NUMERICAL LINEAR ALGEBRA

Prof. Paola Antonietti

MOX - Dipartimento di Matematica

Politecnico di Milano

https://antonietti.faculty.polimi.it

TA: Dr. Michele Botti

P4: Numerical methods for overdetermined linear systems & SVD

The starting point and some examples

Overdetermined systems have many applications, including many fitting problems.

When the problems are linear there is a very clean and simple way to find the optimum, if we adopt the sum-of-squares error metric.

FLOP Counts - MGS

- No distinction between real and complex

- No consideration of memory accesses or other performance aspects
- Flops count $\sim 2mn^2$

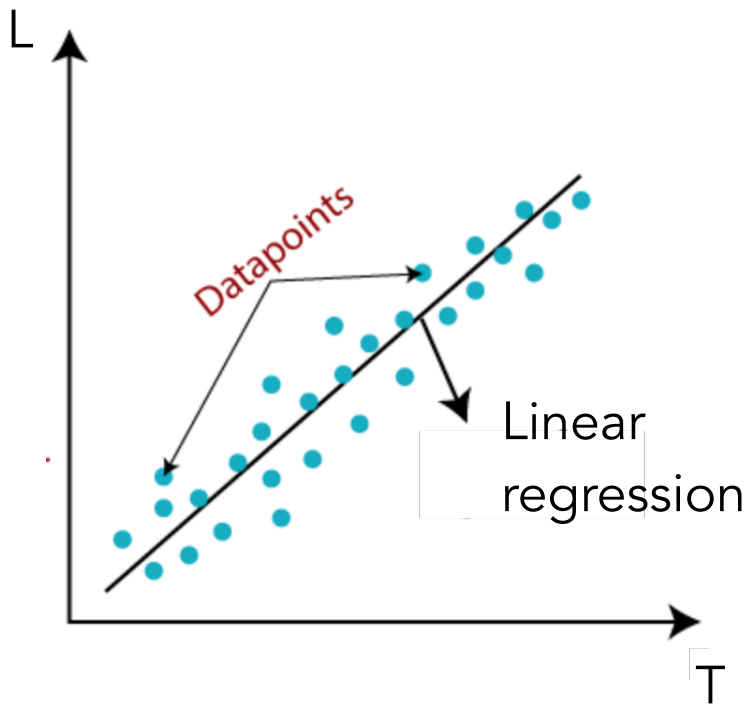
The QR algorithm

Linear regression

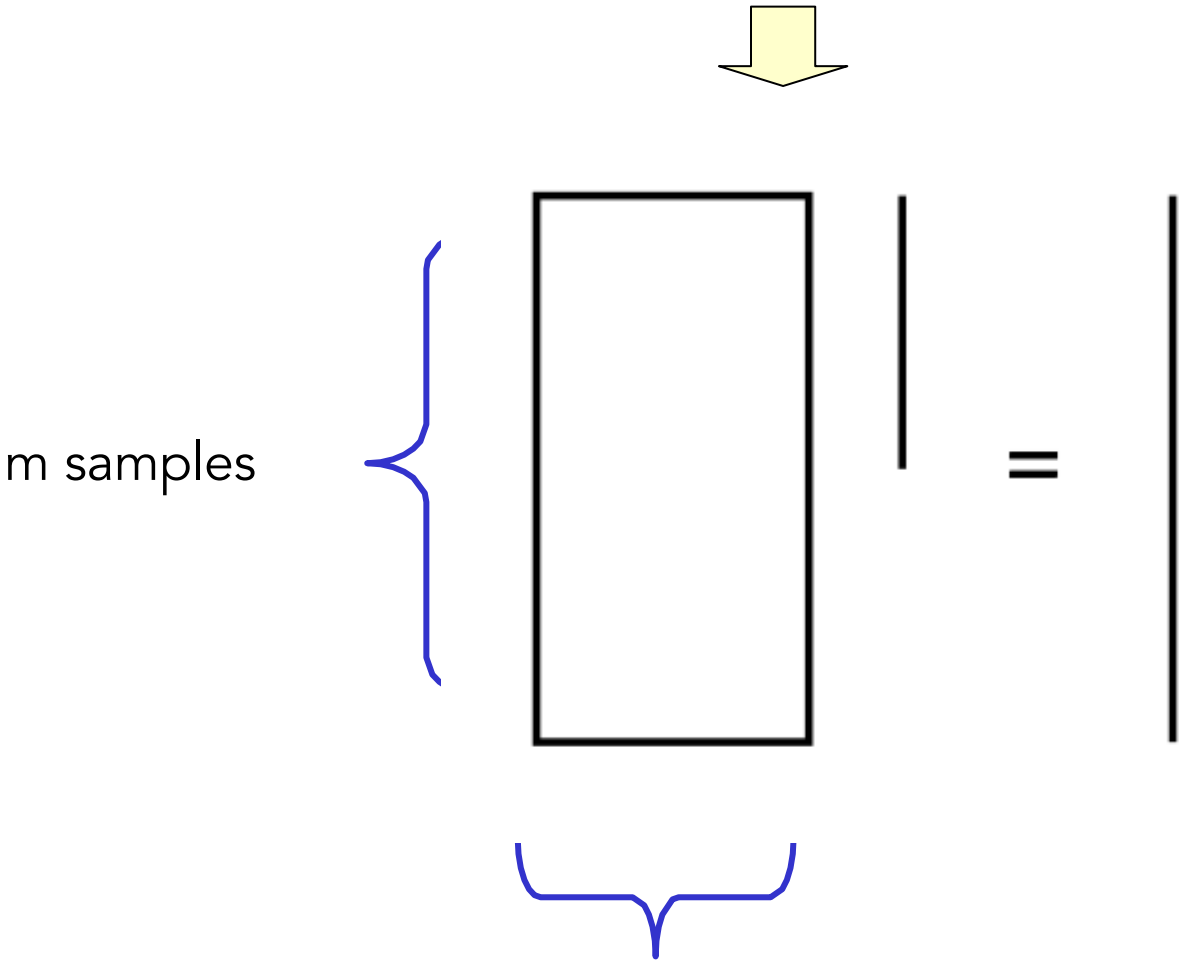
Experiment to find thermal expansion coefficient with metal bar and a torch:

