Singular Value Decomposition

Singular Value Decomposition.

We're actually ready to discretize our inverse problem but having just carried out SVE to understand the fundamentals, I want to continue with its discrete representation SVD. We will use this extensively in solving our numerical inverse problems. (and in fact, any matrix problem.)

Consider an $N \times M$ system of equations

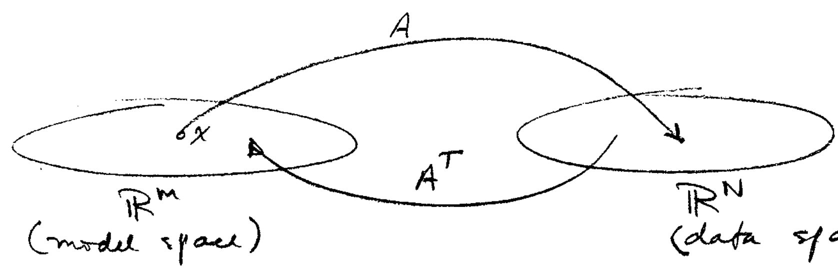
$$\boxed{A\bar{x} = \bar{b}} \qquad x \in \mathbb{R}^M \qquad b \in \mathbb{R}^N \qquad (1)$$

If we want to solve this, then we somehow need to calculate

$$x = (A^{-1})\, b$$

But what is A^{-1}? The matrix A is not even square.

To work with the system (1) we need to consider two matrices A and A^T



A: $N \times M$ matrix $\qquad A: \mathbb{R}^M \to \mathbb{R}^N$

A^T: $M \times N$ matrix $\qquad A^T: \mathbb{R}^N \to \mathbb{R}^M$

From these two non-square matrices we can form two square matrices

(i) AA^T $\quad (N \times N)$

(ii) A^TA $\quad (M \times M)$

Not only are these matrices square but they are also symmetric and semi positive definite. (ie $b^T A A^T b \geq 0 \quad x^T A^T A x \geq 0$) (for any $b \in \mathbb{R}^N$, $x \in \mathbb{R}^M$)

Remark: Both matrices are real and symmetric $\Rightarrow$ real eigenvalues
are semi-positive definite $\Rightarrow$ eigenvalues are ≥ 0

Let λ_i^2 be the eigenvalues of AA^T and let $\vec{u}_i$ $i=1,N$ be the associated eigenvectors (data space)

Let ν_i^2 be the eigenvalues of A^TA and let $\vec{v}_i$ $i=1,M$ be the associated eigenvectors (model space).

So

$$\boxed{\begin{aligned} AA^T \vec{u}_i &= \lambda_i^2 \vec{u}_i \\ A^TA \vec{v}_i &= \nu_i^2 \vec{v}_i \end{aligned}}$$

Although we don't show the details here, the following comments hold.

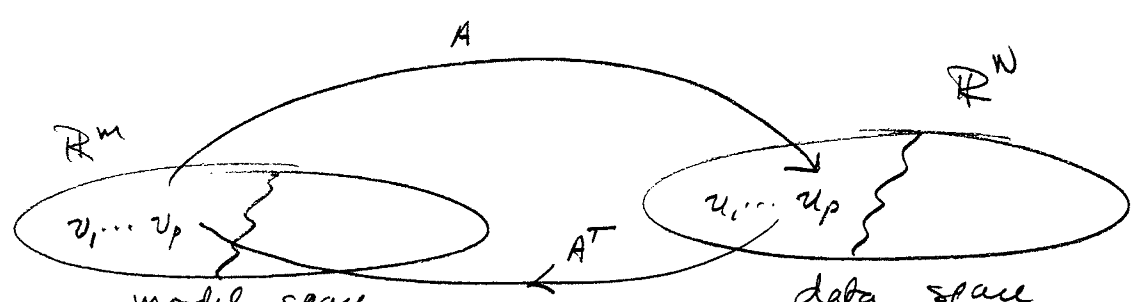
(1) The $M\times M$ matrix A^TA has positive eigenvalues λ_i^2 and these are the same eigenvalues as the positive eigenvalues ν_i^2 for the $N\times N$ matrix AA^T

(2) It follows that the number of non-zero eigenvalues in both square matrices is p where

