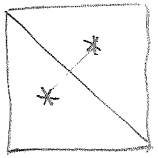
Linear algebra continued.

$H \in \mathbb{R}^{n \times n}$ is called <u>symmetric</u> if $H = H^T$.



A symmetric H is called <u>positive semidefinite</u> (p.s.d.) if

$\forall\, x \in \mathbb{R}^n$, $x^T H x \geqslant 0$.

A symmetric H is called <u>positive definite</u> (p.d.) if

$\forall\, x \in \mathbb{R}^n$ s.t. $x \neq 0$, $x^T H x > 0$.

Every p.d. H is p.s.d.: if $x = 0$, $x^T H x = 0^T H 0 = 0 \geqslant 0$

if $x \neq 0$, $x^T H x > 0$.

Every p.d. H is invertible: if not, then there is a nonzero z s.t.

$Hz = 0$. But then $z^T \underbrace{Hz}_{\substack{\| \\ 0}} = 0 \not> 0$.

Examples

$$H = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$$

$$\begin{aligned} x^T H x &= x_1^2 + 5x_2^2 + 4x_1x_2 \\ &= (x_1 + 2x_2)^2 + x_2^2 \\ &\geq 0 \end{aligned}$$

$$(= 0 \Rightarrow \left\{ \begin{array}{l} x_2 = 0 \\ x_1 + 2x_2 = 0 \end{array} \right\} \Rightarrow x_1 = x_2 = 0)$$

So H is p.d.

(and p.s.d.)

$$H = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$\begin{aligned} x^T H x &= x_1^2 + 4x_2^2 + 4x_1x_2 \\ &= (x_1 + 2x_2)^2 \\ &\geq 0 \end{aligned}$$

$x = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \neq 0$, but

$x^T H x = (-2 + 2(1))^2 = 0^2 = 0.$

So H is p.s.d,

but not p.d

$$H = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$$

$$\begin{aligned} x^T H x &= x_1^2 + 3x_2^2 + 4x_1x_2 \\ &= (x_1 + 2x_2)^2 - x_2^2 . \end{aligned}$$

Take $x = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$.

Then

$$\begin{aligned} x^T H x &= (-2 + 2 \cdot 1)^2 - 1^2 \\ &= 0^2 - 1^2 \\ &= -1 < 0 . \end{aligned}$$

So H is <u>not</u> p.s.d.

Is there a systematic algorithm for checking if a symmetric H is p.s.d. or not?

For diagonal matrices, this is easy. $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$. Then $x^TDx = \lambda_1 x_1^2 + \dots + \lambda_n x_n^2$.

<u>Claim</u>. D is p.s.d. $\Leftrightarrow$ all λ_ks are ≥ 0

D p.s.d. $\Rightarrow$ $x^TDx \geq 0$ $\forall x$, In particular $x = e_k$ gives $\lambda_k \geq 0$ $(k = 1, \dots, n)$.

$\lambda_k \geq 0$ $\forall k$ $\Rightarrow$ $\underbrace{\lambda_1 x_1^2}_{\substack{\vee\!\!/ \\ 0}} + \dots + \underbrace{\lambda_n x_n^2}_{\substack{\vee\!\!/ \\ 0}} \geq 0$ i.e., $x^TDx \geq 0$ $\forall x$, i.e., D is p.s.d.

<u>Claim</u>. D is p.d. $\Leftrightarrow$ all λ_ks are > 0.

D p.d $\Rightarrow$ $x^TDx > 0$ $\forall$ non zero x. Put $x = e_k$ to get $\lambda_k > 0$ $(k = 1, \dots, n)$.

$\lambda_k > 0$ $\forall k$ $\Rightarrow$ $x^TDx = \underset{\substack{\vee \\ 0}}{\lambda_1} x_1^2 + \dots + \underset{\substack{\vee \\ 0}}{\lambda_n} x_n^2 > 0$ $\forall$ nonzero x. $\Rightarrow$ D is p.d

Also, if P is invertible, then

PDP^T is p.sd. $\Leftrightarrow$ D is p.s.d.

PDP^T is p.d. $\Leftrightarrow$ D is p.d.

