Least Squares Solutions

The problem

Let $\mathbf{A}$ denote a $m \times n$ matrix. The equation

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b}$$

shall be solved. In general the solution is not unique. Since $\mathbf{A} \cdot \mathbf{x}$ is a weighted addition of the column vectors of matrix $\mathbf{A}$ vector $\mathbf{b}$ must be in the column-space if a solution exists. In general only an approximation

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b} + \mathbf{r}$$

can be found. $\mathbf{r}$ is the residual vector.

$$\mathbf{r} = \mathbf{A} \cdot \mathbf{x} - \mathbf{b}$$

Because $\mathbf{r}$ is not in the column space of matrix $\mathbf{A}$ the vector must be orthogonal to each column of $\mathbf{A}$. This condition is equivalent to this equation:

$$\mathbf{A}^T \cdot \mathbf{r} = \mathbf{0}$$

or

$$\begin{aligned}
\mathbf{A}^T \cdot (\mathbf{A} \cdot \mathbf{x} - \mathbf{b}) &= \mathbf{0} && (1) \\
\mathbf{A}^T \cdot \mathbf{A} \cdot \mathbf{x} &= \mathbf{A}^T \cdot \mathbf{b} && (2) \\
\underbrace{\left(\mathbf{A}^T \cdot \mathbf{A}\right)^{-1} \cdot \mathbf{A}^T}_{left\ inverse} \cdot \mathbf{A} \cdot \mathbf{x} &= \left(\mathbf{A}^T \cdot \mathbf{A}\right)^{-1} \cdot \mathbf{A}^T \cdot \mathbf{b} && (3) \\
\mathbf{x} &= \underbrace{\left(\mathbf{A}^T \cdot \mathbf{A}\right)^{-1} \cdot \mathbf{A}^T}_{left\ inverse} \cdot \mathbf{b} && (4)
\end{aligned}$$

So if the `left-inverse` can be computed, the equation $\mathbf{A} \cdot \mathbf{x} = \mathbf{b} + \mathbf{r}$ can be solved.

A different approach to the solution of equation

$$\mathbf{A} \cdot \mathbf{x} = \mathbf{b} + \mathbf{r}$$

is to minimise the L2 norm of the residual vector $||\mathbf{r}||$. This is equivalent to the minimisation of $||\mathbf{r}||^2 = \mathbf{r}^T \cdot \mathbf{r}$

$$
\begin{aligned}
\mathbf{r}^T \cdot \mathbf{r} &= (\mathbf{A}\cdot\mathbf{x}-\mathbf{b})^T \cdot (\mathbf{A}\cdot\mathbf{x}-\mathbf{b}) && (5)\\
\mathbf{r}^T \cdot \mathbf{r} &= \left(\mathbf{x}^T\cdot\mathbf{A}^T-\mathbf{b}^T\right) \cdot (\mathbf{A}\cdot\mathbf{x}-\mathbf{b}) && (6)\\
f(\mathbf{x}) = \mathbf{r}^T \cdot \mathbf{r} &= \mathbf{x}^T\cdot\mathbf{A}^T\cdot\mathbf{A}\cdot\mathbf{x} - 2\cdot\mathbf{b}^T\cdot\mathbf{A}\cdot\mathbf{x}+\mathbf{b}^T\cdot\mathbf{b} && (7)
\end{aligned}
$$

Thus the goal is to minimize a scalar function $f(\mathbf{x})$:

Regardless of the shape of matrix $\mathbf{A}$ the matrix $\mathbf{U}$ is square, symmetric and positive definite.

The gradient of $f(\mathbf{x})$ is computed like this:

$$
f'(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_N} f(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1}\left(\mathbf{x}^T\cdot\mathbf{A}^T\cdot\mathbf{A}\cdot\mathbf{x}\right) \\ \frac{\partial}{\partial x_2}\left(\mathbf{x}^T\cdot\mathbf{A}^T\cdot\mathbf{A}\cdot\mathbf{x}\right) \\ \vdots \\ \frac{\partial}{\partial x_N}\left(\mathbf{x}^T\cdot\mathbf{A}^T\cdot\mathbf{A}\cdot\mathbf{x}\right) \end{bmatrix} - 2\cdot \begin{bmatrix} \frac{\partial}{\partial x_1}\mathbf{b}^T\cdot\mathbf{A}\cdot\mathbf{x} \\ \frac{\partial}{\partial x_2}\mathbf{b}^T\cdot\mathbf{A}\cdot\mathbf{x} \\ \vdots \\ \frac{\partial}{\partial x_N}\mathbf{b}^T\cdot\mathbf{A}\cdot\mathbf{x} \end{bmatrix}
$$

$$
\begin{aligned}
f'(\mathbf{x}) &= \mathbf{A}^T\cdot\mathbf{A}\cdot\mathbf{x}+\mathbf{A}\cdot\mathbf{A}^T\cdot\mathbf{x}-2\cdot\mathbf{A}^T\cdot\mathbf{b} && (8)\\
f'(\mathbf{x}) &= 2\cdot\left(\mathbf{A}^T\cdot\mathbf{A}\cdot\mathbf{x}-\mathbf{A}^T\cdot\mathbf{b}\right) && (9)
\end{aligned}
$$

Setting the gradient to 0 results in the *normal* equation for the unknown vector $\mathbf{x}$:

$$
\mathbf{A}^T\cdot\mathbf{A}\cdot\mathbf{x} = \mathbf{A}^T\cdot\mathbf{b}
$$

Note

By inserting $\mathbf{A}\cdot\mathbf{x}=\mathbf{b}+\mathbf{r}$ we get

$$
\begin{aligned}
\mathbf{A}^T\cdot(\mathbf{b}+\mathbf{r}) &= \mathbf{A}^T\cdot\mathbf{b} && (10)\\
\mathbf{A}^T\cdot\mathbf{b}+\mathbf{A}^T\cdot\mathbf{r} &= \mathbf{A}^T\cdot\mathbf{b} && (11)\\
&\rightarrow && (12)\\
\mathbf{A}^T\cdot\mathbf{r} &= \mathbf{0} && (13)
\end{aligned}
$$

This equation shows again that the residual vector $\mathbf{r}$ is orthogonal to all columns of matrix $\mathbf{A}$.

If the inverse $\left(\mathbf{A}^T\cdot\mathbf{A}\right)^{-1}$ exist vector $\mathbf{x}$ is computed by:

$$
\mathbf{x} = \underbrace{\left(\mathbf{A}^T\cdot\mathbf{A}\right)^{-1}\cdot\mathbf{A}^T}_{left\ inverse}\cdot\mathbf{b}
$$

The solution can be found using `Numpy` method `numpy.linalg.lstsq`

https://numpy.org/doc/stable/reference/generated/numpy.linalg.lstsq.html

In [ ]:

In [ ]:

In [ ]: