Joint Inversion of Compact Operators

J. F. Ford[1] and J. L. Mead[2]

[1]Clearwater Analytics, [2]Boise State University, 1910 W University Dr, Boise, ID 8372 USA

E-mail: [2]`jmead@boisestate.edu`

August 2019

Abstract. Joint inversion of multiple data types was studied as early as 1975 in [1], where the authors used the singular value decomposition to determine the degree of ill-conditioning of joint inverse problems. The authors demonstrated in several examples that combining two physical models in a joint inversion, and effectively stacking discrete linear models, improved the conditioning of individual inversions. This work extends the notion of using the singular value decomposition to determine the conditioning of discrete joint inversion to using the singular value expansion to determining the well-posedness of joint operators. We provide a convergent technique for approximating the singular values of continuous joint operators. In the case of self-adjoint operators, we give an algebraic expression for the joint singular values in terms of the singular values of the individual operators. This expression allows us to show that while rare, there are situations where ill-posedness may be not improved through joint inversion and in fact can degrade the conditioning of an individual inversion. We give an example of improving inversion with two moderately ill-posed Green's function solutions, and quantify the improvement over individual inversions. Results from this work show that analysis of singular values of compact operators describing different data types before an inversion helps identify which types of data are advantageous to combine in a joint inversion.

Keywords: joint inversion, compact operators

Submitted to: *Inverse Problems*

1. Introduction

Joint inversion involves inverting one or more data sets that share common parameters x, e.g. $Ax = d_1$ and $Bx = d_1$ where d_1 and d_2 are distinct data sets while A and B are distinct bounded linear operators. Both systems are typically ill-posed, the data contain noise and hence regularization is required to estimate the parameters. For example, using Tikhonov regularization with the first equation we optimize

$$\min_x \left\{ \|Ax - d_1\|_2^2 + \alpha^2 \|Lx\|_2^2 \right\} .$$

Introducing regularization means we solve a nearby well-posed problem that may add bias to the parameter estimates [2].

To make the problem less ill-posed, we could introduce addition data through joint inversion and optimize

$$\min_x \left\{ \|Ax - d_1\|_2^2 + \|Bx - d_2\|_2^2 \right\}$$

with appropriately weighted data and operators. Additional data has the potential to regularize and hence reduce ill-posedness in individual systems. This is advantageous over regularization methods like Tikhonov because the bias introduced by the additional term comes from a physically motivated model, rather than initial estimates of the parameters or their derivative values.

Joint inversion has become common in Geophysical applications. For example, electromagnetic and seismic data can be jointly inverted for geophysical parameters [3, 4]. Even though the physics describing each data set may not share the same parameters, data can be combined in an inversion using petrophysical relationships [5] or by the cross-gradient approach [6]. Cross-gradient regularization assumes the parameters from each data set are structurally similar and has also been used to combine gravity and magnetic data [7], and resistivity and seismic data [8, 9]. In all cases numerical results show that joint inversion improves separate inversions.

In this work we discuss methods to quantify the amount by which joint inversion improves individual inversions. Ill-posedness in each system can be measured by analyzing their singular values. For example, using discrete representations $\mathbf{A} \in \mathbb{R}^{m_1 \times n}$ and $\mathbf{B} \in \mathbb{R}^{m_2 \times n}$ of the linear operators A and B, the singular values of the stacked matrix

$$\mathbf{C} \equiv \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}$$

give the degree of ill-conditioning of joint inverse problem. In particular, if the singular values σ_k decay like k^{-q}, we call q the degree of a mildly or moderately ill-posed problem. Larger values of q indicate larger degrees of ill-posedness and in severely ill-posed problems σ_k decays like e^{-qk} [10].

In this work we consider continuous compact linear operators that represent physics from the data collection process, rather than discretized versions of them. Therefore, in Section 3.1 we extend the notion of a vertically concatenated matrix to the process of combining compact linear operators. This process is understood for Tikhonov regularization [11] and we extend it to the more general direct sum of integral operators on Hilbert spaces [12]. We give a practical approach to calculating the singular values in Section 3.2 using a Galerkin method. If the operators are self-adjoint operators we show in Section 3.3 that it is possible to get a closed form expression for the joint singular values in terms of the singular values of the individual operators. Understanding the ill-posedness of data collection techniques before data is collected opens the door to experimental design. We illustrate this on joint inversion of two simple one-dimensional ordinary differential equations in Section 4.

2. Background

2.1. Singular Value Expansion

The singular value decomposition is the tool of choice for rigorous analysis of least squares solutions to discrete linear inverse problems. The continuous extension of this tool is the singular value expansion (SVE) [13, 11, 14, 15, 16]. It decomposes a compact linear operator into orthogonal functions.

