University of Aarhus
Department of Computer Science

November 9, 2000

Ole Østerby

The SVD and Linear Equations.

The Singular Value Decomposition

Any matrix $A \in \grave{R}^{m\times n}$ possesses a singular value decomposition (SVD), i.e. there exists an orthogonal matrix $U \in \grave{R}^{m\times m}$, an orthogonal matrix $V \in \grave{R}^{n\times n}$ and a diagonal matrix $S \in \grave{R}^{m\times n}$ with non-negative numbers in the diagonal, such that

$$A \;=\; USV^T. \tag{1}$$

The column vectors $\{u_i\}$ of U are called left singular vectors, the column vectors $\{v_i\}$ of V are called right singular vectors, and the diagonal elements $\{\sigma_i\}$ of S are the singular values. These are usually ordered such that

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = 0.$$

The number r is the rank of A. If $r = \min(m,n)$ then matrix A has full rank, i.e. all the columns (if $m \geq n$) or all the rows (if $m \leq n$) of A are linearly independent.

If we know the SVD of A then we also know the SVD of A^T since

$$A^T \;=\; VS^TU^T. \tag{2}$$

From (1) and (2) we have

$$A^TA \;=\; VS^TU^TUSV^T \;=\; V(S^TS)V^T \tag{3}$$

or

$$(A^TA)V \;=\; V(S^TS) \tag{4}$$

or

$$(A^TA)v_i \;=\; \sigma_i^2 v_i, \qquad i = 1,2,\ldots,n \tag{5}$$

which shows that the right singular vectors of A are the eigenvectors of the symmetric matrix A^TA, and that the singular values of A are the square roots of the eigenvalues of A^TA. Recall that A^TA is symmetric and therefore has (an) orthogonal (basis of) eigenvectors, and that it is non-negative definite.

If $m < n$ then there are at most m positive eigenvalues of A^TA corresponding to the m singular values of A.

From (1) we get

$$AV = US \tag{6}$$

or

$$Av_i = \sigma_i u_i \tag{7}$$

which can be used to determine the left singular vectors of A corresponding to the non-zero singular values. Computationally this is not the way to do it, and this goes for formula (5) as well, since we may lose accuracy when σ_1/σ_r is large.

From (7) we also deduce

$$||A||_2 = \sigma_1 \tag{8}$$

The linear transformation

The matrix A represents a linear transformation from $\grave{R}^n$ to $\grave{R}^m$ given the usual cartesian basis vectors in both spaces. The range of A, $\mathcal{R}(A)$, is a subspace of $\grave{R}^m$ consisting of all vectors in $\grave{R}^m$ which can be reached by applying A to some vector in $\grave{R}^n$. The null space of A, $\mathcal{N}(A)$, is a subspace of $\grave{R}^n$, possibly $\{0\}$, consisting of all the vectors in $\grave{R}^n$ which are mapped by A into $0 \in \grave{R}^m$.

Similarly A^T is a mapping from $\grave{R}^m$ to $\grave{R}^n$. The null space of A^T is equal to the orthogonal complement of $\mathcal{R}(A)$:

$$\mathcal{N}(A^T) = \mathcal{R}(A)^\perp \tag{9}$$

because $x \in \mathcal{N}(A^T) \Leftrightarrow A^Tx = 0 \in \grave{R}^n$,
and $c \in \mathcal{R}(A) \Leftrightarrow \exists y \in \grave{R}^n,\ Ay = c$
and therefore $x^Tc = x^TAy = (A^Tx)^Ty = 0$.
Similarly

