Lecture: Quadratic optimization

1. Positive definite och semidefinite matrices
2. $\mathbf{LDL}^\mathsf{T}$ factorization
3. Quadratic optimization without constraints
4. Quadratic optimization with constraints
5. Least-squares problems

Quadratic optimization without constraints

$$\text{minimize} \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{H}\mathbf{x} + \mathbf{c}^{\mathsf{T}}\mathbf{x} + c_0$$

where $\mathbf{H} = \mathbf{H}^{\mathsf{T}} \in \mathbf{R}^{n \times n}$, $\mathbf{c} \in \mathbf{R}^n$, $c_0 \in \mathbf{R}$.

- Common in applications, *e.g.*,
  - linear regression (fitting of models)
  - minimization of physically motivated objective functions such as minimization of energy, variance etc.
  - quadratic approximation of nonlinear optimization problems.

The quadratic term

Let

$$f(x) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$$

where $\mathbf{H} = \mathbf{H}^\mathsf{T} \in \mathbf{R}^{n\times n}$, $\mathbf{c} \in \mathbf{R}^n$, $c_0 \in \mathbf{R}$.

We can assume that the matrix $\mathbf{H}$ is symmetric.

If $\mathbf{H}$ is not symmetric it holds that

$$\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \frac{1}{2}(\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x})^\mathsf{T} = \frac{1}{2}\mathbf{x}^\mathsf{T}(\mathbf{H} + \mathbf{H}^\mathsf{T})\mathbf{x},$$

i.e., $\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} = \mathbf{x}^\mathsf{T}\tilde{\mathbf{H}}\mathbf{x}$ where $\tilde{\mathbf{H}} = \dfrac{1}{2}(\mathbf{H} + \mathbf{H}^\mathsf{T})$ is a symmetric matrix.

Positive definite and positive semi-definite matrices

Definition 1 (Positive semidefinite). *A symmetric matrix* $\mathbf{H} = \mathbf{H}^\mathsf{T} \in \mathbf{R}^{n\times n}$ *is called positive semidefinite if*

$$\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} \geq 0 \quad \textit{for all } \mathbf{x} \in \mathbf{R}^n$$

Definition 2 (Positive definite). *A symmetric matrix* $\mathbf{H} = \mathbf{H}^\mathsf{T} \in \mathbf{R}^{n\times n}$ *is called positive definite if*

$$\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} > 0 \quad \textit{for all } \mathbf{x} \in \mathbf{R}^n - \{\mathbf{0}\}$$

How do you determine if a symmetric matrix

$\mathbf{H} = \mathbf{H}^{\mathsf{T}} \in \mathbf{R}^{n\times n}$ is positive (semi-) definite:

- $\mathbf{LDL}^{\mathsf{T}}$ factorization (suitable for calculations by hand). Chapter 27.6 in the book.
- Use the spectral factorization of the matrix $\mathbf{H} = \mathbf{Q}\Lambda\mathbf{Q}^{\mathsf{T}}$, where $\mathbf{Q}^{\mathsf{T}}\mathbf{Q} = \mathbf{I}$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ where λ_k are the eigenvalues of the matrix. From this it follows that
  - $\mathbf{H}$ is positive semidefinite if and only if $\lambda_k \geq 0$, $k = 1, \ldots, n$
  - $\mathbf{H}$ is positive definite if and only if $\lambda_k > 0$, $k = 1, \ldots, n$

$\mathbf{LDL}^\mathsf{T}$ factorization

Theorem 1. *A symmetric matrix $\mathbf{H} = \mathbf{H}^\mathsf{T} \in \mathbf{R}^{n\times n}$ is positive semidefinite if and only if there is a factorization $\mathbf{H} = \mathbf{LDL}^\mathsf{T}$ where*

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 \\ \vdots & & & \ddots & \\ l_{n1} & l_{n2} & l_{n3} & \dots & 1 \end{bmatrix} \qquad \mathbf{D} = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & \dots & 0 \\ 0 & 0 & d_3 & \dots & 0 \\ \vdots & & & \ddots & \\ 0 & 0 & 0 & \dots & d_n \end{bmatrix}$$

and $d_k \geq 0$, $k = 1, \dots, n$.

$\mathbf{H}$ is positive definite if and only if $d_k > 0$, $k = 1, \dots, n$, in the $\mathbf{LDL}^\mathsf{T}$ factorization.

Proof: *See chapter 27.9 in the book.*

The factorization $\mathbf{LDL}^{\mathsf{T}}$ is determined by one of the following methods

- Elementary row and column operations,
- Completion of squares,

$\mathbf{LDL}^\mathsf{T}$ factorization - Example

Determine the $\mathbf{LDL}^\mathsf{T}$ factorization of

$$\mathbf{H} = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 4 & 9 & 17 & 22 \\ 6 & 17 & 44 & 61 \\ 8 & 22 & 61 & 118 \end{bmatrix}$$

The idea is now to perform symmetric row- and column-operations from the left and right in order to form the diagonal matrix $\mathbf{D}$.

The row-operations can be performed with lower triangular invertible matrices, and the product of lower triangular matrices is lower triangular and also the inverse.