Theorem 2.1 (Singular Value Expansion) *Let H, H_A be Hilbert spaces, and let $A : H \to H_A$ be a compact linear operator. Then there exists orthonormal sequences $\{\phi_k\} \subset H$ and $\{\psi_k\} \subset H_A$ and positive numbers $\sigma_1 \geq \sigma_2 \geq \cdots$ converging to zero, such that*

$$A = \sum_{k=1}^{\infty} \sigma_k \psi_k \otimes \phi_k, \quad \textit{and} \quad A^* = \sum_{k=1}^{\infty} \sigma_k \phi_k \otimes \psi_k.$$

We define $\psi_k \otimes \phi_k$ as

$$(\psi_k \otimes \phi_k)h = \langle h, \phi_k \rangle_H \psi_k,$$

for all $h \in H$. Note that A^* is also a compact linear operator and denotes the adjoint of A. Furthermore,

$$A\phi_k = \sigma_k \psi_k \textit{ for all } k$$

and

$$Ah = \sum_{k=1}^{\infty} \sigma_k \langle \phi_k, h \rangle_H \psi_k \textit{ for all } h \in H.$$

Additionally, $\{\phi_k\}$ is a complete orthonormal set for $\mathcal{N}(A)^\perp$ and $\{\psi_k\}$ is a complete orthonormal set for $\overline{\mathcal{R}(A)}$.

Proof. See [11] or [17].

□

The SVE yields a family of singular function, singular value pairs $\{(\sigma_k, \phi_k)\}_{k=1}^{\infty}$ that satisfy

$$A^* A \phi_k = \sigma_k^2 \phi_k.$$

The operator A^*A may not be invertible and we express the generalized inverse as

$$A^\dagger = \sum_{k=1}^{\infty} \sigma_k^{-1} \phi_k \otimes \psi_k. \tag{1}$$

The least squares solution that minimizes $\|Ah - f\|_{H_A}^2$ is given by

$$h = A^\dagger f = \sum_{k=1}^{\infty} \sigma_k^{-1} (\phi_k \otimes \psi_k) f = \sum_{k=1}^{\infty} \frac{\langle \psi_k, f \rangle_{H_A}}{\sigma_k} \phi_k \text{ for all } f \in D\left(A^\dagger\right).$$

However, in infinite dimensions A has an infinite sequence of singular values decaying towards zero. Therefore, σ_k^{-1} increase in an unbounded manner and $A^\dagger$ is not a compact operator [18, 15].

> Example: Define the compact linear operator $A : H \to H_A$, where $H = H_A = L^2(0,1)$, by $Ah(t) = \int_0^t h(s)ds$. Then the adjoint operator $A^* : H_A \to H$ is $A^* f(t) = \int_t^1 f(s)ds$ while the self-adjoint operator $A^*A : H \to H$ is $A^*Ah(t) = \int_t^1 \left(\int_0^s h(\tau)d\tau\right) ds$. The singular values and right-singular functions of A are $\sigma_k = \frac{2}{(2k-1)\pi}$ and $\phi_k(t) = \sqrt{2}\cos\frac{t}{\sigma_k}; k \in \mathbb{N}$.

The condition number is defined as the ratio of largest to smallest singular values. However, in infinite dimensions is not a sufficient metric by which to measure ill-posedness so we characterize the ill-posedness of the problem in terms of the decay rate of its singular values. As in the discrete case, it is clear that small singular values (relative to σ_1) will disproportionately amplify the contribution from corresponding singular vectors or functions. If there is noise in the data, this too will be amplified, perhaps to an unacceptable level.

2.2. Tikhonov Regularization

The negative effect decaying singular values have on the parameter estimates in an ill-posed problem can be alleviated with regularization. In infinite dimensional Hilbert spaces, a truncated SVE approximation to the operator A requires truncation of infinitely many singular values, and we will not investigate this finite sum approximation. Alternatively, we focus on Tikhonov regularization for compact operators as it relates to joint inversion.

Tikhonov regularization changes the problem to one which has an invertible operator, and therefore has a well-defined inverse solution. This invertible operator will require us to consider the space $H_A \times H = \{(h_A, h) : h_A \in H_A, h \in H\}$ which is a Hilbert space under the inner product

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

The Tikhonov operator $T_\alpha : H \to H_A \times H$ is defined by

$$T_\alpha h = (Ah, \alpha h)$$

and we minimize

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

Theorem 2.2 *Suppose $\alpha > 0$. Then $\mathcal{R}(T_\alpha)$ is closed and $\mathcal{N}(T_\alpha)$ is trivial. Therefore $T_\alpha h = (f,0)$ has a unique least squares solution for all $f \in H_A$, [11].*

Proof. Consider the normal equation for this problem:

$$\begin{aligned}
T^*Th &= T^*(f,0) \\
T^*(Ah,\alpha h) &= T^*(f,0) \\
A^*Ah+\alpha^2 h &= A^*f+\alpha\cdot 0 \\
(A^*A+\alpha^2 I)h &= A^*f.
\end{aligned}$$

For appropriate choice of α $(A^*A+\alpha^2I)$ is invertible with a bounded inverse. Therefore a unique solution to the normal equations exists.

□

Tikhonov regularization replaces the not necessarily invertible operator A^*A with $(A^*A+\alpha^2I)$ in the normal equations. The generalized inverse operator for the modified least squares problem therefore can be written

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

Since

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

the operator $A_\alpha^\dagger$ is bounded, and inverse solutions depend continuously on f.

Solution estimates found with this generalized operator depend strongly on the regularization parameter α, which restricts the space of acceptable solutions. Alternatively, joint inversion uses additional physics and the corresponding observations to restrict the solution space. This allows more physically relevant solutions and restricts the parameters to ones that satisfy two or more mathematical models. Joint inversion does not contain a parameter such as α that guarantees a well posed problem, but it will more likely require less regularization. This is explained in detail in Section 3.

