Math 504: Advanced Linear Algebra

Hugo Woerdeman, with edits by Darij Grinberg*

October 27, 2021 (unfinished!)

Contents

Math 504 Lecture 16

1. Hermitian matrices (cont'd)

Recall: A **Hermitian matrix** is an $n \times n$-matrix $A \in \mathbb{C}^{n \times n}$ such that $A^* = A$.

A Hermitian matrix A is **positive semidefinite** if it satisfies

$$\langle Ax, x \rangle \geq 0 \qquad \text{for all } x \in \mathbb{C}^n.$$

A Hermitian matrix A is **positive definite** if it satisfies

$$\langle Ax, x \rangle > 0 \qquad \text{for all nonzero } x \in \mathbb{C}^n.$$

1.1. The Cholesky decomposition

Theorem 1.1.1 (Cholesky decomposition for positive definite matrices)**.** Let $A \in \mathbb{C}^{n \times n}$ be a positive definite Hermitian matrix. Then, A has a unique factorization of the form

$$A = LL^*,$$

where $L \in \mathbb{C}^{n \times n}$ is a lower-triangular matrix whose diagonal entries are positive reals.

*Drexel University, Korman Center, 15 S 33rd Street, Philadelphia PA, 19104, USA

Example 1.1.2. For $n = 1$, the theorem is trivial: In this case, $A = \begin{pmatrix} a \end{pmatrix}$ for some $a \in \mathbb{R}$, and this a is > 0 because A is positive definite. Thus, setting $L = \begin{pmatrix} \sqrt{a} \end{pmatrix}$, then $A = LL^*$. Moreover, this is clearly the only choice.

Example 1.1.3. Let us manually check the theorem for $n = 2$. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a positive definite Hermitian matrix. We are looking for a lower-triangular matrix $L = \begin{pmatrix} \lambda & 0 \\ x & \delta \end{pmatrix}$ whose diagonal entries λ and δ are positive reals that satisfies $A = LL^*$.

So we need

$$\begin{aligned}
\begin{pmatrix} a & b \\ c & d \end{pmatrix} = A = LL^* &= \begin{pmatrix} \lambda & 0 \\ x & \delta \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ x & \delta \end{pmatrix}^* \\
&= \begin{pmatrix} \lambda & 0 \\ x & \delta \end{pmatrix} \begin{pmatrix} \lambda & \overline{x} \\ 0 & \delta \end{pmatrix} \qquad \text{(since } \lambda, \delta \text{ are real)} \\
&= \begin{pmatrix} \lambda^2 & \lambda \overline{x} \\ \lambda x & x\overline{x} + \delta^2 \end{pmatrix} = \begin{pmatrix} \lambda^2 & \lambda \overline{x} \\ \lambda x & |x|^2 + \delta^2 \end{pmatrix}.
\end{aligned}$$

So we need to solve the system of equations

$$\begin{cases} a = \lambda^2; \\ b = \lambda \overline{x}; \\ c = \lambda x; \\ d = |x|^2 + \delta^2. \end{cases}$$

First, we solve the equation $a = \lambda^2$ by setting $\lambda = \sqrt{a}$. Since A is positive definite, we have $a = \langle Ae_1, e_1 \rangle > 0$, so that $\sqrt{a}$ is well-defined, and we get a positive real λ. Next, we solve the equation $c = \lambda x$ by setting $x = \dfrac{c}{\lambda}$. Next, the equation $b = \lambda \overline{x}$ is automatically satisfied, since the Hermitianness of A entails $b = \overline{c} = \overline{\lambda x} = \lambda \overline{x}$ (since λ is real). Finally, we solve the equation $d = |x|^2 + \delta^2$ by setting $\delta = \sqrt{d - |x|^2}$. Here, we need to convince ourselves that $d - |x|^2$ is a positive real, i.e., that $d > |x|^2$. Why is this the case?

I claim that this follows from applying $\langle Az, z \rangle \geq 0$ to the vector $z = \begin{pmatrix} b \\ -a \end{pmatrix}$. Indeed, setting $z = \begin{pmatrix} b \\ -a \end{pmatrix}$, we obtain

$$Az = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} b \\ -a \end{pmatrix} = \begin{pmatrix} 0 \\ bc - ad \end{pmatrix},$$

so that

$$\langle Az, z \rangle = \left\langle \begin{pmatrix} 0 \\ bc - ad \end{pmatrix}, \begin{pmatrix} b \\ -a \end{pmatrix} \right\rangle = (bc - ad)\,\overline{-a} = a\,(ad - bc)$$

and thus $a(ad-bc) = \langle Az, z\rangle > 0$ (by the positive definiteness of A, since $z \neq 0$). We can divide this inequality by a (since $a > 0$), and obtain $ad - bc > 0$. Now, recall that $x = \dfrac{c}{\lambda}$ and $\lambda = \sqrt{a}$. Hence,

$$\begin{aligned} d - |x|^2 &= d - \left|\frac{c}{\lambda}\right|^2 = d - \frac{c\overline{c}}{\lambda^2} = d - \frac{cb}{a} \qquad \left(\text{since } \overline{c} = b \text{ and } \lambda^2 = a\right) \\ &= \frac{ad-bc}{a} > 0 \qquad (\text{since } ad - bc > 0 \text{ and } a > 0). \end{aligned}$$

This is what we need. So the theorem is proved for $n = 2$.