$\mathrm{LDL}^{\mathrm{T}}$ factorization - Example

Perform row-operations to get zeros in the first column:

$$\ell_1 \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ -4 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 6 & 8 \\ 4 & 9 & 17 & 22 \\ 6 & 17 & 44 & 61 \\ 8 & 22 & 61 & 118 \end{bmatrix} = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 0 & 1 & 5 & 6 \\ 0 & 5 & 26 & 37 \\ 0 & 6 & 37 & 86 \end{bmatrix}$$

and the corresponding column-operations:

$$\ell_1 \mathbf{H} \ell_1^T = \begin{bmatrix} 2 & 4 & 6 & 8 \\ 0 & 1 & 5 & 6 \\ 0 & 5 & 25 & 30 \\ 0 & 6 & 30 & 35 \end{bmatrix} \begin{bmatrix} 1 & -2 & -3 & -4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 5 & 6 \\ 0 & 5 & 26 & 37 \\ 0 & 6 & 37 & 86 \end{bmatrix}$$

$\mathrm{LDL}^{\mathrm{T}}$ factorization - Example

Perform row-operations to get zeros in the second column:

$$\ell_2\ell_1\mathbf{H}\ell_1^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -5 & 1 & 0 \\ 0 & -6 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 5 & 6 \\ 0 & 5 & 26 & 37 \\ 0 & 6 & 37 & 86 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 7 & 50 \end{bmatrix}$$

and the corresponding column-operations:

$$\ell_2\ell_1\mathbf{H}\ell_1^T\ell_2^T = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 7 & 50 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -5 & -6 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 7 & 50 \end{bmatrix}$$

LDL^T factorization - Example

Perform row-operations to get zeros in the third column:

$$\ell_3\ell_2\ell_1\mathbf{H}\ell_1^T\ell_2^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -7 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 7 & 50 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and the corresponding column-operations:

$$\ell_3\ell_2\ell_1\mathbf{H}\ell_1^T\ell_2^T\ell_3^T = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = D$$

$\mathrm{LDL}^\top$ factorization - Example

We finally got the diagonal matrix

$$D = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Since all elements on the diagonal of D are positive the matrix H is positive definite.

The values on the diagonal are not the eigenvalues of H, but they have the same sign as the eigenvalues, and that is all we need to know.

$\mathrm{LDL^T}$ factorization - Example

How do we determine the L matrix ?

$\ell_3\ell_2\ell_1\mathbf{H}\ell_1^T\ell_2^T\ell_3^T = D$, *i.e.*, $\mathbf{H} = LDL^T$ if $L = (\ell_3\ell_2\ell_1)^{-1} = \ell_1^{-1}\ell_2^{-1}\ell_3^{-1}$

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ -4 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -5 & 1 & 0 \\ 0 & -6 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -7 & 1 \end{bmatrix}^{-1}$$

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

Unbounded problems

Let $f(\mathbf{x}) = \dfrac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$.

Assume that $\mathbf{H}$ is not positive semidefinite. Then there is a $\mathbf{d} \in \mathbf{R}^n$ such that $\mathbf{d}^\mathsf{T}\mathbf{H}\mathbf{d} < 0$. It follows that

$$f(t\mathbf{d}) = \frac{1}{2}t^2\mathbf{d}^\mathsf{T}\mathbf{H}\mathbf{d} + t\mathbf{c}^\mathsf{T}\mathbf{d} + c_0 \to -\infty \text{ when } t \to \infty$$

Conclusion: A necessary condition for $f(\mathbf{x})$ to have a minimum is that $\mathbf{H}$ is positive semidefinite.

Assume that $\mathbf{H}$ is positive semidefinite. Then we can use the factorization $\mathbf{H} = \mathbf{L}\mathbf{D}\mathbf{L}^\mathsf{T}$ which gives

$$\begin{aligned}
&\text{minimize } \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0 \\
&\quad = \text{minimize } \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{L}\mathbf{D}\mathbf{L}^\mathsf{T}\mathbf{x} + \mathbf{c}^\mathsf{T}(\mathbf{L}^\mathsf{T})^{-1}\mathbf{L}^\mathsf{T}\mathbf{x} + c_0 \\
&\quad = \text{minimize } \sum_{k=1}^{n} \frac{1}{2}d_k y_k^2 + \tilde{c}_k y_k + c_0 \\
&\quad = c_0 + \sum_{k=1}^{n} \text{minimize } \frac{1}{2}d_k y_k^2 + \tilde{c}_k y_k
\end{aligned}$$

where we have introduced the new coordinates $y = \mathbf{L}^\mathsf{T}\mathbf{x}$ and $\tilde{c} = \mathbf{L}^{-1}\mathbf{c}$.

The optimization problem separates to n independent scalar quadratic optimization problems, which are easy to solve.

Scalar Quadratic optimization without constraints

The scalar quadratic optimization problem

$$\text{minimize}_{x \in \mathbf{R}} \quad \frac{1}{2}hx^2 + cx + c_0$$

has a finite solution, if and only if,

- $h = 0$ and $c = 0$, or
- $h > 0$.

In the first case there is no unique optimum.

In the second case $\frac{1}{2}hx^2 + cx + c_0 = \frac{1}{2}h(x + c/h)^2 + c_0 - c^2/(2h)$, and the optimum is achieved for $x = -c/h$.

