3 Multidimensional search

(P) $\underset{x \in \mathbb{R}^n}{\text{minimize}} \; f(x)$

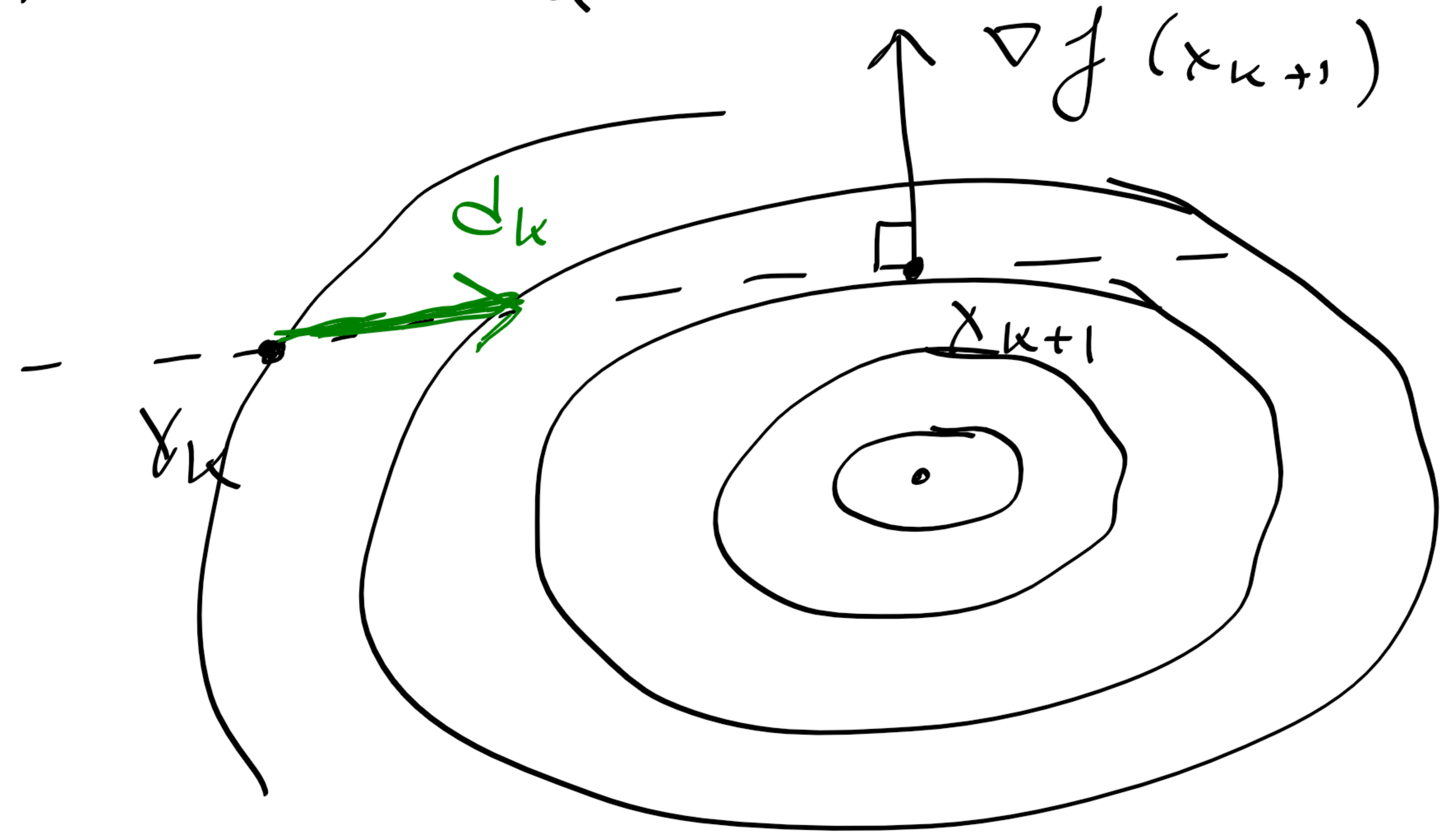
Prototype algorithm

- given x_k, choose search direction d_k
- line search: minimize $F(\lambda) = f(x_k + \lambda d_k)$ exact or inexact to define λ_k
- set $x_{k+1} = x_k + \lambda_k d_k$

Lemma 1: Exact line search for $f \in C^1$ implies that $\nabla f(x_{k+1})^T d_k = 0$.