D p.d. $\Rightarrow$ $y^T D y > 0 \quad \forall\ y \neq 0$.

If $x \neq 0$, then $y := P^T x \neq 0$. ($\because$ P, and so also P^T, are invertible).

So $(P^T x)^T D\, P^T x > 0$ i.e., $x^T (PDP^T)\, x > 0$.

So PDP^T is p.d.

PDP^T p.d. $\Rightarrow$ $x^T PDP^T x > 0 \quad \forall\ x \neq 0$.

Let $y \neq 0$. Then $\exists!\ x \neq 0$ s.t. $P^T x = y$. ($\because$ P^T is inv.)

Thus $(P^T x)^T D P^T x = y^T D y > 0$.

So D is p.d.

How do these observations help us?

For a general symmetric H, if we have a procedure to find an inv. P and a diagonal D s.t. $H = PDP^T$,

then we have
- H is p.s.d. iff all entries of D are ≥ 0
- H is p.d. iff all entries of D are > 0.

LDL^T factorization

Example

$$H = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ 3 & 7 & 11 \end{bmatrix} . \qquad E_1 H = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix} \qquad H E_1^T = \begin{bmatrix} 1 & 0 & 0 \\ 2 & \multicolumn{2}{c}{\boxed{\tilde{H}^T}} \\ 3 & & \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}}_{\tilde{H}}$$

$$E_1 H E_1^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \multicolumn{2}{c}{\boxed{\tilde{H}^T}} \\ 0 & & \end{bmatrix} \qquad E_1 H E_1^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \multicolumn{2}{c}{\boxed{\tilde{H}}} \\ 0 & & \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

$$E_2 E_1 H E_1^T E_2^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \boxed{1} \end{bmatrix} \qquad 1 > 0 .$$

So H is p.d.

$$H = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 11 \end{bmatrix} \qquad E_1 H E_1^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & ⓪ & 0 \\ 0 & 0 & 2 \end{bmatrix} \qquad \begin{matrix} 1 \geq 0 \\ 0 \geq 0 \\ 2 \geq 0 \end{matrix}$$

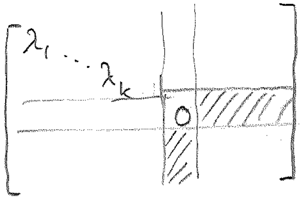
H is p.s.d. and not p.d.

If at some step you find

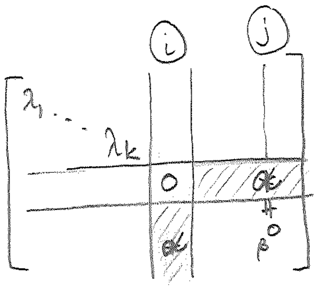
$$PHP^T = \underbrace{\begin{bmatrix} \lambda_1 \cdots \lambda_k & & \\ & 0 & * \quad * \\ & * & \\ & * & \boxed{\tilde{H}} \end{bmatrix}}_{D}$$

, then H is not p.d., since $e_k^T D e_k = 0$ and $e_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow ⓚ \neq 0$.

If at some step you find

$PHP^T =$  with the shaded part containing a nonzero entry,

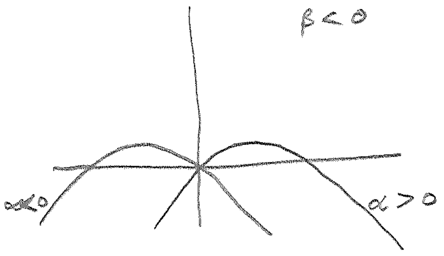
then H is not p.s.d.