3. Joint Inversion

Joint inversion minimizes

$$\|Ah-d_1\|_{H_A}^2+\|Bh-d_2\|_{H_B}^2=\|Ch-(d_1,d_2)\|_{H_A\times H_B}^2.$$

It maps the Cartesian product of two Hilbert spaces A and B with physical spaces H_A and H_B, respectively. H_B can be considered as an alternative to the mathematically defined space H in Tikhonov regularization. Analogous to stacking matrices, the space defined by joint inversion of linear operators A and B consists of all ordered pairs in $H_A\times H_B$.

Example: Define the Hilbert spaces $H=L^2(0,2\pi)$, and $H_A=H_B=\mathbb{R}$. Define the compact operators $A:H\to H_A$ and $B:H\to H_B$ as

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

Then $C : H \to H_A \oplus H_B$ is defined as

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

3.1. Singular Value Expansion

As mentioned previously, the decay rate of the singular values provide a metric for the ill-posedness of an operator. With A, B, and C defined as above, we can compare the singular value decay rates of the three operators to see if the joint operator yielded any improvement. For the purposes of visualization, it is helpful to think of the C as defining a parametric curve in the space $H_A \oplus H_B$.

Theorem 3.1 *Let $A : H \to H_A$ and $B : H \to H_B$ be compact operators from the Hilbert space H to the Hilbert spaces H_A and H_B respectively. Then $C : H \to H_A \oplus H_B$ with $Ch = (Ah, Bh)$ admits a singular value expansion for all $h \in H$.*

Proof. The Hilbert space direct sum

$$H_A \oplus H_B = \{(h_A, h_B) : h_A \in H_A, h_B \in H_B\}.$$

admits the inner product $\langle \cdot, \cdot \rangle$ on $H_A \oplus H_B$ with

$$\langle (h_{A,1}, h_{B,1}), (h_{A,2}, h_{B,2}) \rangle_{H_A \oplus H_B} = \langle h_{A,1}, h_{A,2} \rangle_{H_A} + \langle h_{B,1}, h_{B,2} \rangle_{H_B}.$$

C is thus a compact operator between two Hilbert spaces [19] and admits a SVE.

$\square$

Lemma 3.2 *The family of singular function, singular value pairs $\{(\sigma_k, \phi_k)\}_{k=1}^{\infty}$ that satisfy*

$$C^* C \phi_k = \sigma_k^2 \phi_k$$

also satisfy

$$\sigma_k^2 \phi_k = A^* A \phi_k + B^* B \phi_k.$$

Proof. The adjoint $C^* : H_A \oplus H_B \to H$ is given by

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

Expanding we get

$$\begin{aligned} \sigma_k^2 \phi_k &= C^* C \phi_k \\ &= C^* (A\phi_k, B\phi_k) \\ &= A^* A \phi_k + B^* B \phi_k. \end{aligned}$$

$\square$

Example: Let A and B be Green's function operators

$$Ah = \int_\Omega K_A h \quad \text{and,} \quad Bh = \int_\Omega K_B h,$$

associated with the differential operations L_A and L_B respectively. Then (2) is an integral equation and it can be transformed to an equivalent ODE. In particular,

$$\begin{aligned} L_A^* \left(\sigma^2 \phi\right) &= L_A^* \left(A^* A \phi + B^* B \phi\right), \\ &\text{and} \\ L_A \left(\sigma^2 L_A^* \phi\right) &= L_A \left(A \phi + L_A^* B^* B \phi\right), \\ &\text{or} \\ \sigma^2 L_A L_A^* \phi &= \phi + L_A L_A^* B^* B \phi. \end{aligned}$$

If $L_A L_A^*$ and $L_B L_B^*$ commute and we apply L_B^* and L_B in the same manner we eliminate all integrals and obtain and ODE in ϕ:

$$\sigma^2 \left(L_B L_B^* L_A L_A^*\right) \phi = \left(L_B L_B^* + L_A L_A^*\right) \phi.$$

The approach in this example to finding the singular value, singular function pairs can be challenging. It produces an ODE with much higher order than that of the given differential operators L_A or L_B, and introduces many more boundary conditions. Alternatively, we suggest using a Galerkin method to approximate the singular values as described in the next section.

3.2. Galerkin Method

The singular value expansion of an individual integral kernel $K_A(s,t)$ defined over $\Omega_s \times \Omega_t$, such as a Green's function, can be approximated using the Galerkin method. It has been shown that the singular values derived using the Galerkin method converge to the true singular values [14, 20]. Here, we extend the method to joint operators.

The idea of the Galerkin method is to approximate the integral operator A with an integral operator whose kernel is degenerate. We accomplish this by restricting ϕ and ψ to a the span of finitely many, n, orthonormal basis functions $\{q_i(s)\}_{i=1}^n$ and $\{p_j(t)\}_{j=1}^n$ for $L^2(\Omega_s)$ and $L^2(\Omega_t)$ respectively.

