Singular Value Decomposition (SVD) of General Linear Operators

February 21, 2006

1 Kernel description (Green's functions)

A spatially invariant operator can be described using a convolution kernel which is just the response of the system to an impulse function located at the origin. Thus, if A is a linear operator, representable abstractly by

$$d = Gm,$$

then there is an impulse response $h = G\delta$ so that

$$d(\mathbf{y}) = \int h(\mathbf{y} - \mathbf{x}) f(\mathbf{x})\, d\mathbf{x}$$

In the more general spatially invariant case, where the impulse response is located makes a difference in not only the location of the output, but also it's shape. Therefore, we have to keep track of multiple impulse responses—one for each value of $\mathbf{x}$—in order to describe the linear operator. Specifically, we need $h_{\mathbf{x}} = G\delta_{\mathbf{x}}$ where $\delta_{\mathbf{x}}(\mathbf{y}) = \delta(\mathbf{y} - \mathbf{x})$ is an impulse function located at position $\mathbf{x}$ (but thought of as a function of $\mathbf{y}$). If we have this function then

$$\begin{aligned} m(\mathbf{y}) &= \int \delta(\mathbf{y} - \mathbf{x})\, m(\mathbf{x})\, d\mathbf{x} \\ d(\mathbf{y}) = (Gm)(\mathbf{y}) &= \int h_{\mathbf{x}}(\mathbf{y})\, m(\mathbf{x})\, d\mathbf{x} \end{aligned}$$

where again $h_{\mathbf{x}}(\mathbf{y}) = A\delta(\mathbf{y} - \mathbf{x})$. In perhaps an abuse of (multi-)index notation we can simplify matters by writing these integrations as

$$\begin{aligned} m^y &= \delta^y_x m^x \\ d^y &= G^y_x m^x, \end{aligned}$$

where by definition $G^y_x \equiv h_{\mathbf{x}}(\mathbf{y})$.

Using index notation, we can treat general linear operators in the same way that we treat finite-dimensional linear operators. The only difference is that repeated indices imply an integration. Whether to sum or integrate should be apparent from context.

The kernel function $h_{\mathbf{x}}(\mathbf{y})$ is sometimes called the Green's function of the operator (this terminology is typically used when G is described implicitly as a partial differential equation with some set of boundary conditions). You can see that the Green's function is nothing more than the description of the impulse responses of the operator. Finding the Green's function is a matter of putting in impulse functions at different locations and computing the resulting output. This can be done empirically or analytically depending on the problem.

2 Adjoint

The adjoint of a general linear operator in terms of the kernel can be found from the definition of the adjoint

$$(Gm, d)_{\mathbb{D}} = \left(m, G^A d\right)_{\mathbb{M}}.$$

where the inner products are taken over possibly different spaces. To find the components of G^A we compute:

$$\begin{aligned}(Gm,d)_{\mathbb{D}} &= G^x_z m^z W^{D*}_{xy} d^{y*}\\ \left(m,G^A d\right)_{\mathbb{M}} &= m^x W^{M*}_{xy}\left(G^A\right)^{y*}_z d^{z*},\end{aligned}$$

where W^M_{xy} and W^D_{xy} are (hermititian-symmetric and positive definite) weighting operators for the space (mapping from the space to its dual) and the inner product is defined in terms of the duality product induced by these weigthing operators. Because the result must hold for all m and d, it must be that

$$\begin{aligned}m^{x*}W^M_{xy}\left(G^A\right)^y_z d^z &= G^{x*}_z m^{z*} W^D_{xy} d^y\\ &= m^{x*} G^{y*}_x W^D_{yz} d^z\end{aligned}$$

or

$$W^M_{xy}\left(G^A\right)^y_z = G^{y*}_x W^D_{yz}.$$

This relationship shows explicitly how to get the adjoint operator from the kernel (impulse responses) of the linear operator. Where possible, the weighting operator can be inverted to get the formal covariance operator C^{xy}_M defined so that $C^{xy}_M W^M_{yz} = \delta^x_z$, and then

$$\left(G^A\right)^y_x = C^{yz}_M G^{w*}_z W^D_{wx},$$

or formally $G^A = C_M G^H W^D$. If the weigthing operators are both δ functions then we have the simpler condition

$$\left(G^A\right)^y_x = G^{x*}_y.$$

Notice that $\left(G^A\right)^A = C_D\left(G^A\right)^H W^M = C_D\left(C_M G^H W^D\right)^H W^M = G$.