The data is m pairs $(T_i, L_i), i = 1, \ldots, m.$

The hypothesis is that $L(T) = \alpha_0 + \alpha_1 T$, We want to estimate α_0, α_1.



Linear regression



Linear regression

If there were no experimental uncertainty the model would fit the data exactl
but since there is noise the best we can do is minimise the error .
The problem is

$$\min_{\alpha_0,\alpha_1} \sum_{i=1}^{m} e_i^2 = \min_{\alpha_0,\alpha_1} \sum_{i=1}^{m} (\alpha_0 + \alpha_1 T_i - L_i)^2$$

The above problem is equivalent to the following:

$$\min_{\alpha} \|A\alpha - \mathbf{b}\|_2^2$$

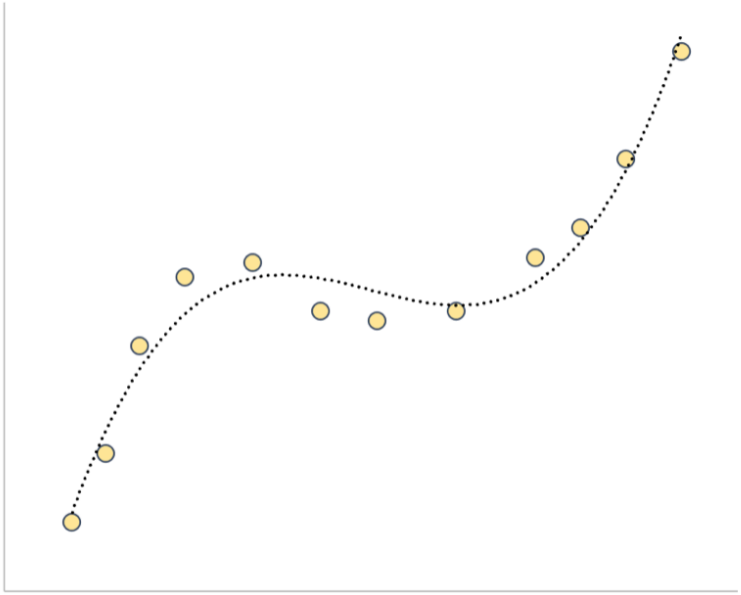
with

$$A = \begin{bmatrix} 1 & T_1 \\ 1 & T_2 \\ \vdots & \vdots \\ 1 & T_m \end{bmatrix} \qquad \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_m \end{bmatrix}$$