The matrix $A^{(n)}$ with entries $a_{ij}^{(n)}$ approximates the operator A, and is defined by

$$\begin{aligned} a_{ij}^{(n)} &= \langle q_i^A, A p_j^A \rangle \\ &= \langle q_i^A, \langle K_A, p_j^A \rangle \rangle \\ &= \int_{\Omega_s} \int_{\Omega_t} q_i^A(s) K_A(s,t) p_j^A(t) dt ds. \end{aligned} \tag{2}$$

The SVD $A^{(n)}$ is denoted $U_A^{(n)} \Sigma_A^{(n)} \left(V_A^{(n)}\right)^T$ with $\Sigma_A^{(n)} = \text{diag}\left(\sigma_1(A^{(n)}), \sigma_2(A^{(n)}), \ldots \sigma_n(A^{(n)})\right)$ containing the discrete singular values $\sigma_k(A^{(n)})$ which approximate the continuous singular values $\sigma_k(A)$.

Definition 3.1 *The singular values of an integral operator A with a real, square integrable kernel K are the stationary values of the functional*

$$F[p,q] = \frac{\langle q, Kp \rangle}{\|p\|\|q\|},$$

with the corresponding left and right singular functions given by $p/\|p\|$ and $q/\|q\|$ respectively.

The singular values of the degenerate kernel

$$\tilde{K}_A = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^{(n)} q_i^A(s) p_j^A(t)$$

are the stationary values of

$$\tilde{F}_A[\phi,\psi] = \frac{\langle q^A, \tilde{K}_A p^A \rangle}{\|p^A\|\|q^A\|}.$$

Using the discretization

$$p_A(t) = \sum_{i=1}^{n} y_i^A p_i^A(t) \quad \text{and} \quad q_A(s) = \sum_{i=1}^{n} z_i^A q_i^A(s)$$

the stationary values of $\tilde{F}_A$ are those of

$$G_A[y^A, z^A] = \frac{(z^A)^T A^{(n)} y^A}{\|y^A\|\|z^A\|}$$

which are also the singular values of A [14].

Theorem 3.3 *Let $C^{(n)}$ be the matrix with entries $c_{ij}^{(n)}$ that approximate the operator C using the Galerkin method, then $\sigma_k(C^{(n)}) \leq \sigma_k(C^{(n+1)}) \leq \sigma_i(C),\ k = 1, 2, \ldots n$.*

Proof. The basis functions $\{p_k^A\}_{i=1}^n$ and $\{q_k^A\}_{i=1}^n$ are orthonormal, and the singular values $\sigma_k(A^{(n)})$ and $\sigma_k(A^{(n+1)})$ are the stationary values of

$$F_A[p_A, q_A] = \frac{\langle q_A, K_A p_A \rangle}{\|p_A\|\|q_A\|},$$

restricted to n-dimensional and $n+1$-dimensional function subspaces respectively. Thus the approximate singular values $\sigma_k(A^{(n)})$, where n is the number of basis functions, are increasingly (with n) better approximations to the true singular values $\sigma_k(A)$. A similar statement holds for $\sigma_k(B^{(n)})$ and $\sigma_k(B^{(n+1)})$ with basis functions $\{p_k^B\}_{i=1}^n$ and $\{q_k^B\}_{i=1}^n$, and functional F_B.

The kernel of the direct sum integral operator $C = A \oplus B$ is $K_A \oplus K_B$ [12]. Thus the singular values of the joint operator C are the stationary values of the functional

$$F_C[p_A, q_A, p_B, q_B] = \left(\frac{\langle q_A, K_A p_A \rangle}{\|p_A\|\|q_A\|}, \frac{\langle q_B, K_B p_B \rangle}{\|p_B\|\|q_B\|} \right).$$

The singular values of the discrete joint operator $C^{(n)}$ are the stationary values of the functional

$$G_C[y^A, z^A, y^B, z^B] = \left(\frac{(z^A)^T A^{(n)} y^A}{\|y^A\| \|z^A\|}, \frac{(z^B)^T B^{(n)} y^B}{\|y^B\| \|z^B\|} \right),$$

which are also the singular values of C. Thus the approximate singular values $\sigma_k(C^{(n)})$, where n is the number of basis functions, are increasingly (with n) better approximations to the true singular values $\sigma_k(C)$.

$\square$

If the discretizations $A^{(n)}$ and $B^{(n)}$ are stacked to form $\left[C^{(n)}\right] = \begin{bmatrix} A^{(n)} \\ B^{(n)} \end{bmatrix}$ we get the following result.

Theorem 3.4 *Define*

$$\begin{aligned} \left(\Delta_C^{(n)}\right)^2 &= \|K_A \oplus K_B\|^2 - \left\| \begin{bmatrix} A^{(n)} \\ B^{(n)} \end{bmatrix} \right\|_F^2 \\ &= \sum_{k=1}^{\infty} \sigma_i(C)^2 - \sum_{k=1}^{n} \sigma_i(\left[C^{(n)}\right])^2. \end{aligned}$$

Then $\left(\Delta_C^{(n)}\right)^2 = \left(\Delta_A^{(n)}\right)^2 + \left(\Delta_B^{(n)}\right)^2$*, i.e. the square of the joint error is the sum of squares of the individual errors. Thus if* $\lim_{n\to\infty} \left(\Delta_A^{(n)}\right)^2 = 0$ *and* $\lim_{n\to\infty} \left(\Delta_B^{(n)}\right)^2 = 0$.

Proof.

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

$\square$

This says that if we stack the Galerkin approximations of the individual operators, as it typically done in a discrete joint inversion, the error in the approximation to the singular values of the joint operator converges with n if the errors in the singular values approximations of the individual operators to go zero.

