Math 416 Lecture Note: Week 15

Daesung Kim

1 Projections and least squares approximation

Lemma 1.1. *Let $A \in \mathcal{M}_{m\times n}(F)$, then*

$$\langle Ax, y\rangle = \langle x, A^*y\rangle$$

for all $x \in F^n$ and $y \in F^m$, where $\langle\cdot,\cdot\rangle$ denotes the standard inner product on F^m and F^n.

Remark 1.2. The standard inner product on F^n can be thought of as

$$\langle x, y\rangle = y^*x$$

where x, y are $(n \times 1)$ matrices (or column vectors).

Lemma 1.3. *Let $A \in \mathcal{M}_{m\times n}(F)$. Then, $\mathcal{N}(A) = \mathcal{N}(A^*A)$.*

Proof. Note that $\mathcal{N}(A) \leq \mathcal{N}(A^*A)$ is trivial. Suppose $x \in \mathcal{N}(A^*A)$, then $A^*Ax = 0$. Thus,

$$0 = \langle x, A^*Ax\rangle = \langle Ax, Ax\rangle = \|Ax\|^2$$

and so $Ax = 0$. Thus, $\mathcal{N}(A^*A) \leq \mathcal{N}(A)$. □

Lemma 1.4. *Let $A \in \mathcal{M}_{m\times n}(F)$. If A has rank r, then A^*A has the same rank.*

Proof. It follows from the Dimension theorem that

$$r = \operatorname{rank}(A) = \dim(\mathcal{R}(A)) = n - \dim(\mathcal{N}(A)) = n - \dim(\mathcal{N}(A^*A)) = \dim(\mathcal{R}(A^*A)) = \operatorname{rank}(A^*A).$$

□

Remark 1.5. If $m \geq n$ and A has rank n, then $A^*A \in \mathcal{M}_{n\times n}(F)$ is invertible.

Lemma 1.6. *Let V be an inner product space over F and $T : V \to V$ linear. Then,*

$$\mathcal{N}(T^*) = \mathcal{R}(T)^\perp.$$

Proof. Suppose $x \in \mathcal{N}(T^*)$. Then,

$$\langle x, T(y)\rangle = \langle T^*(x), y\rangle = 0$$

for all y. Thus, $x \in \mathcal{R}(T)^\perp$, i.e., $\mathcal{N}(T^*) \leq \mathcal{R}(T)^\perp$. Suppose $x \in \mathcal{R}(T)^\perp$, then

$$0 = \langle x, T(y)\rangle = \langle T^*(x), y\rangle$$

for all $y \in V$. In particular, if we choose $y = T^*(x)$, then $\|T^*(x)\| = 0$. Thus, $x \in \mathcal{N}(T^*)$. □

Theorem 1.7. *Let $A \in \mathcal{M}_{m\times n}(F)$ and $m \geq n$. Suppose $\operatorname{rank}(A) = n$ and $W = \mathcal{R}(L_A) = \operatorname{Col}(A)$. Then,*

$$\operatorname{proj}_W(y) = My$$

for all $y \in F^n$, where $M = A(A^*A)^{-1}A^*$.

Proof. Recall that $\mathrm{proj}_W(y)$ is the unique vector such that $y = \mathrm{proj}_W(y) + z$ where $z \in W^\perp$. Thus, it suffices to prove $y - My \in W^\perp$. Since $W = \mathcal{R}(L_A)$, it is enough to show that $y - My \in \mathcal{N}(A^*)$. Indeed,

$$A^*(y - My) = A^*y - A^*A(A^*A)^{-1}A^*y = A^*y - A^*y = 0.$$

$\square$

Example 1.8. Let W be a plain in $\mathbb{R}^3$ given by $x + y + z = 0$. We have seen that $T = L_P$ is the projection on W where

$$P = \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

Since W is a subspace generated by $\{(1, -1, 0), (0, 1, -1)\}$, we have $W = \mathcal{R}(L_A)$ where

$$A = \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix}.$$

One can see that

$$\begin{aligned} A(A^tA)^{-1}A^t &= \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \left(\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} 2 & 1 \\ -1 & 1 \\ -1 & -2 \end{pmatrix} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} = P \end{aligned}$$

Remark 1.9. We call a matrix $A \in \mathcal{M}_{n\times n}(F)$ idempotent if $A^2 = A$. Note that $M = A(A^*A)^{-1}A^*$ and $I - M$ are idempotent. Note that M and $I - M$ are always diagonalizable. Note also that L_M is the projection on $\mathcal{R}(A)$ and $L_{(I-M)}$ is the projection on $\mathcal{N}(A^*) = \mathcal{R}(A)^\perp$.