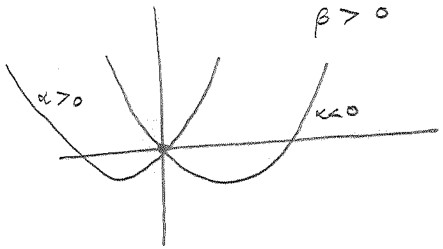
<u>Reason</u>: Suppose that $D :=$ 

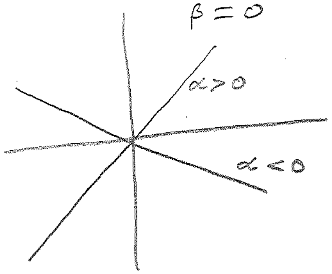
Then with $x := e_i + te_j$, we have

$$x^T D x = \begin{bmatrix} 1 & t \end{bmatrix} \begin{bmatrix} 0 & \alpha \\ \alpha & \beta \end{bmatrix} \begin{bmatrix} 1 \\ t \end{bmatrix}$$

$$= \begin{bmatrix} 1 & t \end{bmatrix} \begin{bmatrix} \alpha t \\ \alpha + t\beta \end{bmatrix} = 2\alpha t + t^2\beta = t(2\alpha + t\beta)$$







Quadratic optimization

Quadratic function. $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ having the form

$$f(x) = \frac{1}{2} x^T H x + c^T x + c_0 \qquad (x \in \mathbb{R}^n)$$

where $H = H^T \in \mathbb{R}^{n \times n}$

$c \in \mathbb{R}^n$

$c_0 \in \mathbb{R}$.

Example

(1) $n = 1$ $\quad f(x) = Ax^2 + Bx + C \; = \; \frac{1}{2} x \, (2A) \, x + Bx + C$

$$\left(H = 2A \; ; \quad c = B \; ; \quad c_0 = C \right)$$

(2) $\quad f(x_1, x_2, x_3) = -3 x_1 x_2 + x_3 + 1$

$$= \frac{1}{2} \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \underbrace{\begin{bmatrix} 0 & -3 & 0 \\ -3 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{H :=} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}}_{c^T :=} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \underbrace{1}_{c_0 :=}$$

$$= \frac{1}{2} x^T H x + c^T x + c_0 .$$

Quadratic optimization

No constraints.

$$\begin{cases} \text{minimite} & \frac{1}{2}x^T H x + c^T x + c_0 \\ \text{s.t.} & \boxed{x \in \mathbb{R}^n} \end{cases}$$

Equality constraints

$$\begin{cases} \text{minimize} & \frac{1}{2}x^T H x + c^T x + c_0 \\ \text{s.t.} & \boxed{Ax = b} \end{cases}$$

Inequality constraints

$$\begin{cases} \text{minimize} & \frac{1}{2}x^T H x + c^T x + c_0 \\ \text{s.t.} & \boxed{Ax \leq b} \end{cases}$$

(study this later
when we learn the
KKT – conditions for
solving nonlinear
problems)

Let $f(x) := \frac{1}{2} x^T H x + c^T x + c_0 \quad (x \in \mathbb{R}^n) \quad$ with $\quad H = H^T$.

Theorem. If $x, y \in \mathbb{R}^n$, then

$$f(y) = f(x) + (Hx + c)^T (y - x) + \frac{1}{2} (y-x)^T H (y-x).$$

(This allows us to compare values $f(y), f(x)$ at two points $x, y \in \mathbb{R}^n$:

$$f(y) - f(x) = (Hx + c)^T (y - x) + \frac{1}{2} (y-x)^T H (y-x).\ )$$

Proof.

$$\begin{aligned} f(y) &= \frac{1}{2} y^T H y + c^T y + c_0 = \frac{1}{2} (\boxed{x} + \underline{y-x})^T H (\boxed{x} + \underline{y-x}) + c^T (\boxed{x} + \underline{y-x}) + \boxed{c_0}. \\ &= \boxed{\frac{1}{2} x^T H x + c^T x + c_0} + \frac{1}{2} (\underline{y-x})^T H (\underline{y-x}) + \underline{\frac{1}{2} x^T H (\underline{y-x})} + \underline{\frac{1}{2} (\underline{y-x})^T H x} + c^T (\underline{y-x}) \\ &= \boxed{f(x)} + \underline{x^T H (y-x)} + c^T (y-x) + \frac{1}{2} (y-x)^T H (y-x) \\ &= f(x) + (Hx + c)^T (y-x) + \frac{1}{2} (y-x)^T H (y-x). \end{aligned}$$

□

Remark. For the quadratic function f above, the gradient of f at x, namely

$$\nabla f(x) := \left[\frac{\partial f}{\partial x_1}(x) \quad \cdots \quad \frac{\partial f}{\partial x_n}(x) \right]$$

turns out to be $\nabla f(x) = (Hx + c)^T$.