To prove the theorem in general, we need a lemma that essentially generalizes our above argument for $d - |x|^2 > 0$:

Lemma 1.1.4. Let $Q \in \mathbb{C}^{n\times n}$ be a invertible matrix. Let $x \in \mathbb{C}^n$ be some column vector. Let $d \in \mathbb{R}$. Let

$$A := \begin{pmatrix} QQ^* & Qx \\ (Qx)^* & d \end{pmatrix} \in \mathbb{C}^{(n+1)\times(n+1)}.$$

Assume that A is positive definite. Then, $||x||^2 < d$.

Proof. Set $Q^{-*} := \left(Q^{-1}\right)^* = \left(Q^*\right)^{-1}$. (This is well-defined, since Q is invertible.) Set $u = \begin{pmatrix} Q^{-*}x \\ -1 \end{pmatrix} \in \mathbb{C}^{n+1}$. (This is in block-matrix notation. Explicitly, this is the column vector obtained by appending the extra entry -1 at the bottom of $Q^{-*}x$.)

Then,

$$\begin{aligned} Au &= \begin{pmatrix} QQ^* & Qx \\ (Qx)^* & d \end{pmatrix} \begin{pmatrix} Q^{-*}x \\ -1 \end{pmatrix} = \begin{pmatrix} QQ^*Q^{-*}x + Qx(-1) \\ (Qx)^* Q^{-*}x + d(-1) \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ x^*Q^*Q^{-*}x - d \end{pmatrix} = \begin{pmatrix} 0 \\ x^*x - d \end{pmatrix} = \begin{pmatrix} 0 \\ ||x||^2 - d \end{pmatrix} \end{aligned}$$

(since $x^*x = \langle x, x\rangle = ||x||^2$). Hence,

$$\langle Au, u\rangle = \left\langle \begin{pmatrix} 0 \\ ||x||^2 - d \end{pmatrix}, \begin{pmatrix} Q^{-*}x \\ -1 \end{pmatrix} \right\rangle = \left(||x||^2 - d\right)\left(\overline{-1}\right) = d - ||x||^2.$$

However, the vector u is nonzero (since its last entry is -1), and the matrix A is positive definite (by assumption). Thus, $\langle Au, u\rangle > 0$. Since $\langle Au, u\rangle = d - ||x||^2$, we thus obtain $d - ||x||^2 > 0$. In other words, $d > ||x||^2$. This proves the lemma. □

Now, let us prove the Cholesky factorization theorem:

Proof of the Theorem. We proceed by induction on n.

The *base cases* $n = 0$ and $n = 1$ are essentially obvious ($n = 1$ was done in an example).

Induction step: Assume that the theorem holds for some n. We must prove that it holds for $n + 1$ as well.

Let $A \in \mathbb{C}^{(n+1)\times(n+1)}$ be a positive definite Hermitian matrix. Write A in the form

$$A = \begin{pmatrix} B & b \\ b^* & d \end{pmatrix},$$

where $B \in \mathbb{C}^{n\times n}$ and $b \in \mathbb{C}^n$. Note that the b^* on the bottom of the right hand side is because A is Hermitian, so all entries in the last row of A are the complex conjugates of the corresponding entries in the last column of A. Also, $d = d^*$ for the same reason, so $d \in \mathbb{R}$. Moreover, B is Hermitian (since A is Hermitian).

Next, we claim that B is positive definite. Indeed, for any nonzero vector $x \in \mathbb{C}^n$, we have $\langle Bx, x\rangle = \langle Ax', x'\rangle$, where $x' = \begin{pmatrix} x \\ 0 \end{pmatrix}$. Thus, positive definiteness of B follows from positive definiteness of A. (More generally, any principal submatrix of a positive definite matrix is positive definite.)

Therefore, by the induction hypothesis, we can apply the theorem to B instead of A. We conclude that B can be uniquely written as a product $B = QQ^*$, where $Q \in \mathbb{C}^{n\times n}$ is a lower-triangular matrix whose diagonal entries are positive reals.

