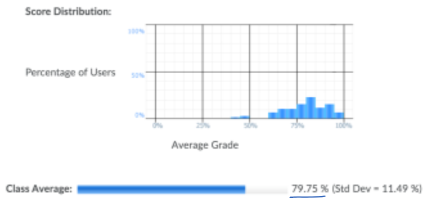
Lecture 23: Positive semidefinite matrices (class)

Admin:



T/F: If $\vec{v}$ is an e-vector of A^2, then $\vec{v}$ is an e-vector of A.

e.g. $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $A^2 = I$

$A^2 \vec{e}_1 = \vec{e}_1 \quad A\vec{e}_1 = \vec{e}_2$

A, B — e-values the same: AB, BA; singular values no relationship

$AB = B^{-1}(BA)B$ if B invertible

$AB = A(BA)A^{-1}$ if A invertible

If $BA\vec{v} = \lambda\vec{v}$, then $AB(A\vec{v}) = A(\lambda\vec{v}) = \lambda(A\vec{v})$

$\vec{v}$ e-vector for BA $\Rightarrow$ $A\vec{v}$ e-vector for AB unless $A\vec{v} = \vec{0}$ ($\lambda = 0$)

Positive Semi-Definite Matrices

Reading: Meyer §7.6, Strang ch. 6

Definition: A Hermitian (or real-symmetric) matrix A with all eigenvalues ≥ 0 is called $\boxed{\text{positive semi-definite}}$ denoted $A \succeq 0$.

If all e-values > 0 $\Rightarrow$ "positive definite" $A \succ 0$

Key property:

Theorem: A Hermitian matrix A is positive semidefinite $\iff$ $x^\dagger A x \geq 0$ for all vectors $\vec{x}$.

Proof:

$\Rightarrow$: Assume A Hermitian ($A = A^\dagger$)

$$A = \sum_j \lambda_j \vec{v}_j \vec{v}_j^\dagger \quad \text{with } \{\vec{v}_j\} \text{ orthonormal}, \ \lambda_j \in \mathbb{R}$$

$A \succeq 0 \Rightarrow \lambda_j \geq 0$ (definition)

$$x^\dagger A x = \sum_j \underbrace{\lambda_j}_{\geq 0} \underbrace{|\vec{v}_j \cdot \vec{x}|^2}_{\geq 0} \geq 0 \ ✓$$

$\Leftarrow$: If Hermitian A is not p.s.d., then some $\lambda_i < 0$

Let $\vec{x}$ be the corr. e-vector $\vec{x} = \vec{v}_i$:

$$\vec{v}_i^\dagger A \vec{v}_i = \lambda_i \cdot 1 = \lambda_i < 0. ✓$$

Example 1: For any real matrix A, $A^T A \succeq 0$.

Proof: for any x,

$$x^\dagger A^T A x = \|Ax\|^2 \geq 0 \underset{\text{Thm.}}{\Longrightarrow} A^T A \succeq 0.$$

Proof 2: e-values of $A^T A$ are the squared singular values of A.

$\Rightarrow$ all e-values $\geq 0 \Rightarrow A^T A \succeq 0$. ✓

Examples:

a)

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix} \succeq 0 \qquad \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -3 \end{pmatrix} \preceq 0$$

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix} \not\geq 0$$

not symmetric

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 0 & 5 \\ 6 & 7 & 3 \end{pmatrix}$$

$$A = \begin{pmatrix} -1 & 2 & 3 \\ 2 & 0 & 5 \\ 3 & 5 & 3 \end{pmatrix} \not\geq 0$$

$e_1^T A \vec{e}_1 = -1 < 0$

$\Rightarrow$ some e-value of A must be < 0

$$A = \begin{pmatrix} 3 & 4 \\ 4 & 0 \end{pmatrix}$$

$$\frac{1}{y^2}(x \;\; y)\begin{pmatrix} 3 & 4 \\ 4 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}$$

$$(x \;\; 1)\begin{pmatrix} a & b \\ b & 0 \end{pmatrix}\begin{pmatrix} x \\ 1 \end{pmatrix} = (x \;\; 1)\begin{pmatrix} ax+b \\ bx \end{pmatrix} = ax^2 + 2b \cdot x$$
$$= x(a\,x + 2b)$$

if $b \neq 0$

$A \not\geq 0$

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \quad A \geq 0 \Leftrightarrow a, c \geq 0 \text{ and } ac - b^2 \geq 0$$

eg. if $x = -\varepsilon$, then

$(-\varepsilon \;\; 1)\, A \begin{pmatrix} -\varepsilon \\ 1 \end{pmatrix} < 0$

$\Rightarrow A \not\geq 0$

Notice:

$$(x \;\; y)\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = (x \;\; y)\begin{pmatrix} ax+by \\ bx+cy \end{pmatrix}$$
$$= ax^2 + bxy + bxy + cy^2 = ax^2 + 2bxy + cy^2$$

b) $$B = \begin{pmatrix} 3 & 4 & 5 \\ 4 & 0 & 0 \\ 5 & 0 & 0 \end{pmatrix} \not\geq 0 \qquad A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 0 & 5 \\ 3 & 5 & 3 \end{pmatrix} \not\geq 0$$

why?