Solving the minimization problems separately shows that

$\hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 & \dots & \hat{y}_n \end{bmatrix}^\mathsf{T}$ is a minimizer if

$(i)'$ $d_k \geq 0, \;\; k = 1, \dots, n$

$(ii)'$ $d_k \hat{y}_k = -\tilde{c}_k$

$\Leftrightarrow$

(i) $\mathbf{H}$ is positive semidefinite

(ii) $\mathbf{D}\mathbf{L}^\mathsf{T}\hat{\mathbf{x}} = -\mathbf{L}^{-1}\mathbf{c}$

Constraint (ii) can be replaced with $\mathbf{L}\mathbf{D}\mathbf{L}^\mathsf{T}\hat{\mathbf{x}} = -\mathbf{c}$ *i.e.*, $\mathbf{H}\hat{\mathbf{x}} = -\mathbf{c}$.

Furthermore, $f(\mathbf{x}) = \dfrac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$ is unbounded (from below) if

$(i)'$ some $d_k < 0$, or

$(ii)'$ one of the equations

$d_k \hat{y}_k = -\tilde{c}_k$

do not have a solution

i.e., $d_k = 0$ and $\tilde{c}_k \neq 0$

$\Leftrightarrow$

(i) $\mathbf{H}$ is not positive semi-definite

(ii) the equation $\mathbf{H}\hat{\mathbf{x}} = -\mathbf{c}$

do not have a solution

The above arguments can be formalized as the following theorem

Theorem 2. *Let* $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$. *Then* $\hat{\mathbf{x}}$ *is a minimizer if*

$$(i)\ \mathbf{H} \text{ is positive semidefinite}$$

$$(ii)\ \mathbf{H}\hat{\mathbf{x}} = -\mathbf{c}$$

If $\mathbf{H}$ *is positive definite then* $\hat{\mathbf{x}} = -\mathbf{H}^{-1}\mathbf{c}$.

If $\mathbf{H}$ *is not positive semidefinite or* $\mathbf{H}\mathbf{x} = -\mathbf{c}$ *do not have a solution, then* f *is not limited from below.*

Comment 1. *The condition that* $\mathbf{H}\mathbf{x} = -\mathbf{c}$ *has a solution is equivalent to* $\mathbf{c} \in \mathcal{R}(\mathbf{H})$. *Therefore,* f *is a minimizer if and only if* $\mathbf{H}$ *is positive semidefinite and* $\mathbf{c} \in \mathcal{R}(\mathbf{H})$.

Quadratic optimization with linear constraints

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{H}\mathbf{x} + \mathbf{c}^{\mathsf{T}}\mathbf{x} + c_0 \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b} \end{array} \tag{1}$$

where $\mathbf{H} = \mathbf{H}^{\mathsf{T}} \in \mathbf{R}^{n\times n}$, $\mathbf{A} \in \mathbf{R}^{m\times n}$, $\mathbf{c} \in \mathbf{R}^{n}$, and $\mathbf{b} \in \mathbf{R}^{m}$, and $c_0 \in \mathbf{R}$.

- We assume that $\mathbf{b} \in \mathcal{R}(\mathbf{A})$ and $n > m$, *i.e.*, $\mathcal{N}(\mathbf{A}) \neq \{\mathbf{0}\}$.
- Assume that $\mathcal{N}(A) = \text{span}\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$ and define the nullspace matrix $\mathbf{Z} = \begin{bmatrix}\mathbf{z}_1 & \dots & \mathbf{z}_k\end{bmatrix}$.
- If $\mathbf{A}\bar{\mathbf{x}} = \mathbf{b}$ it holds that an arbitrary solution to the linear constraint has the form $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Z}\mathbf{v}$, for some $\mathbf{v} \in \mathbf{R}^{k}$. This follows since

$$\mathbf{A}(\bar{\mathbf{x}} + \mathbf{Z}\mathbf{v}) = \mathbf{A}\bar{\mathbf{x}} + \mathbf{A}\mathbf{Z}\mathbf{v} = \mathbf{b} + \mathbf{0} = \mathbf{b}$$

With $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Z}\mathbf{v}$ inserted in $f(\mathbf{x}) = \dfrac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$ we get

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{v}^\mathsf{T}\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}\mathbf{v} + \left(\mathbf{Z}^\mathsf{T}(\mathbf{H}\bar{\mathbf{x}} + \mathbf{c})\right)^\mathsf{T}\mathbf{v} + f(\bar{\mathbf{x}}).$$

The minimization problem (1) is thus equivalent with

$$\text{minimize}\frac{1}{2}\mathbf{v}^\mathsf{T}\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}\mathbf{v} + \left(\mathbf{Z}^\mathsf{T}(\mathbf{H}\bar{\mathbf{x}} + \mathbf{c})\right)^\mathsf{T}\mathbf{v} + f(\bar{\mathbf{x}}).$$

We can now apply Theorem 2 to obtain the following result:

Theorem 3. $\hat{\mathbf{x}}$ *is an optimal solution to (1) if*

(i) $\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}$ *is positive semidefinite.*

(ii) $\hat{\mathbf{x}} = \bar{\mathbf{x}} + \mathbf{Z}\hat{\mathbf{v}}$ *where* $\mathbf{Z}^\mathsf{T}\mathbf{H}\mathbf{Z}\hat{\mathbf{v}} = -\mathbf{Z}^\mathsf{T}(\mathbf{H}\bar{\mathbf{x}} + \mathbf{c})$

Comment 2. *The second condition in the theorem is equivalent with the existence of a* $\hat{\mathbf{u}} \in \mathbf{R}^m$ *such that*

$$\begin{bmatrix} \mathbf{H} & -\mathbf{A}^\mathsf{T} \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} -\mathbf{c} \\ \mathbf{b} \end{bmatrix}$$

