Lecture 23: Positive semidefinite matrices

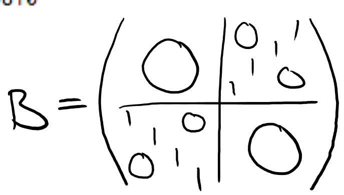
Eigenvalue examples (continued):

a)

```
>> n = 10;
A = -2 * diag(ones(n,1)) + diag(ones(n-1,1),1) + diag(ones(n-1,1),-1)

A =

    -2     1     0     0     0     0     0     0     0     0
     1    -2     1     0     0     0     0     0     0     0
     0     1    -2     1     0     0     0     0     0     0
     0     0     1    -2     1     0     0     0     0     0
     0     0     0     1    -2     1     0     0     0     0
     0     0     0     0     1    -2     1     0     0     0
     0     0     0     0     0     1    -2     1     0     0
     0     0     0     0     0     0     1    -2     1     0
     0     0     0     0     0     0     0     1    -2     1
     0     0     0     0     0     0     0     0     1    -2

>> eig(A)

ans =

   -3.9190
   -3.6825
   -3.3097
   -2.8308
   -2.2846
   -1.7154
   -1.1692
   -0.6903
   -0.3175
   -0.0810
```

why are these centered on -2?

b)



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 $A \succeq 0$.

Key property:

Theorem: A Hermitian matrix A

is positive semidefinite
$$\iff$$
$x^+ A x \geq 0$ for all vectors x.

Proof:

$\Rightarrow$: If $A \succeq 0$,

let $v_1, \dots, v_n$ be an orthonormal basis of eigenvectors, with corresponding e-values $\lambda_1, \dots, \lambda_n \geq 0$.

Any vector x can be expanded as
$$\vec{x} = \sum_{j=1}^{n} \alpha_j \vec{v}_j$$
where $\alpha_j = v_j \cdot \vec{x}$.

$$\Rightarrow x^+ A x = \sum_{i,j} \alpha_i^* \underbrace{v_i^+ A \vec{v}_j}_{\lambda_j v_i^+ v_j} \alpha_j$$

$$\lambda_j v_i^+ v_j = \begin{cases} \lambda_j, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

$$= \sum_{j=1}^{n} |\alpha_j|^2 \cdot \lambda_j$$
$$\geq 0 \checkmark$$

$\Leftarrow$: If $A \not\succeq 0$, i.e., some eigenvalue $\lambda < 0$, let $\vec{x}$ be a corresponding e-vector.

$$\Rightarrow x^+ A x = \lambda x^+ x$$
$$= \lambda \|x\|^2$$
$$< 0 . \checkmark$$

$\square$

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

Proof:
$$x^T A^T A x = (Ax)^T (Ax)$$
$$= \|Ax\|^2$$
$$\geq 0$$

$\Rightarrow$ by the theorem, $A^T A \succeq 0$. ✓ $\square$

Examples:

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

Claim: $A \not\succeq 0$

Proof 1: $\det(A) = 3 \cdot 0 - 4 \cdot 4 = -16$

$$= \lambda_1 \cdot \lambda_2$$

Since $\det(A) < 0$ and eigenvalues are real (as $A = A^T$), one of the e-values is < 0.

Proof 2:

$$(1 \quad x)\, A \begin{pmatrix} 1 \\ x \end{pmatrix} = 3 + 8x$$

This is < 0 if $x = -1$ (for example).

b) $$B = \begin{pmatrix} 3 & 4 & 5 \\ 4 & 0 & 0 \\ 5 & 0 & 0 \end{pmatrix}$$

Claim: $B \not\succeq 0$

Note that $\operatorname{rank}(B) = 2$, so $\det(B) = 0 = \lambda_1 \lambda_2 \lambda_3$.
One of the e-values is 0, so the determinant doesn't tell us anything about the others.

Proof:

$$(1, -1, 0)\, B \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = -5 < 0 \qquad \square$$

Observe: Since A is a submatrix of B along the diagonal

$$\begin{pmatrix} 3 & 4 & 5 \\ 4 & 0 & 0 \\ 5 & 0 & 0 \end{pmatrix}$$

and $A \not\succeq 0$, B can't be $\succeq 0$ either. ✓

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

Note: $C = C^T$ (so e-values are real)

Rank$(C) = 2$ since $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

(so one e-value is 0, and $\det(C) = 0$)

Claim: $C \succeq 0$.

```
>> C = [1 1 1; 1 2 0; 1 0 2];
>> eig(C)

ans =

    0.0000
```

$\longleftarrow$ this is not a proof!

```
ans =

   -0.0000
    2.0000
    3.0000
```

← this is not a proof!

Proof:

$$C = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

$$= \underbrace{\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}}_{=\vec{x}} (1 \ 1 \ 1) + \underbrace{\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}}_{=\vec{y}} (0 \ -1 \ 1)$$

$$= \vec{x}\vec{x}^T + \vec{y}\vec{y}^T$$

Recall: $\boxed{\text{For any real matrix } A,\ A^TA \succcurlyeq 0.}$

$$\Rightarrow xx^T \succcurlyeq 0 \text{ and } yy^T \succcurlyeq 0$$

$$\Rightarrow xx^T + yy^T \succcurlyeq 0$$

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

Proof:

For any vector $\vec{v}$,

$$v^t(A+B)v = \underbrace{v^tAv}_{\geq 0} + \underbrace{v^tBv}_{\geq 0} \quad (\text{since } A \succcurlyeq 0, B \succcurlyeq 0)$$

$$\geq 0$$

$$\Rightarrow A+B \succcurlyeq 0 \qquad \square$$

SUMMARY:

Definition:

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

$\Updownarrow$

$x^TAx > 0$ for all $x \neq \vec{0}$ ← we proved this!

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

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

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

Ways to show that $A \succeq 0$

- compute all e-values, check ≥ 0 (lame)
- write $A = B + C$ with $B \succeq 0$, $C \succeq 0$ (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 $A \not\succeq 0$

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

$\mathrm{Det}(A) = \prod_{j=1}^{n} \lambda_j$

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

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

No! $A^T \neq A$

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

No! $\vec{e}_3^{\,T} B \vec{e}_3 = -1 < 0$

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

No! Observe the submatrix $C' = \begin{pmatrix} 5 & 1 \\ 1 & 0 \end{pmatrix}$ $\quad \mathrm{Det}\, C' = -1 < 0$

or: $(-x \;\; 1)\, C' \begin{pmatrix} -x \\ 1 \end{pmatrix} = 5x^2 - 2x = x(5x-2)$
< 0 for $0 < x < \frac{2}{5}$

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

Yes! $D = D^T$ ✓

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

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

Let $i = \arg\max |v_j|$.
We may assume that $v_i > 0$. (If $v_i < 0$, multiply by -1)

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

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

$$\geq v_i \cdot \left(D_{ii} - \sum_{j \neq i} |D_{ij}|\right) \quad (|v_j| \leq v_i)$$

$$> 0. \quad \text{X contradiction} \quad \square$$

$\lambda v_i < 0$

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

$$\vec{f}(\vec{y}) - \vec{f}(y_0) \approx \frac{\partial \vec{f}}{\partial x_1}(\vec{y}_0) \cdot (\vec{y} - \vec{y}_0)_1 + \cdots + \frac{\partial \vec{f}}{\partial x_n}(\vec{y}_0) \cdot (\vec{y} - \vec{y}_0)_n$$

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

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

But often, an affine approximation isn't enough!
Eg., studying local optima requires a quadratic approx^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 $\longleftrightarrow f'(x_0)=0,\ f''(x_0)>0$

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

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

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

$$f(y) = f(y_0) + \sum_j \frac{\partial f}{\partial x_j}(y_0)\cdot(y-y_0)_j + \frac{1}{2}\sum_{j,k} \frac{\partial^2 f}{\partial x_j \partial x_k}\cdot(y-y_0)_j(y-y_0)_k + O(\|y-y_0\|^3)$$

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 $(y-y_0)_1, \dots, (y-y_0)_n$

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

$x^2+2xy+y^2$ ✓ $\qquad x^2+2x$ ✗

$x^2-2xz+2y^2-2yz+z^2$ ✓ $\qquad x^2+2xy+y$ ✗

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 $(x\ y)\begin{pmatrix}1 & 3/2\\ 1/2 & 1\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix}$, but the symmetric form is nicer.