$$\mathcal{R}(A^T) = \mathcal{N}(A)^\perp. \tag{10}$$

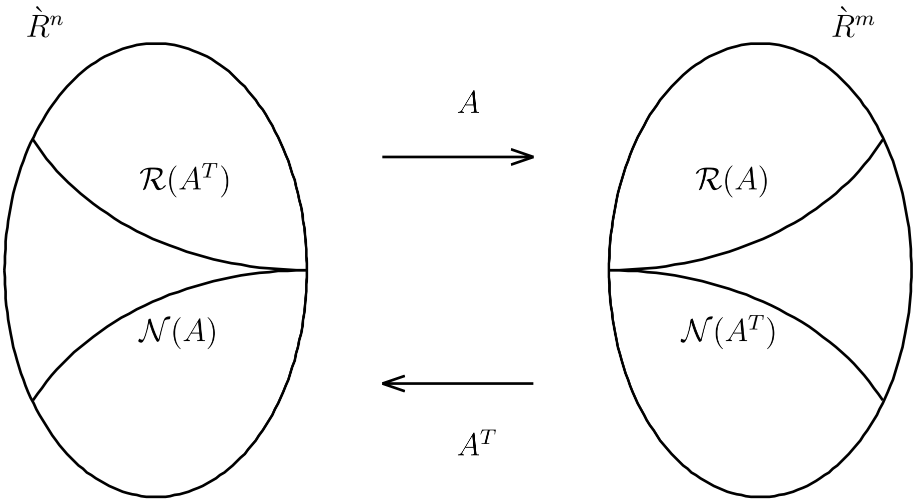

Fig. 1. The linear transformations.

From (6) or (7) we deduce that the first r vectors u_1, u_2, …, u_r span $\mathcal{R}(A)$, and from (9) it follows that the remaining (if any) vectors u_{r+1}, …, u_n form an orthonormal basis of $\mathcal{N}(A^T)$. Any orthonormal basis of $\mathcal{N}(A^T)$ will do, so there is considerable freedom in the choice of these basis vectors (if $n > r+1$).

From (6) or (7) and (10) we see that v_1, v_2, …, v_r span $\mathcal{R}(A^T) = \mathcal{N}(A)^\perp$ and v_{r+1}, …, v_n span $\mathcal{N}(A)$:

$$\begin{array}{lll} u_1,\, u_2,\, \ldots,\, u_r & \text{span} & \mathcal{R}(A) \\ \quad u_{r+1},\, \ldots,\, u_m & \text{span} & \mathcal{R}(A)^\perp = \mathcal{N}(A^T) \\ v_1,\, v_2,\, \ldots,\, v_r & \text{span} & \mathcal{N}(A)^\perp = \mathcal{R}(A^T) \\ \quad v_{r+1},\, \ldots,\, v_n & \text{span} & \mathcal{N}(A) \end{array}$$

We are seldom interested in the (bases of the) null spaces and therefore an alternate form of the SVD is sometimes used:

$$A = U_r S_r V_r^T \tag{11}$$

where $U_r \in \grave{R}^{m\times r}$, $S_r \in \grave{R}^{r\times r}$, $V_r \in \grave{R}^{n\times r}$, and $U_r^T U_r = V_r^T V_r = I_r$. (1) or (11) can also be written

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T \tag{12}$$

expressing A as a sum of rank 1 matrices.

The SVD as expressed in (1) or (6) describes the simple form the linear transformation takes when expressed in terms of the new orthonormal bases v_1, …, v_n of $\grave{R}^n$ and u_1, …, u_m of $\grave{R}^m$. If a vector $x \in \grave{R}^n$ is first expressed in terms of v_1, …, v_n then the linear transformation maps the component after v_i into a component after u_i multiplied by the number σ_i.

