Spectral theorem. Suppose A is real and symmetric. Then there is an orthogonal matrix B such that $B^tAB = B^{-1}AB = $ diagonal.

Proof. We use induction. If $A=[a]$ is 1×1, then $B=[1]$ will do since $B^tAB=[a]$ is diagonal. Assume the result for $n\times n$. Let A be $(n+1)\times(n+1)$ and pick an eigenvector $\vec{v}_1$ with eigenvalue λ_1, say. Extend v_1 to a basis and use Gram-Schmidt to get an orthonormal basis $w_1, w_2, \dots, w_n$ with $w_1 = \frac{1}{\|v_1\|}\vec{v}_1$. Thus $\vec{w}_1$ is an eigenvector as well.

We claim $B^tAB = B^{-1}AB = \left[\begin{array}{c|c} \lambda_1 & 0 \dots 0 \\ \hline 0 & \\ \vdots & P \\ 0 & \end{array}\right]$ when $B=[\vec{w}_1 \dots \vec{w}_n]$.

In fact, we know $B^t = B^{-1}$ by orthonormality and also $(B^tAB)\vec{e}_1 = B^tA\vec{w}_1 = B^t(\lambda_1\vec{w}_1) = B^{-1}(\lambda_1 B\vec{e}_1) = \lambda_1\vec{e}_1$, while B^tAB is symmetric since $(B^tAB)^t = B^tA^tB^{tt} = B^tAB$.

Then P is symmetric as well, so we can use the induction hypothesis to get Q orthogonal with $Q^tPQ = $ diagonal. This implies that

$$\left[\begin{array}{c|c} 1 & \\ \hline & Q \end{array}\right]^t \left[\begin{array}{c|c} \lambda_1 & 0 \\ \hline 0 & P \end{array}\right] \left[\begin{array}{c|c} 1 & \\ \hline & Q \end{array}\right] = \left[\begin{array}{c|c} \lambda_1 & \\ \hline & Q^tPQ \end{array}\right] = \text{diagonal}$$

so $$R^t \; B^t A B \; R = \text{diagonal}$$

so $$(BR)^t \cdot A \cdot (BR) = \text{diagonal}.$$

It remains to check BR is orthogonal. In fact,

$$(BR)^t BR = R^t \underbrace{B^t B}R = R^tR \text{ -- by above}$$

and R is orthogonal (since the columns of R are mutually orthogonal unit vectors). ▨

Application (Quadratic forms) Consider a quadratic function Q in n variables $Q(x_1, \dots, x_n) = \sum_{i \le j} a_{ij} x_i x_j$. This arises from bilinear forms $\langle \vec{x}, \vec{y} \rangle = \sum_{i,j} a_{ij} x_i y_j$ by taking $\vec{x} = \vec{y}$.

We need to be careful with the off-diagonal terms:

$$a_{11} x_1 y_1 + a_{12} x_1 y_2 + a_{21} x_2 y_1 + a_{22} x_2 y_2 \xrightarrow{\vec{x} = \vec{y}} a_{11} x_1^2 + 2 a_{12} x_1 x_2 + a_{22} x_2^2$$

when A is symmetric.

Thus, $Q(x_1, \dots, x_n) = \sum_{i \le j} a_{ij} x_i x_j = \vec{x}^t A \vec{x}$ with

$$a_{ij} = \begin{cases} \frac{1}{2} \text{ coeff. of } x_i x_j & \text{if } i \neq j \\ \text{coeff. of } x_i^2 & \text{if } i = j \end{cases}$$

Example. Consider the quadratic equation $xy = 1$.

We can write $Q(x, y) = xy = \vec{x}^t A \vec{x}$ with $\vec{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $A = \begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}$.

Since A is symmetric, A is diagonalisable by the spectral theorem.

We have ----- eigenvalues $\lambda_1 = 1/2$ and $\lambda_2 = -1/2$

eigenvectors $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

We can diagonalise with $B = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} \cos \pi/4 & -\sin \pi/4 \\ \sin \pi/4 & \cos \pi/4 \end{bmatrix}$.

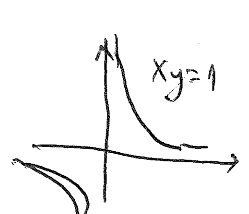
Then $B^t A B = B^{-1} A B = \begin{bmatrix} 1/2 & \\ & -1/2 \end{bmatrix}$ is diagonal.