Now, we want to find a vector $x \in \mathbb{C}^n$ and a positive real δ such that if we set

$$L := \begin{pmatrix} Q & 0 \\ x^* & \delta \end{pmatrix},$$

then $A = LL^*$. If we can find such x and δ, then at least the existence part of the theorem will be settled.

So let us set $L := \begin{pmatrix} Q & 0 \\ x^* & \delta \end{pmatrix}$, and see what conditions $A = LL^*$ places on x and δ. We want

$$\begin{aligned}\begin{pmatrix} B & b \\ b^* & d \end{pmatrix} = A = LL^* &= \begin{pmatrix} Q & 0 \\ x^* & \delta \end{pmatrix} \begin{pmatrix} Q^* & (x^*)^* \\ 0 & \overline{\delta} \end{pmatrix} \\ &= \begin{pmatrix} Q & 0 \\ x^* & \delta \end{pmatrix} \begin{pmatrix} Q^* & x \\ 0 & \delta \end{pmatrix} \qquad \left(\text{since } \delta \in \mathbb{R} \Longrightarrow \overline{\delta} = \delta\right) \\ &= \begin{pmatrix} QQ^* & Qx \\ x^*Q^* & x^*x + \delta^2 \end{pmatrix}.\end{aligned}$$

In other words, we want

$$\begin{cases} B = QQ^*; \\ b = Qx; \\ b^* = x^*Q^*; \\ d = x^*x + \delta^2. \end{cases}$$

The first of these four equations is already satisfied (we know that $B = QQ^*$). The second equation will be satisfied if we set $x = Q^{-1}b$. We can indeed set $x = Q^{-1}b$, since the matrix Q is invertible (since Q is a lower-triangular matrix with positive reals on its diagonal). The third equation follows automatically from the second ($b = Qx \Longrightarrow b^* = (Qx)^* = x^*Q^*$). Finally, the fourth equation rewrites as $d = ||x||^2 + \delta^2$. We can satisfy it by setting $\delta = \sqrt{d - ||x||^2}$, as long as we can show that $d - ||x||^2 > 0$. Fortunately, we can indeed show this, because our Lemma yields that $||x||^2 < d$. Thus, we have found x and δ, and constructed a lower-triangular matrix L whose diagonal entries are positive reals and which satisfies $A = LL^*$.

It remains to show that this L is unique. Indeed, we can basically read our argument above backwards. **If** $L \in \mathbb{C}^{(n+1)\times(n+1)}$ is a lower-triangular matrix whose diagonal entries are positive reals and which satisfies $A = LL^*$, then we can write A in the form $A = \begin{pmatrix} Q & 0 \\ x^* & \delta \end{pmatrix}$ for some $Q \in \mathbb{C}^{n\times n}$ and $x \in \mathbb{C}^n$ and some positive real δ, where Q is lower-triangular with diagonal entries being real. The equation $A = LL^*$ rewrites as

$$\begin{pmatrix} B & b \\ b^* & d \end{pmatrix} = \begin{pmatrix} QQ^* & Qx \\ x^*Q^* & x^*x + \delta^2 \end{pmatrix}.$$

Thus, $B = QQ^*$. By the induction hypothesis, the Q is unique, so this Q is exactly the Q that was constructed above. Moreover, $b = Qx$, so that $x = Q^{-1}b$, so again our new x is our old x. Finally, $d = x^*x + \delta^2$ entails $\delta = \sqrt{d - ||x||^2}$, because δ has to be positive. So our δ is our old δ. Thus, our L is the L that we constructed above. This proves the uniqueness of the L. The theorem is proved. □

Exercise 1.1.1. Let $A \in \mathbb{C}^{n\times n}$ be a positive definite Hermitian matrix. Then, $\det A$ is a positive real.

Exercise 1.1.2. Let A and B be two positive definite Hermitian matrices in $\mathbb{C}^{n\times n}$. Then, $\operatorname{Tr}(AB) \geq 0$.

There is a version of Cholesky decomposition for positive semidefinite matrices, but this will be left to the exercises.

1.2. Rayleigh quotients

Definition 1.2.1. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix, and $x \in \mathbb{C}^n$ be a nonzero vector. Then, the **Rayleigh quotient** for A and x is defined to be the real number

$$R(A, x) := \frac{\langle Ax, x\rangle}{\langle x, x\rangle} = \frac{x^*Ax}{x^*x} = \frac{x^*Ax}{||x||^2} = y^*Ay,$$

where $y = \dfrac{x}{||x||}$.

Let us explore what Rayleigh quotients can tell us about the eigenvalues of a Hermitian matrix.

Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix with $n > 0$. By the spectral theorem, we have $A = UDU^*$ for some unitary U and some real diagonal matrix D. Consider these U and D. We have $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A. We WLOG that

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

(indeed, we can always achieve this by permuting rows/columns of D and integrating the permutation matrices into U). We set

$$\lambda_{\min}(A) := \lambda_1 \qquad \text{and} \qquad \lambda_{\max}(A) := \lambda_n.$$

We will also write $\lambda_{\min}$ and $\lambda_{\max}$ without the "(A)" part.

Let us now pick some vector $x \in \mathbb{C}^n$ of length 1 (that is, $||x|| = 1$). Set $z = U^*x$. Then, writing z as $\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$, we have

$$x^*Ax = \underbrace{x^*U}_{=(U^*x)^*=z^*} D \underbrace{U^*x}_{=z} = z^*Dz = \sum_{k=1}^{n} \lambda_k \overline{z_k} z_k = \sum_{k=1}^{n} \lambda_k |z_k|^2 .$$

We note that $||z|| = 1$ (indeed, since U is unitary, the matrix U^* is also unitary, so $||U^*x|| = ||x|| = 1$, which means $||z|| = 1$). In other words, $\sum\limits_{k=1}^{n} |z_k|^2 = 1$ (since $||z|| = \sqrt{\sum\limits_{k=1}^{n} |z_k|^2}$). Now,

$$x^*Ax = \sum_{k=1}^{n} \underbrace{\lambda_k}_{\leq \lambda_n} |z_k|^2 \leq \sum_{k=1}^{n} \lambda_n |z_k|^2 = \lambda_n \underbrace{\sum_{k=1}^{n} |z_k|^2}_{=1} = \lambda_n.$$

So we have shown that each vector $x \in \mathbb{C}^n$ of length 1 satisfies $x^*Ax \leq \lambda_n$. This inequality becomes an equality at least for one vector x: namely, for the vector $x = Ue_n$ (because for this vector, we have $z = \underbrace{U^*U}_{=I_n} e_n = e_n$, so that $z_k = 0$ for all $k < n$, and therefore the inequality $\sum\limits_{k=1}^{n} \lambda_k |z_k|^2 \leq \sum\limits_{k=1}^{n} \lambda_n |z_k|^2$ becomes an equality). Thus,

$$\begin{aligned}
\lambda_n &= \max\left\{x^*Ax \ \mid\ x \in \mathbb{C}^n \text{ is a vector of length } 1\right\} \\
&= \max\left\{\frac{x^*Ax}{x^*x} \ \mid\ x \in \mathbb{C}^n \text{ is nonzero}\right\} \\
&= \max\left\{R(A, x) \ \mid\ x \in \mathbb{C}^n \text{ is nonzero}\right\}.
\end{aligned}$$

Since $\lambda_n = \lambda_{\max}(A)$, we thus have proved the following fact:

Proposition 1.2.2. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix with $n > 0$. Then, the largest eigenvalue of A is

$$\begin{aligned}\lambda_{\max}(A) &= \max\{x^*Ax \mid x \in \mathbb{C}^n \text{ is a vector of length } 1\}\\ &= \max\{R(A,x) \mid x \in \mathbb{C}^n \text{ is nonzero}\}.\end{aligned}$$

Similarly:

Proposition 1.2.3. Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix with $n > 0$. Then, the smallest eigenvalue of A is

$$\begin{aligned}\lambda_{\max}(A) &= \min\{x^*Ax \mid x \in \mathbb{C}^n \text{ is a vector of length } 1\}\\ &= \min\{R(A,x) \mid x \in \mathbb{C}^n \text{ is nonzero}\}.\end{aligned}$$

What about the other eigenvalues? Can we characterize λ_2 (for example) in terms of Rayleigh quotients?

Theorem 1.2.4 (Courant–Fisher theorem)**.** Let $A \in \mathbb{C}^{n\times n}$ be a Hermitian matrix. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be the eigenvalues of A, with $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Then, for each $k \in [n]$, we have

$$\begin{aligned}\lambda_k &= \min\{\max\{R(A,x) \mid x \in S \text{ nonzero}\} \mid S \subseteq \mathbb{C}^n \text{ is a } k\text{-dimensional subspace}\}\\ &= \min_{\substack{S\subseteq\mathbb{C}^n \text{ is a subspace};\\ \dim S = k}} \ \max_{\substack{x\in S;\\ x\neq 0}} R(A,x)\end{aligned}$$

and

$$\begin{aligned}\lambda_k &= \max\{\min\{R(A,x) \mid x \in S \text{ nonzero}\} \mid S \subseteq \mathbb{C}^n \text{ is a } (n-k+1)\text{-dimensional subspace}\}\\ &= \max_{\substack{S\subseteq\mathbb{C}^n \text{ is a subspace};\\ \dim S = n-k+1}} \ \min_{\substack{x\in S;\\ x\neq 0}} R(A,x).\end{aligned}$$

We will prove this next time.