F. Numerical linear algebra

V. Blobel
Universität Hamburg

April 2001

5. Singular value decomposition

Solving the matrix equation

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \qquad m \times n \text{ matrix } \boldsymbol{A}$$

for an n vector $\boldsymbol{x}$ of unknowns.

Distinguish cases:

- $n = m$ exact solution usually exists
- $m > n$ overdetermined system – obtain best "approximate" solution
- $m < n$ underdetermined system – insufficient data

Overdetermined system $(m > n)$

In general no solution vector $\boldsymbol{x}$ exist.

Vector of residua $\boldsymbol{r}$, defined by

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} ,$$

never zero.

What is the "best" solution?

Minimum value of the square of the vector of residua:

$$\begin{aligned} \boldsymbol{r}^T \boldsymbol{r} &= (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \\ &= \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{b} + \boldsymbol{b}^T \boldsymbol{b} = \text{Minimum} . \end{aligned}$$

The minimum is determined, by setting the derivative of the above expression to zero:

$$\boldsymbol{A}^T \boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^T \boldsymbol{b} = 0 \quad \text{oder} \quad \boldsymbol{A}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) = \boldsymbol{A}^T \boldsymbol{r} = 0 .$$

These equation are called *normal equations*; the vector $\boldsymbol{r}$ of residua is *normal* with respect to the column vectors of matrix $\boldsymbol{A}$.

Solution

In general the system has a unique solution

$$\boldsymbol{x} = \left(\boldsymbol{A}^T \boldsymbol{A}\right)^{-1} \boldsymbol{A}^T \boldsymbol{b}$$

The symmetric matrix $\boldsymbol{W} = \boldsymbol{A}^T \boldsymbol{A}$ and the vector $\boldsymbol{c} = \boldsymbol{A}^T \boldsymbol{b}$ are formed:

$$\boxed{\boldsymbol{W}} = \boxed{\boldsymbol{A}^T} \times \boxed{\boldsymbol{A}} \qquad \boxed{\boldsymbol{c}} = \boxed{\boldsymbol{A}^T} \times \boxed{\boldsymbol{b}}$$

The solution vector $\boldsymbol{x}$ is determined by the solution of the matrix equation $\boldsymbol{W}\boldsymbol{x} = \boldsymbol{c}$:

$$\boxed{\boldsymbol{x}} = \boxed{\boldsymbol{W}^{-1}} \times \boxed{\boldsymbol{c}}$$

A solution in different form . . .

The symmetric matrix $\boldsymbol{W}$ can be diagonalized:

$$\boxed{\boldsymbol{W}} = \boxed{\boldsymbol{U}} \times \boxed{\boldsymbol{D}} \times \boxed{\boldsymbol{U}^T}$$

Matrix $\boldsymbol{U}$ is orthonormal and matrix $\boldsymbol{D}$ is diagonal with the eigenvalues λ of matrix $\boldsymbol{W}$ on the diagonal:

$$\boldsymbol{D} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \qquad \boldsymbol{D}^{-1} = \begin{pmatrix} 1/\lambda_1 & & & \\ & 1/\lambda_2 & & \\ & & \ddots & \\ & & & 1/\lambda_n \end{pmatrix}$$

$$\boxed{\boldsymbol{W}^{-1}} = \boxed{\boldsymbol{U}} \times \boxed{\boldsymbol{D}^{-1}} \times \boxed{\boldsymbol{U}^T}$$

$$\boldsymbol{x} = \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}^T\boldsymbol{c}\,.$$

Case of singular matrix $\boldsymbol{W}$

At least one of the eigenvalues is zero and the solution above is impossible

The columns of matrix $\boldsymbol{U}$ (these are the eigenvectors of matrix $\boldsymbol{W}$), which correspond to the *zero* eigenvalues, form an orthogonal basis of the nullspace of $\boldsymbol{W}$. Any arbitrary vector of the null space can be added to the solution $\boldsymbol{x}$.

How to avoid this kind of arbitrariness?

Better define a solution with *no* contribution from the null space: define a modified matrix $\boldsymbol{D}^{*-1}$, where the diagonal elements for zero eigenvalues of $\boldsymbol{D}$ are set to zero (thus replacing $1/0 = \infty$ by 0). The modified solution

$$\boldsymbol{x} = \boldsymbol{U}\boldsymbol{D}^{*-1}\boldsymbol{U}^T\boldsymbol{c}$$

contains *no* contribution from the null space.

It can be shown, that this is the solution with smallest length of $|\boldsymbol{x}|$.

Case of almost singular matrix $\boldsymbol{W}$

Sometimes a system of equation is not singular, but ill conditioned: some of the eigenvalues are very small and the inversion is inaccurate.

This property is visible in the condition number of the matrix, the ratio of the largest to the smallest eigenvalue: $\kappa = \lambda_{\max}/\lambda_{\min}$.

In such a case a solution with modified matrix Matrix $\boldsymbol{D}^{*-1}$ is preferred, where the small eigenvalues are replaced by zero.

Singular value decomposition

A $m \times n$ Matrix $\boldsymbol{A}$ can be expressed in the form

$$\boxed{\boldsymbol{A}} = \boxed{\boldsymbol{U}} \times \boxed{\boldsymbol{S}} \times \boxed{\boldsymbol{V}^T}$$

$\boldsymbol{U}$ is a column-orthogonal $m \times n$ matrix: $\boldsymbol{U}^T\boldsymbol{U} = \mathbf{1}$
$\boldsymbol{V}$ is an orthogonal $n \times n$ matrix: $\boldsymbol{V}^T\boldsymbol{V} = \mathbf{1} \quad \boldsymbol{V}\boldsymbol{V}^T = \mathbf{1}$

Matrix S is diagonal and has positive or zero elements, called *singular values*.

Modified solution

Matrix S is diagonal and has positive or zero elements, called *singular values*:

$$\boldsymbol{S} = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix} \qquad \boldsymbol{S}^{-1} = \begin{pmatrix} 1/\sigma_1 & & & \\ & 1/\sigma_2 & & \\ & & \ddots & \\ & & & 1/\sigma_n \end{pmatrix}$$

Comparison with previous formalism shows:

$$\boldsymbol{W} = \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}\boldsymbol{S}\boldsymbol{V}^T \qquad \boldsymbol{D} = \boldsymbol{S}\boldsymbol{S}$$

i.e. eigenvalues of $\boldsymbol{D}$ are $\lambda_i = \sigma_i^2$.
The solution vector is the least squares solution with minimum length

$$\boldsymbol{x} = \boldsymbol{V}\boldsymbol{S}^{*-1}\left(\boldsymbol{U}^T\boldsymbol{b}\right) .$$

As above a modified matrix $\boldsymbol{S}^{*-1}$ is defined, where the diagonal elements for small singular values S_{jj} are set to zero.

Comparison

The two methods are equivalent, but there are differences:

- The singular value method is numerically more accurate ($\sigma_i = \sqrt{\lambda_i}$).
- The eigenvalue method works on more compact matrices (small $n \times n$ square matrices instead of large rectangular $m \times n$ matrices).

Both methods avoid those components in the solution vector $\boldsymbol{x}$ from the Null-space, which have no effect on the right hand-side of the initial equations.