The linear algebra perspective

(heavily inspired by Pavel Grinstein)

A): Quadratic Form Minimization:

A function $f : \mathbb{R}^N \to \mathbb{R}$ of the form

$$f(\underline{x}) = \frac{1}{2}\underline{x}^T \underline{\underline{A}}\, \underline{x} - \underline{x}^T \underline{b}$$ with $\underline{\underline{A}}$ square, positive definite

is called the "quadratic form" (with linear extension).

Then, $\min_{\underline{x}} f(\underline{x})$: $\underline{\underline{A}}\,\underline{x} = \underline{b}$

One way to show this: $x \in \mathbb{R}^3$, compute $\nabla f = 0$

$\rightarrow$ solving the linear system $\boxed{\underline{\underline{A}}\,\underline{x} = \underline{b}}$ minimizes the associate quadratic form $\frac{1}{2}\underline{x}^T \underline{\underline{A}}\,\underline{x} - \underline{x}^T \underline{b}$!

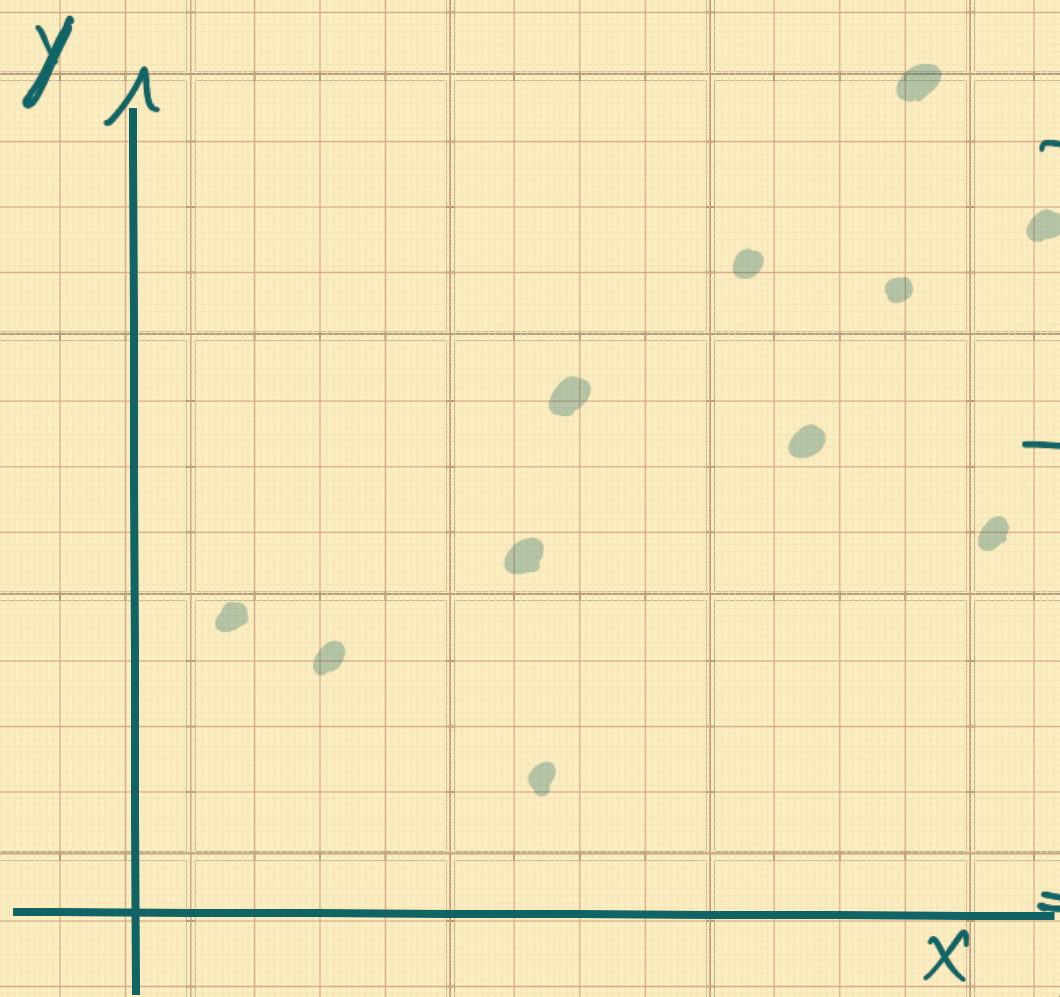
$\rightarrow$ scalar case: $f(x) = \frac{1}{2} a x^2 - bx$,

$$f_x = ax - b \stackrel{!}{=} 0$$

$$\rightarrow \boxed{ax = b}$$

B) Linear least squares

hypothesis space



→ Find a best linear fit to the data $\{(x_i, y_i)\}_{i=1}^{N} =: D$

→ $h_\Theta(x) = w_1 x + w_0 \cdot 1$

$\Theta = (w_1, w_0)^T$

Basis functions

⇒ how to fit the model $h_\Theta(x)$ on the data D?

