Linear Algebra (not only) for computer scientists

Translation Luděk Kučera

Milan Hladík

18. května 2017

Chapter 11 Positive (semi-)definite matrices

Definition 11.1 (Positive (semi-)definite matrices) Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. A is *positive semidefinite*, if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$, and A is *positive definite*, if $x^T Ax > 0$ for all $x \neq o$.

It is clear that a positive definite matrix A is positive definite.

It would be possible to define negative (semi-)definite matrices using a reversed inequality. However, we will not deal with them, because a matrix A is negative (semi-)definite if and only if $-A$ is positive (semi-)definite, which reduces the negative (semi-)definity to the positive one.

Remark 11.2 The definition makes sense for non-symmetric matrices as well, but such matrices can be "symmetrized" by considering $\frac{1}{2}(A + A^T)$, because

$$x^T \frac{1}{2}(A + A^T)x = \frac{1}{2}x^T Ax + \frac{1}{2}x^T A^T x = \frac{1}{2}x^T Ax + (\frac{1}{2}x^T Ax)^T = x^T Ax.$$

Thus, instead of testing the condition for A, we can use a symmetric matrix $\frac{1}{2}(A + A^T)$. It follows that limiting ourselves to symmetric matrices does not affect generality. The reason why we limit our consideration to symmetric matrices is that many test conditions work well for symmetric matrices only.

Example 11.3 An example of a positive semidefinite matrix is 0_n. An example of a positive definite matrix is I_n, because $x^t Ax = x^T I_n x = x^T x = \|x\|_2^2$.

Remark 11.4 (A necessary condition for positive (semi-)definitivity). Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. In order to be positive semidefinite, it must be $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$. By setting $x = e_i$, $i = 1, \ldots, n$, we get $0 \leq x^T Ax = e_i^T Ae_i = a_{ii}$. It follows that a positive semidefinite matrix must have a non-negative diagonal, a positive definite matrix must have a positive diagonal.

Remark 11.5 A matric $A = (a) \in \mathbb{R}^{1\times 1}$ is positive definite if and only if $a \geq 0$, and a positive definite if and only if $a > 0$. Thus, positive semidefiniteness is a generalization of the notion of non-negativity from numbers to matrices. This is why positive semidefiniteness of a matrix A is often denoted as $A \succeq 0$ (to be distinguished from $A \geq 0$, which is used for non-negativity of all elements).

Example 11.6 Properties of positive definite matrices)
(1) If $A, B \in \mathbb{R}^{n\times n}$ are positive definite, then $A + B$ is positive definite,
(2) If $A \in \mathbb{R}^{n\times n}$ is positive definite, and $\alpha > 0$, then αA is positive definite,
(3) If $A \in \mathbb{R}^{n\times n}$ is positive definite, then A is regular and A^{-1} is positive definite.

Proof. The first two properties are trivial, we prove the third one.

Let us first check regularity of A. Let x be a solution of $Ax = o$. Then $x^T Ax = x^T o = 0$. Therefore it must be $x = o$.

Now we prove that A^{-1} is positive definite. Assume that there is $x \neq o$ such that $x^T A^{-1}x \leq 0$. Then

$$x^T A^{-1}x = x^T A^{-1}AA^{-1}x = y^T Ay \leq 0,$$

where $y = A^{-1}x \neq o$ - a contradiction, because A is positive definite. ♣

An analogy is valid for positive semidefinite matrices. (1) is unchanged, (2) holds for all $\alpha \geq 0$, but (3) is not true in general.

The product of positive definite matricesis dealt with in example 12.16.

The next theorem gives an important characterization ofpositive definite matrices both using eigenvalues and using the form $U^T U$.

Theorem 11.7 (Characterization of positive definite matrices). Let $A \in \mathbb{R}^{n\times n}$ be symmetric. The the following conditions are equivalent:
(1) A is positive definite,
(2) all eigenvalues of A are positive,
(3) there is a matrix $U \in \mathbb{R}^{m\times n}$ of the rank n such that $A = U^T U$.

Proof. The implication (1) $\Rightarrow$ (2). By contradiction: assume that there is an eigenvalue $\lambda \leq 0$, and x is the corresponding eigenvector of the Euclidean length 1. Then $Ax = \lambda x$ implies $x^T Ax = \lambda x^T x = \lambda \leq 0$, which contradicts the positive definiteness of A.