Linear Equations

We shall now assume that $m = n$ and $\sigma_n > 0$ such that A is quadratic and non-singular. We shall study the solution to a system of linear equations

$$Ax = b \tag{13}$$

and in particular what happens to the solution if we change the right-hand-side b to $b + \Delta b$. If we write the solution in the form $x + \Delta x$ we have

$$A(x + \Delta x) = b + \Delta b.$$

Subtracting (13) we get

$$A\Delta x = \Delta b$$

or

$$\Delta x = A^{-1}\Delta b$$

and

$$||\Delta x|| \leq ||A^{-1}|| \cdot ||\Delta b|| \tag{14}$$

which gives a bound on the magnitude of the change in the solution. Taking norms in (13) we get

$$||b|| \leq ||A|| \cdot ||x||$$

or

$$\frac{1}{||x||} \leq ||A|| \frac{1}{||b||}$$

which together with (14) gives

$$\frac{||\Delta x||}{||x||} \leq \kappa(A) \frac{||\Delta b||}{||b||} \tag{15}$$

where we have introduced the *condition number* κ of matrix A:

$$\kappa(A) = ||A|| \cdot ||A^{-1}||. \tag{16}$$

The condition number of a matrix depends on the matrix norm used but not very much. Since all norms of a finite dimensional space are equivalent then so are the condition numbers, meaning that if one condition number of matrix A is large then so are all the others.

From (1) we have

$$A^{-1} \;=\; VS^{-1}U^T \tag{17}$$

which is the SVD of A^{-1} except for the ordering of the singular values. We note that

$$||A^{-1}||_2 \;=\; \frac{1}{\sigma_n} \tag{18}$$

and using (8)

$$\kappa_2(A) \;=\; ||A||_2 \cdot ||A^{-1}||_2 \;=\; \frac{\sigma_1}{\sigma_n} \tag{19}$$

so a matrix is ill-conditioned when $\sigma_n \ll \sigma_1$. We can always scale A such that $||A||_2 = \sigma_1$ is 1 or close to 1. In this case a matrix is ill-conditioned if σ_n is small.

The solution to (13) can be expressed as

$$x \;=\; A^{-1}b \;=\; (\sum_{i=1}^{n} \frac{1}{\sigma_i} v_i u_i^T)b \;=\; \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i \tag{20}$$

and we note that if A is ill-conditioned then the solution, x, will be rather sensitive to small changes in b, if these changes have a component in the direction of u_n. This component will be divided by the small number σ_n and reappear in the direction of v_n.

Ill-conditioned equations

There are various types of ill-condition. If σ_{n-1} is reasonably large and only $\sigma_n \ll 1$ then there is only one component to worry about. This is the case that occurs when one has found an eigenvalue, λ, of a matrix, A, and seeks a corresponding eigenvector using inverse iteration. In this case the corresponding coefficient matrix $A - \lambda I$ is singular if λ is accurate and ill-conditioned if λ is an approximation to the eigenvalue. Sometimes it is advocated to deliberately perturb λ to avoid extreme ill-condition, but this is neither necessary nor advisable. From (20) we see that any small $u_n^T b$ will be divided by a very small number making the component after v_n dominate everything else. But in this case we are only interested in the direction, and a subsequent normalization will give us the desired eigenvector. The only modification we might wish to incorporate in our equation solver is to substitute a very small number for a possible exact zero which might appear in the diagonal during the elimination in order to avoid overflow.

A much worse situation occurs if the singular values decay gradually to a very small σ_n. Such a coefficient matrix is really bad news.

Any singular value σ_k indicates the distance, measured in the 2-norm, from A to the nearest matrix of rank $k-1$. If the singular values drop gradually to a value σ_r and then show a considerable decrease to σ_{r+1}, then the matrix might be considered to be effectively of rank r. In particular, if σ_{r+1} is of the same order of magnitude as measuring (or other) errors, then the matrix is for all practical purposes of rank r.

The pseudo-inverse

For a general matrix $A \in \grave{R}^{m\times n}$ with SVD given by (1) we can define the pseudo-inverse by

$$A^\dagger \quad = \quad VS^\dagger U^T \tag{21}$$

where $S^\dagger \in \grave{R}^{n\times m}$ is a diagonal matrix with diagonal elements equal to $1/\sigma_i$ whenever $\sigma_i > 0$ and 0 otherwise.