Proof: $F'(\lambda) = \nabla f(x_k + \lambda d_k)^T d_k \implies$

$0 = F'(\lambda_k) = \nabla f(x_{k+1})^T d_k$ #



Exact line search for a quadratic function

$$q(x) = \frac{1}{2} x^T H x + c^T x \quad \text{with } H \text{ pos. def.}$$

$$d_k^T \nabla q(x_{k+1}) = 0 \quad (\text{Lemma } 1) \iff$$

$$d_k^T (H x_{k+1} + c) = 0 \iff$$

$$d_k^T (H(x_k + \lambda_k d_k) + c) = 0 \iff$$

$$d_k^T H x_k + \lambda_k d_k^T H d_k + d_k^T c = 0 \iff$$

$$\lambda_k = - \frac{d_k^T (H x_k + c)}{d_k^T H d_k} = - \frac{d_k^T \nabla q(x_k)}{d_k^T H d_k}$$

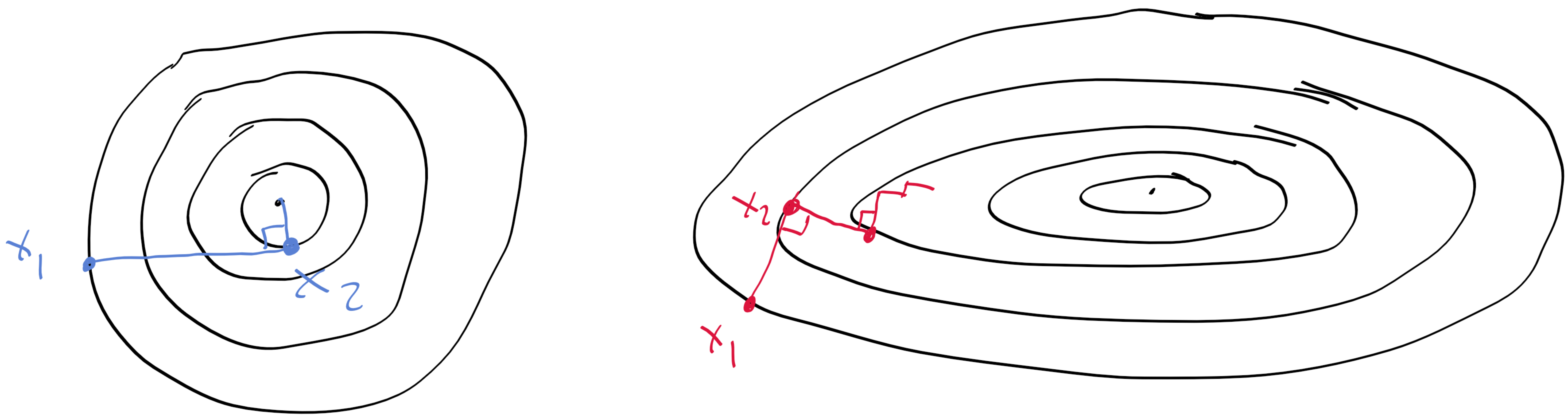
3.3 Steepest descent

search direction $d_k = -\nabla f(x_k)$

with exact line search

$$0 = \nabla f(x_{k+1})^T d_k = -d_{k+1}^T d_k$$

Thus, a zig-zagging behavior:



Property on a quadratic function with Hessian H with eigenvalues $0 < \lambda_{min} \leq \lambda_2 \leq \dots \leq \lambda_{max}$ and exact line search:

Thm 1: $|q(x_{k+1}) - q_{min}| \leq c\,|q(x_k) - q_{min}|$ with

$$c = \left(\frac{\lambda_{max} - \lambda_{min}}{\lambda_{max} + \lambda_{min}}\right)^2 = \left(\frac{\alpha - 1}{\alpha + 1}\right)^2 < 1$$

where the condition number $\alpha = \frac{\lambda_{max}}{\lambda_{min}}$

Fast convergence if $\alpha \approx 1$

Slow convergence if $\alpha \gg 1$

Appendix A. Some matrix theory

THEOREM 1. *Let* $\boldsymbol{A}$ *be a quadratic* $n \times n$ *matrix. The following conditions on* $\boldsymbol{A}$ *are equivalent.*

- *The columns of* $\boldsymbol{A}$ *form a basis for* $\mathbf{R}^n$.
- *The rows of* $\boldsymbol{A}$ *form a basis for* $\mathbf{R}^n$.
- *The rank of* $\boldsymbol{A}$ *is equal to* n.
- $\boldsymbol{A}$ *is invertible.*
- *The homogeneous system of equations* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ *has the zero solution only.*
- *The system of equations* $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ *has a unique solution for any* $\boldsymbol{b}$.
- $\det \boldsymbol{A} \neq 0$.