If we change variables ... $\vec{x} = B \vec{y}$

we get $\vec{x}^t \underline{\underline{A}} \vec{x} = (B\vec{y})^t A (B\vec{y}) = \vec{y}^t \underline{(B^t A B)} \vec{y}$

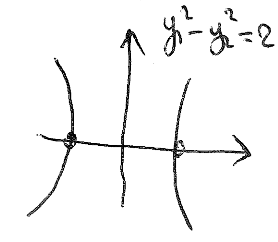
so we get a quadratic with diagonal entries.

In this case $x^t A \vec{x} = \frac{1}{2} y_1^2 - \frac{1}{2} y_2^2$ with $\vec{y} = B^t \vec{x}$.



$\boxed{\text{old variables}}$ $\quad xy = \frac{1}{2} y_1^2 - \frac{1}{2} y_2^2 \quad$ $\boxed{\text{new variables}}$

$$xy = \frac{1}{2}\left(\frac{x+y}{\sqrt{2}}\right)^2 - \frac{1}{2}\left(\frac{-x+y}{\sqrt{2}}\right)^2.$$



More generally, a quadratic $\vec{x}^t A \vec{x}$ with A symmetric can be expressed as $(B\vec{y})^t A (B\vec{y}) = \vec{y}^t (B^t A B)\vec{y} = \sum_{i=1}^{n} \lambda_i y_i^2$, where B is orthogonal and λ_i = the eigenvalues.

Example. Consider $x^2 - y^2 + z^2 - 2xy - 2xz - 2yz = 1$.

In this case, $A = \begin{bmatrix} 1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$. The eigenvalues are

$\lambda = 1$, $\lambda = 2$ and $\lambda = -2$ with eigenvectors

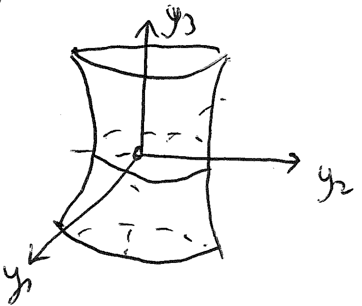
$v = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$, $v = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ and $v = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$. We divide each of those by its length to get

$$B = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ -1/\sqrt{3} & 0 & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \Rightarrow B^t A B = B^{-1} A B = \begin{bmatrix} 1 & & \\ & 2 & \\ & & -2 \end{bmatrix}.$$

This proves: we can change variables by $\vec{x} = B\vec{y}$ to write

$$x^2 - y^2 + z^2 - 2xy - 2xz - 2yz = \sum_{i=1}^{3} \lambda_i y_i^2 = y_1^2 + 2y_2^2 - 2y_3^2.$$

Thus, the original equation becomes $y_1^2 + 2y_2^2 - 2y_3^2 = 1$ ... an 1-sheeted hyperboloid (that does not meet the y_3-axis).



The new variables y_1, y_2, y_3 can be determined as

$$\vec{x} = B\vec{y} \quad \text{or} \quad \vec{y} = B^{-1}\vec{x} = B^t \vec{x}$$

or $\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$. We have actually shown

$$\begin{aligned} x^2 - y^2 + z^2 - 2xy - 2xz - 2yz &= y_1^2 + 2y_2^2 - 2y_3^2 \\ &= \left(\frac{x-y+z}{\sqrt{3}}\right)^2 + 2\left(\frac{-x+z}{\sqrt{2}}\right)^2 - 2\left(\frac{x+2y+z}{\sqrt{6}}\right)^2. \end{aligned}$$

Positive definite symmetric

We say A is positive definite symmetric, if $A^t = A$ (symmetric) and $\langle \vec{x}, \vec{x} \rangle = \vec{x}^t A \vec{x}$ is positive for all $\vec{x} \neq 0$.

<u>Theorem</u> (Tests for positive definiteness) Suppose A is symmetric & real. Then A is positive definite if and only if:

1. <u>Definition</u> .... $\vec{x}^t A \vec{x}$ is positive for all $\vec{x} \neq 0$
2. <u>Eigenvalues</u> .... $\lambda > 0$ for all eigenvalues λ of A
3. <u>Sylvester's criterion</u> the determinants of all $k \times k$ upper left submatrices are all positive.

