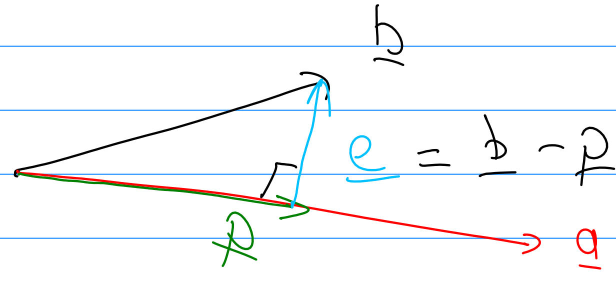
Projections

Start w/ projections onto a vector (line), then generalize onto a subspace.

A vector projection is the determination of which part of one vector lies on another vector.



$\underline{b}$ = Generic vector in $\mathbb{R}^n$

$\underline{a}$ = Another vector in $\mathbb{R}^n$

$\underline{p}$ = the projection of $\underline{b}$ onto $\underline{a}$

$\underline{e}$ = the "error" vector, how far $\underline{b}$ is from $\underline{a}$

To find $\underline{p}$, first define it as

$$\underline{p} = \hat{x}\,\underline{a}$$

$\hat{x}$ = Distance along $\underline{a}$

$$\underline{e} = \underline{b} - \underline{p} = \underline{b} - \hat{x}\,\underline{a}$$

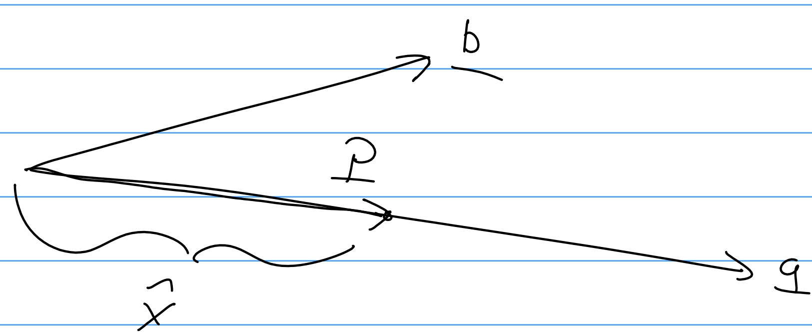
$$\underline{e} \perp \underline{a} \Rightarrow \underline{a} \cdot \underline{e} = 0 \Rightarrow \underline{a}^T \underline{e} = 0$$

$$\underline{a}^T \underline{e} = \underline{a}^T (\underline{b} - \underline{p}) = \underline{a}^T (\underline{b} - \hat{x}\,\underline{a}) = 0$$

$$= \underline{a}^T \underline{b} - \hat{x}\,\underline{a}^T \underline{a} = 0$$

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

$$\Rightarrow \underline{p} = \hat{x}\,\underline{a} = \underbrace{\frac{\underline{a}^T \underline{b}}{\underline{a}^T \underline{a}}}_{\text{Scalar}}\,\underline{a}$$



Issue: $\underline{p}$ is only for that specific $\underline{b}$.

Try to generalize.

Projection Matrix

A matrix $\underline{A}$ such that given any vector $\underline{b}$ can be projected onto $\underline{a}$ via

$$\underline{p} = \underline{A}\,\underline{b}$$

$$\underline{p} = \underbrace{\frac{\underline{a}^T\underline{b}}{\underline{a}^T\underline{a}}}_{\text{Scalar}}\,\underline{a} = \frac{\underline{a}\,\underline{a}^T\underline{b}}{\underline{a}^T\underline{a}} = \underbrace{\frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}}_{\underline{A}\ (\text{matrix})}\,\underline{b}$$

The projection matrix for a vector $\underline{a}$ is given by

$$\underline{A} = \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}} \qquad \underline{a}\,\underline{a}^T = \text{Outer product}$$

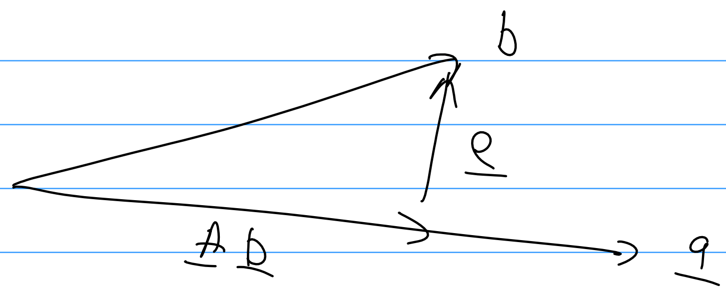
ex.) In 2D $\quad \underline{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$

$$\underline{a}\,\underline{a}^T = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}\begin{bmatrix} a_1 & a_2 \end{bmatrix} = \begin{bmatrix} a_1^2 & a_1 a_2 \\ a_1 a_2 & a_2^2 \end{bmatrix}$$

The nice thing about $\underline{A}$ is that it will project <u>any</u> vector $\underline{b}$ onto $\underline{a}$.