3 Self-adjoint operators

Often we use the adjoint of an operator in combination with itself, and it is useful to have kernel representations of these operators. In other words, define $\bar{G} = G^A G$ (this is a map from $\mathbb{M}$ to $\mathbb{M}$). The kernel of this operator is

$$\begin{aligned}\bar{G}^x_{x'} &= \left(G^A\right)^x_y G^y_{x'}\\ &= C^{xz}_M G^{w*}_z W^D_{wy} G^y_{x'}.\end{aligned}$$

On the other hand, the operator $\tilde{G} = GG^A$ is a map from $\mathbb{D}$ to $\mathbb{D}$ with kernel

$$\tilde{G}^y_{y'} = G^y_x C^{xz}_M G^{w*}_z W^D_{wy'}$$

A self-adjoint operator is one whose adjoint is equal to itself. Look, for example, at the adjoint of $\overline{G}$:

$$\begin{aligned}\left(\overline{G}^A\right)^x_{x'} &= C^{xz}_M\left(\overline{G}\right)^{w*}_z W^M_{wx'}\\ &= C^{xz}_M C^{w\alpha*}_M G^\beta_\alpha W^{D*}_{\beta\gamma} G^{\gamma*}_z W^M_{wx'}\\ &= C^{xz}_M \delta^\alpha_{x'} G^\beta_\alpha W^D_{\gamma\beta} G^{\gamma*}_z\\ &= C^{xz}_M G^{\gamma*}_z W^D_{\gamma\beta} G^\beta_{x'} = \overline{G}^x_{x'}.\end{aligned}$$

To arrive at this result we used the fact that $C^{w\alpha*}_M = C^{\alpha w}_M$ and $W^D_{\beta\gamma} = W_{\gamma\beta}$ (as required for weighting operators). Similarly we can show that $\tilde{G}^A = G$. In addition,

$$\begin{aligned}\left(\overline{G}m,m\right)_{\mathbb{M}} &= \left(G^A Gm,m\right)\\ &= (Gm,Gm)_{\mathbb{D}}\\ &= \|Gm\|^2_{\mathbb{D}} \geq 0.\end{aligned}$$

Similarly,

$$\begin{aligned}\left(\tilde{G}d, d\right) &= \left(GG^A d, d\right)_{\mathbb{D}} \\ &= \left(G^A d, G^A d\right)_{\mathbb{M}} \\ &= \left\|G^A m\right\|_{\mathbb{M}}^2 \geq 0.\end{aligned}$$

and therefore, $\overline{G}$ and $\tilde{G}$ are positive semi-definite.

4 Eigenvalues and eigenvectors (spectral representation)

The (right) eigenvalue problem for discrete square linear operators is well-known as the problem of finding constants (eigenvalues) λ_k and eigenvectors $\mathbf{v}_k$ satisfying

$$\mathbf{A}\mathbf{v}_k = \lambda_k \mathbf{v}_k.$$

In other-words, eigenvectors are vectors that are only scaled by application of the linear operator. Notice that the matrix $\mathbf{A}$ must be square. A similar problem can be defined for linear operators defined on the same space (*i.e.* the linear operator maps $\mathbb{M}$ to $\mathbb{M}$). An eigenvalue, eigenfunction pair of a linear operator satisfies the relationship

$$Av_k = \lambda_k v_k.$$

In index-notation this is

$$A_\alpha^x v_k^\alpha = \lambda_k v_k^x$$

or in integral form:

$$\int h_{\mathbf{x}'}(\mathbf{x})\, v_k(\mathbf{x}')\, d\mathbf{x}' = \lambda_k v_k(\mathbf{x}).$$

Thus, an eigenvector (eigenfunction) of a linear operator is a vector that is only scaled under action of the operator but otherwise unchanged.