Examples:

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

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

$$\lambda_1 + \lambda_2 = Tr(A) = 2$$

$$\lambda_1 \cdot \lambda_2 = Det(A) = 1 - b^2$$

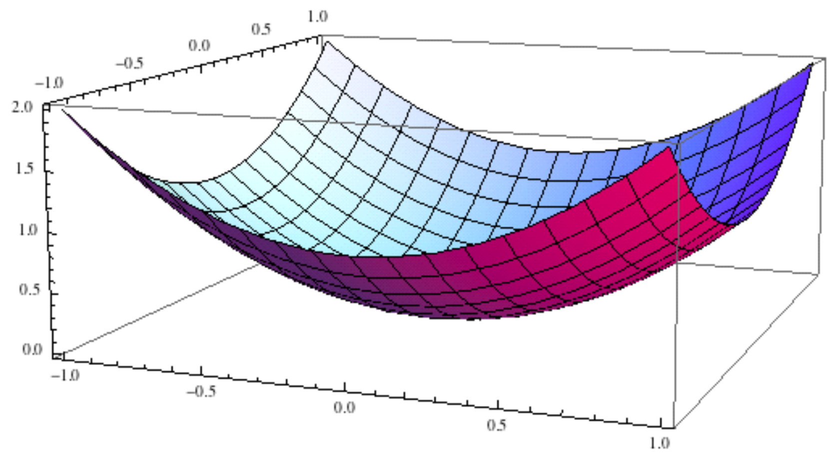
$$\Rightarrow \ A \succ 0 \ \text{if} \ b < 1$$

$$A \succeq 0 \ \text{if} \ b \leq 1$$

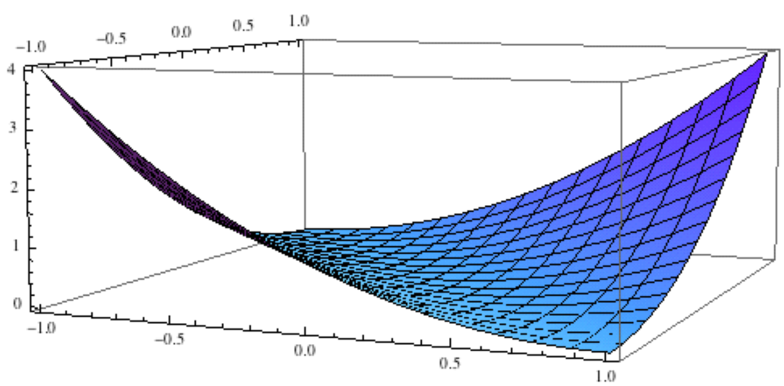
$$A \nsucc 0 \ \text{if} \ b > 1$$

(A is indefinite)

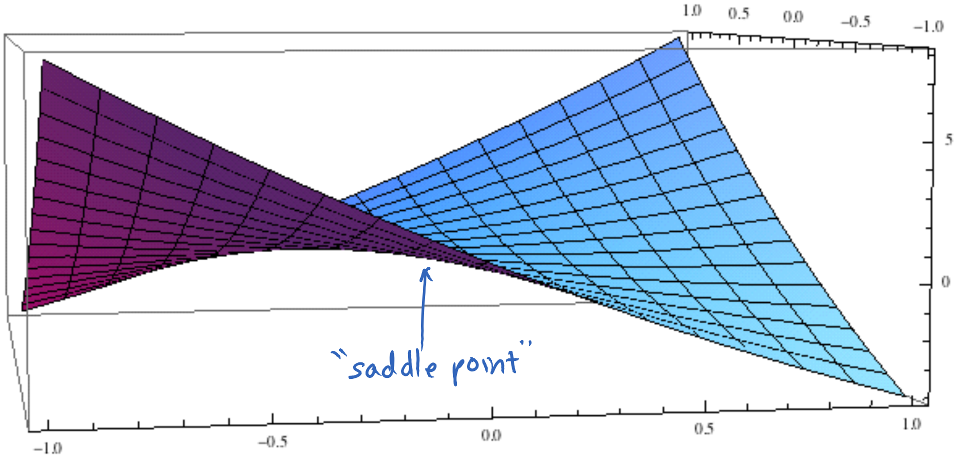
```
b = 0;
Plot3D[x^2 + 2 b x y + y^2, {x, -1, 1}, {y, -1, 1}]
```




