Under-determined System

\[ \begin{bmatrix} n \\ m \end{bmatrix} = \begin{bmatrix} A \\ b \end{bmatrix} \]

Assumption: No noise

There are two possibilities: possibly

1. if \( A \) is degenerate, then there is no solution.

2. if \( A \) is full rank, then there are infinitely many solutions.

Question: If there are infinitely many solutions, then which solution should we pick?

The solution to this question is somewhat tricky, because it depends on what solution we want.

Consider

\[ \min_x J(x) \]

s.t. \( A x = b \).

This function \( J(x) \) defines "what we want".
For example:
\[
J(x) = \|x\|_2^2 = \sum_{n=1}^{N} x_n^2
\]
\[
J(x) = \|x\|_1 = \sum_{n=1}^{N} |x_n|
\]
\[
J(x) = \|x\|_0 = \# \text{ non-zeros}
\]

**\(l^2\)-norm minimization**

Let's first consider

\[
(P_2) \quad \min_x \|x\|_2^2 \\
\text{s.t. } Ax = b.
\]

To solve this problem, we consider the Lagrange multiplier:

\[
L(x, \lambda) = \|x\|_2^2 + \lambda^T(Ax - b)
\]

Take derivatives on both sides

\[
\begin{align*}
\frac{\partial}{\partial x} L(x, \lambda) &= 2x + A^T\lambda = 0 \quad (1) \\
\frac{\partial}{\partial \lambda} L(x, \lambda) &= Ax - b = 0 \quad (2)
\end{align*}
\]

From (1), we have

\[
\hat{x} = -\frac{1}{2} A^T\lambda
\]

Therefore,

\[
A\hat{x} = A\left(-\frac{1}{2} A^T\lambda\right) = -\frac{1}{2} AA^T\lambda
\]
But since $\hat{x}$ is the optimal solution, by (2) we must have $A\hat{x} = b$. So

$$\frac{1}{2}AA^T\lambda = b$$

$$\Rightarrow \lambda = -2(AA^T)^{-1}b \quad (3)$$

Therefore

$$\hat{x} = \frac{1}{2}A^T\lambda$$

$$= \frac{1}{2}A^T(-2(AA^T)^{-1}b)$$

$$= A^T(AA^T)^{-1}b.$$ 

Computationally,

$$\hat{x} = A^T(AA^T)^{-1}b$$

is easy to solve; especially when $n \ll m$

We can also extend the idea to

$$\min_x \|Bx\|^2$$

s.t. $Ax = b$
In this case, the Lagrange multiplier returns

\[ L(x, \lambda) = \| Bx \|^2 + \lambda^T (Ax - b) \]

\[
\begin{align*}
\frac{\partial}{\partial x} L(x, \lambda) &= 2B^T Bx + A^T \lambda = 0 \\
\frac{\partial}{\partial \lambda} L(x, \lambda) &= Ax - b = 0
\end{align*}
\]

Suppose that \( B^T B \) is invertible. Then

\[ \hat{x} = -\frac{1}{2} (B^T B)^{-1} A^T \lambda \]

\[ \Rightarrow \quad A \hat{x} = \frac{1}{2} A (B^T B)^{-1} A^T \lambda = b \]

\[ \Rightarrow \quad \lambda = -2 \left( A (B^T B)^{-1} A^T \right)^{-1} b \]

\[ \Rightarrow \quad \hat{x} = A (B^T B)^{-1} A^T \left[ A (B^T B)^{-1} A^T \right]^{-1} b \]

Good thing about \( \ell^2 \)-norm:

1. Closed-form solution
   - so that you can analyze
   - often much easier/faster compute than iterative methods

2. Unique.
   - although \( Ax = b \) has infinitely many solutions, minimizing \( \| x \|^2 \) returns only one solution.