Polynomial regression

Suppose the model we expect to fit our data pairs $(T_i, L_i), i = 1, \ldots, m$ is a cubic polynomial rather than a linear one. The hypothesis now is

$$L(T) = \alpha_0 + \alpha_1 T + \alpha_2 T^2 + \alpha_3 T^3$$



Polynomial regression

The problem now read

$$\min_{\alpha_0,\alpha_1,\alpha_2,\alpha_4} \sum_{i=1}^{m} e_i^2 = \min_{\alpha_0,\alpha_1,\alpha_2,\alpha_4} \sum_{i=1}^{m} (\alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 T_i^3 - L_i)^2$$

The above problem is equivalent to the following:

$$\min_{\alpha} \|A\alpha - \mathbf{b}\|_2^2$$

with

$$A = \begin{bmatrix} 1 & T_1 & T_1^2 & T_1^3 \\ 1 & T_2 & T_2^2 & T_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & T_m & T_m^2 & T_m^3 \end{bmatrix} \quad \alpha = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_m \end{bmatrix}$$

Economic prediction

So far we have looked at a single independent variable, with complexity arising from the type of model. Some problems have many independent variables. We consider an example of an economic application. We would like to be able to predict total employment from a set of other economic measures:

- x_1: Gross national product (GNP) implicit price deflator
- x_2: Gross National Product
- x_3: Unemployment
- x_4: Size of armed forces
- x_5: Population
- x_6: Year

Economic prediction

We'd like to approximate y, the total employment, as a linear combination of the others, i.e.

$$y \approx \beta_0 + \sum_j \beta_j x_j$$

We have historical data available for many years, and so we can set up a system with a row for each year, each of which reads

$$y = [1, x_1, x_2, \ldots, x_6] = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_6 \end{bmatrix}$$

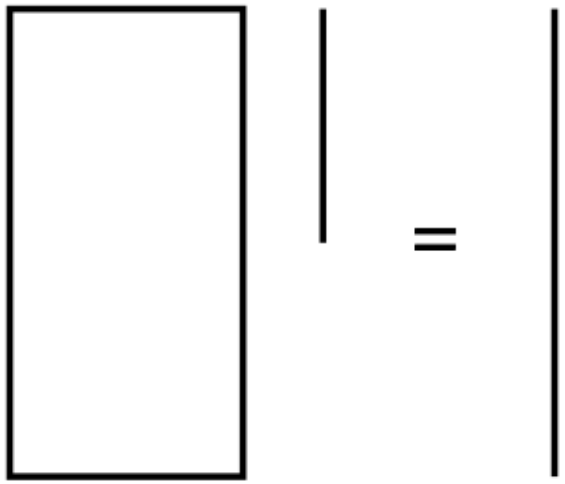
with more than 7 years of data, this will be an overdetermined system that can be solved by least squares. Then y can be predicted in future years for which only the x_i are available.

The mathematical problem

We are interested in the over-determined linear systems of equations (more equations than unknowns).

The mathematical problem reads: given $A \in \mathbb{R}^{mxn}, m \geq n,$ and $\mathbf{b} \in \mathbb{R}^m$ find $\mathbf{x} \in \mathbb{R}^n$ such that

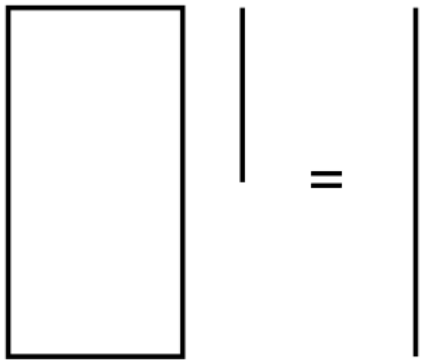
$$A\mathbf{x} = \mathbf{b}$$



Some preliminary remarks

Given $A \in \mathbb{R}^{mxn}, m \geq n,$ and $\mathbf{b} \in \mathbb{R}^m$ find $\mathbf{x} \in \mathbb{R}^n$ such that

$A\mathbf{x} = \mathbf{b}$



We notice that generally the above problems has no solution (in the classical sense) unless the right side $\mathbf{b}$ is an element of range(A).