Least square approximation

Suppose that there is a data set $(y_1, t_1), \cdots, (y_m, t_m)$ where y_i represents the population of a region at time t_i. Our goal is to understand the relationship between y_i and t_i (or the trend of y_i). Specifically, we assume y and t have the relation

$$y = ct + d$$

and find best possible constants c and d. If the model is true, then for each t_i, the population should be $\overline{y}_i = ct_i + d$. What we want to do is to find c and d that minimize the difference between the actual output Y and the expected output $\overline{Y}$, $\|Y - \overline{Y}\|$, where

$$A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{pmatrix}, \qquad x = \begin{pmatrix} c \\ d \end{pmatrix}, \qquad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}, \qquad \overline{Y} = \begin{pmatrix} \overline{y}_1 \\ \overline{y}_2 \\ \vdots \\ \overline{y}_m \end{pmatrix} = Ax.$$

In general, let $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ and $Y \in \mathbb{R}^m$ be given. The question is to find $x_0 \in \mathbb{R}^n$ such that

$$\|Y - Ax_0\| \leq \|Y - Ax\|$$

for all $x \in \mathbb{R}^n$. Let $W = \mathcal{R}(A)$, then we have seen that $Ax_0 = \text{proj}_W(Y)$. Such a vector x_0 always exists because $\text{proj}_W(Y)$ exists and belongs to W, and every vector in W is of the form Ax for some $x \in \mathbb{R}^n$. Recall that Y can be uniquely written as $Y = \text{proj}_W(Y) + Z$ where $Z \in W^\perp$. Thus, $Y - Ax_0 \in W^\perp = \mathcal{R}(A)^\perp = \mathcal{N}(A^*)$ and so $A^*(Y - Ax_0) = 0$. Thus, the solution x_0 satisfies

$$A^*Ax_0 = A^*Y.$$

If A has full rank, then A^*A is invertible and so the solution is unique and

$$x_0 = (A^*A)^{-1}A^*Y.$$

2 Triangularization: Shur's theorem

Definition 2.1. Let V be a vector space over F, $T : V \to V$ linear, and W be a subspace of V. We say W is T-invariant if $T(W) \leq W$.

Remark 2.2. Suppose that W is T-invariant. We define $T_W : W \to W$ by $T_W(v) = T(v)$ for all $v \in W$. Note that T_W is well-defined.

Remark 2.3. Let V be an inner product space over F and W be a finite dimensional subspace of V. Then, we have $V = W \oplus W^\perp$. In particular, $\dim(V) = \dim(W) + \dim(W^\perp)$.

Theorem 2.4 (Schur's theorem)**.** *Let V be an inner product space over $\mathbb{C}$ and $T : V \to V$ linear. Then, there exists an orthonormal basis β for V such that $[T]_\beta$ is upper triangular.*

Proof. Use an induction on n. If $n = 1$, there is nothing to prove. Suppose that $n \geq 2$ and the theorem holds for $n - 1$.

Since every polynomial over $\mathbb{C}$ splits, the characteristic polynomial of T^* splits. In particular, there exist $\lambda \in \mathbb{C}$ and an unit vector $v \in V \setminus \{0\}$ such that $T^*(v) = \lambda v$. Let $W = \mathrm{Span}(\{v\})$. We claim that $W^\perp$ is T-invariant. For $y \in W^\perp$, we want to show that $T(y) \in W^\perp$. If $w \in W$, then $w = cv$ and so

$$\langle T(y), w\rangle = \langle y, cT^*(v)\rangle = \langle y, c\lambda v\rangle = \overline{c\lambda}\,\langle y, v\rangle = 0.$$

Thus, $T(W^\perp) \leq W^\perp$. Since $\dim(W^\perp) = n - 1$, the induction hypothesis provides an orthonormal basis β' for $W^\perp$ such that $[T_{W^\perp}]_{\beta'}$ is upper triangular. Let $\beta = \beta' \cup \{v\}$, then β is orthonormal. Furthermore,

$$[T]_\beta = \begin{pmatrix} [T_{W^\perp}]_{\beta'} & * \\ 0 & * \end{pmatrix}.$$

Thus, $[T]_\beta$ is upper triangular. ∎

3 Normal operators

Definition 3.1. Let V be an inner product space over F and $T : V \to V$ linear. We say that T is normal if $TT^* = T^*T$. A matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is normal if $AA^* = A^*A$.

Example 3.2. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be rotation by angle $\theta \in [0, 2\pi]$, then the matrix representation in terms of the standard basis is

$$A = [T]_\beta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

One can see that $AA^* = A^*A = I$.

Example 3.3. If $A = A^*$ or $A = -A^*$, then A is normal.

Remark 3.4. If β is an orthonormal basis, then T is normal if and only if $[T]_\beta$ is normal.

Theorem 3.5. *Let V be an inner product space over F and $T : V \to V$ normal.*

(i) $\|T(x)\| = \|T^*(x)\|$ *for all* $x \in V$.

(ii) $T - cI$ *is normal for all* $c \in F$.