$\underline{A}$ is also idempotant: Repeated application of $\underline{A}$ have no effect.

ex 1) $\underline{A}(\underline{A}\,\underline{b}) = \underline{A}\,\underline{b}$



$$\underline{A}^2 = \underline{A}\,\underline{A} = \left(\frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}\right)\left(\frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}\right) = \frac{\underline{a}\,\cancel{(\underline{a}^T\underline{a})}\,\underline{a}^T}{\cancel{(\underline{a}^T\underline{a})}\,(\underline{a}^T\underline{a})}$$

$$= \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}} = \underline{A}$$

Another nice thing is that $\underline{I}-\underline{A}$, $\underline{I}$ can now project on to the perpendicular space of $\underline{a}$:

$$(\underline{I}-\underline{A})\,\underline{b} = \underline{b} - \underline{A}\,\underline{b} = \underline{b} - \underline{p} = \underline{e}$$

Also, $(\underline{I}-\underline{A})^2 = (\underline{I}-\underline{A})(\underline{I}-\underline{A}) =$

$$\underline{I} - \underline{A} - \underline{A} + \underline{A}^2 = \underline{I} - \underline{A} - \underline{A} + \underline{A}$$

$$= \underline{I} - \underline{A}$$

ex1) Find the projection matrix of $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ & then project $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ onto $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

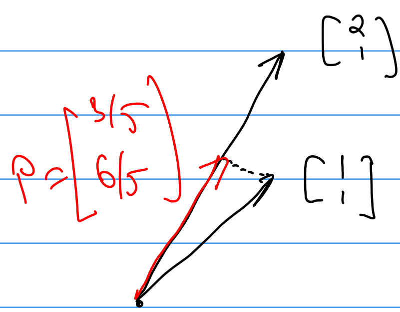
$\underline{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ $\qquad$ $\underline{A} = \dfrac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}$

$\underline{a}^T\underline{a} = \begin{bmatrix} 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1^2 + 2^2 = 5$

$\underline{a}\,\underline{a}^T = \begin{bmatrix} 1 \\ 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$

$\underline{A} = \dfrac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}} = \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}$

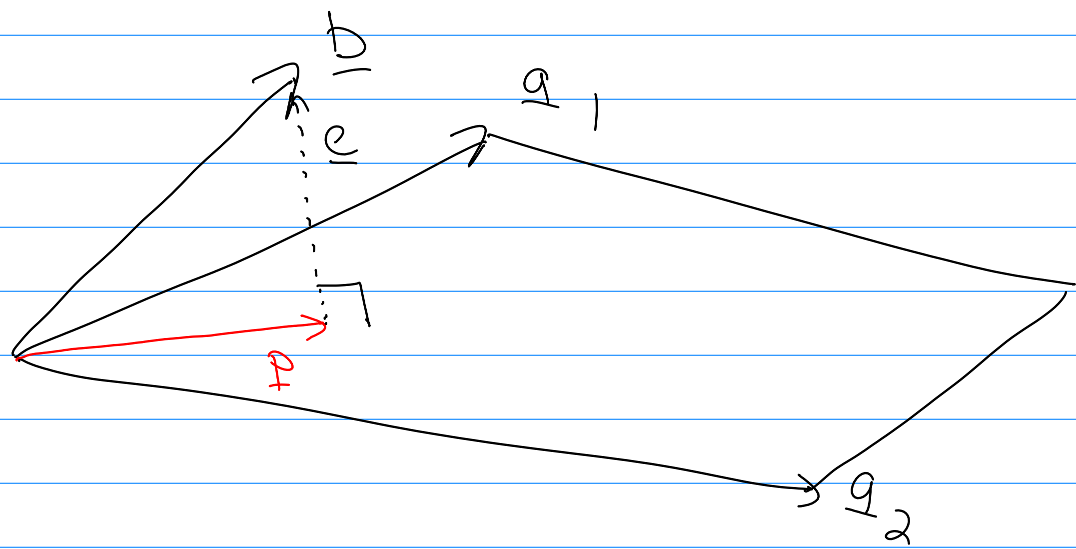
$\underline{p} = \underline{A}\,\underline{b} = \begin{bmatrix} 1/5 & 2/5 \\ 2/5 & 4/5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/5 \\ 6/5 \end{bmatrix}$



Projection onto Subspace

The idea is to project a vector $\underline{b}$ in $\mathbb{R}^m$ onto a subspace S in $\mathbb{R}^n$

Let S be spanned by vectors $\underline{a}_1$ & $\underline{a}_2$



Let the subspace S be spanned by $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$

We want to find $\underline{A}$ such that

$\underline{p} = \underline{A}\,\hat{\underline{x}}$ where the columns of $\underline{A}$ span the subspace and $\hat{\underline{x}}$ is the coordinates ("weights") of the column space of $\underline{A}$.