We need a "new" concept of solution ! The basic approach is to look for an $\mathbf{x}$ that makes $A\mathbf{x}$"close" to $\mathbf{b}$.

Solution in the least-square sense

Given $A \in \mathbb{R}^{mxn}, m \geq n,$ we say that $\mathbf{x}^* \in \mathbb{R}^n$ is a solution of the linear system $A\mathbf{x} = \mathbf{b}$ in the least-squares sense if

$$\Phi(\mathbf{x}^*) = \min_{\mathbf{y} \in \mathbb{R}^n} \Phi(\mathbf{y}),$$

where

$$\Phi(\mathbf{y}) = \|A\mathbf{y} - \mathbf{b}\|_2^2.$$

The problem thus consists of minimising the Euclidean norm of the residual.

The solution $\mathbf{x}^*$ can be found by imposing the condition that the gradient of the function $\Phi(\,\cdot\,)$ must be equal to zero at $\mathbf{x}^*$.

Solution in the least-square sense

From the definition we have

$$\Phi(\mathbf{y}) = (A\mathbf{y} - \mathbf{b})^T(A\mathbf{y} - \mathbf{b})$$
$$= \mathbf{y}^T A^T A\mathbf{y} - 2\mathbf{y}^T A\mathbf{b} + \mathbf{b}^T\mathbf{b}$$

Therefore:

$$\nabla\Phi(\mathbf{y}) = 2A^T A\mathbf{y} - 2A^T\mathbf{b}$$

from which it follows that $\mathbf{x}^*$ must be the solution of the square system

$$A^T A\mathbf{x}^* = A^T\mathbf{b}$$ System of normal equations

Some remarks

The system of normal equations is nonsingular if A has full rank and, in such a case, the least-squares solution exists and is unique.

We notice that $B = A^T A$ is a symmetric and positive definite matrix.

Thus, in order to solve the normal equations, one could first compute the Cholesky factorization $B = R^T R$ and then solve the two systems

$R^T \mathbf{y} = A^T \mathbf{y}$ and $R^T \mathbf{x}^* = A^T \mathbf{y}$.

However, $A^T A$ is very badly conditioned and, due to roundoff errors, the computation of $A^T A$ may be affected by a loss of significant digits, with a consequent loss of positive definiteness or nonsingularity of the matrix!

Example (in matlab)

$$\mathrm{A} = \begin{bmatrix} 1 & 1 \\ 2^{-27} & 0 \\ 0 & 2^{-27} \end{bmatrix}, \; fl(\mathrm{A}^T\mathrm{A}) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

For a matrix A with full rank, the corresponding matrix fl($A^T A$) turns out to be singular

Another example

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-5} \\ 10^{-10} & 10^{-10} \end{bmatrix}$$

$$A^T A = \begin{bmatrix} 1 & 10^{-20} \\ 10^{-20} & 10^{-10} \end{bmatrix} \qquad \hat{R} = \begin{bmatrix} 1 & 10^{-20} \\ 0 & 10^{-5} \end{bmatrix}$$

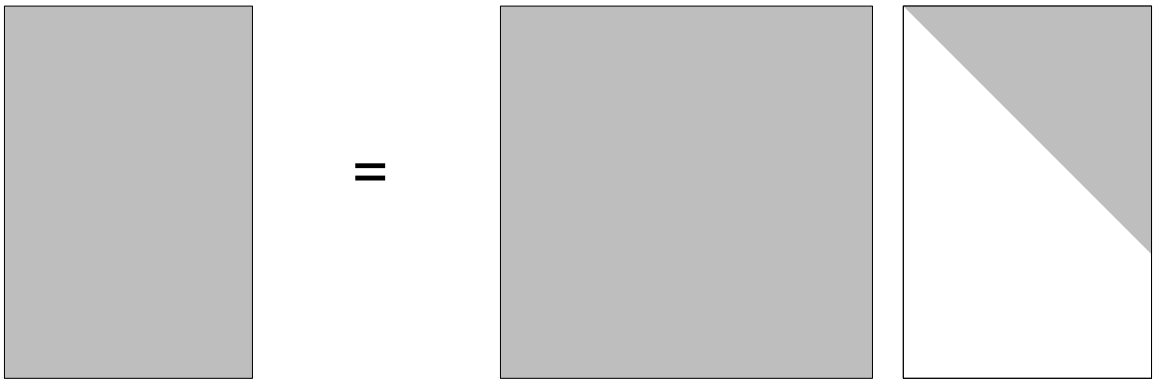
$$\kappa(A^T A) = 10^{10} \qquad \kappa(\hat{R}) = 10^{5}$$