$$A = \begin{pmatrix} & & \\ 3 & 5 & 3 \end{pmatrix}$$

why?

we know $\begin{pmatrix} 3 & 4 \\ 4 & 0 \end{pmatrix} \not\succeq 0$

$$(-1 \quad 1)\begin{pmatrix} 3 & 4 \\ 4 & 0 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 3 - 8 = -5 < 0$$

$$(-1 \quad 1 \quad 0)\begin{pmatrix} 3 & 4 & 5 \\ 4 & 0 & 0 \\ 5 & 0 & 0 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = -5 < 0$$

Corollary: If $A \succeq 0$, then every submatrix along the diagonal must also be $\succeq 0$.

$$\begin{pmatrix} 5 & -1 \\ -1 & 5 \end{pmatrix} \succeq 0 \qquad \begin{pmatrix} 100 & 0 & 3 & 4 \\ 0 & 100 & 4 & 0 \\ 3 & 4 & 100 & 0 \\ 4 & 0 & 0 & 100 \end{pmatrix} \succeq 0$$

Proof in a second...

c) $C = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix} \succeq 0$

rank$(C) = 2$

e-vector	e-value
$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	3
$\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$	2

$C \succeq 0$

$$C = \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}(1 \quad 1 \quad 1)}_{\succeq 0} + \underbrace{\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}(0 \quad 1 \quad -1)}_{\succeq 0}$$

Recall: For any real matrix A, $A^T A \succeq 0$.

Claim: If $A \succeq 0$ and $B \succeq 0$, then $A + B \succeq 0$.

Proof: For any x,

$$x^t(A+B)x = \underbrace{x^t A x}_{\geq 0} + \underbrace{x^t B x}_{\geq 0} \geq 0 \quad \checkmark \square$$

SUMMARY:

Definition: Symmetric matrix A is positive definite $\Longleftrightarrow$ Eigenvalues all > 0

$\not\Uparrow$

Definition:

Symmetric matrix A is **positive definite** $\iff$ Eigenvalues all > 0

$\Updownarrow$ $\qquad$ $\Updownarrow$

$x^T A x > 0$ for all $x \neq \vec{0}$

$A \succ 0$

Symmetric matrix A is **positive semi-definite** $\iff$ Eigenvalues all ≥ 0

$\Updownarrow$ $\qquad$ $\Updownarrow$

$x^T A x \geq 0$ for all $\vec{x}$

$A \succcurlyeq 0$

hard

Ways to show that $A \succcurlyeq 0$

- compute all e-values, check ≥ 0 (lame)
- write $A = B + C$ with $B \succcurlyeq 0$, $C \succcurlyeq 0$ (or $B+C+D$, etc)
- show that $x^T A x \geq 0$ for all $\vec{x}$

Maybe also consider rank(A), ...

Ways to show $A \not\succcurlyeq 0$

- find an e-value < 0
- find $\vec{x}$ so $x^T A x < 0$
- show that a **submatrix** along diagonal of A is not $\succcurlyeq 0$
- check $\det(A) < 0$

$\det(A) = \prod_j \lambda_j$

Exercise: Which of these matrices is p.s.d.?

$$A = \begin{pmatrix} 3 & 4 & 5 & 0 & 1 \\ 4 & 2 & 1 & 0 & 1 \\ 5 & 1 & 10 & 2 & 1 \\ 0 & 0 & 2 & 2 & 1 \\ 1 & 1 & 1 & 2 & 3 \end{pmatrix}$$

$A^T \neq A$ so can't be p.s.d.

$$B = \begin{pmatrix} 3 & 1 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 & 1 \\ 1 & 1 & \boxed{-1} & 2 & 1 \\ 0 & 0 & 2 & 2 & 1 \\ 1 & 1 & 1 & 1 & 3 \end{pmatrix}$$

$e_3^T B e_3 = -1 < 0$

$\Rightarrow B \not\succcurlyeq 0$

$$C = \begin{pmatrix} 5 & 2 & 0 & 2 & 1 \\ 2 & 6 & 1 & 2 & 0 \\ 0 & 1 & 5 & 0 & 0 \\ 2 & 2 & 0 & 6 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$\begin{pmatrix} 5 & 1 \\ 1 & 0 \end{pmatrix} \not\succcurlyeq 0$ $\quad$ ($\det = -1 < 0$)