For example, for spatially-invariant (Toeplitz) operators, you may recall that $v_{\mathbf{s}}(\mathbf{x}) = e^{\mathbf{s}\cdot\mathbf{x}}$ is an eigenfunction with eigenvalue

$$H(\mathbf{s}) = \int h(\mathbf{y})\, e^{-\mathbf{s}\cdot\mathbf{y}} d\mathbf{y}.$$

This can be easily shown because

$$\begin{aligned}\int h(\mathbf{x}-\mathbf{x}')\, e^{\mathbf{s}\cdot\mathbf{x}'} d\mathbf{x}' &= \int h(\mathbf{y})\, e^{\mathbf{s}\cdot(\mathbf{x}-\mathbf{y})} d\mathbf{y} \\ &= \left[\int h(\mathbf{y})\, e^{-\mathbf{s}\cdot\mathbf{y}}\right] e^{\mathbf{s}\cdot\mathbf{x}}.\end{aligned}$$

This fact was used to great advantage to understand spatially invariant operators in terms of their frequency response, $H(j\boldsymbol{\omega})$.

In the spatially-invariant case, the eigenvalues of the spatially invariant operator were a useful domain to work in. This motivates us to consider looking at the eigenvalue spectrum for the general linear-operator case as well. In fact, we call the eigenvalues the spectrum of the operator as a generalization of the fact that for spatially-invariant operators, the eigenvalues are precisely the transfer-function of the impulse response evaluated at different frequencies. The problem with applying this analysis to general linear operators is that these may not map to the same space. Thus, we need something more general then eigenvector-eigenvalue analysis.

However, for general linear operators that are self-adjoint with a square-integrable kernel, we can find the eigenfunctions and eigenvalues of the operator. There is a branch of mathematics known as Hilbert-Schmidt theory which deals with the eigenvalue decomposition of self-adjoint linear operators. According to this theory, a self-adjoint operator with square-integrable kernel has real eigenvalues with finite multiplicity. In addition, eigenfunctions associated with different eigenvalues are orthogonal. The eigenvalues form a countable set and accumulate to zero (decaying spectrum). A finite-rank linear operator is one that has only M non-zero eigenvalues with M some fixed integer. The rank of the operator is M.

This fact that the eigenvalues of a self-adjoint linear operator accumulate to zero is the reason that every problem requiring inversion of a linear operator is an inverse-problem (spectral information is lost). The degree of the inverse problem can be measured in terms of how quickly the eigenvalues of the operator tend toward zero.

Order the eigenvalues of the self-adjoint operator (they are all real) as $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3| \geq \cdots$, each eigenvalue being counted as many times as its multiplicity. Also, let $v_1(\mathbf{x})$, $v_2(\mathbf{x})$, $v_3(\mathbf{x})$, $\cdots$ be the sequence of eigenfunctions associated with these eigenvalues. From Hilbert-Schmidt theory, these eigenvectors form an orthonormal (and complete) set of functions. Because we can represent m by expanding on this basis

$$m = \sum_k m^k v_k = \sum_k (m, v_k)\, v_k$$

we can represent

$$Am = \sum_k \lambda_k m^k v_k = \sum_k \lambda_k (m, v_k)\, v_k.$$

In terms of components

$$A^x_\chi m^\chi = \sum_k \lambda_k v^x_k m^\chi W^{M*}_{\chi\alpha} v^{\alpha*}_k$$

This gives the *spectral representation* of A as

$$A^x_{x'} = \sum_{k=1}^{\infty} \lambda_k v^x_k \hat{v}^{k*}_\chi$$

where

$$\hat{v}^k_\chi \equiv W^M_{\chi\alpha} v^\alpha_k$$

is the corresponding bi-orthogonal dual basis. The series is convergent in the sense of the L^2 norm. Thus, the eigenvalues satisfy

$$\sum_{k=1}^{\infty} \lambda_k^2 < \infty.$$

Note: What does this fact that the sum of squared eigenvalues is finite say about the sequence of eigenvalues? How does this hamper the inverse problem?