→ Interpolation leads to an overdetermined system:

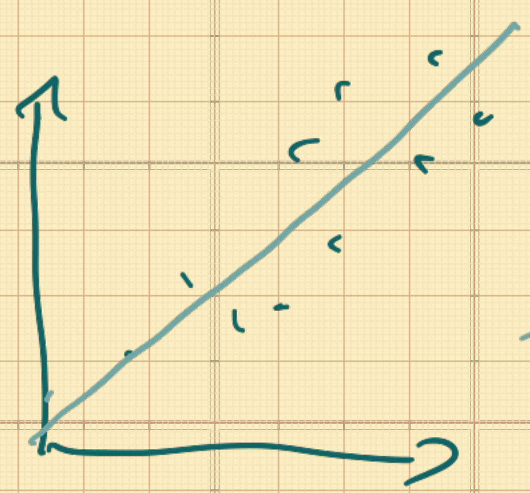
$$y_i \overset{!}{=} \underline{w_1}\, x_i + \underline{w_0} \cdot 1 \qquad \forall i = 1, \dots, \underline{N}$$

2 unknowns, N constraints

in matrix vector form with $x = \phi_1,\ 1 = \phi_0$:

$$\begin{pmatrix} \phi_1(x_1) & \phi_0(x_1) \\ \phi_1(x_2) & \vdots \\ \phi_1(x_3) & \vdots \\ \vdots & \vdots \\ \phi_1(x_N) & \phi_0(x_N) \end{pmatrix} \begin{pmatrix} w_1 \\ w_0 \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ \vdots \\ y_N \end{pmatrix}$$

↳ VanderMonde-Matrix



$$\underline{\underline{A}} \qquad \underline{x} = \underline{b}$$

→ cannot be inverted exactly!

→ What can we say about the error?

$$\underline{r} = \underline{b} - \underline{\underline{A}}\,\underline{x}$$

$\rightarrow$ need norm for $\underline{r}$ $\rightarrow$ inner product!

$$\underline{r}^T\underline{r} = (\underline{b} - \underline{\underline{A}}\,\underline{x})^T(\underline{b} - \underline{\underline{A}}\,x)$$

$$= \underline{x}^T\underline{\underline{A}}^T\underline{\underline{A}}\,\underline{x} - \underline{x}^T\underline{\underline{A}}^T\underline{b} - \underline{b}^T\underline{\underline{A}}\,\underline{x} + \underline{b}^T b$$

$$\downarrow$$

$$\underline{b}^T(\underline{x}^T\underline{\underline{A}}^T)^T = \underline{b}^T(\underline{\underline{A}}\,\underline{x})$$

$$= 2\left(\frac{1}{2}\underline{x}^T\underbrace{\underline{\underline{A}}^T\underline{\underline{A}}}_{\text{"}\tilde{\underline{\underline{A}}}\text{"}}\,\underline{x} - \underline{x}^T\underbrace{\underline{\underline{A}}^T\underline{b}}_{\text{"}\tilde{\underline{b}}\text{"}}\right) + \underbrace{\underline{b}^T\underline{b}}_{= \text{const.}}$$

$$= 2\cdot\left(\frac{1}{2}\underline{x}^T\tilde{\underline{\underline{A}}}\,\underline{x} - \underline{x}^T\tilde{\underline{b}}\right) + \text{const.}$$

→ The term in brackets is a <u>quadratic form</u>!
(the position of the minimum does not change through addition of a constant or scaling)

→ This error becomes minimal if we thus

solve $\underline{\underline{\tilde{A}}}\,\underline{x} = \underline{\tilde{b}}$

or $\underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x} = \underline{\underline{A}}^T \underline{b}$

$$\Rightarrow \underline{x} = \frac{\underline{\underline{A}}^T \underline{b}}{\underline{\underline{A}}^T \underline{\underline{A}}}$$

→ The famous normal equation of linear least squares!

$$\begin{pmatrix} w_1 \\ w_0 \end{pmatrix} = \Theta^T$$

What did we learn from this?

- Solving a linear system minimizes the associated quadratic form
- Thus, for a linear hypothesis, a quadratic error is natural choice
- A closed form solution to the linear least squares problem is given by the normal form
- This is thus a nice test case for learning algorithms!

- The normal form corresponds to a discrete $\mathbb{L}_2$-projection onto $\mathbb{P}_1$ with standard inner product.