$\Rightarrow C \not\succcurlyeq 0$

$$\begin{pmatrix} x & 0 & 0 & 0 & 1 \end{pmatrix} C \begin{pmatrix} x \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = 5x^2 + 2x + 0 \cdot 1^2$$
$$= x(5x+2)$$
$$< 0 \text{ if } x = \frac{-1}{100}$$

$$D = \begin{pmatrix} 8 & -2 & 3 & 1/2 & 0 \\ -2 & 6 & 1 & 1 & -1 \\ 3 & 1 & 10 & 2 & 1 \\ 1/2 & 1 & 2 & 5 & 1 \\ 0 & -1 & 1 & 1 & 7 \end{pmatrix} \succ 0$$

$$E = \begin{pmatrix} 100 & 0 & 3 & 4 \\ 0 & 100 & 4 & 0 \\ 3 & 4 & 100 & 0 \\ 4 & 0 & 0 & 100 \end{pmatrix} = 100 I + \begin{pmatrix} 0 & \begin{smallmatrix} 3 & 4 \\ 4 & 0 \end{smallmatrix} \\ \begin{smallmatrix} 3 & 4 \\ 4 & 0 \end{smallmatrix} & 0 \end{pmatrix}$$

$\hookrightarrow$ norm ≤ 7

$-7 \leq$ evalues ≤ 7

$E \succ 0$

$0 < 93 \leq$ evalues of E ≤ 107

Also D is **diagonally dominant**: $\forall i$, $D_{ii} > \sum_{j \neq i} |D_{ij}|$

Also D is diagonally dominant: $\forall i,\ D_{ii} > \sum_{j\neq i} |D_{ij}|$.

Claim: Any symmetric, diagonally dominant matrix is $\succ 0$.

Proof:

Assume for contradiction $D\vec{v} = \lambda\vec{v}$ with $\lambda < 0$, $\vec{v} \neq \vec{0}$.

Let $i = \operatorname{argmax} |v_j|$.

We may assume that $v_i > 0$. (If $v_i < 0$, multiply by -1)

$$0 > (D\vec{v})_i = \sum_j D_{ij} v_j = D_{ii} v_i + \sum_{j\neq i} D_{ij} v_j$$

$$\overset{\|}{(\lambda\vec{v})_i = \lambda v_i}$$

$$\geq D_{ii} v_i - \sum_{j \neq i} |D_{ij} v_j|$$

$|v_j| \leq v_i$

$$\geq \Big(D_{ii} - \sum_{j\neq i} |D_{ji}|\Big) v_i$$

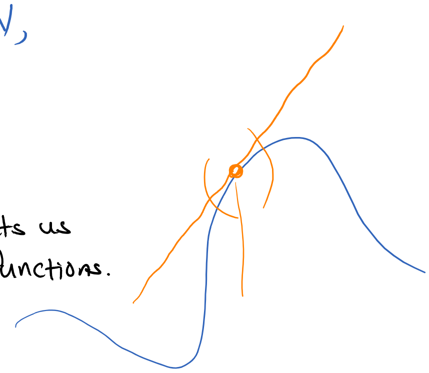
> 0 contradiction. ✓

WHY CARE? one reason: optimization

Recall: A function $T: V \to W$ is linear if

- $T(x+y) = T(x) + T(y)$ for all $x, y \in V$,
- $T(\alpha x) = \alpha T(x)$ for all scalars α.

Linear functions $\xleftrightarrow[\text{for } V \,\&\, W]{\text{w/ fixed bases}}$ Matrices



Not all functions are linear... but calculus lets us give affine approximations to differentiable functions.

Calculus $\longrightarrow$ linearity

$$f(x) - f(x_0) \approx f'(x_0)\cdot(x - x_0)$$

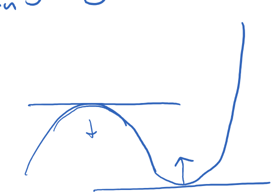
for $f: \mathbb{R} \to \mathbb{R}$ differentiable

Or for $f: \mathbb{R}^n \to \mathbb{R}^m$ differentiable,

$$\vec{f}(\vec{y}) - \vec{f}(\vec{y}_0) = \frac{\partial \vec{f}}{\partial x_1}(\vec{y}_0)\cdot(\vec{y}-\vec{y}_0)_1 + \cdots + \frac{\partial \vec{f}}{\partial x_n}(\vec{y}_0)\cdot(\vec{y}-\vec{y}_0)_n$$

$$= \begin{pmatrix} | & | & & | \\ \frac{\partial \vec{f}}{\partial x_1} & \frac{\partial \vec{f}}{\partial x_2} & \cdots & \frac{\partial \vec{f}}{\partial x_n} \\ | & | & & | \end{pmatrix} \begin{pmatrix} | \\ \vec{y} - \vec{y}_0 \\ | \end{pmatrix}$$

$$\overset{\|}{\sum_{j=1}^{n} \frac{\partial \vec{f}}{\partial x_j} \cdot \vec{e}_j^{\,T}}$$



But often, an affine approximation isn't enough!

Eg., studying local optima requires a quadratic approx$^{\underline{n}}$:

$$f(x) = f(x_0) + f'(x_0)\cdot(x-x_0) + \tfrac{1}{2} f''(x_0)\cdot(x-x_0)^2 + O((x-x_0)^3)$$

Single-variable functions

local minimum $\Leftarrow$ $f'(x_0)=0$, $f''(x_0)>0$

local maximum $\Leftarrow$ $f'(x_0)=0$, $f''(x_0)<0$

(if $f''(x_0)\neq 0$)

Multi-variable functions $f: \mathbb{R}^n \to \mathbb{R}$

$$f(y) = f(y_0) + \sum_j \frac{\partial f}{\partial x_j}(y_0)\cdot(y-y_0)_j$$

$$+\frac{1}{2}\sum_{j,k}\frac{\partial^2 f}{\partial x_j \partial x_k}(y_0)\cdot(y-y_0)_j(y-y_0)_k$$

$$= \frac{1}{2}(y-y_0)^T \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & & \\ & \frac{\partial^2 f}{\partial x_j \partial x_k} & \\ & & \end{pmatrix}(y-y_0)$$

"Hessian matrix" for f

For $f:\mathbb{R}^n\to\mathbb{R}$, y_0 is a local minimum if

$$\frac{\partial f}{\partial x_j}(y_0)=0 \quad \forall j=1,\dots,n$$

and $\underbrace{\frac{1}{2}\sum_{j,k}\frac{\partial^2 f}{\partial x_j \partial x_k}(y_0)\cdot(y-y_0)_j(y-y_0)_k}_{} > 0$

This is a quadratic form in the variables $(y-y_0)_1, \dots, (y-y_0)_n$

$$x^2-2xz+2y^2-4yz+z^2 = (x\ \ y\ \ z)\begin{pmatrix}1&0&-1\\0&2&-2\\-1&-2&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}$$

Definition: A quadratic form is a degree-two homogeneous polynomial.

$x^2+2xy+y^2$ $\qquad$ $x^2+y^2+z^2$ ✓ $\qquad$ $\cancel{x^2+y+1}$

Hessian: $\frac{1}{2}\begin{pmatrix}2&2\\2&2\end{pmatrix}=\begin{pmatrix}1&1\\1&1\end{pmatrix}$

$$x^2+xy+z^2 \checkmark \longleftrightarrow (x\ \ y\ \ z)\begin{pmatrix}1&\frac{1}{2}&0\\\frac{1}{2}&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}$$

Observe: Quadratic forms $\updownarrow$ Symmetric matrices

Corollary: Symmetric matrices let us study homogeneous quadratic functions.

$$x^2+2xy+y^2 = (x\ \ y)\begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$$

It also equals $(x\ \ y)\begin{pmatrix}1&2\\0&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$ and $(x\ \ y)\begin{pmatrix}1&\frac{3}{2}\\\frac{1}{2}&1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$,
but the symmetric form is nicer.

Example

$$f(x,y) = x^2+(2b)xy+y^2$$

$$= (x\ \ y)\underbrace{\begin{pmatrix}1&b\\b&1\end{pmatrix}}_{A\,=\,I+b\begin{pmatrix}0&1\\1&0\end{pmatrix}}\begin{pmatrix}x\\y\end{pmatrix}$$

$f(0,0)=0$

e-values	e-vectors
$1+b$	$\begin{pmatrix}1\\1\end{pmatrix}$
$1-b$	$\begin{pmatrix}1\\-1\end{pmatrix}$

$A\succ 0$ if $-1<b<1$ $\Rightarrow$ $\vec{0}$ is a global minimum

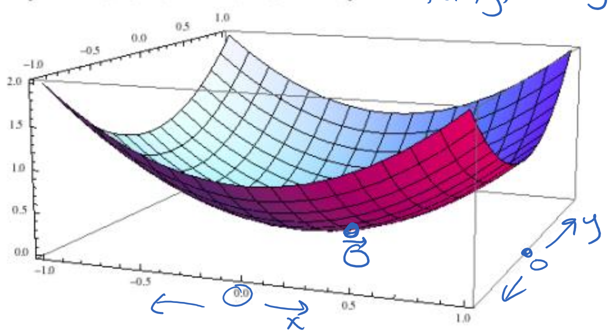
$A \geq 0$ if $-1 \leq b \leq 1$

$A \ngeq 0$ otherwise

```
b = 0;
Plot3D[x^2 + 2 b x y + y^2, {x, -1, 1}, {y, -1, 1}]
```

$A > 0$

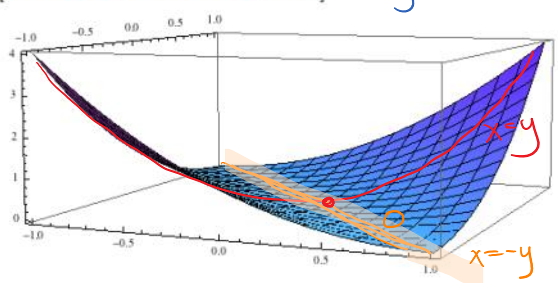
$f(x,y) = x^2 + y^2$




```
b = 1;
Plot3D[x^2 + 2 b x y + y^2, {x, -1, 1}, {y, -1, 1}]
```

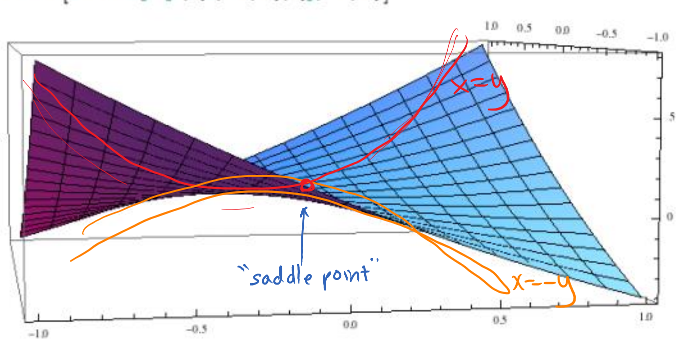
$f(x,y) = x^2 + 2bxy + y^2 = (x+y)^2$

e-values	e-vectors
$1+b = 2$	$(1,1)$
$1-b = 0$	$(1,-1)$



$\mathrm{Span}\left(\begin{pmatrix}1\\-1\end{pmatrix}\right)$

```
b = 3;
Plot3D[x^2 + 2 b x y + y^2, {x, -1, 1}, {y, -1, 1}]
```



$1+b = 4$

$1-b = -2$

critical point neither local max. nor min.

$A \ngeq 0,\ A \nleq 0$

$\begin{pmatrix}1 & b\\ b & 1\end{pmatrix} = \begin{pmatrix}0 & b\\ b & 0\end{pmatrix} + I$	Eigenvalues	Eigenvectors
	$1+b$	$(1,1)$
	$1-b$	$(1,-1)$

$$\Rightarrow \begin{pmatrix}1 & b\\ b & 1\end{pmatrix} = (1+b)\cdot\tfrac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}(1\ \ 1) + (1-b)\cdot\tfrac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}(1\ \ -1)$$

$$\Rightarrow f(x,y) = (x\ \ y)\begin{pmatrix}1 & b\\ b & 1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = x^2 + y^2 + 2b\,xy$$
$$= (1+b)\cdot\left(\tfrac{1}{\sqrt{2}}(1,1)\cdot(x,y)\right)^2$$
$$+ (1-b)\cdot\left(\tfrac{1}{\sqrt{2}}(1,-1)\cdot(x,y)\right)^2$$
$$= \frac{1+b}{2}(x+y)^2 + \frac{1-b}{2}(x-y)^2$$

Observe: Any quadratic form can be expressed as a sum of squares like this, and it is ≥ 0 iff all coefficients are ≥ 0.

By writing $A = \lambda_1 v_1 v_1^T + \cdots + \lambda_n v_n v_n^T$,

$$x^T A x = \sum_j \lambda_j (v_j \cdot x)^2 \quad \checkmark$$

Claim: $A \succeq 0 \iff A = M^T M$ for some matrix M.

Proof:

If $A = M^T M$, $x^T A x = \|Mx\|^2 \geq 0$
$\Rightarrow A \succeq 0$. ✓

If $A \succeq 0$, $A = \sum_j \lambda_j v_j v_j^T$.

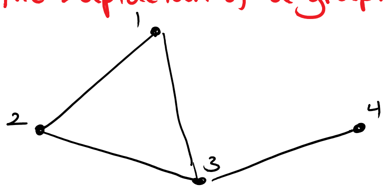
Let $M = \begin{pmatrix} \text{---} \sqrt{\lambda_1}\, v_1^T \text{---} \\ \text{---} \sqrt{\lambda_2}\, v_2^T \text{---} \\ \vdots \end{pmatrix} \Rightarrow M^T M = A$ ✓

□

(**Remark:** This also proves that $\underset{n\times n}{A} \succeq 0$ if and only if A is the **Gram matrix** for a set of n vectors.)

Example:

The Laplacian of a graph



$$L = \begin{pmatrix} \text{degrees along diagonal} & \text{-edges} \\ \text{-edges} & \end{pmatrix} = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 1 & 2 & -1 & -1 & 0 \\ 2 & -1 & 2 & -1 & 0 \\ 3 & -1 & -1 & 3 & -1 \\ 4 & 0 & 0 & -1 & 1 \end{array}$$

$L = E E^T$ for the incidence matrix E

$$E = \sum_{\substack{\text{edges} \\ (u,v)}} (e_u - e_v)\, e_{(u,v)}^T$$

$\Rightarrow L \succeq 0$

$$x^T L x = \sum_{\substack{\text{edges} \\ (u,v)}} (x_u - x_v)^2$$

Question: Is L positive definite? (not just $\succeq 0$)

Answer: No! $L\begin{pmatrix}1 \\ \vdots \\ 1\end{pmatrix} = 0$.