<u>Example 1.</u> Let $A = \begin{bmatrix} 3 & 1 \\ 1 & 4 \end{bmatrix}$. To check it is pos. def. we note that $3 > 0$ and $\det A = 3 \cdot 4 - 1 > 0$. Alternatively, we can compute eigenvalues

$$\lambda^2 - 7\lambda + 11 = 0 \implies \lambda = \frac{7 \pm \sqrt{5}}{2} \text{ are both positive.}$$

<u>Example 2.</u> Let $A = \begin{bmatrix} 3 & 1 & 0 \\ 1 & 2 & 5 \\ 0 & 5 & x \end{bmatrix}$ with x some parameter.

The eigenvalues are difficult to find explicitly. Using Sylvester's criterion .... $3 > 0$ and $\det \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} = 6 - 1 > 0$

so we need $\det A > 0$ as well.

In this case $\det A = 5x - 5 \cdot 15 = 5(x - 15)$

so A is positive definite $\iff x > 15$.

<u>Proof: ① - ② equivalent.</u> We need $\vec{x}^t A \vec{x}$ positive $\forall \vec{x} \neq 0$.

Let $\vec{x} = B\vec{y}$ with B orthogonal and $B^t A B = B^{-1} A B = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$.

We get $\vec{x}^t A \vec{x} = \vec{y}^t (B^t A B) \vec{y} = \sum \lambda_i y_i^2$. This is positive $\forall \vec{y} \neq 0$ if and only if $\lambda_i > 0 \ \forall i$. ∎

Example 3. Let $A = \begin{bmatrix} 2 & x & 0 \\ x & x+4 & x-4 \\ 0 & x-4 & x+1 \end{bmatrix}$. In this case,

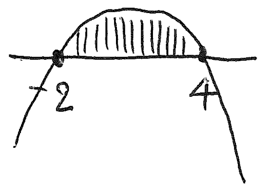
ⓐ $\det[2]$ ... is positive ✓

ⓑ $\det \begin{bmatrix} 2 & x \\ x & x+4 \end{bmatrix} = 2x + 8 - x^2$ should be positive.

Let's factor ... $2x+8-x^2 = -(x^2-2x-8)$

$$= -(x-4)(x+2)$$

$-2 < x < 4$



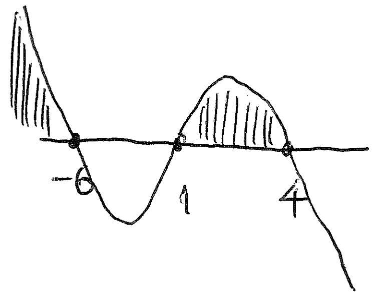
ⓒ $\det A > 0$

$$\det A = -x^3 - x^2 + 26x - 24$$

Roots are ... $x = 1, 4, -6$

So $\det A = -(x-1)(x-4)(x+6)$

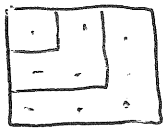
$x < -6$ or $1 < x < 4$



Conclusion: A pos. def. $\iff 1 < x < 4$.

Theorem (Tests for being positive definite)
If A is symmetric, then A is pos. def. if and only if

1. **Definition:** We have $\vec{x}^t A \vec{x} > 0$ for all $\vec{x} \neq 0$.
2. **Eigenvalues:** We have $\lambda > 0$ for each eigenvalue λ.
3. **Sylvester's criterion:** We have $\det A_k > 0$ for each $k \times k$ upper left submatrix A_k.



Proof of ① ⟺ ②. Since A is symmetric, there exists B orthogonal with $B^t A B = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$ and $\vec{x} = B\vec{y}$ gives $\vec{x}^t A \vec{x} = \vec{y}^t B^t A B \vec{y} = \sum \lambda_i y_i^2$. ▨

Proof of ① ⇒ ③. Suppose $\vec{x}^t A \vec{x} > 0$ for all $\vec{x} \neq 0$.
This means $\sum a_{ij} x_i x_j > 0$ for nonzero $\vec{x}$.
We look at the case $x_{k+1} = x_{k+2} = \dots = x_n = 0$.
We get $\sum_{i,j \leq k} a_{ij} x_i x_j > 0$ for nonzero $\vec{x}$.
This means $k \times k$ upper left submatrix is pos. def.
Its eigenvalues are positive by ② $\Rightarrow \det A_k > 0$. ▨

Proof of ③ ⇒ ①. Assume $\det A_k > 0$ $\forall k$. We need to show A pos. def.
We use induction on the size of A. If A is 1×1, $x^t A x = A x^2$ is a product of scalars and it's positive $\Leftrightarrow A > 0$. Suppose the result holds for $n \times n$. We prove it for $(n+1) \times (n+1)$ matrices.