The full QR Factorisation

Let A be an $m \times n$ matrix. The full QR factorisation of A is the factorisation $A = QR$, where

Q is $m \times m$ orthogonal ($QQ^T = I$)

R is $m \times n$ upper-trapezoidal

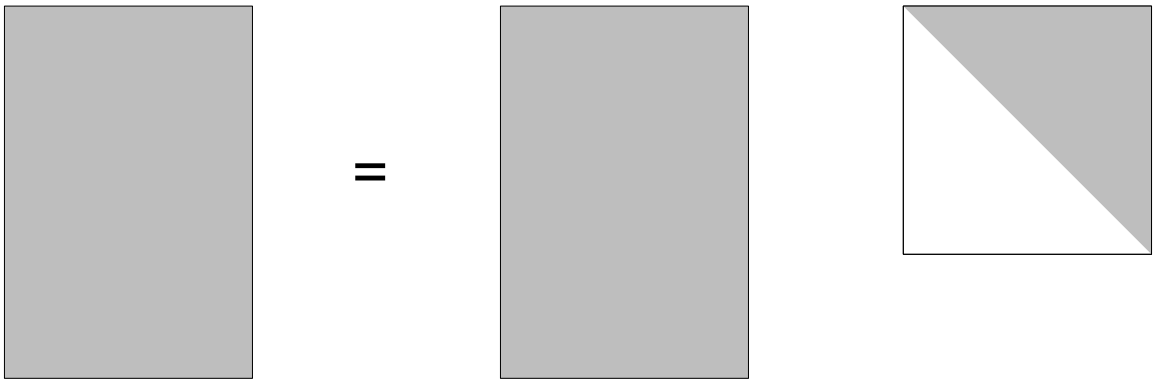


The reduced QR factorisation

Let A be an $m \times n$ matrix. The reduced QR factorisation of A is the factorisation$A = \hat{Q}\hat{R}$, where

$\hat{Q}$ is $m \times n$

$\hat{R}$ is $n \times n$ upper-triangular



A $\hat{Q}$ $\hat{R}$

Solution in the least-square sense

Instead of considering the system of normal equations, we can use the QR factorisation! The following result holds.

Theorem: Let $A \in \mathbb{R}^{m \times n}$, with $m \geq n$, be a full rank matrix. Then the unique solution in the least-square sense $\mathbf{x}^*$ of $A\mathbf{x}^* = \mathbf{b}$ is given by $\mathbf{x}^* = \hat{R}^{-1}\hat{Q}^T\mathbf{b}$, where $\hat{R} \in \mathbb{R}^{n \times n}$ and $\hat{Q} \in \mathbb{R}^{m \times n}$ are the matrices of the reduced QR factorisation of A. Moreover, the minimum of $\Phi(\cdot)$ is given by $\Phi(x^*) = \sum_{i=n+1}^{m} [(Q^T b)_i]^2$

Solution in the least-square sense - proof

Step 1. The QR factorization of A exists and is unique since A has full rank.

Thus, there exist two matrices, $Q \in \mathbb{R}^{m\times m}$ and $R \in \mathbb{R}^{m\times n}$ such that $A = QR$, where Q is orthogonal and R is an upper trapezoidal matrix.

Step 2. We observe that since Q is an orthogonal matrix ($Q^TQ = QQ^T = I$) it preserve the Euclidean scalar product, i.e.,

$$\|Q\mathbf{z}\|_2^2 = (Q\mathbf{z})^TQ\mathbf{z} = \mathbf{z}^TQ^TQ\mathbf{z} = \mathbf{z}^T\mathbf{z} = \|\mathbf{z}\|_2^2 \qquad \forall \mathbf{z} \in \mathbb{R}^m$$

$$\|Q^T\mathbf{z}\|_2^2 = (Q^T\mathbf{z})^TQ^T\mathbf{z} = \mathbf{z}^TQQ^T\mathbf{z} = \mathbf{z}^T\mathbf{z} = \|\mathbf{z}\|_2^2 \qquad \forall \mathbf{z} \in \mathbb{R}^m$$

it follows that

$$\|A\mathbf{x} - b\|_2^2 = \|Q^T(A\mathbf{x} - b)\|_2^2 = \|Q^T(QR\mathbf{x} - b)\|_2^2 = \|R\mathbf{x} - Q^Tb\|_2^2.$$