If $m = n$ and A is non-singular the $A^\dagger$ coincides with the regular inverse. In all other cases $A^\dagger$ has some but not all the properties of an inverse matrix. For example $A^\dagger A$ acts as the identity when restricted to $\mathcal{R}(A^T)$:

$$\begin{aligned} y \in \mathcal{R}(A^T) &\Leftrightarrow \exists c \in \grave{R}^m,\ A^T c = y \\ A^\dagger A y &= A^\dagger A A^T c \\ &= VS^\dagger(U^T U)S(V^T V)S^T U^T c \\ &= V(S^\dagger S)S^T U^T c = VS^T U^T c = A^T c = y \end{aligned}$$

and similarly $AA^\dagger$ acts as the identity when restricted to $\mathcal{R}(A)$.

Least squares

When $m > n$ the linear system

$$Ax \quad = \quad b \tag{22}$$

is overdetermined. We have more equations than unknowns and cannot hope to fulfil all equations. Geometrically speaking the vector $b \in \grave{R}^m$ will usually be outside of $\mathcal{R}(A)$ so there is no vector $x \in \grave{R}^n$ which can be mapped by A into b.

Instead we choose to satisfy (22) in the least squares sense, i.e. we seek the vector $x \in \grave{R}^n$ which minimizes $||Ax - b||_2$. Geometrically we find the vector $c \in \mathcal{R}(A)$

which minimizes $||c - b||_2$. The vector c is characterized as the orthogonal projection of b onto $\mathcal{R}(A)$. Then we find a vector $x \in \grave{R}^n$ such that $Ax = c$. If $\mathcal{N}(A) \neq \{0\}$ then x is not unique since we can add any vector in $\mathcal{N}(A)$ to x without changing Ax. In this case one often seeks the *minimal solution*, which is the vector of smallest 2-norm which maps into c. If y is any solution then the minimal solution is the orthogonal projection of y onto $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$.

$$
\begin{aligned}
& c \text{ is the orthogonal projection of } b \text{ onto } \mathcal{R}(A) \\
\Leftrightarrow\ & c - b \in \mathcal{R}(A)^\perp = \mathcal{N}(A^T) \\
\Leftrightarrow\ & A^T(c - b) = 0.
\end{aligned}
$$

But $c \in \mathcal{R}(A) \Leftrightarrow \exists y \in \grave{R}^n$ such that $Ay = c$, therefore

$$(A^T A)y \;=\; A^T b. \tag{23}$$

This set of equations is called the *normal equations* (normal = orthogonal). If A has rank n then $\mathcal{N}(A) = \{0\}$ and the normal equations will determine the least squares solution to (22).

In the general case the minimal solution to (22) can be found as

$$x \;=\; A^\dagger b \;=\; VS^\dagger U^T b. \tag{24}$$

$U^T b$ expresses b in terms of the new basis $\{u_i\}$. The multiplication by $S^\dagger$ projects b down to $\mathcal{R}(A)$ and maps to $\mathcal{R}(A^T)$ where the result is represented in terms of the basis $\{v_i\}$.

Scaling

We return now to the ordinary $n \times n$ system of linear equations (13) and the inequality (15). This bound is relative in the sense that it measures the norm of the change Δx relative to the norm of x in terms of the norm of the change Δb relative to the norm of the right-hand-side b. But when we talk about relative changes we often think of the change Δx_i in one component relative to the magnitude, $|x_i|$, of that component. And this is not (always) captured by the inequality (15). The usual norms $||\cdot||_1$, $||\cdot||_2$ and $||\cdot||_\infty$ all have the property that they put equal emphasis on all components of a vector. But if one component is large it will govern the norm of the vector, also when the large numerical value is mainly due to a poor choice of measuring units. Choosing different units will have a significant effect on the norms and the relative importance of the various components. Similarly the common matrix norms and the condition numbers derived from them are rather sensitive to the scaling of the matrix.