And the Hessian of f at x, namely the square matrix

$$F(x) := \begin{array}{cc} & \textcircled{j} \\ \textcircled{i} & \begin{bmatrix} & & \\ & \frac{\partial^2 f(x)}{\partial x_i \partial x_j} & \\ & & \end{bmatrix} \end{array}$$

turns out to be $F(x) = H$.

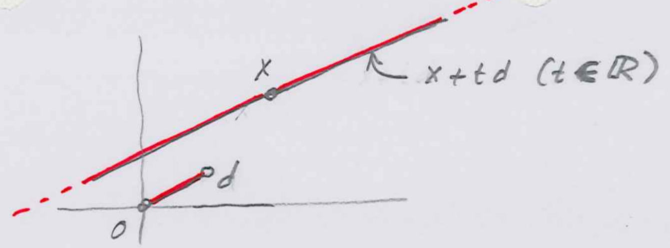
So the above result says:

$$f(y) = f(x) + \nabla f(x)\,(y-x) + \frac{1}{2}(y-x)^T F(x)\,(y-x),$$

which is the usual Taylor expansion. of f around x.

Note that since the Hessian is a constant, all derivatives of f of order $\geqslant 3$ are 0.

Useful consequence.



How do the values of f compare with $f(x)$ along a line passing through x?

<u>Corollary</u>. Let $x, d \in \mathbb{R}^n$, $t \in \mathbb{R}$. For our quadratic function f,

$$f(x+td) = \boxed{f(x)} + t\,\boxed{(Hx+c)^T d} + \boxed{\frac{1}{2}}\,t^2\,\boxed{d^T H d}.$$

(Just take $y = x + td$).

So in the direction d, $t \longmapsto f(x+td)$ is simply a quadratic function of the variable t.

Q. How do we solve $\left\{\begin{array}{ll}\text{minimize} & \frac{1}{2}x^THx + c^Tx + c_0 \\ \text{s.t.} & x \in \mathbb{R}^n\end{array}\right\}$?

A. 1° If H is not p.s.d., then there is no optimal solution

2° If H is p.s.d., then an optimal solution exists <u>if and only if</u> $-c \in \operatorname{ran} H$, and is given by any $\hat{x}$ satisfying $H\hat{x} = -c$.

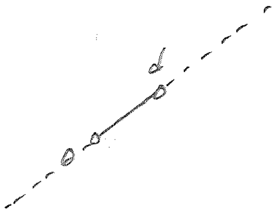
1°. Theorem.

If H is not p.s.d., then there is no optimal solution to

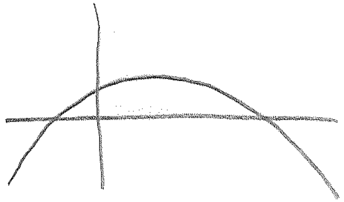
$$(Q): \left\{\begin{array}{ll}\text{minimize} & \frac{1}{2}x^THx + c^Tx + c_0 \\ \text{s.t.} & x \in \mathbb{R}^n .\end{array}\right.$$

Proof Since H is not p.s.d., there is a $d \in \mathbb{R}^n$ s.t. $d^THd < 0$.

Now look at the values of f along td, where $t \in \mathbb{R}$.



$$f(td) = f(0) + (H0+c)^T(td) + \frac{1}{2}t^2d^THd$$
$$= c_0 + tc^Td + \frac{1}{2}t^2\underbrace{d^THd}_{\wedge \atop 0}$$
$$\xrightarrow{t\to\infty} -\infty$$



So $\{f(x) : x \in \mathbb{R}^n\}$ is not bounded below. So (Q) has no opt. solution. □

Example. $(Q): \begin{cases} \text{minimize} & x_1^2 + x_2^2 + 3x_1x_2 \\ \text{s.t.} & x \in \mathbb{R}^2 \end{cases}$

$$x_1^2 + x_2^2 + 3x_1x_2 = \frac{1}{2} x^T \begin{bmatrix} 2 & +3 \\ +3 & -2 \end{bmatrix} x$$

$$H = \begin{bmatrix} 2 & 3 \\ 3 & +2 \end{bmatrix} \qquad E_1 H E_1^T = \begin{bmatrix} 2 & 0 \\ 0 & +2 - \frac{3}{2} \cdot 3 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -\frac{5}{2} \end{bmatrix} \text{ is not p.s.d.}$$

So by the above result (Q) has no optimal solution.

