k^2+2^2}\right)^2}$$

Example - Joint Singular Values



$$\sigma_k(C^{(200)}) = \sqrt{\left(\frac{1}{k^2}\right)^2 + \left(\frac{1}{k^2+10^2}\right)^2}$$



$$\sigma_k(C^{(200)}) = \sqrt{\left(\frac{1}{k^2}\right)^2 + \left(\frac{1}{k^2+0.1^2}\right)^2}$$

Conclusions

We suggest using Green's function solutions of differential equations to quantify how combining different types of data in a joint inversion improves conditioning of individual inverse problems.

- This analysis required us to extend the following to joint inversion:
  - Tikhnov regularization in a continuous domain
  - Singular Value Expansion (SVE)
  - Convergence of the Galerkin method to approximate the SVE
- If the individual operators are both self-adjoint, we found an expression for the joint singular values in terms of the individual singular values.

Thank you!

Jodi Mead
Mathematics Department
Boise State University
jmead@boisestate.edu