Proof: *The second constraint in the theorem can be written*

$$\mathbf{Z}^\mathsf{T}(\mathbf{H}(\bar{\mathbf{x}} + \mathbf{Z}\hat{\mathbf{v}}) + \mathbf{c}) = \mathbf{0}$$

Since $\mathcal{N}(\mathbf{A}) = \mathcal{R}(\mathbf{Z})$ *this is equivalent with*

$$\mathbf{H}(\underbrace{\bar{\mathbf{x}} + \mathbf{Z}\hat{\mathbf{v}}}_{\hat{\mathbf{x}}}) + \mathbf{c} \in \mathcal{N}(\mathbf{A})^\perp = \mathcal{R}(\mathbf{A}^\mathsf{T})$$

$$\Leftrightarrow \qquad \mathbf{H}\hat{\mathbf{x}} + \mathbf{c} = \mathbf{A}^\mathsf{T}\hat{\mathbf{u}}, \text{ for some } \hat{\mathbf{u}} \in \mathbf{R}^m$$

Linear constrained QP: Example

Let $f(x) = \frac{1}{2}x^T Hx + c^T x + c_0$, where

$$H = \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix}, \quad c = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad c_0 = 0.$$

Consider minimization of f over the feasible region $\mathfrak{F} = \{x \mid Ax = b\}$ where

$$A = \begin{bmatrix} 2 & 1 \end{bmatrix}, \quad b = 2.$$

I.e. the minimization problem

$$\begin{bmatrix} \text{min} & -x_1^2 + \frac{1}{2}x_2^2 + x_1 + x_2 \\ \text{s.t.} & 2x_1 + x_2 = 2 \end{bmatrix}$$

Linear constrained QP: Example - Nullspace method

The nullspace of A is spanned by Z and the reduced Hessian is given by

$$Z = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad \bar{H} = Z^T H Z = 2$$

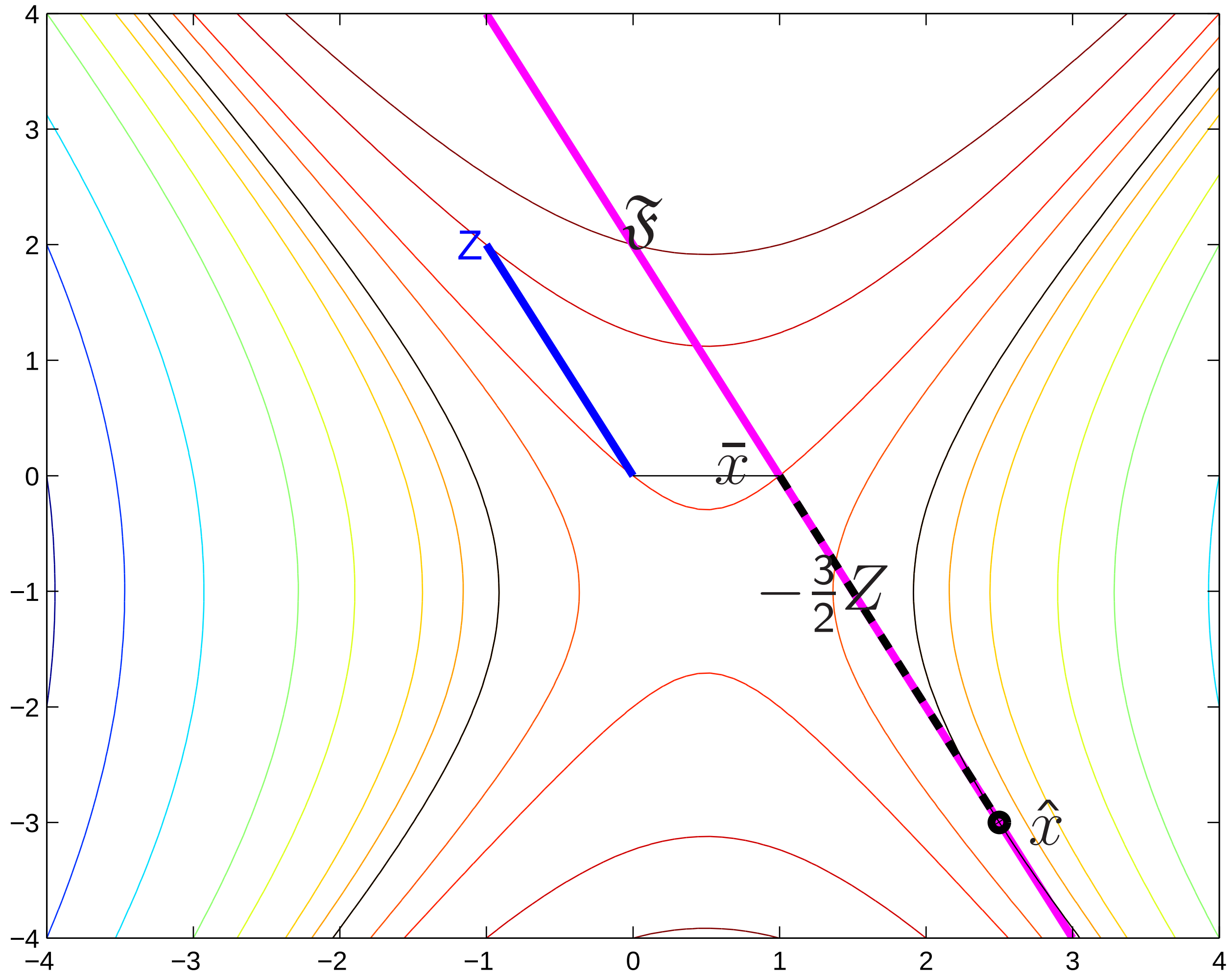
Since the reduced Hessian is positive definite the problem is strictly convex. The unique solution is then given by taking an arbitrary $\bar{x} \in \mathfrak{F}$, *e.g.* $\bar{x} = (1\ 0)^T$ and solving

$$\bar{H}\hat{v} = -\bar{c} = -Z^T(H\bar{x} + c) = -3 \quad \Rightarrow \hat{v} = -\frac{3}{2}$$