Observe: For $f: \mathbb{R}^n \to \mathbb{R}$, a "critical point" (where all derivatives $\frac{\partial f}{\partial x_j} = 0$) is a local minimum if

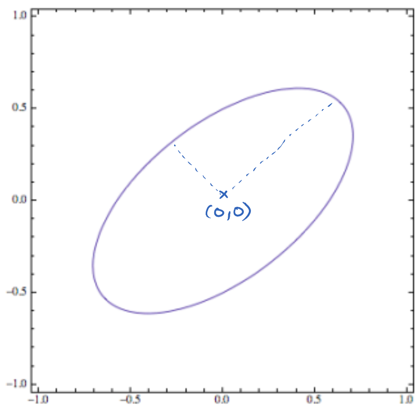
$$\begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \\ \vdots & & \ddots \end{pmatrix}$$ is positive definite.

Geometrically:

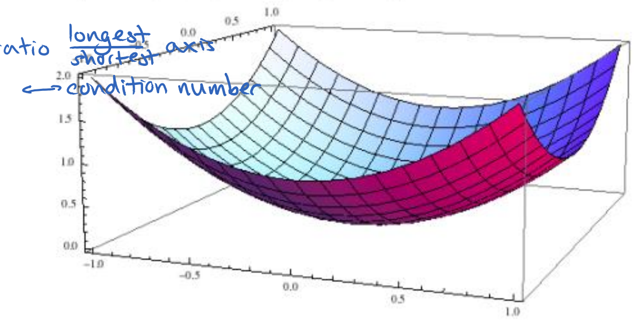
Quadratic forms assoc. to positive definite matrices $\longleftrightarrow$ Ellipses and ellipsoids

```
In[1]:= A = (3  -2
             -2  4);
        Eigenvalues[A // N]
        Eigensystem[A] // N // Transpose // MatrixForm
        ContourPlot[{x, y}.A.{x, y} == 1, {x, -1, 1}, {y, -1, 1}]

Out[2]= {5.56155, 1.43845}

Out[3]//MatrixForm=
        (5.56155  {-0.780776, 1.}
         1.43845  {1.28078, 1.})
```

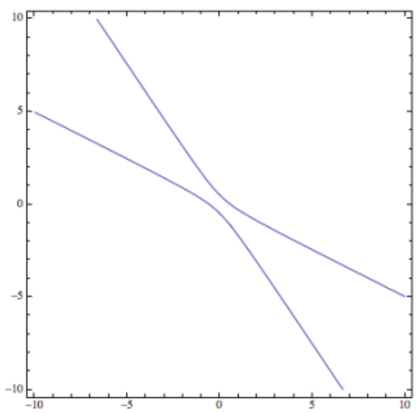
Out[4]=



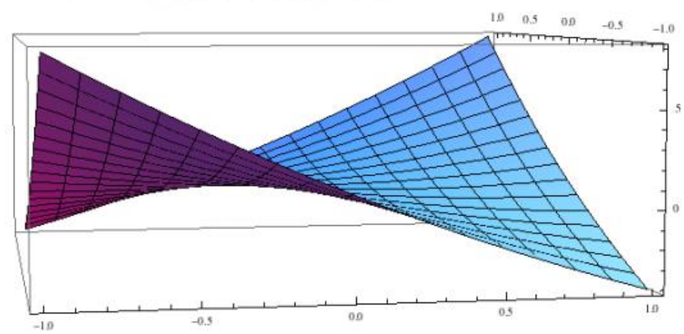

```
b = 0;
Plot3D[x^2 + 2 b x y + y^2, {x, -1, 1}, {y, -1, 1}]
```




```
A = (3  4
     4  4);
Eigenvalues[A // N]
ContourPlot[{x, y}.A.{x, y} == 1, {x, -10, 10}, {y, -10, 10}]

{7.53113, -0.531129}
```



```
b = 3;
Plot3D[x^2 + 2 b x y + y^2, {x, -1, 1}, {y, -1, 1}]
```



Ellipsoids also correspond to inner products/norms where some directions are weighted differently than others.

Example: GRADIENT DESCENT

Goal: Solve $A\vec{x} = \vec{b}$.

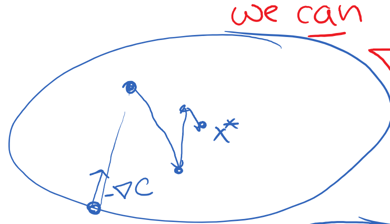
Recall: If A is $n \times n$, Gaussian elimination solves for $\vec{x}$ using $O(n^3)$ basic operations.
Can we approximately solve for $\vec{x}$ faster?

if $A \succ 0$
use $C(x) = (x - x^*) A (x - x^*)$
↑ we can't compute it!
we can compute $\nabla C = 2(Ax - b)$

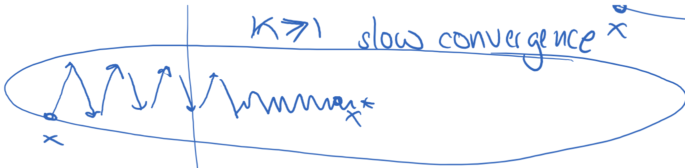
Define the cost function
$C(\vec{x}) = \|Ax - b\|^2$ if $Ax^* = b$

