Preliminaries for Courant Minimax Principle

Let $\vec{c}_1, \vec{c}_2, \vec{c}_3, \dots, \vec{c}_{k-1} \in \mathbb{R}^n$. If we form the matrix

$$C = \begin{bmatrix} \text{---} & \vec{c}_1^{\,T} & \text{---} \\ \text{---} & \vec{c}_2 & \text{---} \\ & \vdots & \\ \text{---} & \vec{c}_{k-1} & \text{---} \end{bmatrix} \in \mathbb{R}^{(k-1)\times n}$$

Then $C\vec{x} = \vec{0} \Rightarrow$ the vector $\vec{x}$ is orthogonal to each row of C. That is $\langle \vec{c}_i, x \rangle = 0$ for $i = 1, 2, \dots k-1$.

- If we let $S = \text{span}\{\vec{c}_1, \vec{c}_2, \dots, \vec{c}_{k-1}\}$, then the vector $\vec{x}$ from above that satisfies $C\vec{x} = \vec{0}$ is contained in $S^\perp$. That is, $x \in S^\perp$.

- If $\{\vec{c}_1, \vec{c}_2, \dots \vec{c}_{k-1}\}$ forms a linearly independent set of vectors, then
$$\dim(S) = k-1 \text{ and } \dim(S^\perp) = n-(k-1) = n-k+1$$

- Requiring $C\vec{x} = \vec{0}$ is requiring $x \in S^\perp$, and $S^\perp$ is an $n-k+1$-dimensional hyperplane in $\mathbb{R}^n$!

- Thm 1.7 finds λ_k by first maximizing $g(x)$ with $\|x\| = 1$ on this $n-k+1$-dimensional hyperplane and then varying the hyperplane until the maximum is the smallest it can be.

Thm 1.7 Courant Minimax Principle

For any real symmetric matrix A,

$$\lambda_k = \min_{C} \left(\max_{\substack{\|x\|=1 \\ Cx=0}} \langle Ax, x\rangle \right)$$

where C is any $(k-1)\times n$ matrix.

Some Comments:

Ⓐ Thm 1.7 builds on the ideas of Thm. 1.6.
For example, if I already know $x_1, x_2, \ldots, x_{k-1}$
where (λ_i, x_i) are e-pairs of A for $i=1,2,\ldots k-1$
(with x_i so that $\|x_i\|=1$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$), then using the hyperplane defined by $Cx=0$, where

$$C = \begin{bmatrix} \text{---} & \vec{x}_1^T & \text{---} \\ \text{---} & \vec{x}_2^T & \text{---} \\ & \vdots & \\ \text{---} & \vec{x}_{k-1}^T & \text{---} \end{bmatrix} \qquad (*)$$

is equivalent to the inductive step of Thm 1.6 #2,
where we require $\langle x, x_j\rangle = 0$ for $j=1,2,\ldots,k-1$.
So, if C is chosen as given in $(*)$, then Thm 1.6 tells us that

$$\lambda_k = \max_{\substack{\|x\|=1 \\ Cx=0}} \langle Ax, x\rangle$$

Ⓑ Now, if we remove $\vec{x}_1^T$ from the first row of C and replace it with some other vector in $\mathbb{R}^n$, say $\vec{y}_1^T$. Then x_1 will be contained in the $n-k+1$-dim'nal hyperplane that we are using to $\langle Ax, x\rangle$. If I

connect it with the lingo on page 1, then
$S = \text{span}\{\vec{y}_1, \vec{x}_2, \ldots \vec{x}_{k-1}\}$ and $x_1 \in S^\perp$.
(Note, here we assume that $\vec{y}_1$ is not a scalar multiple of $\vec{x}_1$.)

So, if $x_1 \in S^\perp$, then we get $\langle Ax_1, x_1 \rangle = \lambda_1 \geq \lambda_{k-1}$ and the value of $\displaystyle\max_{\substack{\|x\|=1 \\ Cx=0}} \langle Ax, x\rangle$ will / may increase

↳ this is our new C

A similar situation arises when any of the others $x_2, x_3, \ldots, x_{k-1}$ are replaced by other vectors from $\mathbb{R}^n$. The value of $\langle Ax, x\rangle$ will have contributions from $\langle Ax_i, x_i\rangle$ where $i = 1, 2, \ldots, k-1$, and these values will potentially be larger than λ_k. Hence, we allow C to range over all possible choices of $(k-1)\times n$ matrices, and λ_k will occur when we take the minimum value of

$$\max_{\substack{\|x\|=1 \\ Cx=0}} \langle Ax, x\rangle$$

over all possible choices of C.

- We will not go through the proof of Thm 1.7 but press forward to a point where we can see how this thm helps us.

Some Preliminaries to Help us Understand Thm 1.8

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 6 \\ 3 & 6 & 8 \end{pmatrix}$$

Spse we impose the constraint that $x_1 = 0$, then for any $\vec{x} \in \mathbb{R}^3$, we can describe $x_1 = 0$ as

$$(1 \quad 0 \quad 0)\,\vec{x} = (1 \quad 0 \quad 0)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \vec{0}$$