Then

$$\hat{x} = \bar{x} + Z\hat{v} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix} (-\frac{3}{2}) = \begin{bmatrix} \frac{5}{2} \\ -3 \end{bmatrix}$$

Linear constrained QP: Example - Geometric illustration



Linear constrained QP: Example - Lagrange method

Assume that we know the problem is convex, then

$$\begin{bmatrix} H & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} -c \\ b \end{bmatrix}, \qquad \left[\begin{array}{cc|c} -2 & 0 & -2 \\ 0 & 1 & -1 \\ \hline 2 & 1 & 0 \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ \hline u \end{array}\right] = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

Solving

$$\begin{aligned} -2x_1 - 2u &= -1 \\ x_2 - u &= -1 \end{aligned} \quad \Rightarrow \quad \begin{aligned} x_1 &= \tfrac{1}{2} - u \\ x_2 &= -1 + u \end{aligned},$$

then $2x_1 + x_2 = 2$ implies $2(\frac{1}{2} - u) + (-1 + u) = -u = 2$, so $u = -2$ and then $x_1 = 5/2$ and $x_2 = -3$.

As for the nullspace method.

Linear constrained QP: Example - Sensitivity analysis

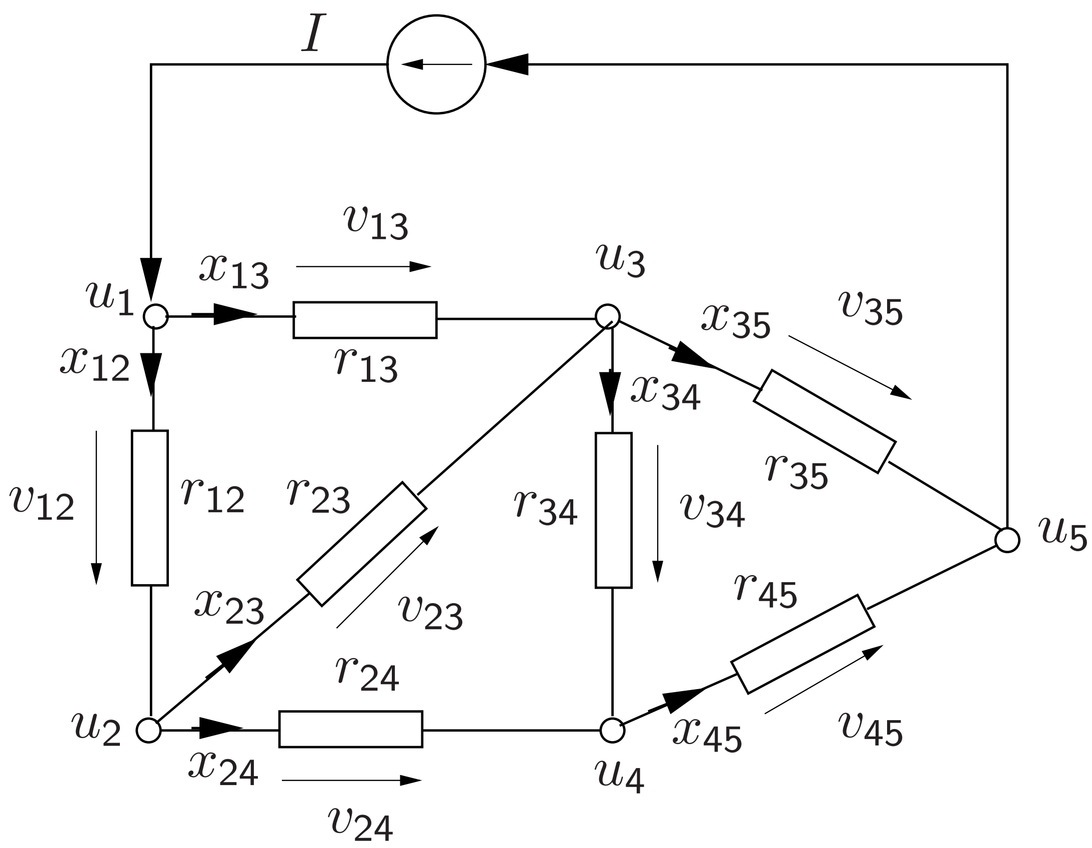
From the Lagrange method we know that $u = -2$, it tells us how the objective function changes due to variations in the right hand side b of the linear constraint, in first approximation.

Assume that $b := b + \delta$, then from the previous calculations $u = -(2 + \delta)$ and then $x_1 = 5/2 + \delta$ and $x_2 = -3 - \delta$. Then

$$f(x) = \frac{1}{2} \begin{bmatrix} \frac{5}{2} + \delta \\ -3 - \delta \end{bmatrix}^T \begin{bmatrix} -2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{5}{2} + \delta \\ -3 - \delta \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} \frac{5}{2} + \delta \\ -3 - \delta \end{bmatrix}$$

$$= -\frac{9}{4} - 2\delta - \frac{1}{2}\delta^2$$

Example 2: Analysis of a resistance circuit



- Let a current I go through the circuit from node 1 to node 5.
- What is the solution to the circuit equation?

The current x_{ij} goes from node i to node j if $x_{ij} > 0$. Kirchoff's currentlaw gives

$$
\begin{aligned}
x_{12} + x_{13} &= i \\
x_{12} - x_{23} - x_{24} &= 0 \\
x_{13} + x_{23} - x_{34} - x_{35} &= 0 \\
x_{24} + x_{34} - x_{45} &= 0 \\
x_{35} + x_{45} &= -i
\end{aligned}
$$

This can be written $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{x} = \begin{bmatrix} x_{12} & x_{13} & x_{23} & x_{24} & x_{34} & x_{35} & x_{45} \end{bmatrix}^{\mathsf{T}}$,

$$
\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \\ -I \end{bmatrix}
$$

- The Node potentials are denoted u_j, $j = 1, \ldots, 5$
- The voltage drop over the resistor r_{ij} is denoted v_{ij} and is given by the equations

$$v_{ij} = u_i - u_j$$

$$v_{ij} = r_{ij} x_{ij} \quad \text{(Ohms law)}$$

This can also be written

$$\mathbf{v} = \mathbf{A}^\mathsf{T} \mathbf{u}$$

$$\mathbf{v} = \mathbf{D} \mathbf{x}$$

where

$$\mathbf{u} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 \end{bmatrix}^\mathsf{T} \quad \mathbf{v} = \begin{bmatrix} v_{12} & v_{13} & v_{23} & v_{24} & v_{34} & v_{35} & v_{45} \end{bmatrix}^\mathsf{T}$$

$$\mathbf{D} = \text{diag}(r_{12}, r_{13}, r_{23}, r_{24}, r_{34}, r_{35}, r_{45})$$

The circuit equations:

$$
\begin{aligned}
\mathbf{A}\mathbf{x} &= \mathbf{b} \\
\mathbf{v} &= \mathbf{A}^{\mathsf{T}}\mathbf{u} \\
\mathbf{v} &= \mathbf{D}\mathbf{x}
\end{aligned}
\quad \Leftrightarrow \quad
\begin{bmatrix} \mathbf{D} & -\mathbf{A}^{\mathsf{T}} \\ \mathbf{A} & 0 \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{0} \\ \mathbf{b} \end{bmatrix}
$$

Since $\mathbf{D}$ is positive definite $(r_{ij} > 0)$ this is, see comment 2, equivalent to the optimality conditions for the following optimization problem

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{D}\mathbf{x} \\
\text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{b}
\end{aligned}
$$

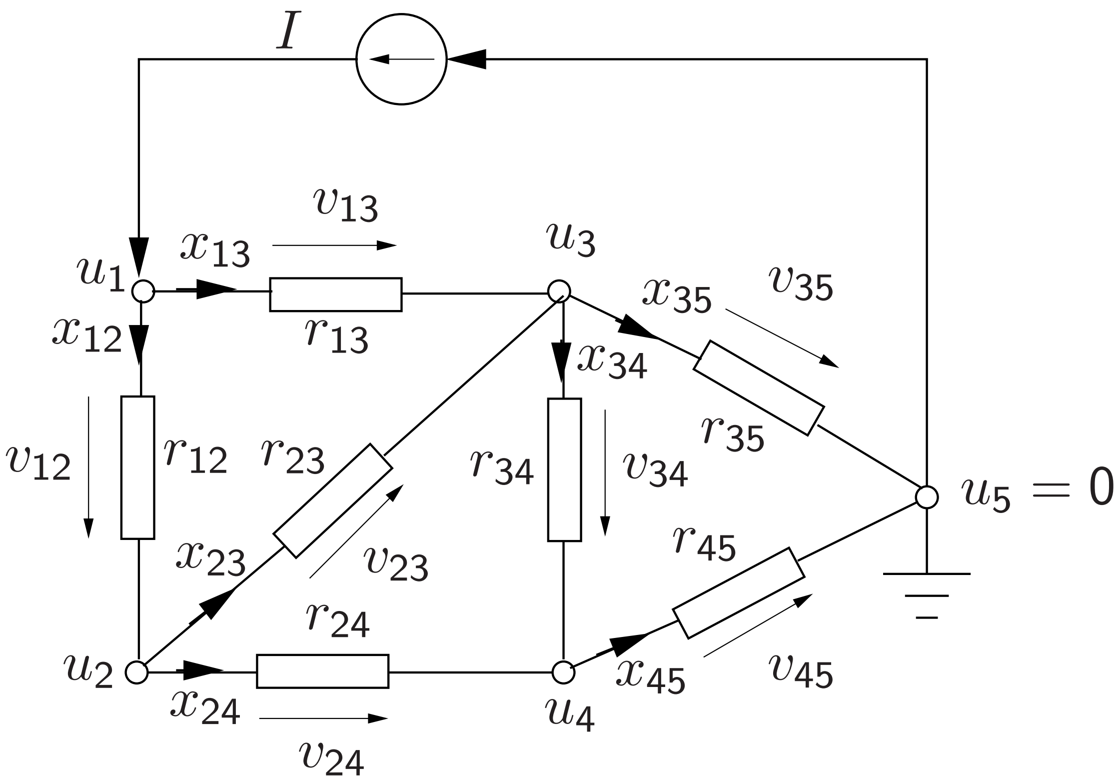
Comment

In the lecture on network optimization we saw that the matrix $\mathcal{N}(\mathbf{A}) \neq \mathbf{0}$ (then the matrix was called $\tilde{\mathbf{A}}$) and that the nullspace corresponded to cycles in the circuit.