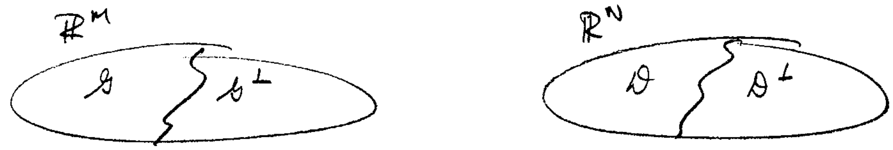
$$\begin{aligned} p &\leq M \quad \text{if} \quad M < N \\ p &\leq N \quad \text{if} \quad N < M \end{aligned}$$

(3) The relationship between the basis vectors in model space $\{v_i\}$ and the basis vectors in data space $\{u_i\}$ is given by

$$\boxed{\begin{aligned} A^T \vec{u}_i &= \lambda_i \vec{v}_i \\ A \vec{v}_i &= \lambda_i u_i \end{aligned}} \qquad i=1,p$$



So we see that model space and data space are each divided into two parts. There is an activated portion (of dimension p) and an unactivated portion



$\mathcal{G}$: $sp\{v_i\}$ is the activated portion of model space $\dim\{\mathcal{G}\} = p$
$\mathcal{D}$: $sp\{u_i\}$ is the activated portion of data space $\dim\{\mathcal{D}\} = p$
$\mathcal{G}^\perp$: portion of $\mathbb{R}^M$ that is $\perp$ to $\mathcal{G}$. It is the annihilator portion of model space $\dim \mathcal{G}^\perp = M - p$. If $\vec{y} \in \mathcal{G}^\perp$ then $A\vec{y} = 0$. (It does not lie in the row space of A)
$\mathcal{D}^\perp$: portion of $\mathbb{R}^N$ that is $\perp$ to $\mathcal{D}$. It does not lie in the column space of A and hence cannot be reproduced by any vector x. $\dim\{\mathcal{D}^\perp\} = N - p$.

Singular Value Decomposition

Let A be an $N \times M$ matrix. Let $\Lambda_p = \mathrm{diag}(\lambda_1, \lambda_2, \dots \lambda_p)$ be the positive square roots of the p nonzero eigenvalues of AA^T or A^TA. Let

$$U_p = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_p \\ | & & | \end{pmatrix}$$ be the $N \times p$ matrix containing the activated basis of data space $\{\mathcal{D}\}$

$$V_p = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_p \\ | & & | \end{pmatrix}$$ be the $M \times p$ matrix containing the activated basis of model space $\{\mathcal{G}\}$

The $N \times M$ matrix A has the unique decomposition

$$\boxed{A = U_p \Lambda_p V_p^T} \qquad (N\times p)(p \times p)(p \times M) \rightarrow N \times M$$

The components of Λ_p are referred to as the singular values of the matrix A.

Remark: Having generated the decomposition of A, how do we use it to solve the matrix system of equations?

$$A x = b$$
$$U_p \Lambda_p V_p^T x = b$$

but $U_p^T U_p = I_p$ is the $p \times p$ identity matrix. (Columns of U_p are eigenvectors; they are orthonormal)

$$\Lambda_p V_p^T x = U_p^T b$$

Λ_p^{-1} $$V_p^T x = \Lambda_p^{-1} U_p^T b$$

The next step generally would be to get rid of V_p^T. We multiply by V_p

$$\boxed{V_p V_p^T x = V_p \Lambda_p^{-1} U_p^T b}$$

But $V_p V_p^T$ is not necessarily the identity matrix. It will be only if $p=m$

Let us call

$$\boxed{\bar{x}_c = V_p V_p^T \bar{x} = V_p \Lambda_p^{-1} U_p^T b} \qquad (1)$$

The generalized inverse. Thus let the matrix

$$\boxed{A^+ = V_p \Lambda_p^{-1} U_p^T} \qquad (2)$$

denote the generalized inverse of A.