The spectral theorem for symmetric matrices:

THEOREM 2. *Let* $\boldsymbol{H}$ *be a symmetric matrix. Then*

a) there exists an orthogonal basis of eigenvectors of $\boldsymbol{H}$,

b) there is an orthogonal matrix $\boldsymbol{Q}$ *and a diagonal matrix* Λ *such that*

$$\boldsymbol{Q}^{-1}\boldsymbol{H}\boldsymbol{Q} = \boldsymbol{Q}^T\boldsymbol{H}\boldsymbol{Q} = \Lambda.$$

In $\boldsymbol{Q}$ *the columns are eigenvectors of* $\boldsymbol{H}$, *in* Λ *the diagonal elements are the eigenvalues.*

A.1 Positive definite matrices

DEFINITION 1. *Let* $\boldsymbol{H}$ *be a symmetric* $n \times n$ *matrix.* $\boldsymbol{H}$ *is called*[2]

positive definite *if* $\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} > 0$ *for all* $\boldsymbol{x} \neq \boldsymbol{0}$,

positive semidefinite *if* $\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} \geq 0$ *for all* $\boldsymbol{x}$,

indefinite *if* $\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} > 0$ *for some* $\boldsymbol{x} \in \mathbf{R}^n$ *and* $\boldsymbol{x}^T\boldsymbol{H}\boldsymbol{x} < 0$ *for some* $\boldsymbol{x} \in \mathbf{R}^n$,

negative definite *if* $-\boldsymbol{H}$ *is positive definite,*

negative semidefinite *if* $-\boldsymbol{H}$ *is positive semidefinite.*

Proposition. H pos. def. $\Rightarrow$ all diagonal elements are > 0.

Proof: Take $e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ ← position i. Then $0 < e_i^T H e_i = h_{ii}$ #

Thm 3: H pos. def. $\iff$ eigenvalues $\lambda_i > 0$, $i = 1, \dots, n$

H semidef $\iff$ — " — $\lambda_i \geq 0$ — " —

Proof: Given $x^T H x$, change variables according to the spectral thm: $x = Q\hat{x}$ with Q orthogonal. Then

$$x^T H x = \hat{x}^T Q^T H Q \hat{x} = \hat{x}^T \Lambda \hat{x} = \sum_{i=1}^{n} \lambda_i \hat{x}_i^2 \geq 0$$

Thus pos. semidef. proved.

H pos. def. $\iff$ $x^T H x > 0 \quad \forall x \neq 0 \iff$

$\lambda_i > 0 \quad \forall \hat{x} \neq 0 \iff$

$\lambda_i > 0 \quad \forall Q^{-1} x \neq 0 \iff$

$\lambda_i > 0 \quad \forall x \neq 0$ #

Cor. H pos. def. $\implies \det H = \lambda_1 \lambda_2 \dots \lambda_n > 0$

Let A_k denote the upper left $k \times k$ submatrix of A:

$$A = \begin{pmatrix} A_k & \vdots & \\ \dots & & \\ & & \end{pmatrix}$$

Lemma 1: H pos. def. $\implies \det H_k > 0$, $k = 1, \dots, n$

Proof: Given $x^T H x > 0 \quad \forall x \neq 0$. Take

$$x = \begin{pmatrix} \tilde{x} \\ 0 \end{pmatrix} \begin{matrix} \leftarrow \text{size } k \times 1 \\ \leftarrow \text{size } (n-k) \times 1 \end{matrix}$$

$0 < x^T H x = \tilde{x}^T H_k \tilde{x} \quad \forall \tilde{x} \neq 0 \iff$

H_k pos. def. $\implies \det H_k > 0$ #

Appendix A.2 LU factorization

Gaussian elimination:

$$Ax = b \iff \underbrace{\begin{pmatrix} * & \cdots & * \\ 0 & * & \\ \vdots & & \ddots \\ 0 & \cdots \; 0 & * \end{pmatrix}}_{U \text{ upper triangular}} x = \tilde{b}$$

If the system needs to be solved many times for different b, then we keep track of row operation in a matrix L:

$$Ax = b \iff L\underline{Ux} = b \iff \begin{cases} L\tilde{b} = b \\ Ux = \tilde{b} \end{cases}$$

THEOREM 4. *Assume that* $\det \mathbf{A}_k \neq 0$ *for all* k.