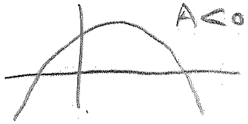
(**Remark:** Note that with $d := E_1^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -3/2 & 1 \end{bmatrix}^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & -3/2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/2 \\ 1 \end{bmatrix}$

we have $d^T H d = \begin{bmatrix} -3/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -3/2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -5/2 \end{bmatrix} \begin{bmatrix} -3/2 \\ 1 \end{bmatrix} = -5/2 < 0$

Then
$$\begin{aligned} f(td) = f\left(\begin{bmatrix} -3/2 t \\ t \end{bmatrix} \right) &= \frac{9}{4} t^2 + t^2 + 3 \cdot \left(-\frac{3}{2} t\right)(t) \\ &= \frac{9}{4} t^2 + t^2 - \frac{9}{2} t^2 \\ &= -\frac{5}{2} t^2 \\ &\xrightarrow{t \to \infty} -\infty. \end{aligned}$$

So the problem has no optimal solution).

Example $(\tilde{Q}): \left\{ \begin{array}{ll} \text{minimize} & Ax^2 + Bx + C \\ \text{s.t.} & x \in \mathbb{R} \end{array} \right\}$ If $A < 0$, then $(\tilde{Q})$ has no optimal solution:



What if H is p.s.d?

2° **Theorem** Let H be p.s.d. Then:

an optimal solution to

$$(Q): \begin{cases} \text{minimize} & \frac{1}{2}x^T H x + c^T x + c_0 \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

exists if and only if $\boxed{-c \in \operatorname{ran} H}$.

If $-c \in \operatorname{ran} H$, then any $\hat{x}$ satisfying $H\hat{x} = -c$ is an optimal solution.

Proof. **If**: Let $-c \in \operatorname{ran} H$. Then $\exists\, \hat{x}$ s.t. $H\hat{x} = -c$, i.e., $H\hat{x} + c = 0$.

Want: $\hat{x}$ is optimal for (Q).

Let $x \in \mathbb{R}^n$. Then

$$f(x) = f(\hat{x}) + (\underbrace{H\hat{x} + c}_{=0})^T (x - \hat{x}) + \frac{1}{2}(x - \hat{x})^T H (x - \hat{x})$$

$$= f(\hat{x}) + \frac{1}{2} \underbrace{(x - \hat{x})^T H (x - \hat{x})}_{\geq 0 \quad (\because H \text{ is p.s.d})}$$

$$\geq f(\hat{x}).$$

Done!

<u>Only if</u>: Suppose $-c \notin \operatorname{ran} H$.

Take $S = \operatorname{ran} H$ (subspace of $\mathbb{R}^n$).

Then $S^\perp = (\operatorname{ran} H)^\perp = \ker(H^T) = \ker H$. $(\because H = H^T)$

Every vector in $\mathbb{R}^n$ has a decomposition into a sum of vectors from S and $S^\perp$.

In particular, $-c = \underset{\underset{\operatorname{ran} H}{\overset{\|}{S}}}{\overset{z}{\in}} + \underset{\underset{\ker H}{\overset{\|}{S^\perp}}}{\overset{d}{\in}}$, with $\underbrace{d \neq 0}$. otherwise $-c$ would belong to $\operatorname{ran} H$!

Then $f(td) = f(0) + (H0 + c)^T (td) + \frac{1}{2} t^2 d^T \underbrace{Hd}_{=0}$

$= c_0 + t\, c^T d$

Can $c^T d = 0$? <u>Answer</u>: No, since $-c^T d = (z+d)^T d = \underbrace{z^T d}_{\overset{\|}{0}} + d^T d > 0$.