3.3. Self Adjoint Operators

If the operator A is self-adjoint, then $\langle Av, w \rangle = \langle v, Aw \rangle$ for all v and w. This means the singular functions are the eigenfunctions of the operator i.e. $A\phi_k = \sigma_k \phi_k$.

Lemma 3.5 *Let A be a self-adjoint, compact operator, then $A^{(n)} = \Sigma^{(n)}$ where $A^{(n)}$ is formed by the Galerkin method and $\Sigma^{(n)}$ is diagonal with entries $\sigma_i(A^{(n)})$ that approximate the singular values $\sigma_i(A)$.*

Proof. Since A is self-adjoint, in the Galerkin method use orthonormal bases equal to the eigenfunctions ϕ_k to form $A^{(n)}$:

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

□

When joint inversion involves self-adjoint operators, we can form the discrete singular values for the joint stacked operator directly from the individual operators as shown in the following theorem.

Theorem 3.6 *If A and B are compact self adjoint operators, and $\sigma_k(A^{(n)})$ and $\sigma_k(B^{(n)})$ are discrete approximations of the singular values of A and B, respectively, then the discrete approximation of the singular values of the joint operator C are*

$$\sigma_k(\left[C^{(n)}\right]) = \sqrt{\sigma_k(A^{(n)})^2 + \sigma_k(B^{(n)})^2}.$$

Proof. Since A and B are self-adjoint, the Galerkin method produces the matrix of approximate singular values

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

The joint operator $\left[C^{(n)}\right]$ is thus

$$\begin{bmatrix} A^{(n)} \\ B^{(n)} \end{bmatrix} = \begin{bmatrix} \Sigma_A^{(n)} \\ \Sigma_B^{(n)} \end{bmatrix}.$$

The singular values of $\left[C^{(n)}\right]$ are the square roots of the eigenvalues of

$$\left(\left[C^{(n)}\right]\right)^T \left[C^{(n)}\right] = \left[\Sigma_A^{(n)} \ \Sigma_B^{(n)}\right] \begin{bmatrix} \Sigma_A^{(n)} \\ \Sigma_B^{(n)} \end{bmatrix} = \left[\left(\Sigma_A^{(n)}\right)^2 + \left(\Sigma_B^{(n)}\right)^2\right].$$

□

It is very useful to have an analytical expression for the singular values of the joint operator as a function of the singular values of the individual operators. It allows us to determine the decay rate of the joint operator before the joint inversion and characterize the joint problem as as mildly, moderately or severely ill-conditioned.

Corollary 3.7 *The characterization of the ill-conditioning (i.e. mild, moderate or severe) of the discrete stacked joint problem $\left[C^{(n)}\right]$ is the same as that of the least ill-posed problem A or B, as $k \to \infty$. In this case, the conditioning of the joint problem will never be worse than that of the individual problems.*

Proof. We show this by considering different cases for the decay rates of the singular values of $A^{(n)}$ and $B^{(n)}$ and applying Theorem 3.6. A problem is mildly or moderately ill-conditioned if the singular values σ_k decay like $\mathcal{O}(k^{-q})$ and severely ill-conditioned if they decay like $\mathcal{O}(e^{-qk})$ for $q > 0$ [10].

(i) If $\sigma_k(A^{(n)}) = \mathcal{O}(k^{-q_A})$ and $\sigma_k(B^{(n)}) = \mathcal{O}(k^{-q_B})$ then
$$\sigma_k(\left[C^{(n)}\right])^2 \leq c_{_A} k^{-2q_A} + c_{_B} k^{-2q_B} \leq c k^{-2q}$$
where $q = \min(q_A, q_B)$.

(ii) If $\sigma_k(A^{(n)}) = \mathcal{O}(e^{-q_A k})$ and $\sigma_k(B^{(n)}) = \mathcal{O}(e^{-q_B k})$ then
$$\sigma_k(\left[C^{(n)}\right])^2 \leq c_{_A} e^{-2q_A k} + c_{_B} e^{-2q_B k} \leq c e^{-2qk}$$
where $q = \min(q_A, q_B)$.

(iii) If $\sigma_k(A^{(n)}) = \mathcal{O}(k^{-q_A})$ and $\sigma_k(B^{(n)}) = \mathcal{O}(e^{-q_B k})$ then
$$\sigma_k(\left[C^{(n)}\right])^2 \leq c_{_A} k^{-2q_A} + c_{_B} e^{-2q_B k} \leq c k^{-2q_A}.$$
The last inequality holds because as $k \to \infty$ $q_{_A} \leq q_B \frac{k}{\ln(k)}$ for any $q_{_A}$ or q_B and hence $e^{-q_B k} \leq k^{-q_A}$.

$\square$

Corollary 3.8 *For a finite set of singular values (e.g. the truncated SVD) if $\sigma_k(A^{(n)}) = \mathcal{O}(k^{-q_A})$ and $\sigma_k(B^{(n)}) = \mathcal{O}(e^{-q_B k})$ with $q_B \leq e^1 q_A$ there is k for which $k^{-q_A} \leq e^{-q_B k}$ and hence the conditioning of the joint problem $\left[C^{(n)}\right]$ will be worse than the conditioning of the better posed problem A.*

Proof. If $q_{_B} \leq \frac{\ln(k)}{k} q_A$ then there is k for which $k^{-q_A} \leq e^{-q_B k}$ and
$$\sigma_k(\left[C^{(n)}\right])^2 \leq c_{_A} k^{-2q_A} + c_{_B} e^{-2q_B k} \leq c e^{-2q_B k}.$$
The result follows with $q_{_B} \leq \frac{\ln(k)}{k} q_{_A} \leq e^1 q_{_A}$.