Recalling that R is upper trapezoidal, we have

$$\|A\mathbf{x} - b\|_2^2 = \|R\mathbf{x} - Q^Tb\|_2^2 = \|\hat{R}\mathbf{x} - \hat{Q}^Tb\|_2^2 + \sum_{i=n+1}^{m} [(Q^Tb)_i]^2$$

If A does not have full rank?

If A does not have full rank, the above solution techniques above fail.

In this case if $\mathbf{x}^*$ is a solution in the least square sense, the vector $\mathbf{x}^* + \mathbf{z}$, with$\mathbf{z} \in \ker(A)$, is a solution too.

We must therefore introduce a further constraint to enforce the uniqueness of the solution. Typically, one requires that $\mathbf{x}^*$ has minimal Euclidean norm, so that the least-squares problem can be formulated as:

find $\mathbf{x}^* \in \mathbb{R}^n$ **with minimal Euclidean** norm such that
$\|A\mathbf{x}^* - \mathbf{b}\|_2^2 \leq \min_{\mathbf{x}\in\mathbb{R}^n}\|A\mathbf{x} - \mathbf{b}\|_2^2$ (1)

This problem is consistent with our formulation. If A has full rank, since in this case the solution in the least-square sense exists and is unique it necessarily must have minimal Euclidean norm.
The tool for solving (1) is the singular value decomposition (SVD)

Singular Value Decomposition (SVD)

Any matrix can be reduced in diagonal form by a suitable pre and post-multiplication by unitary matrices.

Theorem. Let $A \in \mathbb{R}^{m\times n}$. There exist two **orthogonal** matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ such that

$$U^T A V = \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{m\times n} \qquad (2)$$

with

$p = \min(m, n)$ and $\sigma_1 \geq \ldots \geq \sigma_p \geq 0$. Formula (2) is called **Singular Value Decomposition** or (SVD) of A and the numbers σ_i are called singular values of A.

Singular Value Decomposition

Singular Value Decomposition

Singular Value Decomposition (SVD) is a robust mathematical tool commonly employed in machine learning for tasks such as
-dimensionality reduction,
- data compression,
-feature extraction.

It is especially effective in handling high-dimensional datasets, helping to lower computational complexity and enhance the efficiency of machine learning algorithms.

Remarks on SVD

- If A is a real-valued matrix, U and V will also be real-valued and in (2) U^T must be written instead of U^H.

- The following characterisation of the singular values holds

$$\sigma_i(A) = \sqrt{\lambda_i(A^T A)}, \qquad i = 1,\ldots,p\,.$$

Indeed, from (2) it follows that $A = U\Sigma V^T$, $A^T = V\Sigma^T U^T$, so that, since U and V are unitary,

$$A^T A = V\Sigma^T \Sigma V^T,$$

that is,

(3)

$$\lambda_i(A^T A) = \lambda_i(\Sigma^T \Sigma) = (\sigma_i(A))^2.$$

Remarks on SVD

- Since AA^T and A^TA are symmetric matrices, the columns of U, called the left singular vectors of A, turn out to be the eigenvectors of AA^T and, therefore, they are not uniquely defined. The same holds for the columns of V, which are the right singular vectors of A.

- As far as the rank(A) is concerned, if

$$\sigma_1 \geq \sigma_2 \ldots \geq \sigma_r > 0 \text{ and } \sigma_{r+1} = \ldots = \sigma_p = 0$$

then the rank of A is r, the kernel of A is the span of the column vectors of V, $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$, and the range of A is the span of the column vectors of U, $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$

Generalized inverse of A

Definition. Suppose that $A \in \mathbb{R}^{m \times n}$ has rank equal to r and that it admits a SVD of the type $U^T A V = \Sigma$. The matrix

$A^\dagger = V \Sigma^\dagger U^T$

is called the Moore-Penrose pseudo-inverse matrix, being

$$\Sigma^\dagger = \text{diag}\left\{ \frac{1}{\sigma_1}, \ldots, \frac{1}{\sigma_p}, 0, \ldots, 0 \right\}$$

