Linear Algebra

Petros Maragos

School of Electrical & Computer Engineering,
National Technical University of Athens,
Zografou 15773, Athens, Greece.
Email: maragos@cs.ntua.gr

0

We assume that the basic concepts and tools of linear algebra are known at the analytical level as in Strang [2] and hopefully also at the numerical level as in Golub & Van Loan [1]. Herein we only summarize a few ideas and results on (i) linear equations using least squares and (ii) matrix factorizations emphasizing SVD, as a supplement of the linear methods in this book. Our exposition follows Strang [2].

1 Linear Equations and Least Squares

To solve a system of m linear equations in n unknowns, represented by the matrix equation

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

it greatly helps if we consider the four fundamental spaces related to matrix $\boldsymbol{A}$ and their geometry, depicted in Fig. 1(a). We can assume that all coefficients and unknowns are real numbers. Thus, $\boldsymbol{A}$ represents a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, whose range space is $\mathcal{R}(\boldsymbol{A})$ and null space is $\mathcal{N}(\boldsymbol{A})$. Its rank is

$$r \triangleq \text{rank}(\boldsymbol{A}) \leq \min(m, n)$$

Then the following theorem summarizes some algebraic and geometric properties of the range and null spaces of $\boldsymbol{A}$ and its transpose.

Theorem 1 (Fundamental Theorem of Linear Algebra)
(a) The dimensions of the four fundamental spaces of a real $m \times n$ matrix $\boldsymbol{A}$ are:

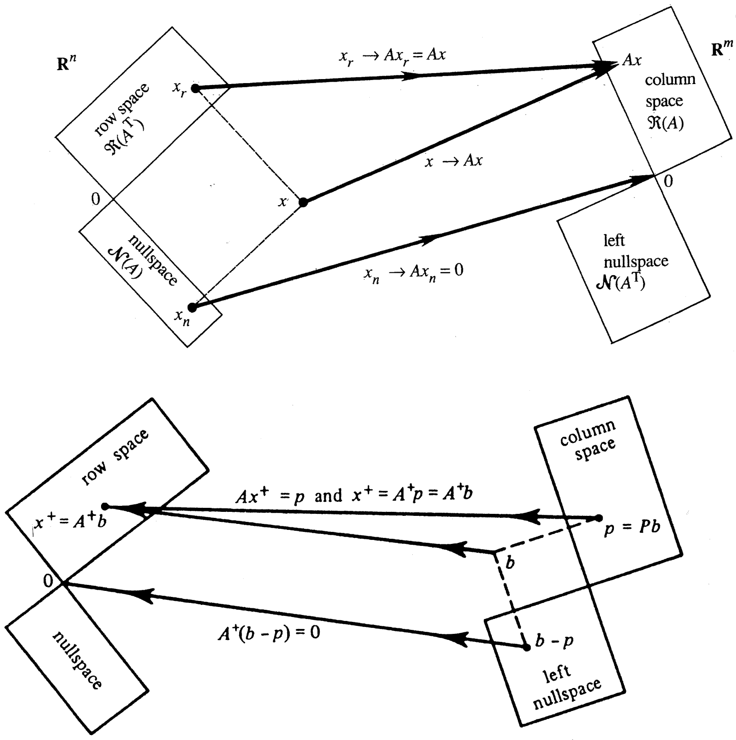
$$\begin{aligned} \text{column space} &: \dim \mathcal{R}(\boldsymbol{A}) = r \\ \text{row space} &: \dim \mathcal{R}(\boldsymbol{A}^T) = r \\ \text{null space} &: \dim \mathcal{N}(\boldsymbol{A}) = n - r \\ \text{left null space} &: \dim \mathcal{N}(\boldsymbol{A}^T) = m - r \end{aligned} \tag{1}$$

(b) In $\mathbb{R}^m$, the orthogonal complement of the column space of $\boldsymbol{A}$ is the null space of its transpose. In $\mathbb{R}^n$, the orthogonal complement of the null space of $\boldsymbol{A}$ is the row space of its transpose.

It can be shown that the above theorem can also be extended to a complex matrix $\boldsymbol{A}$, by using the Hermitian of the matrix instead of its transpose and by viewing the matrix as a linear map from $\mathbb{C}^n$ to $\mathbb{C}^m$.

0 Copyright © by Petros Maragos 2005-2009. These lecture notes are a part of a textbook on image analysis and signal processing to be published. *Working version:* 31 Dec. 2009. It is for the sole and exclusive use of the students enrolled in university courses taught by P. Maragos. It is not to be sold, reproduced, or generally distributed without the author's consent.