As before, the error vector $\underline{e}$ is perpendicular to subspace S

$\Rightarrow$ $\underline{e} \cdot$ any vector in $S = 0$

Thus, since $\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_n$ are in S, then

$$\underline{a}_1 \cdot \underline{e} = \underline{a}_1^T \underline{e} = 0$$

$$\underline{a}_2 \cdot \underline{e} = \underline{a}_2^T \underline{e} = 0$$

$$\vdots$$

$$\underline{a}_n \cdot \underline{e} = \underline{a}_n^T \underline{e} = 0$$

Write as
$$\begin{bmatrix} — \underline{a}_1^T — \\ — \underline{a}_2^T — \\ \vdots \\ — a_n^T — \end{bmatrix} \underline{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\Rightarrow \quad \underline{A}^T \underline{e} = \underline{0}$$

$$\underline{e} = \underline{b} - \underline{p} = \underline{b} - \underline{A}\hat{\underline{x}}$$

$$\underline{A}^T \underline{e} = \underline{A}^T (\underline{b} - \underline{A}\hat{\underline{x}}) = 0$$

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

or, $\hat{\underline{x}} = (\underline{A}^T\underline{A})^{-1}\underline{A}^T\underline{b}$

We wanted a matrix such

$\underline{b}$ times that matrix gives $\underline{p}$,

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

$$\Rightarrow \boxed{\underline{p} = \underbrace{\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T}_{\text{The matrix that will project any vector } \underline{b} \text{ in } \mathbb{R}^m \text{ onto the space spanned by the columns of } \underline{A}. \text{ (in } \mathbb{R}^n)}\underline{b}}$$

What if S is spanned by one vector, $\underline{a}$?

$$\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T = a(\underline{a}^T\underline{a})^{-1}\underline{a}^T = \frac{\underline{a}\,\underline{a}^T}{\underline{a}^T\underline{a}}$$

Question: Why is this not valid?

$$\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T = \underline{A}\,\underline{A}^{-1}\underline{A}^{-T}\underline{A}^T = \underline{I}\,\underline{I} = \underline{I}$$

You do not know if $\underline{A}^{-1}$ or $\underline{A}^{-T}$ exist!

But why is $(\underline{A}^T\underline{A})^{-1}$ ok?

$\underline{A} \in M_{mn}$ $\qquad$ $\underline{A}^T \in M_{nm}$

$\underline{A}^T\underline{A} \Rightarrow (n \times m)(m \times n) = n \times n$

What if $\underline{A}^{-1}$ does exist?

(1) $\underline{A} \in M_{nn}$

(2) The columns of $\underline{A}$ span $\mathbb{R}^n$,

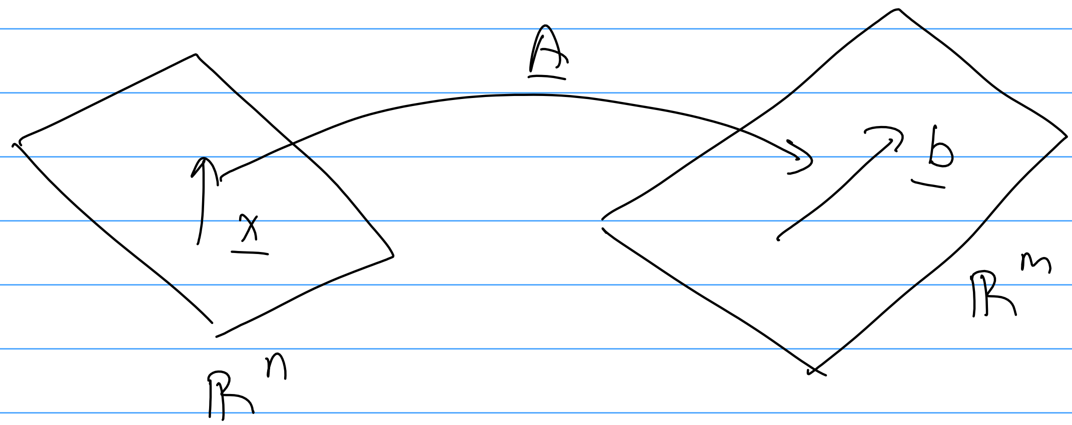
$\Rightarrow$ A vector $\underline{b}$ in $\mathbb{R}^n$ projected onto $\mathbb{R}^n$ is nothing but itself,

$\Rightarrow$ If $\underline{A}^{-1}$ exists, then

$\underline{A}(\underline{A}^T\underline{A})^{-1}\underline{A}^T$ must equal $\underline{I}$.