The laws of electricity (Ohms law + Kirchoffs law) determines the current that minimizes the loss of power. This follows since the objective function can be written

$$\frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{D}\mathbf{x} = \frac{1}{2}\sum_{ij} r_{ij}x_{ij}^2$$

Comment 3. *It is common to connect one of the nodes to earth. Let us, e.g., connect node 5 to earth, then* $u_5 = 0$*.*



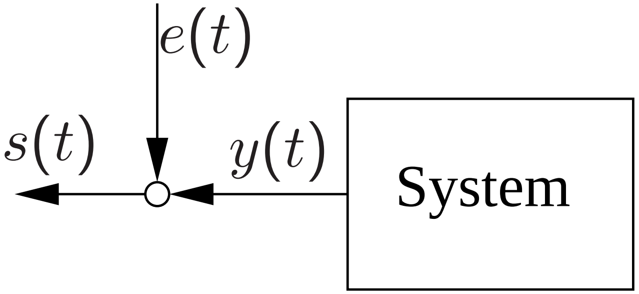
With the potential in node 5 fixed at $u_5 = 0$ the last column in the voltage equation $\mathbf{v} = \mathbf{A}^\mathsf{T}\mathbf{u}$ is eliminated. The Voltage equation and the current balance can be replaced with $\mathbf{v} = \bar{\mathbf{A}}^\mathsf{T}\bar{\mathbf{u}}$ and $\bar{\mathbf{A}}\mathbf{x} = \bar{\mathbf{b}}$ where

$$\bar{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 1 \end{bmatrix}, \quad \bar{\mathbf{b}} = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \bar{\mathbf{u}} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix}$$

The advantage with this is that the rows of $\bar{\mathbf{A}}$ are linearly independent. This facilitates the solution of the optimization problem.

Least-Squares problems

Application: Linear regression (model fitting)



Problem: Fit a linear regressionsmodel to measured data.

$$\text{Regressionmodel :} \quad y(t) = \sum_{j=1}^{n} \alpha_j \psi_j(t)$$

- $\psi_k(t)$, $k = 1, \ldots, n$ are the regressors (known functions)
- $\alpha_k = k = 1, \ldots, n$ are the model parameters (to be determined)
- $e(t)$ measurement noise (not known)
- $s(t)$ observations.

Idéa for fitting the model: Minimize the quadratic sum of the prediction errors

$$\begin{aligned}
&\text{minimize } \frac{1}{2}\sum_{i=1}^{m}\Big(\sum_{j=1}^{n}\alpha_j\psi_j(t_i) - s(t_i)\Big)^2 \\
=&\text{minimize } \frac{1}{2}\sum_{i=1}^{m}(\psi(t_i)^{\mathsf{T}}\mathbf{x} - s(t_i))^2 \\
=&\text{minimize } \frac{1}{2}(\mathbf{A}\mathbf{x}-\mathbf{b})^{\mathsf{T}}(\mathbf{A}\mathbf{x}-\mathbf{b}) \qquad (2)
\end{aligned}$$

$$\psi(t) = \begin{bmatrix}\psi_1(t)\\ \vdots \\ \psi_n(t)\end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix}\alpha_1\\ \vdots \\ \alpha_n\end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix}\psi(t_1)^{\mathsf{T}}\\ \vdots \\ \psi(t_n)^{\mathsf{T}}\end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix}s(t_1)\\ \vdots \\ s(t_n)\end{bmatrix}$$

The solution of the Least-Squares problem (LSQ)

The Least-Squares problem (2) can equivalently be written

$$\text{minimize } \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{H}\mathbf{x} + \mathbf{c}^\mathsf{T}\mathbf{x} + c_0$$