!h



(a)

(b)

Figure 1: (a) The four fundamental subspaces in solving $\boldsymbol{Ax} = \boldsymbol{b}$ of m linear equations in n unknowns. (b) The geometry of the pseudo-inverse in the least-squares solution. (Figure from Strang [2].)

As Fig. 1(a) shows, the mapping between the row and column spaces is always invertible. The existence and uniqueness of the solution depend on the relationships among r, m, n and on $\boldsymbol{b}$. We distinguish two cases.

Case I (Full Rank): $r = \min(m, n)$:
If $r = m \leq n$ (independent rows), then there exists *at least one solution.* Namely, every vector $\boldsymbol{b}$ belongs to the column space and comes from a unique vector $\boldsymbol{x}_r$ in the row space such that $\boldsymbol{A}\boldsymbol{x}_r = \boldsymbol{b}$. The full solution will be $\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_n$, where $\boldsymbol{x}_n$ belongs to the null space. If $r = m = n$, the solution is unique, i.e. $\boldsymbol{x}_n = 0$, and can be found using the *matrix inverse*: $\boldsymbol{x} = \boldsymbol{x}_r = \boldsymbol{A}^{-1}\boldsymbol{b}$. If $r = m < n$, there is an infinite set of solutions created by selecting any nonzero $\boldsymbol{x}_n$ from the $(n - r)$-dimensional null space.

If $r = n < m$ (independent columns), we have *at most one solution.* Let us consider first the most frequent case where $\boldsymbol{b} \notin \mathcal{R}(\boldsymbol{A})$ and we have an *inconsistent* system of equations which has no solution. However, we can search for a **least squares solution** that minimizes the Euclidean norm of the approximation error:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{Ax} - \boldsymbol{b}\| \tag{2}$$

This approximate solution is obtained by solving the $n \times n$ system of *normal equations*

$$\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$$

Note that $\boldsymbol{A}^T\boldsymbol{A}$ is invertible since $\boldsymbol{A}$ has independent columns. Hence, the least squares solution is

$$\hat{\boldsymbol{x}} = \boldsymbol{A}^\dagger\boldsymbol{b}, \quad \boldsymbol{A}^\dagger \triangleq (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T \tag{3}$$

where $\boldsymbol{A}^\dagger$ is the *Moore-Penrose pseudo-inverse* of the matrix $\boldsymbol{A}$. A geometrical insight can be gained if we realize that the orthogonal projection of $\boldsymbol{b}$ onto the column space is the vector

$$\boldsymbol{p} = \boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{A}^\dagger\boldsymbol{b}$$

Now, in the rare case where $\boldsymbol{b}$ belongs to the column space of $\boldsymbol{A}$ the result (3) becomes a unique solution of the original system $\boldsymbol{Ax} = \boldsymbol{b}$. This is an exact solution with zero approximation error.

Case II (Low Rank): $r < \min(m, n)$:
Now both the rows and the columns of $\boldsymbol{A}$ are linearly dependent. Since the most general and interesting case is when $\boldsymbol{b}$ does not belong to the column space, consider the projection $\boldsymbol{p}$ of $\boldsymbol{b}$ onto the column space; see Fig. 1(b). We are generally interested in *least squares solutions* $\hat{\boldsymbol{x}}$ as in (2). (Of course, if $\boldsymbol{b}$ belongs to the column space, these solutions become exact and yield zero error.) There exists at least one such solution; it is the unique vector $\hat{\boldsymbol{x}}_r$ in the row space with $\boldsymbol{A}\hat{\boldsymbol{x}}_r = \boldsymbol{p}$. Unfortunately, this solution is not unique because $r < n$. Specifically, we can obtain an infinite number of least squares solutions $\hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}_r + \hat{\boldsymbol{x}}_n$ by adding orthogonal vectors $\hat{\boldsymbol{x}}_n$ from the null space. However, if we select $\hat{\boldsymbol{x}}_n = \boldsymbol{0}$, this will give us a unique solution with minimum norm. Thus, by adding the constraint that the least squares solution $\hat{\boldsymbol{x}}$ should also have minimum norm $\|\hat{\boldsymbol{x}}\|$, we find that $\hat{\boldsymbol{x}}_r$ is the *unique least squares solution with minimum length*, denoted henceforth by $\boldsymbol{x}^+$. This can be found as $\boldsymbol{x}^+ = \boldsymbol{A}^+\boldsymbol{b}$ where $\boldsymbol{A}^+$ is the most general *pseudo-inverse* matrix of $\boldsymbol{A}$ and can be computed using its singular value decomposition (SVD), as explained next.