Write $A = \left[\begin{array}{c|c} A_n & \vec{v} \\ \hline \vec{v}^t & \lambda \end{array}\right]$ for some $\vec{v} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

Change variables to eliminate $\vec{v}$. Take $B = \left[\begin{array}{c|c} I_n & \vec{x} \\ \hline 0 & 1 \end{array}\right]$ with the vector $\vec{x}$ to be found. Then

$$B^t A B = \left[\begin{array}{c|c} I_n & 0 \\ \hline x^t & 1 \end{array}\right] \left[\begin{array}{c|c} A_n & v \\ \hline v^t & \lambda \end{array}\right] \left[\begin{array}{c|c} I_n & x \\ \hline 0 & 1 \end{array}\right] = \left[\begin{array}{c|c} A_n & \vec{v} \\ \hline * & * \end{array}\right] \left[\begin{array}{c|c} I_n & x \\ \hline 0 & 1 \end{array}\right]$$

$$= \left[\begin{array}{c|c} A_n & A_n\vec{x}+\vec{v} \\ \hline * & * \end{array}\right]$$

so we can make $A_n\vec{x}+\vec{v}=0$ by taking $\vec{x} = -A_n^{-1}\vec{v}$. We can do this since $\det A_n > 0$.

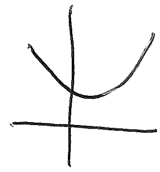
Then $B^t AB = \left[\begin{array}{c|c} A_n & 0 \\ \hline 0 & \mu \end{array}\right]$ for some scalar μ. However,

$$\det(B^tAB) = \mu \cdot \det A_n \Rightarrow \cancel{(\det B)}^{2\,\nearrow 1}(\det A) = \mu \cdot \det A_n \Rightarrow \mu > 0$$

so
$$\vec{x}^t \underline{\underline{A}}\vec{x} = (B\vec{y})^t \underline{\underline{A}}(B\vec{y}) = \vec{y}^t \left[\begin{array}{c|c} A_n & 0 \\ \hline 0 & \mu \end{array}\right] \vec{y}$$

$$= \left[\, y_{up}^t \mid y_{n+1} \right] \begin{bmatrix} A_n & 0 \\ 0 & \mu \end{bmatrix} \left[\begin{array}{c} y_{up} \\ \hline y_{n+1} \end{array}\right] = y_{up}^t A_n y_{up} + \mu y_{n+1}^2 > 0. \quad ▨$$

Application 1. Min of quadratic functions



Consider $\underline{f(x) = ax^2+bx+c}$ with $a>0$. By calculus, a min exists and the min occurs when $2ax+b=0$, namely $x = -b/2a$.

The analogue for n variables would be:

$$f(x_1, x_2, \ldots, x_n) = \sum_{i,j} a_{ij} x_i x_j + \sum_k b_k x_k + c$$

or
$$f(\vec{x}) = \underline{\underline{\vec{x}^t A \vec{x} + \vec{b}^t \vec{x} + c}}$$

with $\vec{b}$ = vector & c = scalar.

Theorem. Suppose A is positive definite. Then $f(\vec{x}) = \vec{x}^t A\vec{x} + \vec{b}^t\vec{x} + c$ has a min that occurs when $\vec{x} = -\frac{1}{2}A^{-1}\vec{b}$.

Proof. We claim that $f(\vec{x}_0)$ is a minimum, $\vec{x}_0 = -\frac{1}{2}A^{-1}\vec{b}$.