Notice that this series expansion allows writing

$$Am = \sum_{k=1}^{\infty} \lambda_k (m, v_k)\, v_k.$$

5 Singular Value Decomposition

The singular value decomposition (SVD) provides the generalization of eigenvector-eigenvalue analysis for general linear operators. It can be applied to any linear operator with a square-integrable kernel. The idea for the SVD comes from analyzing the operators $\tilde{G} = GG^A$ and $\bar{G} = G^A G$. Both of these operators are self-adjoint and square-integrable (if the kernel of G is square integrable) as shown previously. Therefore, they have a spectral decomposition. Because they are also positive semi-definite, the eigenvalues are all positive. Let σ_k^2 be the ordered (positive) eigenvalues of $\overline{G}$ with v_k being the normalized associated eigenfunctions. Note that $\sigma_k^2 \to 0$ as $k \to \infty$ and $(v_k, v_j)_{\mathbb{M}} = \delta_{kj}$.

For every non-zero singular value, we can associate a vector u_k with every vector v_k of $\overline{G}$ for which $\sigma_k \neq 0$, we can associate a function u_k in $\mathbb{D}$ defined by

$$u_k = \frac{1}{\sigma_k} G v_k.$$

Notice that

$$\tilde{G}u_k = \frac{1}{\sigma_k}GG^AGv_k = \frac{1}{\sigma_k}G\bar{G}v_k = \frac{\sigma_k^2}{\sigma_k}Gv_k = \sigma_k^2 u_k.$$

In addition,

$$\begin{aligned}(u_k, u_j)_{\mathbb{D}} &= \frac{1}{\sigma_k\sigma_j}(Gv_k, Gv_j)_{\mathbb{D}} \\ &= \frac{1}{\sigma_k\sigma_j}\left(v_k, G^AGv_j\right)_{\mathbb{M}} \\ &= \frac{\sigma_j}{\sigma_k}(v_k, v_j)_{\mathbb{M}} = \delta_{kj}.\end{aligned}$$

This shows that u_k is an eigenfunction of the operator $\tilde{G}$ and σ_k^2 is also an eigenvalue of $\tilde{A}$. In addition, the $\{u_k\}$ constitute an orthonormal system in $\mathbb{D}$. Note that it is not necessarily true that $\mathbb{D} = \text{span}\{u_k\}$ but only that $\text{span}\{u_k\} = \mathcal{R}(G) = \mathcal{N}\left(G^A\right)^{\perp} \subseteq \mathbb{D}$. The square-root of the positive eigenvalues of $\tilde{G}$ and $\bar{G}$ are called the *singular values* of G.

We can reverse this procedure and show that if u_k and σ_k^2 denote the eigenfunctions and (non-zero) eigenvalues of $\tilde{G}$, then

$$v_k = \frac{1}{\sigma_k}G^Au_k$$

give orthonormal eigenfunctions of $\bar{G}$ with eigenvalues σ_k^2. In addition $\text{span}\{v_k\} = \mathcal{N}(G)^{\perp} = \mathcal{R}\left(G^A\right) \subseteq \mathbb{A}$.Thus, we have the pair of shifted eigenvalue problems,

$$\begin{aligned}\sigma_k v_k &= G^A u_k \\ \sigma_k u_k &= Gv_k.\end{aligned}$$

Notice that by limiting the span of the basis set $\{v_k\}$ to be the orthogonal complement of the null space of G we have restricted all singular values to be positive. This ensures that the expression $\frac{1}{\sigma_k}$ always makes sense.