2 Matrix Factorizations

For greater generality, we shall assume complex matrices $\boldsymbol{A}$. The two most frequent differences from the real case are: (i) the transpose $\boldsymbol{M}^T$ of a real matrix $\boldsymbol{M}$ is replaced by the conjugate transpose $\boldsymbol{A}^H$, and (ii) real symmetric matrices $\boldsymbol{M} = \boldsymbol{M}^T$ are replaced by Hermitian matrices $\boldsymbol{A} = \boldsymbol{A}^H$.

2.1 Triangular factorizations

From Gauss elimination, assuming that row exchanges are not required, any square matrix $\boldsymbol{A}$ can be factored as $\boldsymbol{A} = \boldsymbol{LDU}$, where $\boldsymbol{L}$ and $\boldsymbol{U}$ are lower and upper triangular, respectively, with unit diagonals, and $\boldsymbol{D}$ is the diagonal matrix of pivots. If $\boldsymbol{A}$ is invertible, this factorization is unique. If $\boldsymbol{A}$ is Hermitian, then $\boldsymbol{U} = \boldsymbol{L}^H$ and we obtain the

$$\boldsymbol{A} = \boldsymbol{LDL}^H \tag{4}$$

If $\boldsymbol{A}$ is Hermitian and positive semidefinite, then $\boldsymbol{D}$ has nonnegative diagonal; hence, the triangular factorization becomes

$$\boldsymbol{A} = \boldsymbol{L}\sqrt{\boldsymbol{D}}\sqrt{\boldsymbol{D}}\boldsymbol{L}^H = \boldsymbol{L}\sqrt{\boldsymbol{D}}(\boldsymbol{L}\sqrt{\boldsymbol{D}})^H \tag{5}$$

which is a product of a lower triangular matrix with its conjugate transpose. Often, the above factorization, called *Cholesky decomposition*, is more compactly written as in (4).

2.2 Spectral decomposition

Any $n \times n$ matrix $\boldsymbol{A}$ that has n linearly independent eigenvectors accepts an *eigenvalue decomposition*

$$\boldsymbol{A} = \boldsymbol{V\Lambda V}^{-1} \tag{6}$$

where $\boldsymbol{V}$ contains as columns the eigenvectors $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ and $\boldsymbol{\Lambda}$ is a diagonal matrix that contains the eigenvalues $\lambda_1, ..., \lambda_n$. This factorization is also called *diagonalization* of $\boldsymbol{A}$.

Normal matrices $\boldsymbol{A}$, i.e. matrices with the property $\boldsymbol{AA}^H = \boldsymbol{A}^H\boldsymbol{A}$, are exactly those square matrices that possess a complete set of orthonormal eigenvectors and hence can be diagonalized by a unitary matrix $\boldsymbol{Q} = [\boldsymbol{q}_1, ..., \boldsymbol{q}_n]$. Special cases of normal matrices are the Hermitian matrices. Thus, the spectral theorem of linear algebra states that, any Hermitian matrix $\boldsymbol{A}$ accepts a *harmonic decomposition* as

$$\boldsymbol{A} = \boldsymbol{Q\Lambda Q}^{-1} = \boldsymbol{Q\Lambda Q}^H = \sum_{i=1}^{n} \lambda_i \boldsymbol{q}_i \boldsymbol{q}_i^H \tag{7}$$

where all the eigenvalues are real. If $\boldsymbol{A}$ is Hermitian and positive semidefinite, then the eigenvalues are nonnegative. The general harmonic decomposition obviously applies to any real symmetric matrix $\boldsymbol{A}$, with the simplification that $\boldsymbol{Q}$ is a real orthogonal matrix.