1. *There is a unique factorization* $\mathbf{A} = \mathbf{LU}$, *with* $\mathbf{L}$ *a lower triangular matrix with all diagonal elements equal to 1, and* $\mathbf{U}$ *an upper triangular matrix;*

$$\mathbf{L} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ l_{21} & 1 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ l_{n1} & l_{n2} & \cdots & 1 \end{pmatrix}, \qquad \mathbf{U} = \begin{pmatrix} d_1 & u_{12} & \cdots & u_{1n} \\ 0 & d_2 & \cdots & u_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}.$$

2. *When using Gaussian elimination to solve the linear system* $\mathbf{A}\boldsymbol{x} = \boldsymbol{b}$ *these matrices appear as follows:*

   - *The columns* $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_n$ *of* $\mathbf{L}$ *contain the multipliers needed in the elimination; specifically, column* $\boldsymbol{l}_j$ *contains those needed to eliminate* x_j *from rows* $j+1, \ldots, n$.
   - *The result of the eliminations is* $\mathbf{U}$.
   - *The diagonal elements in* $\mathbf{U}$ *(the* **pivot elements***) are*

$$d_k = \frac{\det \mathbf{A}_k}{\det \mathbf{A}_{k-1}}, \qquad k = 1, \ldots, n \tag{1}$$

   (where $\det \mathbf{A}_0$ *should be interpreted as 1).*

3. *In the special case of* $\mathbf{A}$ *being symmetric,* $\mathbf{U} = \mathbf{DL}^T$, *where* $\mathbf{D}$ *is the diagonal matrix whose diagonal entries are* $d_1, \ldots, d_n$. *Hence*

$$\mathbf{A} = \mathbf{LDL}^T$$

when $\mathbf{A}$ *is symmetric.*

$$A = \begin{pmatrix} ② & 3 & 4 \\ 4 & 9 & 7 \\ 2 & -3 & 5 \end{pmatrix} \longrightarrow \begin{pmatrix} 2 & 3 & 4 \\ 0 & \boxed{3} & -1 \\ 0 & -6 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 2 & 3 & 4 \\ 0 & 3 & -1 \\ 0 & 0 & \triangle{-1} \end{pmatrix} = U$$

row operations $r_1 := r_1$, $r_2 := r_2 - 2r_1$, $r_3 := r_3 - 1r_1$; $r_1 := r_1$, $r_2 := r_2$, $r_3 := r_3 - (-2)r_2$

Alternatively

$$A = \begin{pmatrix} 2 & 3 & 4 \\ 4 & 9 & 7 \\ 2 & -3 & 5 \end{pmatrix} \to \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & -1 \\ 0 & -6 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$r_1 := r_1 - r_1$, $r_2 := r_2 - 2r_1$, $r_3 := r_3 - 1r_1$; $r_1 := r_1 - 0r_2$, $r_2 := r_2 - r_2$, $r_3 := r_3 - (-2)r_2$

set $l_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ $\quad l_2 = \begin{pmatrix} 0 \\ 1 \\ -2 \end{pmatrix}$

$u_1^T = (2 \quad 3 \quad 4)$ $\quad u_2^T = (0 \quad 3 \quad -1)$

$l_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, $u_3^T = (0 \quad 0 \quad -1)$

$$A - l_1 u_1^T - l_2 u_2^T - l_3 u_3^T = 0 \iff$$

$$A - (l_1 \quad l_2 \quad l_3) \begin{pmatrix} u_1^T \\ u_2^T \\ u_3^T \end{pmatrix} = 0 \iff A - LU = 0$$

This can only be done if all pivot elements but the last one are nonzero: $d_1 = ②$, $d_2 = \boxed{3}$, $d_3 = \triangle{-1}$