The implication (2) $\Rightarrow$ (3). Since A is symmetric, it has a spectral decomposition $A = Q\Lambda Q^T$, where Λ is a diagonal matrix with the diagonal elements $\lambda_1, \ldots, \lambda_n > 0$. Let us define a diagonal matrix Λ' as having the diagonal elements $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n} > 0$. The the matrix U is $U = \Lambda' Q^T$, because $U^T U = Q\Lambda'\Lambda' Q^T = Q\Lambda Q^T = A$. Note that U nas the rank n andit is regular, because it is a product of two regular matrices.

The implication (3) $\Rightarrow$ (1). By contradiction. Let $x^T Ax \leq 0$ for some $x \neq o$. Then $0 \leq x^T Ax = x^T U^T Ux = (Ux)^T(Ux) = \langle Ux, Ux\rangle = \|Ux\|_2^2$. This means that $Ux = o$, but since the rank of U is n, we get $x = o$, a contradiction. ♣

A similar characterization of positive semidefiniteness follows without proof.

Theorem 11.8 (Characterization of positive semidefinite matrices). The following statements are equivalent:
(1) A is positive semidefinite,
(2) all eigenvalues of A are non-negative,
(3) there is a matrix $U \in \mathbb{R}^{m\times n}$ such that $A = U^T U$.

11.1 Methods of testing of positive definiteness

Now we will look for particular methods of positive definiteness testing. A number of them follows from the following recurrent formula. Note that it can be neither used nor easily modified to test positive semidefiniteness.

Theorem 11.9 (A recurrent formula for positive definiteness testing) A symmetric matrix $A = \begin{pmatrix} \alpha & a^T \\ a & \tilde{A} \end{pmatrix}$, where $\alpha \in \mathbb{R}$, $a \in \mathbb{R}^{n-1}$, $\tilde{A} \in \mathbb{R}^{(n-1)\times(n-1)}$ is positive definite if and only if $\alpha > 0$ and $\tilde{A} - \frac{1}{\alpha}aa^T$ is positive definite.

Proof. The implication "$\Rightarrow$". Let A be positive definite. Then $x^T Ax > 0$ for all $x \neq 0$; especiall for $x = e_1$ we get $\alpha = e_1^T Ae_1 > 0$. Moreover, let $\tilde{x} \in \mathbb{R}^{n-1}$, $\tilde{x} \neq o$. Then

$$\tilde{x}^T \left(\tilde{A} - \frac{1}{\alpha}aa^T\right)\tilde{x} = \tilde{x}^T\tilde{A}\tilde{x} - \frac{1}{\alpha}(a^T\tilde{x})^2 = \left(-\frac{1}{\alpha}a^T\tilde{x}\tilde{x}^T\right)\begin{pmatrix} \alpha & a^T \\ a & \tilde{A} \end{pmatrix}\begin{pmatrix} -\frac{1}{\alpha}a^T\tilde{x} \\ \tilde{x} \end{pmatrix} > 0.$$

The implication "$\Leftarrow$". Let $x = \begin{pmatrix} \beta \\ \tilde{x} \end{pmatrix} \in \mathbb{R}^n$. Then

$$x^T Ax = \begin{pmatrix} \beta & \tilde{x}^T \end{pmatrix}\begin{pmatrix} \alpha & a^T \\ a & \tilde{A} \end{pmatrix}\begin{pmatrix} \beta \\ \tilde{x} \end{pmatrix} = \alpha\beta^2 + 2\beta a^T\tilde{x} + \tilde{x}^T\tilde{A}\tilde{x} =$$

$$= \tilde{x}^T\left(\tilde{A} - \frac{1}{\alpha}aa^T\right)\tilde{x} + \left(\sqrt{\alpha}\beta + \frac{1}{\sqrt{\alpha}}a^T\tilde{x}\right)^2 \geq 0.$$

The equality holds only if $\tilde{x} = o$ and the second square is zero, i.e., $\beta = 0$. ♣

Even though the recurent formula can be used to test positive definiteness, the following theorem, the Cholesky[1] decomposition, is much more important.

Theorem 11.10 (Cholesky decomposition). Given a positive definite matrix $A \in \mathbb{R}^{n\times n}$, there is the unique lower triangular matrix $L \in \mathbb{R}^{n\times n}$ with a positive diagonal, such that $A = LL^T$.

Proof. By mathematical induction on n. For $n = 1$ it is $A = (a_{11})$ and $L = (\sqrt{11})$.