The SVD (spectral) decomposition of the operator A uses the singular values and eigenfunctions to represent the operator. Because the u_k can form a basis for (a subspace of) $\mathbb{D}$ we can write

$$\begin{aligned}Gm &= \sum_{k=1}^{\infty}(Gf, u_k)_{\mathbb{D}}\, u_k \\ &= \sum_{k=1}^{\infty}\left(m, G^Au_k\right)_{\mathbb{M}} u_k \\ &= \sum_{k=1}^{\infty}\frac{1}{\sigma_k}\left(m, G^AGv_k\right)_{\mathbb{M}} u_k \\ &= \sum_{k=1}^{\infty}\frac{\sigma_k^2}{\sigma_k}(m, v_k)_{\mathbb{M}}\, u_k.\end{aligned}$$

Therefore,

$$Gm = \sum_{k=1}^{\infty}\sigma_k\,(m, v_k)_{\mathbb{M}}\, u_k.$$

In a similar fashion we can show that

$$G^Ad = \sum_{k=1}^{\infty}\sigma_k\,(d, u_k)_{\mathbb{D}}\, v_k.$$

These are equivalent to representing G and G^A as the sum of separable functions:

$$\begin{aligned}G_x^y &= \sigma_k u_k^y \hat{v}_x^{k*} \\ \left(G^A\right)_x^y &= \sigma_k v_k^y \hat{u}_x^{k*},\end{aligned}$$

where

$$\hat{u}_x^k = W^D_{x\alpha} u_k^\alpha.$$

The singular values show which basis vectors are important to the The infinite sums are convergent with respect to the L^2-norm.

Notice that the Frobenius norm is

$$\begin{aligned}
\|G\|_F^2 &= G_x^y \left(G^A\right)_y^x \\
&= \sigma_k \sigma_l u_k^y \hat{v}_x^{k*} \hat{u}_y^{k*} v_l^x \\
&= \sum_{k,l} \sigma_k \sigma_l \left(u_k, u_l\right)_{\mathbb{D}} \left(v_k, v_l\right)_{\mathbb{M}} \\
&= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \sigma_k \sigma_l \delta_{kl} \delta_{kl} \\
&= \sum_{k=1}^{\infty} \sigma_k^2.
\end{aligned}$$

The norm of a linear operator is

$$\|G\| = \sup_{f \in \mathcal{X}} \frac{\|Gm\|_{\mathbb{D}}}{\|m\|_{\mathbb{M}}}.$$

Using the SVD we can see that

$$\begin{aligned}
\|Gm\|_{\mathbb{D}}^2 &= \left(Gm, Gm\right), \\
&= \sum_{k=1}^{\infty} \sigma_k \sigma_l \left(m, v_l\right)^* \left(m, v_k\right) \left(u_k, u_l\right) \\
&= \sum_{k=1}^{\infty} \sigma_k^2 \left|\left(m, v_k\right)\right|^2.
\end{aligned}$$

Also,

$$m = \sum_{k=1}^{\infty} \left(m, v_k\right) v_k$$

because the v_k form an orthonormal set, thus,

$$\|m\|^2 = \sum_{k=1}^{\infty} \left|\left(m, v_k\right)\right|^2,$$

and it can then be shown that

$$\|A\| = \sigma_1,$$

the largest singular value.

6 Inverse Problem with general linear operators

We can loosely define an inverse for spatially varying linear operators as

$$\left(G^\dagger d\right)(\mathbf{x}) = \int \frac{D(\mathbf{f})}{H_G(\mathbf{f})} e^{j2\pi \mathbf{f}\cdot\mathbf{x}} d\mathbf{f},$$

where we understand that the integration is limited to regions in the frequency domain such that $H_G(\mathbf{f}) \neq 0$. We can describe a similar inverse for general linear operators with an SVD.