2.3 Factorization of Symmetric Positive-definite Matrices

A square matrix $\boldsymbol{A}$ is Hermitian and positive semidefinite iff it can be factored as

$$\boldsymbol{A} = \boldsymbol{R}^H \boldsymbol{R} \tag{8}$$

where $\boldsymbol{R}$ is any matrix. Further, $\boldsymbol{A}$ is positive definite iff $\boldsymbol{R}$ has independent columns. Three choices for $\boldsymbol{R}$ are:
(1) From the Cholesky decomposition of $\boldsymbol{A}$, we can choose $\boldsymbol{R}$ to be the upper triangular matrix $\sqrt{\boldsymbol{D}}\boldsymbol{L}^H$.
(2) A different choice results from the harmonic decomposition of $\boldsymbol{A}$, by setting $\boldsymbol{R} = \sqrt{\boldsymbol{\Lambda}}\boldsymbol{Q}^H$.
(3) Another factorization based on the harmonic decomposition is:

$$\boldsymbol{A} = \boldsymbol{R}^2, \quad \boldsymbol{R} = \boldsymbol{Q}\sqrt{\boldsymbol{\Lambda}}\boldsymbol{Q}^H \tag{9}$$

The above choice for $\boldsymbol{R}$ is called the Hermitian positive semidefinite *square root* of $\boldsymbol{A}$.

2.4 Singular Value Decomposition (SVD)

Any (real or complex) $m \times n$ matrix A can be factored as

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H = \sum_{i=1}^{r} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^H \tag{10}$$

where the $m \times m$ matrix $\boldsymbol{U}$ is unitary and its columns $\boldsymbol{u}_1, ..., \boldsymbol{u}_m$ are the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^H$, the $n \times n$ matrix $\boldsymbol{V}$ is unitary and its columns $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$ are the eigenvectors of $\boldsymbol{A}^H\boldsymbol{A}$, and the $m \times n$ matrix $\boldsymbol{S}$ is real diagonal whose only nonzero elements are its r diagonal terms $\sigma_1, \sigma_2, ..., \sigma_r > 0$, called *singular values*, with

$$r \triangleq \operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{A}\boldsymbol{A}^H) = \operatorname{rank}(\boldsymbol{A}^H\boldsymbol{A})$$

The singular values are the square roots of the common nonzero eigenvalues σ_i^2, $i = 1, ..., r$, of both $\boldsymbol{A}\boldsymbol{A}^H$ and $\boldsymbol{A}^H\boldsymbol{A}$.

Thus, the SVD of $\boldsymbol{A}$ is related to the spectral decomposition of the Hermitian $\boldsymbol{A}\boldsymbol{A}^H$ as follows:

$$\boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{U}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{U}^H = \sum_{i=1}^{r} \sigma_i^2 \boldsymbol{u}_i \boldsymbol{u}_i^H \tag{11}$$

Similarly for the other Hermitian product:

$$\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{V}^H = \sum_{i=1}^{r} \sigma_i^2 \boldsymbol{v}_i \boldsymbol{v}_i^H \tag{12}$$

If $\boldsymbol{A}$ is real, the only difference in its SVD (compared to the complex case) is that $\boldsymbol{U}$ and $\boldsymbol{V}$ are real orthogonal matrices. If $\boldsymbol{A}$ is Hermitian and positive semidefinite, its SVD is identical to its spectral decomposition $\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^H$. If $\boldsymbol{A}$ is indefinite, then any negative eigenvalue in $\boldsymbol{\Lambda}$ becomes positive in $\boldsymbol{S}$.

The columns of $\boldsymbol{U}$ and $\boldsymbol{V}$ provide orthonormal bases for all four fundamental subspaces of $\boldsymbol{A}$:

$$\begin{array}{rclcl}
\mathcal{R}(\boldsymbol{A}) & = & \text{column space} & = & \operatorname{span}(\{\boldsymbol{u}_1, ..., \boldsymbol{u}_r\}) \\
\mathcal{R}(\boldsymbol{A}^H) & = & \text{row space} & = & \operatorname{span}(\{\boldsymbol{v}_1, ..., \boldsymbol{v}_r\}) \\
\mathcal{N}(\boldsymbol{A}) & = & \text{null space} & = & \operatorname{span}(\{\boldsymbol{v}_{r+1}, ..., \boldsymbol{v}_n\}) \\
\mathcal{N}(\boldsymbol{A}^H) & = & \text{left null space} & = & \operatorname{span}(\{\boldsymbol{u}_{r+1}, ..., \boldsymbol{u}_m\})
\end{array}$$

These bases are interrelated since

$$\boldsymbol{AV} = \boldsymbol{US} \tag{13}$$

Hence, $\boldsymbol{Av}_i = \sigma_i \boldsymbol{u}_i$ for $i = 1, ..., n$. For $i > r$, we set $\sigma_i = 0$.

Applications of SVD:

(1) *Effective Rank*: Keep only the singular values above a threshold that determines the numerical precision.