The induction step $n \leftarrow n-1$. Suppose $A = \begin{pmatrix} \alpha & a^T \\ a & \tilde{A} \end{pmatrix}$. In view of Theorem 11.9, $\alpha > 0$ and $\tilde{A} - \frac{1}{\alpha}aa^T$ is positively definite. The induction hypothesis implies the existence of a lower triangular matrix $\tilde{L} \in \mathbb{R}^{(n-1)\times(n-1)}$ with a positive diagonal such that $\tilde{A} - \frac{1}{\alpha}aa^T = \tilde{L}\tilde{L}^T$. We will prove that $L = \begin{pmatrix} \sqrt{\alpha} & o^T \\ \frac{1}{\sqrt{\alpha}}a & \tilde{L} \end{pmatrix}$, because

$$LL^T = \begin{pmatrix} \sqrt{\alpha} & o^T \\ \frac{1}{\sqrt{\alpha}}a & \tilde{L} \end{pmatrix}\begin{pmatrix} \sqrt{\alpha} & \frac{1}{\sqrt{\alpha}}a^T \\ o & \tilde{L}^T \end{pmatrix} = \begin{pmatrix} \alpha & a^T \\ a & \frac{1}{\alpha}aa^T + \tilde{L}\tilde{L}^T \end{pmatrix} = A.$$

[1] André-Louis Cholesky, a French officer (most likely of Polish origin), developped the method in 1910 for triangulation and creation of more precise maps (by solving systems of normal equations by the least square method), the method was published in 1924, after his death in the 1st World War.

To prove uniqueness, let $A = L'L'^T$ be another such decomposition, where $L' = \begin{pmatrix} \beta & o^T \\ b & \tilde{L}' \end{pmatrix}$. Then

$$\begin{pmatrix} \alpha & a^T \\ a & \tilde{A} \end{pmatrix} = A = L'L'^T = \begin{pmatrix} \beta^2 & \beta b^T \\ \beta b & bb^T + \tilde{L}'\tilde{L}'^T \end{pmatrix}.$$

By comparing matrices, we get $\beta = \sqrt{\alpha}$, $b = \frac{1}{\sqrt{\alpha}} a$ and $\tilde{A} = bb^T + \tilde{L}'\tilde{L}'^T$, which means $\tilde{L}'\tilde{L}'^T = \tilde{A} - \frac{1}{\alpha} aa^T$. But the induction hypothesis says that the decomposition $\tilde{A} - \frac{1}{\alpha} aa^T = \tilde{L}\tilde{L}^T$ is unique, and hence $\tilde{L}' = \tilde{L}$, which implies $L' = L$. ♣

Cholesky decomposition exists for positive semidefinite matrices as well, but it is not unique.

The theorem is more or less of existential character, but the construction of Cholesky decomposition is quite simple. The basic idea is to compare in the top-down manner the elements of the first column of the matrix $A = LL^T$, then the elements in the second column, etc. The final method is given below. If A is positive definite, the algorithm gives the decomposition, in the negative case it announces that A is not positive definite.

1: $L := 0_n$,

2: **for** $k := 1$ **to** n **do** // in the k-th cycle we determine the values L_{*k}

3: **if** $a_{kk} - \sum_{j=1}^{k-1} \ell_{kj}^2 \leq 0$ **then return** "A is not positive definite",

4: $\ell_{kk} := \sqrt{a_{kk} - \sum_{j=1}^{k-1} \ell_{kj}^2}$,

5: **for** $i := k + 1$ **to** n **do**

6: $\ell_{ik} := \frac{1}{\ell_{kk}} \left(a_{ik} - \sum_{j=1}^{k-1} \ell_{ij}\ell_{kj} \right)$,

7: **end for**

8: **end for**

9: **return** $A = LL^T$.

Proof of correctness of the Cholesky algorithm. Suppose we have computed the first up to $(k-1)$-th column of the matrix L. From the equation $A = LL^T$ we get

$$a_{kk} = \sum_{j=1}^{n} L_{kj}(L^T)_{jk} = \sum_{j=1}^{n} \ell_{kj}^2 = \sum_{j=1}^{k} \ell_{kj}^2.$$

The only unknown in this equation is the value of ℓ_{kk}, and if it is expressed, we get the formula of the step 4.