$\square$

While combining multiple data sets in an inversion should produce a better conditioned problem most of the time, a severely posed problem may degrade a mildly or moderately ill-posed problem if the singular values are truncated in a joint inversion.

4. Green's Functions Example

We show results from combining data from two distinct boundary value problems
$$\begin{aligned} &-u''(x) = h(x), \quad u(0) = u(\pi) = 0 \\ &u''(x) + b^2 u(x) = h(x), \quad u(0) = u(\pi) = 0, \quad b \notin \mathbb{Z} \end{aligned}$$
with $L_{_A} u = -u''$ and $L_B u = u'' + b^2 u$. The Green's functions for both differential operators are given in [21]. In particular for $A : L^2[0, \pi] \to L^2[0, \pi]$ we have
$$Ah(x) = \int_0^\pi K_{_A}(x, y)\, h(y) dy,$$
with
$$K_{_A} = \begin{cases} \frac{1}{\pi}(\pi - x)\, y, & 0 \leq y \leq x \leq \pi, \\ \frac{1}{\pi}(\pi - y)\, x, & 0 \leq x \leq y \leq \pi \end{cases}$$

and $Ah(x) = u(x)$. Similarly for $B : L^2[0,\pi] \to L^2[0,\pi]$ we have

$$Bh(x) = \int_0^\pi K_B(x,y)\,h(y)dy,$$

with

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

and $Bh(x) = u(x)$.

A is a self-adjoint compact operator, it admits an eigenvalue expansion and the singular values of A are the absolute value of its eigenvalues. The equation $A\phi = \lambda\phi$ is equivalent to

$$\int_0^\pi K_A(x,y)\,\phi(y)dy = \lambda\phi(x)$$

with $\phi(0) = 0$ and $\phi(\pi) = 0$. Differentiating both sides twice with respect to x and applying the Leibniz integral rule gives

$$\begin{aligned}\lambda\phi''(x) &= \frac{d}{dx}\left(\int_0^x -\frac{1}{\pi}y\phi(y)dy + \int_x^\pi \frac{1}{\pi}(\pi-y)\,\phi(y)dy\right) \\ &= -\frac{1}{\pi}x\phi(x) - \frac{1}{\pi}(\pi-x)\,\phi(x) \\ &= -\phi(x).\end{aligned}$$

This yields the eigenvalue-eigenfunction pairs $(\lambda_k, \phi_k(x))$:

$$\phi_k(x) = \sqrt{\frac{2}{\pi}}\sin(kx), \qquad \lambda_k = \frac{1}{k^2} \quad k = 1,2,\ldots,\infty.$$

The singular values for A are thus $\sigma_k(A) = \frac{1}{k^2}$ for $k = 1,2,\ldots,\infty$. This means the decay rate of the singular values are $\mathcal{O}(k^{-2})$ and the problem is moderately ill-posed [10].

B is also a self-adjoint operator and its eigenvalues are

$$\lambda_k = \frac{1}{k^2+b^2}, \quad \text{for } k = 0,1,\ldots,\infty.$$

The singular values for B are thus $\sigma_k(B) = \frac{1}{k^2+b^2}$. We omit the eigenfunctions since the decay rate of the singular values is the focus of this work. The singular values have the same decay rate as that for A and this problem is also moderately ill-posed.

4.1. Joint Singular Values

The joint operator $C : L^2[0,\pi] \to L^2[0,\pi] \oplus L^2[0,\pi]$ is

$$Ch(x) = \int_0^\pi K_A(x,y)\,h(y)dy \;\oplus \int_0^\pi K_B(x,y)\,h(y)dy.$$

If we were to use the same approach to finding the singular values of A and B to now find the joint singular values of C, the result would be a linear constant coefficient ODE with a eighth order characteristic polynomial. Alternatively we use the Galerkin method presented in Section 3.1.

The discretizations $A^{(n)}$ and $B^{(n)}$ approximate the operators A and B with orthonormal bases. Since A and B are self-adjoint we use the eigenfunction bases and apply Theorem 3.6. The singular values of the stacked joint operator are thus

$$\sigma_k(\left[C^{(n)}\right]) = \sqrt{\left(\frac{1}{k^2}\right)^2 + \left(\frac{1}{k^2+b^2}\right)^2}, \quad k = 1, \ldots, n.$$

This shows that the moderately ill-posed problems A and B are combined to form a joint moderately ill-posed problem. While the overall conditioning of the problem has not changed through joint inversion, we will quantify the benefits of the joint problem by determining the number singular values in a truncated singular value expansion (TSVD) for each problem.

The TSVD is a regularization method whereby small singular values are discarded so that the problem is well-conditioned. However, information is lost when singular values are discarded and therefore we wish to keep as many as possible. Let the number of singular values in the TSVD be denoted by r with r chosen by requiring that $\sigma_r \geq T$ for small T. The further T is from zero the better conditioned the problem, however the solution will also be less accurate. If the number of singular values in the TSVD for A are denoted by r_A, then $r_A \leq \sqrt{T^{-1}}$. Similarly, the number of singular values in the TSVD for B satisfy $r_B \leq \sqrt{T^{-1} - b^2}$ with $T^{-1} > b^2$.