It is insightful to write (1) in component form.

$$\boxed{\bar{x}_c = \sum_{i=1}^{p} \frac{(\bar{u}_i, b)}{\lambda_i} \vec{v}_i = \sum_{i=1}^{p} \left(\frac{\bar{u}_i^T \bar{b}}{\lambda_i} \right) \bar{v}_i}$$

* Note the potential instability that arises if $\lambda_i \to 0$. This is similar to that cited earlier

we see explicitly that the constructed model is a linear combination of the vectors $\vec{v}_i \in \mathcal{S}$. The amplitudes of each vector is given by the inner product of the corresponding vector in $\mathcal{D}$ with the data vector. and then amplified by $1/\lambda_i$

coefficient for $\vec{v}_i$ is $\quad \bar{u}_i^T \bar{b} \quad$ (projection of $\bar{b}$ onto $\vec{u}_i$)

then divided by $\quad \lambda_i$

$$A^T u_i = \lambda_i v_i$$

$$v_i = \frac{1}{\lambda_i} A^T u_i$$

Remark: If we divide $x = x^{\parallel} + x^{\perp}$ where $x^{\parallel} \in \mathcal{S}$ and $x^{\perp} \in \mathcal{S}^{\perp}$ then the above expression shows clearly that

$$x_c = x^{\parallel}$$

So generalized inverse gives only the projection of the model onto the restricted space.

It is worthwhile to show this explicitly (and hence illustrate why SVD is so useful)

Consider the solution of the system of equations $Ax = b$ when the system is overdetermined.

In that case we want to find a best fitting solution

minimize $\phi = \|Ax - b\|^2$

$$\frac{\partial \phi}{\partial x} = 2A^T(Ax - b) = 0$$

so $A^T A x = A^T b$

or $$\boxed{\bar{x} = (A^T A)^{-1} A^T \bar{b}}$$

What does this solution look like if we introduce the SVD decomposition of A

$$\bar{x} = (V\Lambda U^T U \Lambda V^T)^{-1} V \Lambda U^T b$$
$$= (V \Lambda^2 V^T)^{-1} V \Lambda U^T b$$

for an overdetermined system $VV^T = V^T V = I$

so $V^T = V^{-1}$

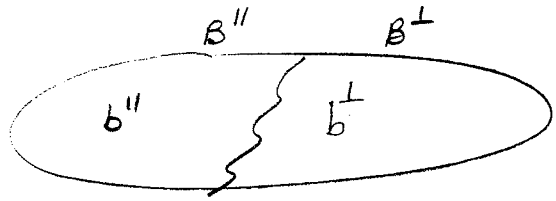
substituting $$\bar{x} = (V^T)^{-1} \Lambda^{-2} V^{-1} V \Lambda U^T b$$
$$= (V^T)^{-1} \Lambda^{-1} U^T b$$
$$\boxed{\bar{x}_c = V \Lambda^{-1} U^T b}$$ which is the SVD solution.

Note that the computed data for this solution are

$$b_c = A x_c$$
$$= U \Lambda V^T x_c$$
$$= U \Lambda V^T V \Lambda^{-1} U^T b$$
$$= U U^T b$$

So $$\boxed{b_c = U U^T b}$$

So for an over-determined system the best fitting solution yields data that are the projection of the observed data onto the activated portion of data space.



UU^T is the projection operator

$$UU^T : B \to B^{\parallel}$$

To see this explicitly, consider any arbitrary vector $b \in B$

$$b = b^{\parallel} + b^{\perp} \qquad \text{(unique decomposition)}$$

$$UU^T b = UU^T(b^{\parallel} + b^{\perp})$$
$$= UU^T b^{\parallel} + UU^T b^{\perp}$$

but $U = \begin{pmatrix} | & & | \\ u_1 & \cdots & u_p \\ | & & | \end{pmatrix}$ $\quad u_i \in B^{\parallel} \Rightarrow U^T b^{\perp} = 0$

So $\quad UU^T b = UU^T b^{\parallel} = b^{\parallel}$

So $\quad \boxed{b_c = UU^T b = b^{\parallel}}$

To prove the latter; Consider any vector $b^{\parallel} = \sum_{i=1}^{p} \beta_i \vec{u}_i \qquad b \in B^{\parallel}$

Then $\quad U^T b^{\parallel} = \beta$
$\quad U\beta = b^{\parallel}$

So $\quad UU^T b^{\parallel} = b^{\parallel}$

Now consider an underdetermined system.