Now, suppose that we know the first $i - 1$ elements of the k-th column of the metrix L. The equation $A = LL^T$ gives for $i > k$ that

$$a_{ik} = \sum_{j=1}^{n} L_{ij}(L^T)_{jk} = \sum_{j=1}^{n} \ell_{ij}\ell_{kj} = \sum_{j=1}^{k} \ell_{ij}\ell_{kj}.$$

The only unknown in this equation is the value of ℓ_{ik}, and if it is expressed, we get the formula of the step 6.

The algorithm can also be used as an alternative proof of the uniqueness of Cholesky decomposition. The elements of the matrix A have determined the elements of the matrix L in a unique way, nowhere there was a choice from greater number of possibilities.

Example 11.12 Cholesky decomposition of a matrix $A = LL^T$:

$$\begin{pmatrix} 4 & -2 & 4 \\ -2 & 10 & 1 \\ 4 & 1 & 6 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ -1 & 3 & 0 \\ 2 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 2 & -1 & 2 \\ 0 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Example 11.13 The use of Cholesky decomposition for a solution of a system $Ax = b$ with positive definite matrix A. If we know the decomposition $A = LL^T$, the system has the form $L(L^T x) = b$. We solve first the system $Ly = b$, then $L^T x = y$ and we get the vector x. The method:

1. Find Cholesky decomposition $A = LL^T$,
2. Find the solution of $Ly = b$ by the forward substitution,
3. Find the solution of $L^T x = y$ by the backward substitution.

This method is by about 50 % faster than the Gaussian elimination.

Cholesky decomposition can also be used to invert positive definite matrices, because $A^{-1} = (LL^T)^{-1} = (L^{-1})^T L^{-1}$ and inverting of the lower triangular matrix L is easy.

The recurrent formula has other consequences, that show how to test positive definiteness using Gauss-Jordan elimination and using determinants.

Theorem 11.14 (Gaussian elimination and positive definiteness). A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive definite if and only if the Gaussian elimination transforms it to row echelon form (an upper triangular matrix) with positive diagonal using only the elementary operation of adding a multiple of the row with a pivot k to another row below it.

Proof. Let $A = \begin{pmatrix} \alpha & a^T \\ a & \tilde{A} \end{pmatrix}$ be positive definite. The first step of Gaussian elimination transforms the matrix to the form $\begin{pmatrix} \alpha & a^T \\ o & \tilde{A} - \frac{1}{\alpha} a a^T \end{pmatrix}$. In view of Theorem 11.9 it is $\alpha > 0$ and $\tilde{A} - \frac{1}{\alpha} a a^T$ is positive definite as well, so we can proceed inductively on. ♣

Theorem 11.15 (Sylvester[2] criterion of positive definiteness). A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive definite if and only if the determinants of all

[2] James Joseph Sylvester, an English mathematician, a co-founder of the matrix theory. He was the first to use the term "matrix" in a work from the year 1850. The criterion is from the year 1852.

main leading submatrices $A_1, \dots, A_n$ are positive, where A_i is the left upper submatrix of A of the size i (which means that it is obtained from A by removing the last $n-i$ rows and columns).

Proof. The implication "⇒". Let $A \in \mathbb{R}^{n\times n}$ is positive definite. Then for each $i = 1, \dots, n$ the matrix A_i is positive definite, because is $x^T A_i x \leq 0$ for some $x \neq o$, then $\begin{pmatrix} x^T & o^T \end{pmatrix} A \begin{pmatrix} x \\ o \end{pmatrix} = x^T A_i x \leq 0$. This implies that the eigenvalues of A_i are positive and therefore its determinant is positive as well, being equal to the product of eigenvalues.

The implication "⇐". During Gaussian elimination of the matrix A, all pivots are positive, because if the i-th pivot is non-positive, then $\det(A_i) \leq 0$. In view of 11.14, A is positive definite. ♣

The non-negativity of all main leading submatrices does not imply positive semidefiniteness (find such an example!). An analogy of Sylvester condition for positive semidefinite matrices is the following

Theorem 11.14 (Sylvester criterion of positive semidefiniteness). A symmetric matrix $A \in \mathbb{R}^{n\times n}$ is positive semidefinite if and only if the determinants of all main submatrices are non-negative, where a main submatrix is a matrix that it is obtained from A by removing certain (includin zero) number of rows and columns with the same indices.

Proof. If A is positive semidefinite, then the main submatrices are obviously positive semidefinite as well, and hence they have non-negative determinants (= products of eigenvalues).

The proof of the implication in the other direction is by mathematical induction. For $n = 1$ the statement is obvious.