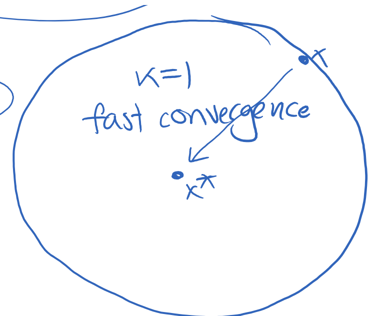
$= \|A(x - x^*)\|^2$

$= (x - x^*)^T \underbrace{A^T A}_{\text{p.s.d. matrix}} (x - x^*)$



$\kappa \gg 1$ slow convergence





Algorithm: Repeat until convergence:

$$x^t = x^{t-1} - \alpha\left(\frac{\partial C}{\partial x_1}(x^{t-1}), \dots, \frac{\partial C}{\partial x_n}(x^{t-1})\right)$$

i.e., step in the direction opposite the largest increase in C.

$$\nabla C = \left(\frac{\partial C}{\partial x_1}, \dots, \frac{\partial C}{\partial x_n}\right)$$

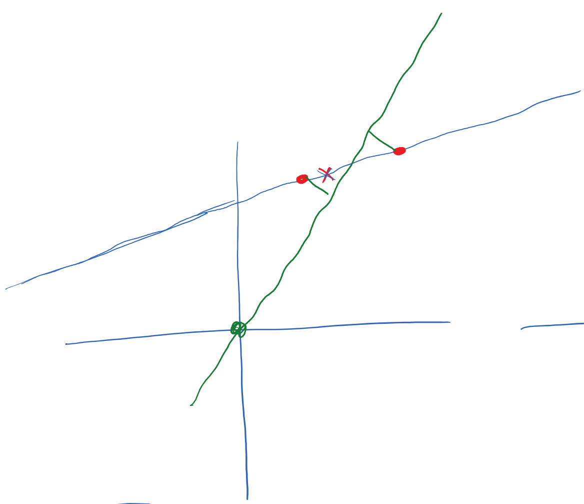
$$\frac{\partial C}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\sum_{ik}(x-x^*)_i (A^TA)_{ik}(x-x^*)_k\right)$$

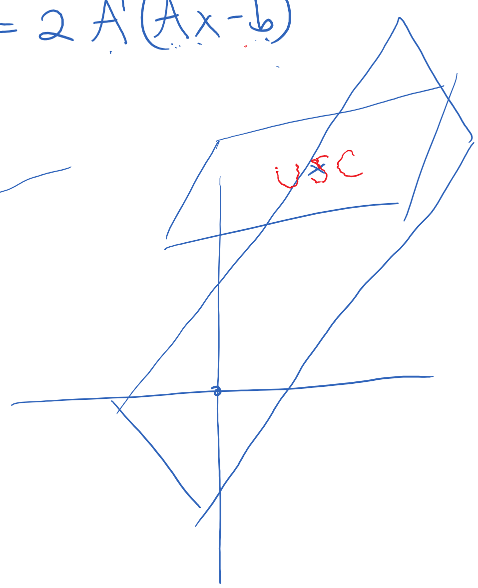
$$= 2(A^TA)_{jj}(x-x^*)_j + \sum_{k\neq j}(A^TA)_{jk}(x-x^*)_k + \sum_{i\neq j}(x-x^*)_i(A^TA)_{ij}$$

$$= 2\left[(A^TA)(x-x^*)\right]_j$$

$$\Rightarrow \nabla C = 2A^TA(x-x^*) = 2A^T(Ax-b)$$

Next time:
Convergence speed as a function of $\kappa(A)$ or $\kappa(A^TA) = \kappa(A)^2$.





Example:

```
A = (  2    -1.5
     -1.5    4  );
b = A.{3, 3};
Eigenvalues[A]

α = .01;
x = {0, -4};
list = {x};
For[k = 1, k ≤ 1000, k++,
  x = x - 2 α (A.x - b).A;
  AppendTo[list, x];
 ];
"The solution is:"
x

{4.80278, 1.19722}

The solution is:
```

= Midterm + (1 - Midterm) * (Midterm 2)

Example:

```
A = ( 2    -1.5 )
    ( -1.5   4  );
b = A.{3, 3};
Eigenvalues[A]

α = .01;
x = {0, -4};
list = {x};
For[k = 1, k ≤ 1000, k++,
  x = x - 2 α (A.x - b).A;
  AppendTo[list, x];
 ];
"The solution is:"
x
```