Example 1. Consider the system $Ax = b$ with

$$A = \left\{ \begin{array}{cc} 1 & 1 \\ .001 & -.001 \end{array} \right\}, \qquad b = \left\{ \begin{array}{c} 1 \\ .003 \end{array} \right\}.$$

The SVD of A is

$$A = USV^T = \left\{ \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right\} \left\{ \begin{array}{cc} \sqrt{2} & 0 \\ 0 & \sqrt{2}/1000 \end{array} \right\} \left\{ \begin{array}{cc} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{array} \right\}$$

such that $\sigma_1 = \sqrt{2}$, $\sigma_2 = \sqrt{2}/1000$, and $\kappa_2 = 1000$.
The solution is $x^T = \{2, -1\}$.
We now make a relative perturbation of the components of the right-hand-side:

$$b + \Delta b = \left\{ \begin{array}{c} 1 + \delta_1 \\ .003(1 + \delta_2) \end{array} \right\} = b + \left\{ \begin{array}{c} \delta_1 \\ .003 \cdot \delta_2 \end{array} \right\}.$$

With $\delta_1 = \delta_2 = 0.001$ the solution becomes

$$x + \Delta x = x + \frac{1}{2} \left\{ \begin{array}{c} \delta_1 + 3\delta_2 \\ \delta_1 - 3\delta_2 \end{array} \right\} = \left\{ \begin{array}{c} 2 \\ -1 \end{array} \right\} + \left\{ \begin{array}{c} .002 \\ -.001 \end{array} \right\}.$$

The bound (15) reads

$$0.001 = \frac{||\Delta x||_2}{||x||_2} \leq \kappa_2 \cdot \frac{||\Delta b||_2}{||b||_2} = 1000 \cdot 0.001 = 1.$$

This inequality is correct although rather pessimistic.

Example 2. Now consider

$$A' = \left\{ \begin{array}{cc} 1 & 1 \\ 1 & -1 \end{array} \right\}, \qquad b' = \left\{ \begin{array}{c} 1 \\ 3 \end{array} \right\}, \qquad x' = x = \left\{ \begin{array}{c} 2 \\ -1 \end{array} \right\}.$$

A' has singular values $\sigma'_1 = \sigma'_2 = \sqrt{2}$ and $\kappa'_2 = 1$.
Again we make a relative perturbation of the components of the right-hand-side:

$$b' + \Delta b' = \left\{ \begin{array}{c} 1 + \delta_1 \\ 3 \cdot (1 + \delta_2) \end{array} \right\} = b' + \left\{ \begin{array}{c} \delta_1 \\ 3 \cdot \delta_2 \end{array} \right\}.$$

With $\delta_1 = \delta_2 = 0.001$ as before the solution becomes

$$x' + \Delta x' = x + \frac{1}{2} \left\{ \begin{array}{c} \delta_1 + 3\delta_2 \\ \delta_1 - 3\delta_2 \end{array} \right\} = x + \Delta x.$$

The bound (15) now reads

$$0.001 = \frac{||\Delta x'||_2}{||x'||_2} \leq \kappa'_2 \cdot \frac{||\Delta b'||_2}{||b'||_2} = 1 \cdot 0.001 = 0.001.$$

We note that the solution is unchanged by the scaling and, just as important, so is the reaction to a relative perturbation in the right-hand-side. But the bound (15) which tells us about the sensitivity w.r.t. perturbations becomes more realistic with a better scaled matrix.

An even more extreme (and trivial) example is the matrix

$$\left\{ \begin{array}{cc} 1000000000 & 0 \\ 0 & 1 \end{array} \right\}$$

with a condition number of 10^9 but just a simple scaling away from the perfectly conditioned identity matrix.

We stress that scaling does not necessarily make the solution any better but it improves what we can say about the quality of the solution in terms of singular values, condition numbers, etc. If we for instance calculate the singular values of a matrix in order to determine the (numerical) rank it is of some importance that the matrix is properly scaled. A scaling will not change the essential properties of the underlying linear transformation, but it will make it easier to deduce reliable information from the relative magnitude of the singular values.