The inductio step $n \leftarrow n-1$ by contradiction. Assume that $\lambda < 0$ is an eigenvalue of A, and let x be the corresponding eigenvector such that $\|x\|_2 = 1$. If all other eigenvalues are positive, then $\det(A) < 0$, a contradiction.

In the other case let $\mu \leq 0$ beanother eigenvalue of A and let y be the corresponding eigenvector such that $\|y\|_2 = 1$. We will find $'alpha \in \mathbb{R}$ such that the vector $z = x + \alpha y$ has at least one element equal to 0; suppose that the element is i-th in the vector. Since $x \perp y$, we have

$$z^T A z = (x+\alpha y)^T A (x+\alpha y) = (x+\alpha y)^T (Ax+\alpha Ay) = (x+\alpha y)^T (\lambda x+\alpha \mu y) = \lambda x^T x+\alpha^2 \mu y^T y = \lambda+\alpha^2\mu < 0.$$

Let A' be obtained from A by removing the i-th row and column, and z' be obtained from z by removing the i-th element. Then $z'^T A' z' = z^T A z < 0$, and hence the main submatrix A' is not positive demidefinite, and we can apply the induction hypothesis. ♣

While Sylvester criterion of positive definiteness requires computing n determinants, the criterion for positive semidefiniteness involves computing of up to $2^n - 1$ determinant, and therefore it is not a very practical method. Better method will be shown in Section 12.2 (Consequence 12.13).

Even though we showed several methods of positive definiteness testing, some of them are quite similar. The proog of Theorem 11.14 shows that a recurrent formula and Gaussian elimination work essentiall in the same way. And if determinants are computed by Gaussian elimination, Sylvester criterion is a variant of the former two tests. However, Cholesky decomposition is a method that is principially different.

11.2 Applications

Theorem 11.17 (Scalar product and positive semidefiniteness). An operation $\langle x, y\rangle$ is a scalar product in $\mathbb{R}^n$ if and only if it has the form $\langle x, y\rangle = x^T Ay$ for some positive definite matrix $A \in \mathbb{R}^{n\times n}$.

Proof. The implication "$\Rightarrow$". Let us define a matrix $\in \mathbb{R}^{n\times n}$ by $A_{ij} = \langle e_i, e_j\rangle$. The matrix is obviously symmetric. Then

$$\langle x, y\rangle = \left\langle \sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{n} y_j e_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i y_j A_{ij} = x^T Ay.$$

The implication "$\Leftarrow$". Let A be positive definite. Then $\langle x, y\rangle = x^T Ay$ is a scalar product, because:
$\langle x, x\rangle = x^T Ax \geq 0$ and it is equal to 0 only for $x = o$,
the function is linear in the first coordinate,
and it is symmetric, because

$$\langle x, y\rangle = x^T Ay = (x^T Ay)^T = y^T A^T x = y^T Ax = \langle y, x\rangle. \qquad \clubsuit$$

We know that a scalar product induces a norm (Definition 8.4). A norm induces by the above scalar product is $\|x\| = \sqrt{x^T Ax}$. In this norm the unit sphere is an ellipsoid (see example 12.18). For $A = I_n$, we get the standard scalar product and the euclidean norm.

Even though a non-standard scalar product $\langle x, y\rangle = x^T Ay$ can look strange, its relation to the standard one is very close. Since the matrix is positive definite, it can be decomposed as $R^T R$, where R is regular. Let B be the basis formed by the columns of the matrix R^{-1}, i.e., $R =_B [id]_{\text{kan}}$ is the matrix of the transformation from the canonical basis to B. Now, $x^T Ay = x^T R_T Ry = (Rx)^T(Ry) = [x]_B^T[y]_B$. This shows that a non-standard scalar product is the standard scalar product with respect to certain basis.

The next application is a square root of a matrix. For positive semidefinite matrices we cand define a positive definite square root $\sqrt{A}$. The square root is even unique, see Rohn (2003).

Theorem 11.18 (Square root of a matrix). For every symmetric positive semidefinite matrix $A \in \mathbb{R}^{n\times n}$ there is a positive semidefinite matrix $B \in \mathbb{R}^{n\times n}$ such that $B^2 = A$.

