Linear Analysis
Lecture 15

The Projection Theorem and Linear Least Squares

Let $A \in \mathbb{C}^{m\times n}$, $b \in \mathbb{C}^m$ and $\|\cdot\|$ be the Euclidean norm. Then a solution to the problem

$$\underset{x\in\mathbb{C}^n}{\text{minimize}} \;\; \|b - Ax\|^2 \qquad (*)$$

exists. Moreover, $\bar{x} \in \mathbb{C}^n$ solves (*) if and only if $\bar{x}$ is a solution to the normal questions $A^{\mathrm{H}} A\bar{x} = A^H b$.

Proof: Observe that

$$\underset{x\in\mathbb{C}^n}{\text{minimize}} \;\; \|b - Ax\|^2 = \underset{y\in\mathbb{R}(A)}{\text{minimize}} \;\; \|b - y\|^2 \; . \qquad (**)$$

By the Projection Theorem there exists a $\bar{y} \in \mathcal{R}(A)$ solving (**). Therefore, there exists $\bar{x}$ such that $\bar{y} = A\bar{x}$ and $\bar{x}$ must solve (*). Furthermore,

$$\bar{y} \text{ solves } (**) \iff b - \bar{y} \in \mathcal{R}(A)^{\perp} = \mathcal{N}(A^T),$$

or equivalently,

$$\bar{x} \text{ solves } (*) \iff b - A\bar{x} \in \mathcal{N}(A^T).$$

Finally,

$$\begin{aligned} b - A\bar{x} \in \mathcal{N}(A^T) \quad &\Leftrightarrow \quad A^T(b - A\bar{x}) = 0 \\ &\Leftrightarrow \quad A^T A\bar{x} = A^T b \; . \qquad \square \end{aligned}$$

Remarks.

(i) The minimizing element $\bar{y}$ in $(**)$ is unique. Moreover, if $\bar{x} \in \mathbb{C}^n$ solves $\bar{y} = A\bar{x}$, then $\bar{x}$ solves (*). But $\bar{x}$ may not be unique. Indeed, if $z \in \mathcal{N}(A)$, then $\hat{x} = \bar{x} + z$ must also solve (*) since $\bar{y} = A(\bar{x} + z) = A\hat{x}$.

(ii) If $\operatorname{rank}(A) = n$, then $\mathcal{N}(A) = \{0\}$ and so there is a unique $\bar{x} \in \mathbb{C}^n$ for which $A\bar{x} = \bar{y}$.
This $\bar{x}$ is the unique minimizer of $\|b - Ax\|^2$ over $x \in \mathbb{C}^n$ as well as the unique solution of the normal equations $A^{\mathrm{H}} Ax = A^H b$.

(iii) If $\operatorname{rank}(A) = r < n$, then the minimizing vector $\bar{x}$ is not unique; $\bar{x}$ can be modified by adding any element of $\mathcal{N}(A)$. For example, we might choose the solution of least norm.

Least Norm Solutions to Linear Least Squares Problems

Again consider the problem

$$\min_x \frac{1}{2} \|Ax - b\|_2^2 . \qquad (\mathcal{LLS})$$

We have shown that the set to solutions to $(\mathcal{LLS})$ is given by the set of solutions to the normal equations, i.e.,

$$\left\{x \in \mathbb{C}^n \,:\, A^H A = A^H b\right\} = \bar{x} + \mathcal{N}(A^H A) = \bar{x} + \mathcal{N}(A),$$

where $\bar{x}$ is any element of the set $\mathrm{argmin}_x \frac{1}{2} \|Ax - b\|_2^2$. Hence, this set is affine. By projecting the origin onto an affine set, we obtain the least-norm element of that set. So $\hat{x} \in \bar{x} + \mathcal{N}(A)$ satisfies

$$\|\hat{x}\| \le \|x\| \;\; \forall \; x \in \bar{x} + \mathcal{N}(A)$$

iff $\hat{x} \perp \mathcal{N}(A)$. In summary, $\hat{x}$ is the least norm solution to $(\mathcal{LLS})$ iff

$$(A\hat{x} - b) \perp \mathcal{R}(A) \;\text{ and }\; \hat{x} \perp \mathcal{N}(A)$$

$$\Longleftrightarrow$$

$$A^H A\hat{x} \;=\; A^H b \;\text{ and }\; \hat{x} \in \mathcal{R}(A^H) \;.$$

The Moore-Penrose Pseudo-Inverse

The mapping $A^\dagger : \mathbb{C}^m \to \mathbb{C}^n$ taking $b \in \mathbb{C}^m$ into the unique minimizer $\hat{x}$ of $\|b - Ax\|^2$ of minimum norm is called the Moore-Penrose pseudo-inverse of A.

We show that $A^\dagger$ is linear, so it is represented by an $n \times m$ matrix which we also denote by $A^\dagger$.