(iii) *If x is an eigenvector for T, then it is also an eigenvector for T^*. Moreover, if $T(x) = \lambda x$, then $T^*(x) = \overline{\lambda}x$.*

(iv) *If λ_1 and λ_2 are distinct eigenvalues for T corresponding to x_1 and x_2 respectively, then $\langle x_1, x_2\rangle = 0$.*

Proof. (i) It follows that

$$\|T(x)\|^2 = \langle T(x), T(x)\rangle = \langle x, T^*T(x)\rangle = \langle x, TT^*(x)\rangle = \langle T^*(x), T^*(x)\rangle = \|T^*(x)\|^2.$$

(ii) We have

$$(T - cI)(T - cI)^* = (T - cI)(T^* - \overline{c}I) = TT^* - cT^* - \overline{c}T + |c|^2$$

and

$$(T - cI)^*(T - cI) = (T^* - \overline{c}I)(T - cI) = T^*T - cT^* - \overline{c}T + |c|^2.$$

(iii) Suppose $T(x) = \lambda x$, then $(T - \lambda I)(x) = 0$. Thus, by Part (i) and (ii),

$$0 = \|(T - \lambda I)(x)\| = \|(T - \lambda I)^*(x)\| = \|(T^* - \overline{\lambda}I)(x)\|$$

and so $T^*(x) = \overline{\lambda}x$.

(iv) It follows from Part (iii) that

$$\lambda_1 \langle x_1, x_2\rangle = \langle \lambda_1 x_1, x_2\rangle = \langle T(x_1), x_2\rangle = \langle x_1, T^*(x_2)\rangle = \left\langle x_1, \overline{\lambda_2}x_2\right\rangle = \lambda_2 \langle x_1, x_2\rangle.$$

Thus, we have $(\lambda_1 - \lambda_2)\langle x_1, x_2\rangle = 0$. Since $(\lambda_1 - \lambda_2) \neq 0$, we conclude that $\langle x_1, x_2\rangle = 0$.

□

Theorem 3.6. *Let V be an inner product space over $\mathbb{C}$ and $T : V \to V$ linear. Then, T is normal if and only if there exists an orthonormal basis β for V consisting of eigenvectors of T.*

Proof. Suppose that there exists an orthonormal basis β for V consisting of eigenvectors of T. Then, $[T]_\beta$ is diagonal and $[T^*]_\beta = ([T]_\beta)^*$ is also diagonal. Since diagonal matrices commute each other, we get

$$[TT^*]_\beta = [T]_\beta [T^*]_\beta = [T^*]_\beta [T]_\beta = [T^*T]_\beta.$$

Thus, T is normal.

Suppose that T is normal. By Schur's theorem, there exists an orthonormal basis $\beta = \{v_1, \cdots, v_n\}$ such that $[T]_\beta$ is upper triangular. Note that v_1 is an eigenvalue for T because $[T]_\beta$ is upper triangular. Suppose $2 \leq k \leq n$ and $v_1, \cdots, v_{k-1}$ are eigenvectors for T. We claim that v_k is also an eigenvector. Let λ_j be an eigenvalue for T corresponding to v_j, $1 \leq j \leq k-1$. Since $A = [T]_\beta$ is upper triangular,

$$T(v_k) = A_{1k}v_1 + A_{2k}v_2 + \cdots + A_{kk}v_k.$$

Since β is orthonormal,

$$A_{jk} = \langle T(v_k), v_j \rangle = \langle v_k, T^*(v_j) \rangle = \lambda_j \langle v_k, v_j \rangle = 0$$

for all $j = 1, 2, \cdots, k-1$. Thus, v_k is an eigenvector. $\square$

Definition 3.7. Let V be a vector space over F and $T : V \to V$ linear. A subspace $W \leq V$ is T-invariant if $T(W) \leq W$. We define the restriction $T_W : W \to W$ by $T_W(x) = T(x)$ for all $x \in W$.

Definition 3.8. A matrix $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is unitary if $A^*A = AA^* = I$. A matrix $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ is orthogonal if $A^tA = AA^t = I$.

Remark 3.9. Let $V = F^n$, β be an orthonormal basis, and γ the standard basis, then one can see that $Q = [I]_\beta^\gamma$ is unitary if $F = \mathbb{C}$, and orthogonal if $F = \mathbb{R}$. Thus, the theorem can be restated as follows: if $A \in \mathcal{M}_{n\times n}(\mathbb{C})$ is normal, then there exists a unitary matrix Q such that Q^*AQ is diagonal.

References

[FIS] Freidberg, Insel, and Spence, *Linear Algebra*, 4th edition, 2002.

[Bee] Beezer, *A First Course in Linear Algebra*, Version 3.5, 2015.

Department of Mathematics, University of Illinois at Urbana-Champaign
E-mail address: daesungk@illinois.edu