In the practical world scaling does have an effect on the choice of pivots in a Gaussian elimination with pivoting and thus may have an effect on the computed solution.

To do the scaling we shall consider transformations

$$A' = D_1 A D_2 \qquad (25)$$

where D_1 and D_2 are diagonal matrices with positive entries. D_1 performs row-scaling and corresponds to a change of units in the measured quantities, b_i. D_2 performs column-scaling and corresponds to a change of units in the unknowns or parameters, x_i. It is difficult to propose an optimal scaling method which to any matrix will produce the best choice for D_1 and D_2 in the sense that $\kappa_2(A')$ is smallest possible. But a less ambitious project might do.

As a rule-of-thumb we would prefer A to have elements all of the same order of magnitude, or at least that all column vectors have (roughly) the same norm, and that all row vectors have (roughly) the same norm. There are, however, several ways of accomplishing this with varying effects on the condition number as the following example shows.

Example 3. (cf. Fox, p. 169) In matrix A

$$A = \left\{ \begin{array}{rrr} 1 & -1 & 1 \\ -1 & \varepsilon & \varepsilon \\ 1 & \varepsilon & \varepsilon \end{array} \right\}$$

all row and column vectors have norms of the same order of magnitude when ε is small. But the singular values are $\sigma_1 = 2$, $\sigma_2 = 1$ and $\sigma_3 = 2\varepsilon$, so $\kappa = 1/\varepsilon$. Using $D_1 = diag\{1, 1/\varepsilon, 1/\varepsilon\}$ and $D_2 = diag\{\varepsilon, 1, 1\}$ we get

$$A' = D_1 A D_2 = \left\{ \begin{array}{rrr} \varepsilon & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \end{array} \right\}$$

with $\sigma_1' = 2$, $\sigma_2' \approx \sigma_3' \approx \sqrt{2}$ and therefore $\kappa_2' \approx \sqrt{2}$, and A' is much better scaled.

Eliminating a dependent parameter

Let $A \in \grave{R}^{m \times n}$ with $m > n$ and assume that σ_n is small relative to measuring or other errors that might affect the components of A. This indicates that up to the relevant accuracy the column vectors of A are linearly dependent, i.e. one of them can be expressed as a linear combination of the others. Each column may correspond to a parameter in a mathematical model and the linear dependence means that these parameters are not independent. One of them can be excluded, but which one?

First of all we must realize that linear dependence is not the fault of one particular vector. Two vectors may point in almost the same direction and either one could be retained. Or three vectors may almost lie in a plane and any two will define the plane reasonably well. On the other hand it might very well happen that one vector is almost orthogonal to the rest, so this vector should certainly be retained. So it appears that it might be easier to produce a positive list rather than a negative list.

Since we are interested in the linear (in-)dependence of the column vectors of A it is advisable to start with scaling these column vectors such that they have the same norm, say $|| \cdot ||_2 = 1$. Next we compute the singular values in order to get reliable information about the (numerical) rank of A.

The best set of columns of A to retain is the one which has the smallest condition number. So what we could do was to remove $n - r$ columns of A, calculate σ_1 and σ_r and $\kappa = \sigma_1/\sigma_r$. Now repeat this for all possible combinations of $n - r$ columns and select the one with the smallest κ. This could be a tedious and time-consuming process. A more efficient procedure is given in [1, section 12.2.2]. Take the transpose of matrix V_r consisting of the first r columns of V and perform a QR-factorisation with column pivoting:

$$QR = (V_r^T)P.$$

The first r columns of the permutation matrix, P, gives the sequence of columns of A to be retained in order to get the smallest condition number.

Reference

[1] Gene H. Golub and Charles F. Van Loan:
Matrix Computations 2. ed. Johns Hopkins 1989.

DAIMI, 27.11.2000, Ole Østerby