For joint inversion using $\left[C^{(n)}\right]$ we have that the number of singular values r_C satisfy

$$\frac{1}{r_C^4} + \frac{1}{(r_C^2 + b^2)^2} \geq T^2.$$

Solving we get

$$r_C = \sqrt{\frac{\sqrt{4\sqrt{T^2b^4+1} + T^2b^4 + 4} - Tb^2}{2T}}.$$

Now for $T \ll 1$ if $b \approx 1$ we approximate $\sqrt{T^2b^4+1} \approx \sqrt{1}$ and $\sqrt{4\sqrt{T^2b^4+1} + T^2b^4 + 4} \approx \sqrt{8}$ so that $r_C \approx \sqrt{\frac{\sqrt{8}-Tb^2}{2T}}$. The percent increase in number of singular values we keep with jointly inverting A and B rather than just A is

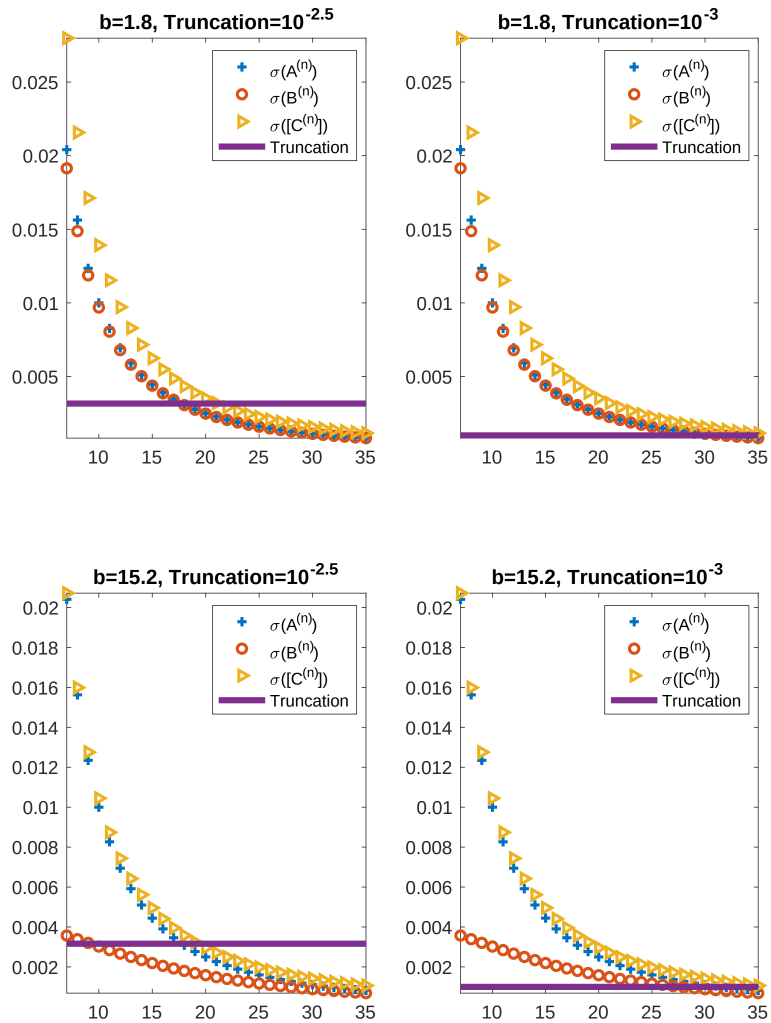
$$\frac{r_C}{r_A} = \sqrt{\frac{\sqrt{8} - Tb^2}{2}} \approx \sqrt{\frac{\sqrt{8}}{2}} \approx 19\%.$$

A similar statement can be made for jointly inverting A and B rather than just B.

The singular values for $A^{(n)}$, $B^{(n)}$ and $\left[C^{(n)}\right]$ with $n = 35$ are given in Figure 4.1. When $b = 1.8$ there is not much difference between the singular values of $A^{(n)}$ and $B^{(n)}$, which is to be expected. Truncation often occurs about the point where the singular values stop changing and we've indicated two values at which to truncate, one on the left column of the Figure and another on the right column.

In Table 1 we give the number of singular values that are kept after truncating. For $b = 1.8$ the TSVD for $A^{(n)}$ and $B^{(n)}$ results in the same number of singular values, for both values of truncation. The number of singular values kept for TSVD with the same

Figure 1. Singular values for the individual inversions with A and B and joint inversion C for $b = 1.8$ (top row) and $b = 15.2$ (bottom row). Two different thresholds for truncation are also represented, $T = 10^{-2.5}$ (left) and $T = 10^{-3}$ (right).



threshold increases with joint inversion. This increase in number of singular values is 23% when singular values are truncated at $10^{-2.5}$ and 19% when truncated at $10^{-3.0}$.

When $b = 15.2$ the singular values of B change behavior from those of A. The singular values of B drop off more quickly and hence fewer are kept in a TSVD. In Figure 4.1 the singular values σ_k start at $k = 7$ in all cases to make the graph more readable.