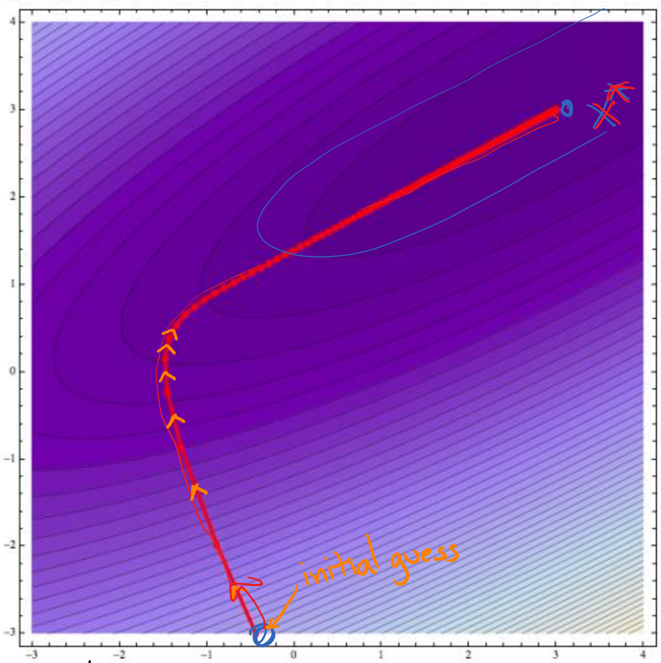
{4.80278, 1.19722}

The solution is:

{3., 3.}

$= \text{Midterm} + (1-\text{Midterm}) * (\text{Midterm } 2)$

```
"Display the results:";
iter = ListPlot[list, PlotRange → 3 {{-1, 1}, {-1, 1}}, Joined → True,
   PlotStyle → {Red, Thickness[.005]}, PlotMarkers → Automatic];
contour = ContourPlot[Norm[A.{x, y} - b]^2, {x, -3, 4}, {y, -3, 4}, Contours → 50];
Show[contour, iter]
```



For small enough α, this has to converge, since there is a unique local and global minimum.

Observe: Smaller condition number
$\Rightarrow$ Rounder ellipse $\Rightarrow$ Faster convergence

Remark: Although not generally the fastest way of solving a set of linear equations, the gradient descent approach is very robust, and **many** variants are used in **many** applications.

Convergence analysis:

Let x^* solve $Ax^* = b$.

$$x_t = x_{t-1} - 2\alpha A^T(Ax_{t-1} - b)$$

$$\Rightarrow x_t - x^* = (x_{t-1} - x^*) - 2\alpha A^T A (x_{t-1} - x^*)$$

$$= (I - 2\alpha A^T A)(x_{t-1} - x^*)$$

$\Rightarrow$ error magnitude

$$\|\vec{x}_t - \vec{x}^*\| \le \|I - 2\alpha A^T A\| \cdot \|\vec{x}_{t-1} - \vec{x}^*\|$$

If the e-values of $A^T A$ are

$$\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n \ge 0$$

$\uparrow$ these are (singular values of $A)^2$

$I - 2\alpha A^T A$ has e-values

$$1 - 2\alpha\lambda_1 \le \cdots \le 1 - 2\alpha\lambda_n$$

$$\Rightarrow \|I - 2\alpha A^T A\| = \max_i |1 - 2\alpha\lambda_i|$$

$$\max\{|1-2\alpha\lambda_1|, |1-2\alpha\lambda_n|\}$$

The best choice of α will make these equal in magnitude, centered on 0:

$$\Rightarrow -(1-2\alpha\lambda_1) = +(1-2\alpha\lambda_n)$$

$$\Rightarrow \alpha = \frac{1}{\lambda_1 + \lambda_n}$$

With this choice for α,

$$\|I - 2\alpha A^T A\| = 1 - \frac{2\lambda_n}{\underbrace{\lambda_1 + \lambda_n}_{\le 2\lambda_1}} \le 1 - \frac{\lambda_n}{\lambda_1}$$

$$= 1 - \frac{1}{\text{condition \# of } A^T A}$$

$$= 1 - \frac{1}{(\text{condition \# of } A)^2}$$

$$\Rightarrow \|x^t - x^*\| \le \left(1 - \frac{1}{\kappa^2}\right)^t \cdot \|x^0 - x^*\|$$

$$\Rightarrow t = \kappa^2\left(\log\frac{1}{\varepsilon} + \log\frac{1}{\|x^0 - x^*\|}\right) \qquad \text{since } \left(1 - \frac{1}{\kappa^2}\right)^{\kappa^2} \le e^{-1}$$

ensures $\|x^t - x^*\| \le \varepsilon$.

Remark: The conjugate gradient algorithm has $O(\sqrt{\kappa})$ dependence instead of $O(\kappa^2)$ dependence, and

$$x^t = x^{t-1} - \alpha\underbrace{(Ax^{t-1} - b)}_{\text{error}}$$

has $O(\kappa)$ dependence if $A \succeq 0$.

One more motivation for positive semi-definite matrices:

Semi-definite programming

This was touched on in the homework....

SDPs $\quad \max \ \mathrm{Trace}(AB)$ $\quad$ are everywhere!

$\quad\quad\quad$ s.t. $A \succeq 0$

They can be solved efficiently and have a notion of duality roughly because the set of p.s.d. matrices is convex.