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



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



$$\begin{pmatrix} 1 & b \\ b & 1 \end{pmatrix} \qquad \underline{\text{Eigenvalues}} \qquad \underline{\text{Eigenvectors}}$$

$$\shortparallel \qquad 1+b \qquad (1,1)$$

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

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

$$\Rightarrow f(x,y) = (x \ \ y)\begin{pmatrix} 1 & b \\ b & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = x^2 + y^2 + 2bxy$$

$$= (1+b)\cdot\left(\frac{1}{\sqrt{2}}(1,1)\cdot(x,y)\right)^2$$

$$+ (1-b)\cdot\left(\frac{1}{\sqrt{2}}(1,-1)\cdot(x,y)\right)^2$$

$$= \frac{1+b}{2}(x+y)^2 + \frac{1-b}{2}(x-y)^2$$

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

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

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 $M = \begin{pmatrix} \text{—} \sqrt{\lambda_1} v_1^T \text{—} \\ \vdots \end{pmatrix} \Rightarrow M^T M = A$ ✓

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

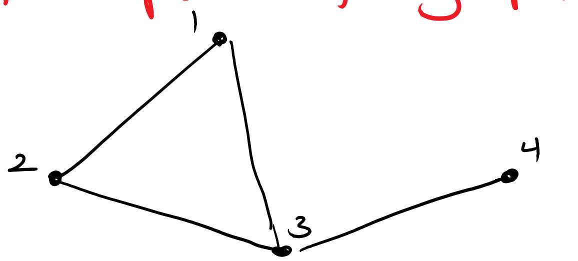
Let $M = \begin{pmatrix} — \sqrt{\lambda_1}\, v_1^T — \\ — \sqrt{\lambda_2}\, v_2^T — \\ \vdots \end{pmatrix} \Rightarrow M^T M = A$ ✓

$\square$

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

Example:

The Laplacian of a graph



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

$L = E E^T$ 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.

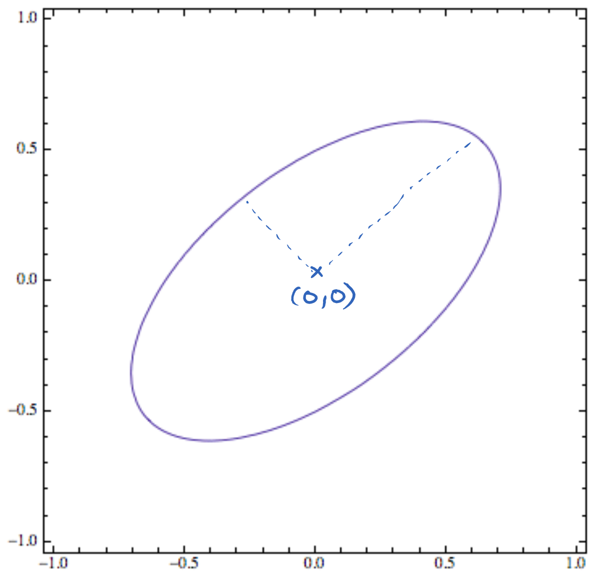
Geometrically:

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]=

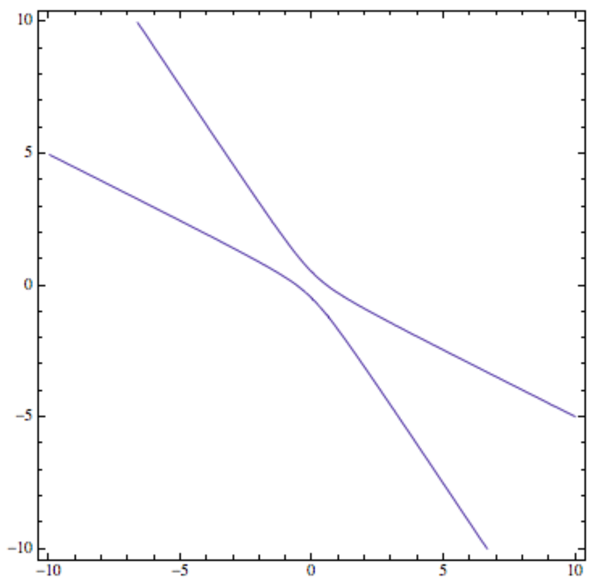


ratio $\frac{\text{longest}}{\text{shortest}}$ axis

$\longleftrightarrow$ condition number

```
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}



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?

Define the cost function

$$C(\hat{x}) = \|Ax - b\|^2$$

$$= \|A(x - x^*)\|^2 \quad \text{where } Ax^* = b$$

$$= (x - x^*)^T \underbrace{A^T A}_{\text{positive semi-definite}} (x - x^*)$$

Algorithm: Repeat until convergence:

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

ie., step in the direction opposite the largest increase in C.

$$\frac{\partial C}{\partial x_j} = \frac{\partial}{\partial x_j}\left[(Ax - b)^T (Ax - b)\right]$$

$$= \frac{\partial}{\partial x_j} \sum_i (Ax-b)_i^2$$

$$= 2\sum_i (Ax-b)_i \cdot \frac{\partial}{\partial x_j}\left[(Ax-b)_i\right]$$

$$= 2\sum_i (Ax-b)_i \, a_{ij}$$

$$\Rightarrow \left(\frac{\partial C}{\partial x_1}, \dots, \frac{\partial C}{\partial x_n}\right) = 2\sum_i \underbrace{(Ax-b)_i}_{(Ax-b)^T e_i} \cdot \underbrace{(a_{i1}, \dots, a_{in})}_{e_i^T A}$$

$$= 2(Ax-b)^T\left(\sum_i e_i^T e_i\right) A$$

$$= 2(Ax-b)^T A$$

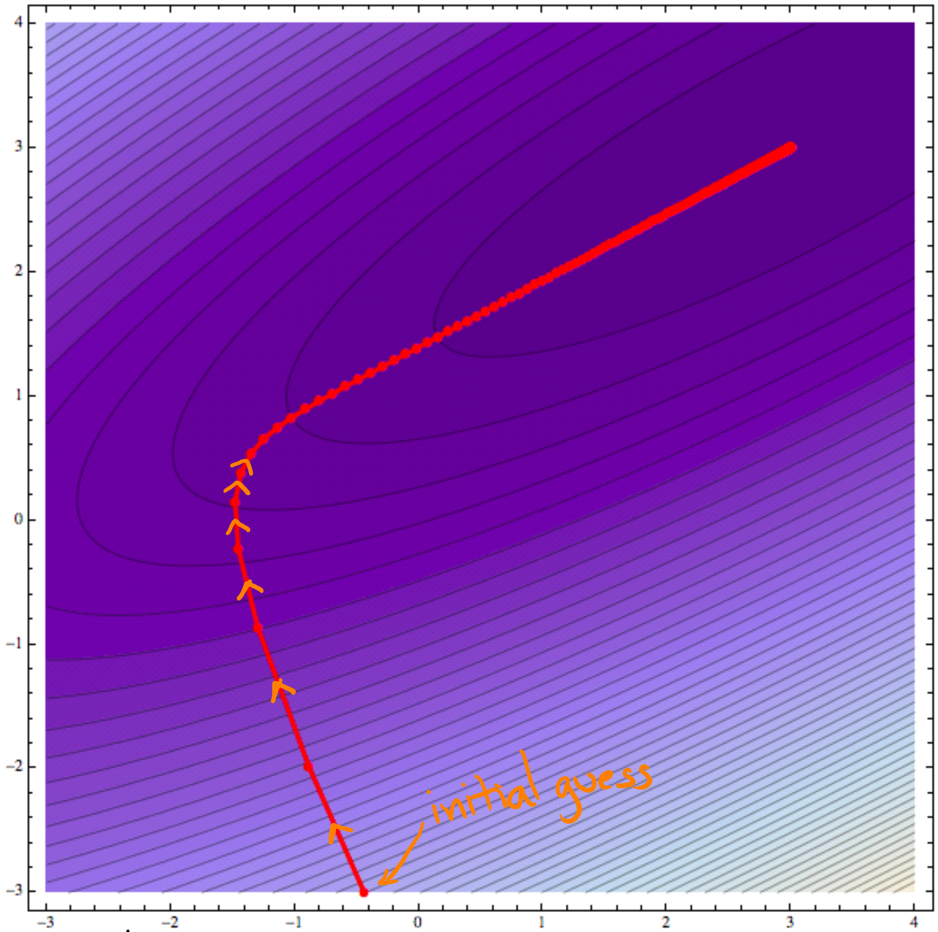
(As you might guess!)

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.}
```

```
"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.

<u>Observe</u>: Smaller condition number
$\Rightarrow$ Rounder ellipse $\Rightarrow$ Faster convergence

<u>Remark</u>: 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.

<u>Convergence analysis</u>:

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

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

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

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$$

these are (singular values of A)^2

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

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

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

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

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,

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

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

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

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

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

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

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

$$x^t = x^{t-1} - \alpha(Ax^{t-1} - b)$$

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

One more motivation for positive semi-definite matrices:

Semi-definite programming

This was touched on in the homework....

SDPs $\quad \max \operatorname{Trace}(AB) \quad$ 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.