Here
$$f(\vec{x}_0) = x_0^t A x_0 + b^t x_0 + c$$
$$= x_0^t\left(-\tfrac{1}{2}b\right) + b^t x_0 + c$$

with $x_0^t b = b^t x_0 = -2x_0^t A x_0$. This gives

$$f(\vec{x}_0) = -\frac{1}{2} x_0^t b + b^t x_0 + c = -x_0^t A x_0 + c.$$

We then compute

$$\underbrace{f(x)} - f(x_0) = \underbrace{x^t A x + b^t x \cancel{+c}} + x_0^t A x_0 \cancel{-c}$$

$$= x^t A x - x_0^t A x - x_0^t A x_0 + x_0^t A x_0$$

$$= (x - x_0)^t A (x - x_0) \geq 0$$

with equality if and only if $x = x_0$. Here, the middle equality holds because $x_0^t A x$ is 1×1 so

$$x_0^t A x = (x_0^t A x)^t = x^t A x_0 = -\frac{1}{2} x^t b = -\frac{1}{2} b^t x. \quad \blacksquare$$

Application 2	Min/max of quadratics over the unit sphere $\|x\| = 1$.

Suppose A is symmetric and consider $f(\vec{x}) = \vec{x}^t A \vec{x}$. Then its min/max values over the sphere $\|x\| = 1$ are the min/max <u>eigenvalues</u> of A and those are attained along <u>eigenvectors</u>.

<u>Proof</u>. We change variables as usual, $\vec{x} = B\vec{y}$ with $B^t A B = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$.
Order these as $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ for simplicity. Then

$$f(\vec{x}) = x^t \underline{\underline{A}} x = y^t (\underline{\underline{B^t A B}}) y = \sum_{i=1}^n \lambda_i y_i^2 \quad \text{and} \quad 1 = \|x\| = \|By\| = \|y\|.$$

This implies

$$f(x) = \sum \lambda_i y_i^2 \leq \sum \lambda_n y_i^2 = \lambda_n$$

and

$$f(x) = \sum \lambda_i y_i^2 \geq \sum \lambda_1 y_i^2 = \lambda_1.$$

Moreover, equality holds when $\vec{y} = \vec{e}_n$ and $x = B\vec{e}_n =$ nth eigenvector
or $\vec{y} = \vec{e}_1$ and $\vec{x} = \vec{v}_1$, respectively. $\blacksquare$

<u>Example</u>. Min/max of $\underline{f(x,y) = 3x^2 + 4xy + 6y^2}$, over $x^2 + y^2 = 1$.
This is $f(x,y) = \vec{x}^t A \vec{x}$ with $A = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}$. Eigenvectors are
$v_1 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$ with $\lambda_1 = 2$ and $v_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ with $\lambda_2 = 7$.

The min value is $f\left(\frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}}\right) = 3 \cdot \frac{4}{5} - \frac{8}{5} + \frac{6}{5} = 2 = \lambda_1$.

The max value is $f\left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) = 3 \cdot \frac{1}{5} + \frac{8}{5} + \frac{24}{5} = 7 = \lambda_2$.

Application 3. | Second derivative test

- Consider $f(x_1, \ldots, x_n)$, a function of n variables.

Its directional derivative is

$$D_u f = \text{rate at which } f \text{ changes in the direction of (unit) vector } \vec{u}$$
$$= \nabla f \cdot \vec{u} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot u_i$$

and the second derivative is

$$D_u D_u f = \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left[D_u f\right] \cdot u_i$$
$$= \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left[\sum_{j=1}^{n} \frac{\partial f}{\partial x_j} u_j\right] \cdot u_i = \sum_{i,j} u_i u_j \boxed{\frac{\partial^2 f}{\partial x_i \partial x_j}}$$
$$= \vec{u}^t A \vec{u}$$

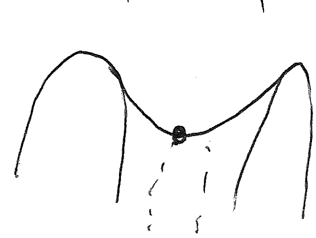
with A = Hessian matrix and $a_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

- We can classify critical points as follows.

1. If A has pos. eigenvalues, $u^t A u \geq 0$ for all $\vec{u}$
   so $D_u D_u f \geq 0$ for all $\vec{u}$
   so f is convex in u-direction
   so we have a local min.

2. If A has neg. eigenvalues, we get a local max

3. If A has some pos. & some neg. eigenvalues,
   we get a saddle point.



When $n=2$, the matrix is $A = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$.