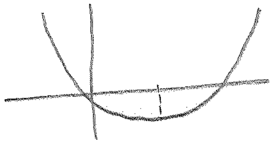
So $c^T d < 0$.

Thus $f(td) = c_0 + t c^T d \xrightarrow{t \to +\infty} -\infty$.

So $\{ f(x) : x \in \mathbb{R}^n \}$ is not bounded below. Hence (Q) has no optimal solution. $\square$

<u>Example</u> (1) $n=1$. $f(x) = A x^2 + B x + C$, with $H = 2A \geqslant 0$.

<u>1°</u> $A > 0$.



Unique optimal solution given by

$$H\hat{x} = -c$$

i.e., $2A\hat{x} = -B$

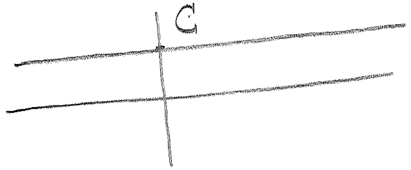
i.e., $\hat{x} = -\frac{B}{2A}$.

<u>2°</u> $A = 0$. Then $H\hat{x} = -c$

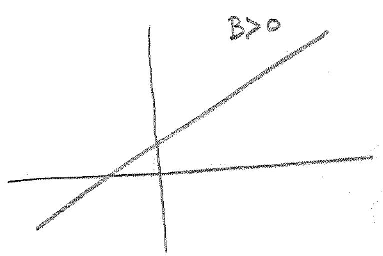
becomes $2A\hat{x} = -B$

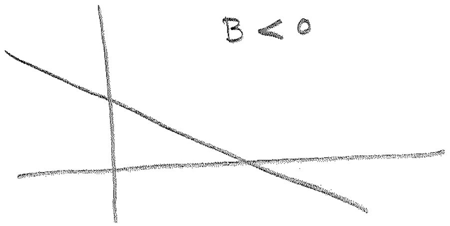
i.e., $0 = -B$.

If $B = 0$, then every $\hat{x} \in \mathbb{R}$ is an optimal solution:



If $B \neq 0$, then there does not exist an optimal solution





8.17

(2) $$\begin{cases} \text{minimize} & x_1^2 + x_2^2 - x_1 x_2 \\ \text{s.t.} & x_1, x_2 \in \mathbb{R} \end{cases}$$

$$x_1^2 + x_2^2 - x_1 x_2 = \frac{1}{2} x^T \underbrace{\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}}_{H} x \,.$$

$$H = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \qquad E_1 H E_1^T = \begin{bmatrix} 2 & 0 \\ 0 & 2 + \frac{1}{2}(-1) \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & \frac{3}{2} \end{bmatrix}.$$

So H is positive definite.

$-c = 0$ belongs to ran H, since $H0 = 0$.

Also $\hat{x} = 0$ is the unique solution to $H\hat{x} = -c$, since H is invertible.

So the unique optimal solution is given by $\hat{x} = 0$.

(Note that $x_1^2 + x_2^2 - x_1 x_2 = \frac{1}{2} x_1^2 + \frac{1}{2} x_2^2 + \left(\frac{x_1 - x_2}{\sqrt{2}}\right)^2 \geqslant 0$

with equality if and only if $x_1 = x_2 = 0$.)

Summary

$$(Q): \begin{cases} \text{minimize} & f(x) := \frac{1}{2} x^T H x + c^T x + c_0 \quad (\text{where } H = H^T) \\ \text{s.t.} & x \in \mathbb{R}^n \end{cases}$$

H	$\{ f(x) : x \in \mathbb{R}^n \}$	(Q)
not p.s.d.	not bounded below	no optimal solution
p.s.d.	bounded below $\Longleftrightarrow$ $-c \in \text{ran}\, H$	$\exists$ optimal solution $\Longleftrightarrow$ $-c \in \text{ran}\, H$. If $-c \in \text{ran}\, H$, then $\hat{x}$ s.t. $H\hat{x} = -c$ is optimal.
p.d.	bounded below	$\exists!$ optimal solution $\hat{x}$ given by $\hat{x} = -H^{-1}c$.