Proof. Let the spectral decomposition o A be $A = Q\Lambda Q^T$, where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, where $\lambda_1, \ldots, \lambda_n \geq 0$. Let us define a diagonal matrix $\Lambda' = \operatorname{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$ and a matrix $B = Q\Lambda' Q^T$. Then $B^2 = Q\Lambda' Q^T Q\Lambda' Q^T = Q\Lambda'^2 Q^T = Q\Lambda Q^T = A$. $\clubsuit$

Remark 11.19 Positive (semi)definiteness is also important in optimalization when determining a minimum of a function. The matrix that is used here, so called hessian, is the matrix of the second partial derivatives. Supposing we are in a point x^* with the zero gradient, then positive definitiveness is a sufficient condition for x^* to be a local minimum, while positive semidefinitiveness is a necessary condition.

The hessian is used similarly to determine convesity of a function. Positive semidefinitiveness at some open convex set implies convexity of he function f. See Rohn (1997).

Positive definite matrices play another important role in optimalization. A semidefinite program is such an optimalization problem, where we are looking for the minimum of a linear function under a condition of positive semidefiniteness of a matrix, whose elements are linear functions of variables (see Gärtner and Matoušek, 2012). Formally, the problem is

$$\min c^T \text{ under the condition } A_0 + \sum_{i-1}^{m} A_i x_i \text{ is positive definite,}$$

where $c \in \mathbb{R}^m$, $A_0, A_1, \ldots, A_m \in \mathbb{R}^{n \times n}$ are given and $x = (x_1, \ldots, x_m)^T$ is the vector of variables. Semidefinite programs model bigger class of problems than linear programs (see Remark 7.18), but they are still solvable efficiently in reasonable time. They make it possible to obtain great progress in combinatorial optimization, because many computationally difficult problems can be tightly approximated in a short time using appropriate semidefinite programs.

However, the occurence of (semi)definite matrices is even broader. E.g., in mathematical statistics we find so called covariant and correlation matrices. Both give certain information about dependence among n ramdom variables and, not by chance, all of them are positeve semidefinite.

Example 11.20 Skipped.

Insert

The following text is not in Hladik's text, but it gives another very important application of positive definite matrices. Suppose we are going to solve a system of linear equations $Ax = b$. It happens very often in numerial mathematics that such system is very sparse, which means that the matrix A has only a very small number of non-zero elements in each row, typically less that one hundered, but often even in the range 10-30. On the other hand, the dimension of such a matrix can be extremely large. If using double precision numbers (i.e., 8 bytes per one number), one row in a sparse matrix could occupy less than one kilobyte (1 kB = 10^3 byte) of the computer memory. However, the biggest computing sytems nowadays use memory of the capacity about one petabyte (1 PB = 10^{15} bytes (!!!)). In this way we are able to store matrices of the dimension $n \approx 10^{12}$ (million of millions of rows and columns).

If we try to solve such a system using direct method like the Gaussian elimination, the matrix A gets dense, but it is absolutely impossible to store a dense matrix of the size $10^{12} \times 10^{12}$! Using Cholesky decomposition fails for the same reason. This is why such systems can only be solved by indirect iterative methods. The by far the most simple case is when the matrix of the system is symmetric and positive definite, which, fortunately, is qute frequent case.

Theorem Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive matrix, $b \in \mathbb{R}^n$. Define a functional f by:

$$f(x) = \frac{1}{2}x^T Ax - x^T b \qquad \text{for each } x \in \mathbb{R}^n.$$

Then the (unique) solution x_{sol} of the system $Ax = b$ is the (unique) minimum of the functional f.

Proof. Consider a vector $x \in E_n$ and a solution x_{sol} of the system $Ax = b$. Put $e = x - x_{sol}$. Then

$$f(x) = f(x_{sol} + e) = \frac{1}{2}(x_{sol} + e)^T A(x_{sol} + e) - (x_{sol} + e)^T b =$$

$$= \frac{1}{2}(x_{sol}^T Ax_{sol} + x_{sol}^T Ae + e^T Ax_{sol} + e^T Ae) - x_{sol}^T b - e^T b =$$

$$= (\frac{1}{2}x_{sol}^T Ax_{sol} - x_{sol}^T b) + e^T Ae + e^T (Ax_{sol} - b) = f(x_{sol}) + e^T Ae,$$

because $x_{sol}^T Ae = e^T Ax_{sol}$, and therefore $\frac{1}{2}(x_{sol}^T Ae + e^T Ax_{sol}) = e^T Ax_{sol}$.

Since A is pozitive definite, the value of the expression $e^T Ae$ is always non-negative, and it is 0 (the smallest possible value) if and only if $e = o$, that is if and only if $x = x_{sol}$.