We have $\det A_1 = d_1$

Since row operations don't change the determinant, we have

$$\det A_2 = \det \begin{pmatrix} d_1 & * \\ 0 & d_2 \end{pmatrix} = d_1 d_2 = d_2 \det A_1$$

$$\det A_3 = d_1 d_2 d_3 = d_3 \det A_2$$

$\vdots$

$$\det A_k = d_1 \cdots d_{k-1} d_k = d_k \det A_{k-1}$$

We can compose

$$U = \begin{pmatrix} d_1 & & * \\ 0 & \ddots & \\ \vdots & & \\ 0 & \cdots\, 0 & d_n \end{pmatrix} = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix} \begin{pmatrix} 1 & & * \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = DM$$

If A is symmetric, then we use uniqueness of LU factorization (see book):

$$LDM = \underbrace{L}\underline{U} = A = A^T = \underbrace{M^T}_{\text{lower}} \underbrace{DL^T}_{\text{upper}} \Rightarrow L = M^T$$

so $A = LDL^T$ #

(we may still have $d_n = 0$)

Appendix A.3

Thm 6. H pos. def. $\iff$ $H = LDL^T$ with L lower triangular with ones on the diagonal and $D = \text{diag}(d_1, \dots, d_n)$ with $d_i > 0 \;\; \forall i$

Proof: ($\Rightarrow$) Lemma 1 and Thm 4.

($\Leftarrow$) Take an arbitrary $x \neq 0$

$$x^T H x = x^T L D L^T x = (\underbrace{L^T x}_{\hat{x}})^T D (L^T x) = \hat{x}^T D \hat{x}$$

$$= \sum d_i \hat{x}_i^2 > 0 \quad \text{iff all } d_i > 0 \qquad \#$$

This is related to completing the squares:

$$\hat{x} = L^T x = \begin{pmatrix} 1 & & 0 \\ l_{21} & 1 & \\ \vdots & & \ddots \\ l_{n1} & \cdots & 1 \end{pmatrix}^T x = \begin{pmatrix} 1 & l_{21} & l_{n1} \\ & 1 & \\ 0 & & \ddots \quad 1 \end{pmatrix} x = \begin{pmatrix} x_1 + l_{21} x_2 + \dots + l_{n1} x_n \\ \vdots \\ x_n \end{pmatrix}$$

$$\hat{x}_1^2 = (x_1 + l_{21} x_2 + \dots + l_{n1} x_n)^2$$

See Ex. 4.

Thm 7 (Sylvester's criterion)

H pos. def. $\iff$ $\det H_k > 0$, $k = 1, \dots, n$

Proof: ($\Rightarrow$) Lemma 1 ($\Leftarrow$) Thm 4 gives

$$H = LDL^T \text{ with } d_k = \frac{\det H_k}{\det H_{k-1}} > 0$$

$\Rightarrow$ H pos. def. by Thm 6. $\qquad \#$

Prop. H pos. semidef. $\iff$ $H + \varepsilon I$ pos. def. $\forall \varepsilon > 0$

$\iff$ $\det (H + \varepsilon I)_k > 0 \quad \forall k \;\; \forall \varepsilon > 0$

$\Rightarrow$ $\det H_k \geq 0 \quad \forall k$

↑ converse is not true

Thm 8. $\begin{cases} \det H_k > 0, & k = 1, \dots, n-1 \\ \det H_n = \det H = 0 \end{cases}$

$\Rightarrow$ H pos. semidef.

Proof: Thm 4 $\Rightarrow$ $d_1, \dots, d_{n-1} > 0$ and $d_n = 0$.
Use Thm 6 to conclude the statement. #

Thm 9: (Cholesky factorization)

H pos. def. $\Leftrightarrow$ $H = \hat{L}\hat{L}^T$ where $\hat{L}$ is lower triangular with positive diagonal elements

Proof: ($\Rightarrow$) $H = LDL^T$ with $L = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ * & & 1 \end{pmatrix}$, $D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$ with all $d_i > 0$. Define $\sqrt{D} = \begin{pmatrix} \sqrt{d_1} & & 0 \\ & \ddots & \\ 0 & & \sqrt{d_n} \end{pmatrix}$. Then

$$H = LDL^T = L\sqrt{D}\sqrt{D}^T L^T = \underbrace{L\sqrt{D}}_{\hat{L}}(L\sqrt{D})^T$$

($\Leftarrow$) Exercise!