Least square method

Cholesky decomposition

- There are two versions of Cholesky decomposition theorem, one for positive semi-definite and one for positive definite symmetric matrices.
- Let $\mathbf{A} \in \mathbb{R}^{m \times m}$ be a symmetric matrix, i.e., $\mathbf{A} = \mathbf{A}^T$.
- The matrix $\mathbf{A}$ is positive semi-definite iff there exists a lower triangular matrix $\mathbf{L}$ such that $\mathbf{A} = \mathbf{L}\mathbf{L}^T$.
- The matrix $\mathbf{A}$ is positive definite iff there exists a unique lower triangular matrix $\mathbf{L}$ with positive diagonal elements such that $\mathbf{A} = \mathbf{L}\mathbf{L}^T$.
- It follows that when we apply the procedure to obtain $\mathbf{L}$ for a symmetric matrix $\mathbf{A}$ there are three possibilities:
  - The procedure terminates and finds a matrix $\mathbf{L}$ whose diagonal elements are all positive. In this case $\mathbf{A}$ is positive definite.
  - The procedure terminates and finds a matrix $\mathbf{L}$ whose diagonal elements are not all positive. In this case $\mathbf{A}$ is positive semi-definite.
  - The procedure fails (because we get a complex number as an entry for $\mathbf{L}$). In this case $\mathbf{A}$ is not positive semi-definite.

The least square method

- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$.
- Suppose $\mathbf{Ax} = \mathbf{b}$ has no solution for $\mathbf{x} \in \mathbb{R}^n$, i.e., $\mathbf{b} \notin \text{range}(\mathbf{A})$.
- The least square method finds $\mathbf{x} \in \mathbb{R}^n$ such that $\|\mathbf{Ax} - \mathbf{b}\|_2$, or equivalently, $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ is minimized.
- Since $\text{range}(\mathbf{A}) + \text{null}(\mathbf{A}^T) = \mathbb{R}^m$ and $\text{range}(\mathbf{A}) \perp \text{null}(\mathbf{A}^T)$ (i.e., are orthogonal subspaces), there is a unique decomposition:
- $\mathbf{b} = \mathbf{b}_r + \mathbf{b}_n$, where $\mathbf{b}_r \in \text{range}(\mathbf{A})$ and $\mathbf{b}_n \in \text{null}(\mathbf{A}^T)$.
- Thus,

$$\|\mathbf{Ax} - \mathbf{b}\|_2^2 = \|\mathbf{Ax} - \mathbf{b}_r - \mathbf{b}_n\|_2^2$$
$$= (\mathbf{Ax} - \mathbf{b}_r - \mathbf{b}_n) \cdot (\mathbf{Ax} - \mathbf{b}_r - \mathbf{b}_n)$$
$$= (\mathbf{Ax} - \mathbf{b}_r) \cdot (\mathbf{Ax} - \mathbf{b}_r) + (\mathbf{b}_n) \cdot (\mathbf{b}_n)$$
$$\|\mathbf{Ax} - \mathbf{b}_r\|_2^2 + \|\mathbf{b}_n\|_2^2 \geq \|\mathbf{b}_n\|_2^2.$$

- Therefore, $\|\mathbf{Ax} - \mathbf{b}\|_2^2$ is minimised when $\|\mathbf{Ax} - \mathbf{b}_r\|_2^2 = 0$, i.e., if $\mathbf{Ax} = \mathbf{b}_r$.

Normal equations

- **Claim:** $\mathbf{Ax} = \mathbf{b}_r$ iff $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$.
- (i) Suppose $\mathbf{Ax} = \mathbf{b}_r$.
  - Then $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}_r$ and since $\mathbf{A}^T\mathbf{b}_n = \mathbf{0}$ we have:

$$\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T(\mathbf{b}_r + \mathbf{b}_n) = \mathbf{A}^T(\mathbf{b}).$$

- (ii) Suppose $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$.
  - Then, $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}) = 0$ and thus, $\mathbf{A}^T(\mathbf{Ax} - \mathbf{b}_r) = \mathbf{0}$ as $\mathbf{A}^T\mathbf{b}_n = \mathbf{0}$.
  - It follows that $\mathbf{Ax} - \mathbf{b}_r \in \mathrm{null}(\mathbf{A}^T)$.
  - Since we also have $\mathbf{Ax} - \mathbf{b}_r \in \mathrm{range}(\mathbf{A})$ and because $\mathrm{range}(\mathbf{A}) \cap \mathrm{null}(\mathbf{A}^T) = \mathbf{0}$, we obtain:

$$\mathbf{Ax} - \mathbf{b}_r = \mathbf{0}.$$

- $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ is called the **normal equations**, whose solution gives the solution to the least square problem.
- Since $\mathbf{A}^T\mathbf{A}$ is positive semi-definite, we can use Cholesky factorisation to get a lower triangular matrix $\mathbf{L}$ with $\mathbf{A}^T\mathbf{A} = \mathbf{LL}^T$ to solve the normal equations.