The matrix $A^\dagger$ is also called the **generalized inverse of A**

If $n = m = rank(A)$, then $A^\dagger = A^{-1}$

Going back to our problem

find $\mathbf{x}^* \in \mathbb{R}^n$ **with minimal Euclidean** norm such that
$\|A\mathbf{x}^* - \mathbf{b}\|_2^2 \leq \min_{\mathbf{x}\in\mathbb{R}^n}\|A\mathbf{x} - \mathbf{b}\|_2^2$ (1)

Theorem. Let $A \in \mathbb{R}^{m\times n}$ with SVD given by $A = U\Sigma V^T$. Then the unique solution to (1) is

$$\mathbf{x}^* = A^\dagger \mathbf{b},$$

where $A^\dagger$ is the pseudo-inverse of A.

Proof

Proof. Using the SVD of A, problem (1) is equivalent to finding $\mathbf{w} = V^T\mathbf{x}$ such that $\mathbf{w}$ has minimal Euclidean norm and

$$\|\Sigma\mathbf{w} - U^T\mathbf{b}\|_2^2 \leq \|\Sigma\mathbf{y} - U^T\mathbf{b}\|_2^2, \qquad \forall \mathbf{y} \in \mathbb{R}^n .$$

If r is the number of nonzero singular values σ_i of A, then

$$\|\Sigma\mathbf{w} - U^T\mathbf{b}\|_2^2 = \sum_{i=1}^{r} (\sigma_i w_i - (U^T\mathbf{b})_i)^2 + \sum_{i=r+1}^{p} ((U^T\mathbf{b})_i)^2$$

which is minimum if $\sigma_i w_i - (U^T\mathbf{b})_i = 0 \quad \forall i = 1,\ldots,r$. Moreover, it is clear that among the vectors $\mathbf{w}$ of $\mathbb{R}^n$ having the first r components fixed, the one with minimal Euclidean norm has the remaining $n - r$ components equal to zero. Thus the solution vector is $\mathbf{w}^* = \Sigma^\dagger U^T\mathbf{b}$, that is,

$$\mathbf{x}^* = V\Sigma^\dagger U^T\mathbf{b} = A^\dagger\mathbf{b},$$

where $A^\dagger$ is the pseudo-inverse of A.

Computing the SVD - 1

The SVD can be computed by performing an eigenvalue computation for the normal matrix $A^T A$. Indeed, let U and V have column partitions

$$U = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \qquad V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$$

From the relations

$$A\mathbf{v_j} = \sigma_j \mathbf{u_j}, \qquad A^T \mathbf{u_j} = \sigma_j \mathbf{v_j}$$

it follows that

$$A^T A \mathbf{v_j} = \sigma_j^2 \mathbf{v_j}$$

It follows that one approach to computing the SVD of A is to apply the symmetric QR algorithm. Then, the relations $A\mathbf{v_j} = \sigma_j \mathbf{u_j}, j = 1, \ldots, p$ can be used in conjunction with the QR factorisation with column pivoting to obtain U.

This squares the condition number for small singular values and is not numerically-stable.

Computing the SVD - 2

A possible remedy is to proceed in two steps.
First, we can use (Householder) reflections to reduce A to upper bidiagonal form:

$$U_1^T A V_1 = B = \begin{bmatrix} d_1 & f_1 & \dots & \dots & \\ 0 & d_2 & f_2 & & \dots \\ \vdots & \ddots & \vdots & & \vdots \\ \vdots & \ddots & \vdots & d_{n-1} & f_{n-1} \\ 0 & 0 & \dots & 0 & d_n \end{bmatrix}$$

It follows that $T = B^T B$ is symmetric and tridiagonal.
We could then apply the (symmetric) QR algorithm directly to B.

Conclusions/Summary

- The singular value decomposition (SVD) is an alternative to the eigenvalue decomposition that is better for rank-deficient and ill-conditioned matrices in general.
- Computing the SVD is always numerically stable for any matrix, but is typically more expensive than other decompositions.
- The SVD can be used to compute low-rank approximations to a matrix via the principal component analysis (PCA) (not discussed)
- PCA has many practical applications, and usually large sparse matrices appear.