Using the SVD we have the representation

$$
\begin{aligned}
Gm &= \sum_{k=1}^{\infty} \sigma_k \, (m, v_k)_{\mathbb{M}} \, u_k \\
d &= \sum_{k=1}^{\infty} (d, u_k)_{\mathbb{D}} \, u_k .
\end{aligned}
$$

Therefore the equation $d = Gm$ can be written component-wise as

$$
(d, u_k)_{\mathbb{D}} = \sigma_k \, (m, v_k)_{\mathbb{M}} \, ,
$$

and we can formally write

$$
\begin{aligned}
A^{\dagger} d &= \sum_{k=1}^{\infty} (m, v_k)_{\mathbb{M}} \, v_k \\
&= \sum_{k=1}^{\infty} \frac{1}{\sigma_k} \, (d, u_k)_{\mathbb{D}} \, v_k .
\end{aligned}
$$

The reason this is only a formal definition is that the sum may not (and very often does not) converge.

For a finite sum there is no convergence difficulty and the formula provides an explicit construction of the inverse. For the infinite sum there are some formal difficulties as the infinite sum may not converge for all d. Convergence is assured (in the sense of the norm of $\mathbb{M}$) if

$$
\sum_{k=1}^{\infty} \frac{1}{\sigma_k^2} \left| (d, u_k)_{\mathbb{D}} \right|^2 < \infty .
$$

This is called the *Picard criterion* for the existence of solutions of the linear inverse problem. Note, that if the data is noise-free then

$$
d = Gm = \sum_{k=1}^{\infty} \sigma_k \, (m, v_k)_{\mathbb{M}} \, u_k
$$

and the Picard criterion sum reduces to

$$
\sum_{k=1}^{\infty} \left| (m, v_k) \right|^2 < \infty ,
$$

which is true if m has finite norm.

This technical jargon may be very interesting to mathematicians, but practically it does little for us at this point in our understanding because even if the sum technically converges, a little noise added to the data will cause the converged value to be unusable. We cannot use this formal definition of the the inverse of a linear operator very often. To make it useful we have to use a-priori information to modify the result.

7 Error sensitivity in the generalized inverse (condition number)

Even though the generalized inverse presents no theoretical difficulty in the case of a finite number of non-zero singular values, inversion using finite-precision arithmetic can still lead to trouble due to the potential sensitivity of the generalized inverse to deviations of the data. To quantify this sensitivity notice that a change in the data, δd, results in an alteration of the inverse of

$$
\delta m = \sum_{k=1}^{p} \frac{1}{\sigma_k} \, (\delta d, u_k) \, v_k ,
$$

where p is the number of non-zero singular values. The norm of this alteration is

$$\|\delta m\| = \sum_{k=1}^{p} \frac{1}{\sigma_k^2} \left|(\delta m, u_k)\right|^2 \leq \frac{1}{\sigma_p^2} \left\|\delta d\right\|^2.$$

The norm of the generalized inverse is

$$\|m\| = \sum_{k=1}^{p} \frac{1}{\sigma_k^2} \left|(d, u_k)\right|^2 \geq \frac{1}{\sigma_1^2} \left\|d\right\|^2.$$

Combining these inequalities shows that the relative error in the generalized inverse is

$$\frac{\|\delta m\|}{\|m\|} \leq \alpha \frac{\|\delta d\|}{\|d\|}$$

where $\alpha = \frac{\sigma_1}{\sigma_p}$ is called the condition number of the operator G. It is the ratio of the largest to the smallest non-zero singular value. It is a measure of the numerical instability of the inverse problem.

8 Practical inversion

Despite the theoretical power of the SVD, there is no general fast method for computing it. As a result, unlike use of the Fourier transform in implementing reconstruction algorithms involving spatially invariant operators, use of the SVD in actual reconstruction algorithms is limited. We typically use the SVD for understanding what least-squares solutions are doing.