(2) *Image/Signal Compact Representation*: Use only a few large singular values to approximately represent $\boldsymbol{A}$ using a truncated version of (10).

(3) *Polar Decomposition*: Factorize a (real or complex) square matrix $\boldsymbol{A}$ as $\boldsymbol{QC}$ where $\boldsymbol{Q}$ is unitary and $\boldsymbol{C}$ is Hermitian positive semidefinite. (If $\boldsymbol{A}$ is invertible, $\boldsymbol{C}$ is positive definite.)

$$\boldsymbol{A} = \boldsymbol{QC}, \quad \boldsymbol{Q} = \boldsymbol{UV}^H, \;\; \boldsymbol{C} = \boldsymbol{VSV}^H \tag{14}$$

This has applications in robotics where $\boldsymbol{Q}$ represents rotation or reflection, and $\boldsymbol{C}$ represents coordinate stretching or compression by the factors $\sigma_1, ...\sigma_r$. Actually, $\boldsymbol{C}$ is the Hermitian positive definite square root of $\boldsymbol{A}^H\boldsymbol{A}$.

(4) *Least Squares*: The **minimum length least squares solution** to the set of linear equations $\boldsymbol{Ax} = \boldsymbol{b}$ is the vector $\boldsymbol{x}^+$ such that

$$\boldsymbol{x}^+ \triangleq \arg\min_{\hat{\boldsymbol{x}}} \|\hat{\boldsymbol{x}}\|, \quad \hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \|\boldsymbol{Ax} - \boldsymbol{b}\| \tag{15}$$

Equivalently, the vector $\boldsymbol{x}^+$ is the minimum-norm solution of $\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{p}$ where $\boldsymbol{p}$ is the orthogonal projection of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$; see Fig. 1(b). Using the SVD of $\boldsymbol{A}$, let us define the general **pseudo-inverse** of $\boldsymbol{A}$ by the $n \times m$ matrix

$$\boldsymbol{A}^+ \triangleq \boldsymbol{VS}^+\boldsymbol{U}^H \tag{16}$$

where $\boldsymbol{S}^+$ is a $n \times m$ diagonal matrix with $1/\sigma_1, ..., 1/\sigma_r$ as its only nonzero diagonal terms. Note a few properties of the pseudo-inverse: $(\boldsymbol{A}^+)^+ = \boldsymbol{A}$. Further, if $r = n < m$, then $\boldsymbol{A}^+$ becomes equal to the Moore-Penrose pseudo-inverse $(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H$. Finally, if $r = m = n$, then $\boldsymbol{A}^+$ coincides with the standard matrix inverse $\boldsymbol{A}^{-1}$.

Now, based on the pseudo-inverse, the optimal solution of (15) can easily be found as

$$\boldsymbol{x}^+ = \boldsymbol{A}^+\boldsymbol{b} = \boldsymbol{VS}^+\boldsymbol{U}^H\boldsymbol{b} \tag{17}$$

Decomposing the action of $\boldsymbol{A}^+$ by looking at its three factors, the multiplication $\boldsymbol{U}^H\boldsymbol{b}$ creates m components of $\boldsymbol{b}$ in the orthonormal basis $(\boldsymbol{u}_1, ..., \boldsymbol{u}_m)$, from which the first r account for its projection $\boldsymbol{p}$ onto the column space whereas the last $m - r$ components account for its projection $\boldsymbol{b} - \boldsymbol{p}$ onto the left null space. Then, the multiplication with $\boldsymbol{S}^+$ zeros the components of $\boldsymbol{b} - \boldsymbol{p}$ and inverts the components of $\boldsymbol{p}$ along the r orthogonal directions. Finally the multiplication with $\boldsymbol{V}$ brings the resulting vector into the row space spanned by the orthonormal basis $(\boldsymbol{v}_1, ..., \boldsymbol{v}_r)$. The above total action of the pseudo-inverse can be summarized by

$$\begin{array}{lcl} \boldsymbol{Ax}^+ = \boldsymbol{p} & , & \boldsymbol{p} \perp \boldsymbol{b} - \boldsymbol{p} \\ \boldsymbol{A}^+\boldsymbol{p} = \boldsymbol{x}^+ & , & \boldsymbol{A}^+(\boldsymbol{b} - \boldsymbol{p}) = \boldsymbol{0} \end{array}$$

In general the SVD has excellent performance for numerical matrix computations.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins Univ. Press, 1989.

[2] G. Strang, *Linear Algebra and Its Applications*, Harcourt Brace Jovanovich, 1988.