We see in Table 1 that for $b = 15.2$ and with a lower threshold, only 9 of the singular values are kept in the TSVD for B. In a joint inversion with A this number is doubled

Table 1. Number of singular values in TSVD for $A^{(n)}$ (r_A), $B^{(n)}$ (r_B) and $\left[C^{(n)}\right]$ (r_C). The last column gives the increase in number of singular values that are kept in the TSVD for the given threshold T.

Threshold T	r_A	r_B	r_C	Increase in number of σ kept for C over (A,B)
b=1.8				
$10^{-2.5}$	17	17	21	(23%,23%)
$10^{-3.0}$	31	31	37	(19%,19%)
b=15.2				
$10^{-2.5}$	17	9	18	(5%,100%)
$10^{-3.0}$	31	27	36	(16%,33%)

and hence the increase in number of values kept with joint inversion over those kept in B is 100%. The increase in the number of singular values over those in A is much smaller and we see that data from this model better informs the joint inversion. For a lower threshold with $b = 15.2$ more singular values are kept in TSVD and the contributions from A and B are better balanced. This could be at the cost of amplifying noise, and the conclusion of which threshold to use in the TSVD is done the context of the noise of the data.

5. Conclusions and Future Work

We have extended singular value analysis of discrete joint inversion to joint inversion of compact linear operators. The analogous operation to stacking discrete matrices is the direct sum of operators and we give results regarding the singular value expansion of the joint operator. Joint inversion can be computationally expensive and in some instances it is not clear if it improves inversions of individual operators. Therefore, we quantify improvement in jointly inverting two operators by comparing the decay rate of the singular values of the joint operator to those from the individual operators.

Tikhonov regularization with compact linear operators is also the direct sum of operators. The regularization parameter can always be chosen so that an ill-posed problem is made well-posed. However, the parameter restricts the solution space in an ad-hoc manner. Alternatively, joint inversion restricts the solution space using additional data. We suggest analyzing the singular values of joint operators to determine which types of data are best to combine before data are collected. This analysis effectively determines which data "regularize" each other and can inform experimental design.

We also developed a method for approximating the singular values of the joint compact operator. The infinitely many singular values are approximated with a Galerkin method. For self-adjoint operators we obtained an analytic formula for the joint operator as a function of the singular values of the individual operators. We calculated singular values arising from joint inversion of Green's function solutions of two simple ordinary differential equations and compared them to the singular values of the individual operators. In this example the conditioning of the joint problem is not significantly better that that of the best conditioned problem. However, the conditioning of both moderately ill-posed problems are improved through joint inversion. These conclusions confirm what was proved as the typical case of joint inversion of self adjoint operators.

6. Acknowledgments

The authors thank John Bradford and Diego Domenzain whose development of joint inversion of ground penetrating radar and electrical resistitvity to image the near subsurface of the earth motivated this work.

They would also like to acknowledge DMS NSF for supporting this work.

[1] Vozoff K and Jupp D 1975 *Geophysical Journal International* **42** 977–991

[2] Evans S N and Stark P B 2002 *Inverse Problems* **18** R55–R97

[3] Lan T, Liu H, Liu N, Li J, Han F and Liu Q H 2018 *IEEE Transactions on Geoscience and Remote Sensing* **56** 436–445 ISSN 0196-2892

[4] Hu W, Abubakar A and Habashy T M 2009 *GEOPHYSICS* **74** R99–R109

[5] Abubakar A, Gao G, Habashy T M and Liu J 2012 *Inverse Problems* **28** 055016

[6] Haber E and Oldenburg D 1997 *Inverse Problems* **13** 63–77

[7] Fregoso E and Gallardo L A 2009 *GEOPHYSICS* **74** L31–L42

[8] Gallardo L A and Meju M A 2004 *Journal of Geophysical Research: Solid Earth* **109**

[9] Saunders J H, Herwanger J V, Pain C C, Worthington M H and De Oliveira C R E 2005 *Geophysical Journal International* **160** 785–796

[10] Aster R C, Borchers B and Thurber C H 2018 *Parameter estimation and inverse problems* (Elsevier Academic Press, Amsterdam, The Netherlands)

[11] Gockenbach M 2015 *Linear Inverse Problems and Tikhonov Regularization* (The Mathematical Association of America)

[12] Halmos P R and Sunder V S 1978 *Bounded integral operators on L2 spaces* Ergebnisse Mathematik und GrenzGeb. (Berlin: Springer)

[13] Stewart G 1993 *SIAM Review* **35** 551–566

[14] Hansen P C 1988 *Computing* **40** 185–199

[15] Kirsch A 2011 *An introduction to the mathematical theory of inverse problems* vol 120 (Springer Science & Business Media)

[16] Kress R, Maz'ya V and Kozlov V 1989 *Linear integral equations* vol 82 (Springer)

[17] Gockenbach M S 2016 *SIAM Journal on Numerical Analysis* **54** 2517–2540

[18] Hori M 2001 *Geophysical Journal International* **147** 77–87

[19] Conway J 1994 *A Course in Functional Analysis* Graduate Texts in Mathematics (Springer New York) ISBN 9780387972459

[20] Renaut R A, Horst M, Wang Y, Cochran D and Hansen J 2016 *BIT Numerical Mathematics* 1–31

[21] Duffy D G 2015 *Green's Function with Applications* (CRC Press)