If $m = n$ and A is invertible, then every $b \in \mathbb{C}^n$ is in $\mathcal{R}(A)$, so $\bar{y} = b$, and the solution of $Ax = b$ is unique, given by $x = A^{-1}b$. In this case $A^\dagger = A^{-1}$. So the pseudo-inverse is a generalization of the inverse to possibly non-square, non-invertible matrices.

The pseudo-inverse of A can be expressed easily using the SVD.

The Moore-Penrose Pseudo-Inverse and the SVD

Let $A = U\Sigma V^H$ be an SVD of A, let $r = \operatorname{rank}(A)$

$$(\text{so} \quad \sigma_1 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots),$$

let U_r and V_r in $\mathbb{C}^{m\times r}$ be the first r columns of U, V, respectively. Let $\widetilde{U} \in \mathbb{C}^{m\times(m-r)}$, $\widetilde{V} = \mathbb{C}^{n\times(n-r)}$ be the remaining columns of U, V, respectively, and let $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$. Then the reduced SVD for A is

$$A = [U_r \; \widetilde{U}] \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^H \\ \widetilde{V}^H \end{bmatrix} = U_r \Sigma_r V_r^H .$$

Note that

$$\mathcal{R}(U_r) = \mathcal{R}(A), \quad \mathcal{R}(\widetilde{U}) = \mathcal{R}(A)^\perp,$$
$$\mathcal{R}(\widetilde{V}) = \mathcal{N}(A), \quad \text{and} \quad \mathcal{R}(V_r) = \mathcal{N}(A)^\perp .$$

Therefore,

$$\begin{aligned} U_r U_r^H &= \text{the orthogonal projector onto } \mathcal{R}(A) \\ \widetilde{U}\widetilde{U}^H &= \text{the orthogonal projector onto } \mathcal{R}(A)^\perp \\ \widetilde{V}\widetilde{V}^H &= \text{the orthogonal projector onto } \mathcal{N}(A) \\ V_r V_r^H &= \text{the orthogonal projector onto } \mathcal{N}(A)^\perp \end{aligned}$$

Linear Least-Squares and the SVD

$$
\begin{aligned}
\|b - Ax\|^2 &= \|b - U\Sigma V^H x\|^2 \\
&= \|U^H b - \Sigma V^H x\|^2 \\
&= \left\| \begin{bmatrix} U_r^H \\ \widetilde{U}^H \end{bmatrix} b - \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^H \\ \widetilde{V}^H \end{bmatrix} x \right\|^2 \\
&= \left\| \begin{bmatrix} U_r^H b - \Sigma_r V_r^H x \\ \widetilde{U}^H b \end{bmatrix} \right\|^2 ,
\end{aligned}
$$

so

$$
\|b - Ax\|^2 = \|U_r^H b - \Sigma_r V_r^H x\|^2 + \|\widetilde{U}^H b\|^2 .
$$

Thus

$$
\begin{aligned}
&\left[\bar{x} \text{ solves } \underset{x \in \mathbb{C}^n}{\text{minimize}} \ \|b - Ax\|^2 \right] \\
\Leftrightarrow \ & \Sigma_r V_r^H \bar{x} = U_r^H b \ \Leftrightarrow \ V_r^H \bar{x} = \Sigma_r^{-1} U_r^H b \ .
\end{aligned}
$$

Linear Least-Squares and the SVD

In addition, $x = \hat{x}$ is the unique minimizer of $\|b - Ax\|^2$ of minimum norm if and only if $x \in \mathcal{N}(A)^\perp = \mathcal{R}(V_r)$, i.e., $\widetilde{V}^H x = 0$.
So $x = \hat{x}$ if and only if

$$V^H x = \begin{bmatrix} V_r^H x \\ \widetilde{V}^H x \end{bmatrix} = \begin{bmatrix} \Sigma_r^{-1} U_r^H b \\ 0 \end{bmatrix}$$

$$\Longleftrightarrow$$

$$x = V \begin{bmatrix} \Sigma_r^{-1} U_r^H b \\ 0 \end{bmatrix}$$

$$\Longleftrightarrow$$

$$x = [\, V_r \;\; \widetilde{V}] \begin{bmatrix} \Sigma_r^{-1} U_r^H b \\ 0 \end{bmatrix} = V_r \Sigma_r^{-1} U_r^H b \ .$$

So $\hat{x} = V_r \Sigma_r^{-1} U_r^H b$. Thus, $\hat{x}$ is a linear function of b, so $A^\dagger$ is linear, with

$$A^\dagger = V_r \Sigma_r^{-1} U_r^H = [\, V_r \;\; \widetilde{V}] \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_r^H \\ \widetilde{U}^H \end{bmatrix} = V \Sigma^\dagger U^H,$$

where $\Sigma^\dagger = \operatorname{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0) \in \mathbb{C}^{n \times m}$. It is appropriate to call this matrix $\Sigma^\dagger$ as it is the pseudo-inverse of $\Sigma \in \mathbb{C}^{m \times n}$.

Non-Spectral Factorizations

The matrix factorizations we have studied so far are *spectral factorizations* in the sense that eigenvalues and eigenvectors are required for these factorizations.

We now discuss two non-spectral matrix factorizations. These factorizations can be determined directly from the entries of the matrix, and are computationally less expensive than spectral factorizations.

Each of these factorizations amounts to a reformulation of a procedure you are already familiar with.

The *LU factorization* is a reformulation of Gaussian Elimination.

The *QR factorization* is a reformulation of Gram-Schmidt orthogonalization.

The LU Factorization and Gaussian Elimination

Recall the method of Gaussian Elimination for solving a system

$$Ax = b$$

of linear equations, where $b \in \mathbb{C}^m$ and either

$$b \in \mathcal{R}(A)$$

or $A \in \mathbb{C}^{m \times n}$ has full row rank,

$$\mathsf{rank}(A) = m \leq n.$$

If the coefficient of x_1 in the first equation is nonzero, one eliminates all occurrences of x_1 from all the other equations by adding appropriate multiples of the first equation.

This operation does not change the set of solutions to the equation.

Now if the coefficient of x_2 in the new second equation is nonzero, it can be used to eliminate x_2 from the further equations, etc...

The LU Factorization

In matrix terms, suppose

$$A = \begin{bmatrix} a_1 & v_1^T \\ u_1 & \widetilde{A}_1 \end{bmatrix} \in \mathbb{C}^{n\times m},$$

with $0 \neq a_1 \in \mathbb{C}$, $u_1 \in \mathbb{C}^{m-1}$, $v_1 \in \mathbb{C}^{n-1}$, and $\widetilde{A}_1 \in \mathbb{C}^{(m-1)\times(n-1)}$.
Using the first row to zero out u_1 amounts to left multiplication of the matrix A by the matrix

$$\begin{bmatrix} 1 & 0 \\ -\frac{u_1}{a_1} & I \end{bmatrix}$$

to get

$$\begin{bmatrix} 1 & 0 \\ -\frac{u_1}{a_1} & I \end{bmatrix} \begin{bmatrix} a_1 & v_1^T \\ u_1 & \widetilde{A}_1 \end{bmatrix} \in \mathbb{C}^{n\times m} = \begin{bmatrix} a_1 & v_1^T \\ 0 & A_1 \end{bmatrix}, \qquad (*)$$

where $A_1 = \widetilde{A}_1 - u_1 v_1^T / a_1$. Define

$$L_1 = \begin{bmatrix} 1 & 0 \\ \frac{u_1}{a_1} & I \end{bmatrix} \in \mathbb{C}^{m\times m} \quad \text{and} \quad U_1 = \begin{bmatrix} a_1 & v_1^T \\ 0 & A_1 \end{bmatrix} \in \mathbb{C}^{m\times n} .$$

The LU Factorization

Observe that

$$L_1^{-1} = \begin{bmatrix} 1 & 0 \\ -\frac{u_1}{a_1} & I \end{bmatrix} .$$

Hence (*) becomes

$$L_1^{-1} A = U_1, \text{ or equivalently, } \quad A = L_1 U_1 \ .$$

Note that L_1 is *unit* lower triangular (ones on the mail diagonal) and U_1 is block upper-triangular with one 1×1 block and one $(m-1) \times (n-1)$ block on the block diagonal. The elements of $u/a \in \mathbb{C}^{m-1}$ are called **multipliers**, they are the multiples of the first row subtracted from subsequent rows. The multipliers are usually denoted

$$u/a = \left[\mu_{21}, \ \mu_{31}, \ \ldots, \ \mu_{m1}\right]^T .$$

The LU Factorization

By (*), we have $L_1^{-1}A = U_1 = \begin{bmatrix} a_1 & v_1^T \\ 0 & A_1 \end{bmatrix}$. If the $(1,1)$ entry of A_1 is not 0, we can apply the same procedure to A_1: if

$$A_1 = \begin{bmatrix} a_2 & v_2^T \\ u_2 & \widetilde{A}_2 \end{bmatrix} \in \mathbb{C}^{(m-1)\times(n-1)}$$

with $a_2 \neq 0$, letting

$$\widetilde{L}_2 = \begin{bmatrix} I & 0 \\ \frac{u_2}{a_2} & I \end{bmatrix} \in \mathbb{C}^{(m-1)\times(m-1)},$$

and forming

$$\widetilde{L}_2^{-1}A_1 = \begin{bmatrix} 1 & 0 \\ -\frac{u_1}{a_2} & I \end{bmatrix} \begin{bmatrix} a_2 & v_2^T \\ u_2 & \widetilde{A}_1 \end{bmatrix} = \begin{bmatrix} a_2 & v_2^T \\ 0 & A_2 \end{bmatrix} \equiv \widetilde{U}_2 \in \mathbb{C}^{(m-1)\times(n-1)},$$

where $A_2 \in \mathbb{C}^{(m-2)\times(n-2)}$, amounts to using the second row to zero out elements of the second column below the diagonal.

The LU Factorization

Setting

$$L_2 = \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{L}_2 \end{bmatrix} \quad \text{and} \quad U_2 = \begin{bmatrix} a & v^T \\ 0 & \widetilde{U}_2 \end{bmatrix},$$

we have

$$L_2^{-1} L_1^{-1} A = \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{L}_2^{-1} \end{bmatrix} \begin{bmatrix} a & v^T \\ 0 & A_1 \end{bmatrix} = U_2,$$

or equivalently, $A = L_2 L_1 U_2$. Here U_2 is block upper triangular with two 1×1 blocks and one $(m-2) \times (n-2)$ block on the diagonal, and again L_2 is unit lower triangular. Notice that the multipliers appear in L_2 in the **second** column, below the diagonal.

We can continue in this fashion at most $\tilde{m} - 1$ times, where $\tilde{m} = \min\{m, n\}$.

The LU Factorization

If we *can* proceed $\tilde{m}-1$ times, then

$$L_{\tilde{m}-1}^{-1} \cdots L_2^{-1} L_1^{-1} A = U_{\tilde{m}-1} = U$$

is upper triangular provided that along the way that the $(1,1)$ entries of

$$A,\ A_1,\ A_2,\ \ldots,\ A_{\tilde{m}-2}$$

are nonzero so the process can continue.

Define

$$L = (L_{\tilde{m}-1}^{-1} \cdots L_1^{-1})^{-1} = L_1 L_2 \cdots L_{\tilde{m}-1}.$$

The matrix L is square unit lower triangular, and so is invertible. Moreover, $A = LU$, where the matrix U is the so called **row echelon form** of A. In general, a matrix $T \in \mathbb{C}^{m\times n}$ is said to be in row echelon form if for each $i = 1, \ldots, m-1$ the first non-zero entry in the $(i+1)^{\text{st}}$ row lies to the right of the first non-zero row in the i^{th} row.

The LU Factorization

Let us now suppose that $m = n$ and $A \in \mathbb{C}^{n \times n}$ is invertible. Writing $A = LU$ as a product of a unit lower triangular matrix $L \in \mathbb{C}^{n \times n}$ (necessarily invertible) and an upper triangular matrix $U \in \mathbb{C}^{n \times n}$ (also nessecarily invertible in this case) is called the **LU factorization** of A.

Remarks:

(1) If $A \in \mathbb{C}^{n \times n}$ is invertible and has an LU factorization, it is unique.

(2) One can show that $A \in \mathbb{C}^{n \times n}$ has an LU factorization iff for $1 \leq j \leq n$, the upper left $j \times j$ principal submatrix

$$\begin{bmatrix} a_{11} & \cdots & a_{ij} \\ \vdots & & \\ a_{j1} & \cdots & a_{jj} \end{bmatrix}$$

is invertible.

The LU Factorization

(3) Not every invertible $A \in \mathbb{C}^{n \times n}$ has an LU-factorization.

Example: $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Typically, one must permute the rows of A to move nonzero entries to the appropriate spot for the elimination to proceed.
Recall that a permutation matrix $P \in \mathbb{C}^{n \times n}$ is the identity I with its rows (or columns) permuted: so

$$P \in \mathbb{R}^{n \times n} \text{ is orthogonal, and } P^{-1} = P^T.$$

Permuting the rows of A amounts to left multiplication by a permutation matrix P^T; then $P^T A$ has an LU factorization, so $A = PLU$ (called the PLU factorization of A).

(4) Fact: Every invertible $A \in \mathbb{C}^{n \times n}$ has a (not necessarily unique) PLU factorization.

The LU Factorization

(5) It turns out that

$$L = L_1 \cdots L_{n-1} = \begin{bmatrix} 1 & \ddots & \\ \mu_{21} & \ddots & \\ \vdots & & \\ \mu_{n-1} & \cdots & 1 \end{bmatrix}$$

has the multipliers m_{ij} below the diagonal.

(6) The LU factorization can be used to solve linear systems $Ax = b$ (where $A = LU \in \mathbb{C}^{n \times n}$ is invertible).

The system $Ly = b$ can be solved by forward substitution (1^{st} equation gives x_1, etc.),

and $Ux = y$ can be solved by back-substitution (n^{th} equation gives x_n, etc.),
giving the solution to $Ax = LUx = b$.