$M > N$

$A : N \times M$

Problem: minimize $\phi = \| x \|^2$

such that $Ax = b$

$$\left\{ \begin{array}{c} \lambda^T A x = x^T A^T \lambda \\ \frac{\partial}{\partial x}(x^T y) = y \end{array} \right\}$$

minimize $\phi = \|x\|^2 + 2\lambda^T(Ax - b)$

$$\frac{\partial \phi}{\partial x} \qquad 0 = 2x + 2A^T\lambda \qquad (1)$$

$$\frac{\partial \phi}{\partial \lambda} \qquad 0 = Ax - b \qquad (2)$$

From (1) $\quad A^T\lambda = -x$

$\times A \quad AA^T\lambda = -Ax$

$= -b \implies \boxed{\lambda = -(AA^T)^{-1} b}$

So $\quad x = -A^T\lambda$

$$\boxed{\vec{x} = A^T(AA^T)^{-1}\vec{b}}$$

This is the usual formula. Moreover, substituting directly into the formula for x using SVD yields

$$\begin{aligned} x &= A^T(AA^T)^{-1} b \\ &= V\Lambda U^T(U\Lambda V^T V\Lambda U^T)^{-1} b \\ &= V\Lambda U^T(U\Lambda^2 U^T)^{-1} b \end{aligned} \qquad V^TV = I_M$$

but we have an underdetermined system. All of data space is activated $\quad UU^T = U^TU = I \quad$ or $\quad U^T = U^{-1}$

So
$$\begin{aligned} x &= V\Lambda U^T (U^T)^{-1}\Lambda^{-2}U^{-1} b \\ &= V\Lambda U^T U \Lambda^{-2} U^T b \end{aligned}$$

$$\boxed{x = V\Lambda^{-1}U^T b}$$

which is the SVD solution.

Mathematics for 2-norm minimization

Remark: It is useful to solve optimization problems by differenti w.r.t. a vector quantity. The rules are generated by working with individual components

Given:

$$\vec{x} = (x_1, x_2, \cdots x_n)^T$$
$$\bar{y} = (y_1, y_2, \cdots y_n)^T$$
$$W = n \times n \text{ matrix}$$
$$A = m \times n \text{ matrix}$$
$$\bar{b} = (b_1, \cdots b_m)$$

matrix system $A\bar{x} = \bar{b}$

(i) Vectors:

$$\frac{\partial}{\partial \bar{x}}(y^T x) = \bar{y}$$

$$\frac{\partial}{\partial \bar{x}}(x^T y) = \bar{y}$$

$$\frac{\partial}{\partial \bar{x}}(y^T W x) = (y^T W)^T = W^T \bar{y}$$

$$\frac{\partial}{\partial \bar{x}}(x^T W y) = W \bar{y}$$

(ii) Quadratic forms.

$$\frac{\partial}{\partial \bar{x}}(x^T W x) = W\bar{x} + W^T \bar{x}$$

if $W = W^T$

$$\frac{\partial}{\partial \bar{x}}(\bar{x}^T W \bar{x}) = 2W\bar{x}$$

(iii) Equations: Let Q be an $m \times m$ matrix. Minimize

$$\frac{\partial}{\partial \bar{x}}\left\{(A\bar{x} - \bar{b})^T Q (Ax - \bar{b})\right\} = A^T (Q + Q^T)(Ax - b)$$

$Q = Q^T$

$$\frac{\partial}{\partial \bar{x}}\left\{(Ax - b)^T Q (Ax - b)\right\} = 2A^T Q (Ax - b)$$

(iv) Composite functions.

$$\phi = x^T W_m \bar{x} + (A\bar{x} - \bar{b})^T W_e (A\bar{x} - \bar{b})$$ W_m, W_e symm

$$\frac{\partial \phi}{\partial \bar{x}} = 2 W_m \bar{x} + 2 A^T W_e (A\bar{x} - b)$$
$$\quad = 2(W_m + A^T W_e A)\bar{x} - 2A^T W_e \bar{b}$$

If $\frac{\partial \phi}{\partial \bar{x}} = 0$ we solve $\boxed{(A^T W_e A + W_m)\bar{x} = A^T W_e \bar{b}}$

Matrix properties.

(1) $(W^{-1})^T = (W^T)^{-1}$