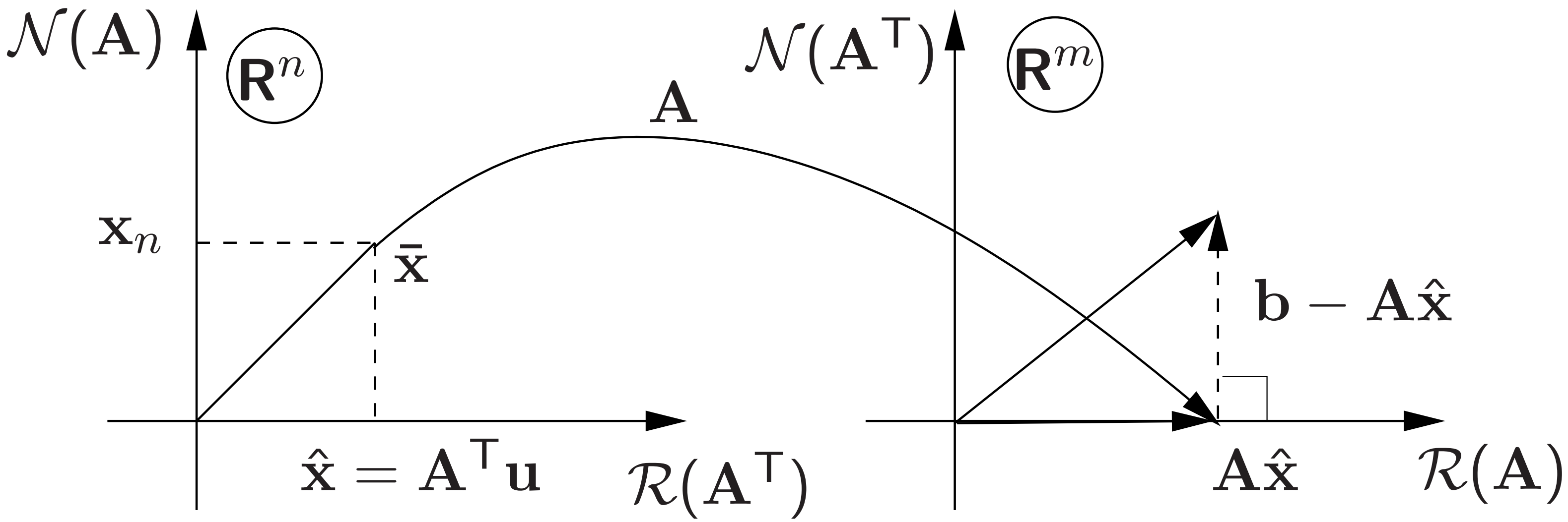
where $\mathbf{H} = \mathbf{A}^\mathsf{T}\mathbf{A}$, $\mathbf{c} = -\mathbf{A}^\mathsf{T}\mathbf{b}$, $c_0 = \frac{1}{2}\mathbf{b}^\mathsf{T}\mathbf{b}$. We note that

- $\mathbf{H} = \mathbf{A}^\mathsf{T}\mathbf{A}$ is positive semidefinite since $\mathbf{x}^\mathsf{T}\mathbf{A}^\mathsf{T}\mathbf{A}\mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 \geq 0$.
- $\mathbf{c} \in \mathcal{R}(\mathbf{H})$ since $\mathbf{c} = -\mathbf{A}^\mathsf{T}\mathbf{b} \in \mathcal{R}(\mathbf{A}^\mathsf{T}) = \mathcal{R}(\mathbf{A}^\mathsf{T}\mathbf{A}) = \mathcal{R}(\mathbf{H})$.

The conditions in Theorem 2 and Comment 1 are satisfied and it follows that the LSQ-estimate is given by

$$\mathbf{H}\hat{\mathbf{x}} = -\mathbf{c} \quad \Leftrightarrow \quad \mathbf{A}^\mathsf{T}\mathbf{A}\hat{\mathbf{x}} = \mathbf{A}^\mathsf{T}\mathbf{b} \quad \text{(The Normal equation)}$$

Linear algebra interpretation



- The Normal equation can be written

$$\mathbf{A}^\mathsf{T}(\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}) = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{b} - \mathbf{A}\hat{\mathbf{x}} \in \mathcal{N}(\mathbf{A}^\mathsf{T})$$

  This orthogonality property has a geometric interpretation which is illustrated in the figure above.

- The Fundamental Theorem of Linear algebra explains the LSQ solution geometrically.

Uniqueness of the LSQ solution

Case 1 If the columns in $\mathbf{A}$ are linearly independent,*i.e.*, $\mathcal{N}(\mathbf{A}) = \{\mathbf{0}\}$, it holds that $\mathbf{H} = \mathbf{A}^\top\mathbf{A}$ is positive definite and hence invertible. The Normal equations then has the unique solution

$$\hat{\mathbf{x}} = (\mathbf{A}^\top\mathbf{A})^{-1}\mathbf{A}\mathbf{b}$$

Case 2 If $\mathbf{A}$ has linearly dependent columns, *i.e.*, $\mathcal{N}(\mathbf{A}) \neq \{\mathbf{0}\}$ then it is natural to chose the solution of smallest length. From the figure on the previous slide it is clear that we should let $\hat{\mathbf{x}} = \mathbf{A}^\top\hat{\mathbf{u}}$, where $\mathbf{A}\mathbf{A}^\top\hat{\mathbf{u}} = \mathbf{A}\bar{\mathbf{x}}$, and where $\bar{\mathbf{x}}$ is an arbitrary solution to the LSQ problem.

Our choice of $\hat{\mathbf{x}}$ in case 2 can equivalently be interpreted as the solution to the quadratic optimization problem

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{A}\mathbf{x} = \mathbf{A}\bar{\mathbf{x}} \end{aligned}$$

According to Comment 2 the solution to this optimization problem is given by the solution to

$$\begin{bmatrix} I & -\mathbf{A}^\mathsf{T} \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{x}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{A}\bar{\mathbf{x}} \end{bmatrix}$$

i.e. $\hat{\mathbf{x}} = \mathbf{A}^\mathsf{T}\hat{\mathbf{u}}$, where $\mathbf{A}\mathbf{A}^\mathsf{T}\hat{\mathbf{u}} = \mathbf{A}\bar{\mathbf{x}}$

Läsanvisningar

- Quadratic constraints without constraints: Chapter 9 and 27 in the book.
- Quadratic optimization with constraints: Chapter 10 in the book.
- Least-Squares method: Chapter 11 in the book.