Let $B = (1 \quad 0 \quad 0) \in \mathbb{R}^{1\times 3}$. Then if we consider the quadratic form $q(x) = \langle Ax, x\rangle$ subject to the constraint that $x_1 = 0$ ie. $B\vec{x} = \vec{0}$, then $q(x)$ reduces to

$$\left\langle \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 6 \\ 3 & 6 & 8 \end{pmatrix}\begin{pmatrix} 0 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} 0 \\ x_2 \\ x_3 \end{pmatrix} \right\rangle$$

$$= (0 \quad x_2 \quad x_3)\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 6 \\ 3 & 6 & 8 \end{pmatrix}\begin{pmatrix} 0 \\ x_2 \\ x_3 \end{pmatrix}$$

$$= (0 \quad x_2 \quad x_3)\begin{pmatrix} 2x_2 + 3x_3 \\ 3x_2 + 6x_3 \\ 6x_2 + 8x_3 \end{pmatrix}$$

$$= 3x_2^2 + 12x_2x_3 + 8x_3^2$$

$$= (x_2 \quad x_3)\begin{pmatrix} 3 & 6 \\ 6 & 8 \end{pmatrix}\begin{pmatrix} x_2 \\ x_3 \end{pmatrix} = \hat{q}(x)$$

where $x = \begin{pmatrix} x_2 \\ x_3 \end{pmatrix}$

$\hat{q} = \langle \hat{A}x, x\rangle$ with $\hat{A} = \begin{pmatrix} 3 & 6 \\ 6 & 8 \end{pmatrix}$

Similarly, if we impose $Bx=0$ where

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix},$$

then $Bx=0 \iff \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$$\begin{pmatrix} x_1 \\ x_2 - x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$x_1 = 0 \text{ and } x_2 = x_3$$

Then $q(x) = \langle Ax, x\rangle$ reduces to

$$\left\langle \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 6 \\ 3 & 6 & 8 \end{pmatrix}\begin{pmatrix} 0 \\ x_2 \\ x_2 \end{pmatrix}, \begin{pmatrix} 0 \\ x_2 \\ x_2 \end{pmatrix} \right\rangle = 3x_2^2 + 12x_2^2 + 8x_2^2 = 23x_2^2$$

So, $\hat{q}(x) = \langle \hat{A}\begin{pmatrix} x_2 \\ x_2 \end{pmatrix}, \begin{pmatrix} x_2 \\ x_2 \end{pmatrix}\rangle = 23x_2^2$

- BIG IDEA → If we impose k constraints described by $Bx=0$, then the original quadratic form $q(x)$ reduces to a quadratic form $\hat{q}(x)$ in $n-k$ variables!

[Note this assumes that each constraint enforced by $Bx=0$ is different, i.e. $B = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$ is not considered.]

Theorem 1.8

Spse the quadratic form $q(x) = \langle Ax, x\rangle$ is constrained by k linear constraints $Bx=0$ and reduced to the quadratic form $\hat{q}(x)$ in $n-k$ variables. The relative extremal values of q, denoted $\lambda_1, \lambda_2, \dots, \lambda_n$ and the relative extremal values of $\hat{q}(x)$, denoted $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_{n-k}$, satisfy the interlacing inequality

$$\lambda_j \geq \hat{\lambda}_j \geq \lambda_{j+k} \quad , \quad j = 1, 2, \dots n-k$$

(We will not cover the proof of this theorem. Instead, we will try to see how it can be used.)

Back to our previous example (on pg. 21-22).

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 6 \\ 3 & 6 & 8 \end{pmatrix} \qquad q(x) = \langle Ax, x\rangle = x^T A x$$

If we impose the constraint that $x_1 = 0$; i.e. 1 constraint described by $Bx=0$ with $B = (1 \ 0 \ 0)$, then we already saw that

$$\hat{q}(x) = \langle \hat{A}x, x\rangle = x^T \hat{A} x \quad \text{with } \vec{x} = \begin{pmatrix} x_2 \\ x_3 \end{pmatrix}$$

and

$$\hat{A} = \begin{pmatrix} 3 & 6 \\ 6 & 8 \end{pmatrix}$$

So Relative Extremal Values of $\hat{q}$ are e-values of $\hat{A}$

$\hat{A}$ e-pairs are

$$(\hat{\lambda}_1, \hat{X}_1) = \left(12, \tfrac{1}{\sqrt{13}}\begin{pmatrix} 2 \\ 3 \end{pmatrix}\right) \quad \text{and} \quad (\hat{\lambda}_2, \hat{X}_2) = \left(-1, \tfrac{1}{\sqrt{13}}\begin{pmatrix} 3 \\ -2 \end{pmatrix}\right)$$

So, if we denote the e-values of A as $\lambda_1 \geq \lambda_2 \geq \lambda_3$, then Thm 1.8 tells us that

$$\lambda_1 \geq 12 \geq \lambda_2 \geq -1 \geq \lambda_3$$

Note that the example in the text quotes a result assigned as an exercise which states that:

(Problem 1.3.22.) "If the diagonal elements of a symmetric matrix are increased, then the e-values are increased (or, not decreased) as well."

And this result is used to further characterize the nature of the e-values of A. In particular, define

$$\tilde{A} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

<u>Note</u>: $\tilde{A}$ was obtained by increased the lower 2 diagonal elements of A by 1 unit. Otherwise, A and $\tilde{A}$ are the same. By Problem (1.3.22). the e-values of A are increased (or at least not decreased) by the formation of $\tilde{A}$.
But $\operatorname{rank}(\tilde{A}) = 1$ since
$R(A) = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}$.
And $\tilde{A}$ has one positive e-value & two zero e-values. From your textbook, the e-values of $\tilde{A}$ are $\tilde{\lambda}_1 = 14$, $\tilde{\lambda}_2 = \tilde{\lambda}_3 = 0$.
If we compare this info with e-values of A, →

This means that $14 \geq \lambda_1 \geq 12$ and $\lambda_2 \leq 0$.
Hence, from both of our analyses,

$$12 \leq \lambda_1 \leq 14 \quad \text{and} \quad \lambda_3 \leq -1 \leq \lambda_2 \leq 0.$$

So, our original matrix A has 1 positive e-value between 12 + 14 and two negative e-values.... GOOD TO KNOW!!