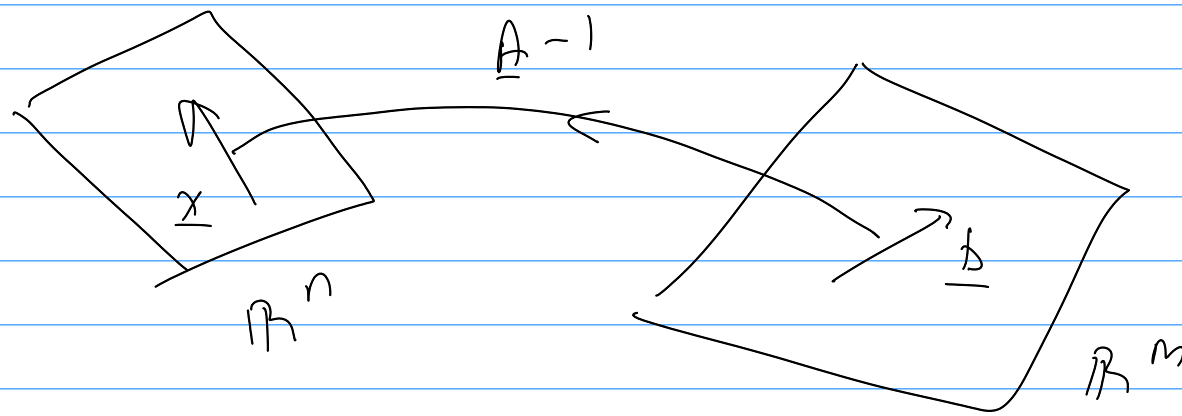
Least Square Approximations

Look at what a linear operator $\underline{A}$ does to $\underline{x}$: $\underline{A}\,\underline{x} = \underline{b}$



Given $\underline{A}$ & $\underline{x}$, there is always a vector $\underline{b}$.

Is the reverse true?



Only exists if $\underline{A}^{-1}$ exists, not always true.

If $\underline{A}^{-1}$ does not exist, can we find an approximate solution, called $\hat{\underline{x}}$, that does lie in $\mathbb{R}^n$?

Try to minimize $\underline{e} = \underline{b} - \underline{A}\,\hat{\underline{x}}$

This is a projection onto a subspace!

The solution that minimizes $\underline{e}$ is

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

$\hat{\underline{x}}$ is the solution that minimizes

$\underline{e} = \underline{A}\hat{\underline{x}} - \underline{b}$, or $\|\underline{e}\|_2 = \|\underline{A}\hat{\underline{x}} - \underline{b}\|_2$

Side: If $\underline{A}^{-1}$ does not exist, then

$\underline{A}\,\underline{x} = \underline{b}$ has a solution iff

$\underline{b}$ is in the $C(\underline{A})$

If $\underline{b}$ not in $C(\underline{A})$, project it onto $C(\underline{A})$

Don't forget that $\bar{x}$ does not solve

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

Note: The columns of $\underline{A}$ must still be independent for this to work.

The set of equations given by

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

are called the Normal Equations

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

$$\Rightarrow \underline{x} = \underbrace{(\underline{A}^T \underline{A})^{-1}} \underline{A}^T \underline{b}$$

This is very difficult to compute.

Why? Look at the condition number of $\underline{A}^T\underline{A}$.

$$\kappa(\underline{A}^T \underline{A}) = ??$$

First, look at $\kappa(\underline{A}\,\underline{B})$

$$\kappa(\underline{A}\,\underline{B}) = \|\underline{A}\,\underline{B}\|\,\|(\underline{A}\,\underline{B})^{-1}\| = \|\underline{A}\,\underline{B}\|\,\|\underline{B}^{-1}\underline{A}^{-1}\|$$

$$\leq \|\underline{A}\|\,\|\underline{B}\|\,\|\underline{B}^{-1}\underline{A}^{-1}\|$$

$$\leq \|\underline{A}\|\,\|\underline{B}\|\,\|\underline{B}^{-1}\|\,\|\underline{A}^{-1}\| = \kappa(\underline{A})\,\kappa(\underline{B})$$

$$\Rightarrow \quad \kappa(\underline{A}\,\underline{B}) \sim \kappa(\underline{A})\,\kappa(\underline{B})$$

Also, $\kappa(\underline{A}^T) = \kappa(\underline{A})$ (Proof later)

$$\Rightarrow \kappa(\underline{A}^T\underline{A}) = \kappa(\underline{A}^T)\,\kappa(\underline{A}) = \kappa^2(\underline{A})$$

$\Rightarrow$ If $\underline{A}$ is not well-conditioned,

$\underline{A}^T\underline{A}$ is even worse!

$\kappa(\underline{A}) \sim 10^2$, but $\kappa(\underline{A}^T\underline{A}) \sim 10^4$

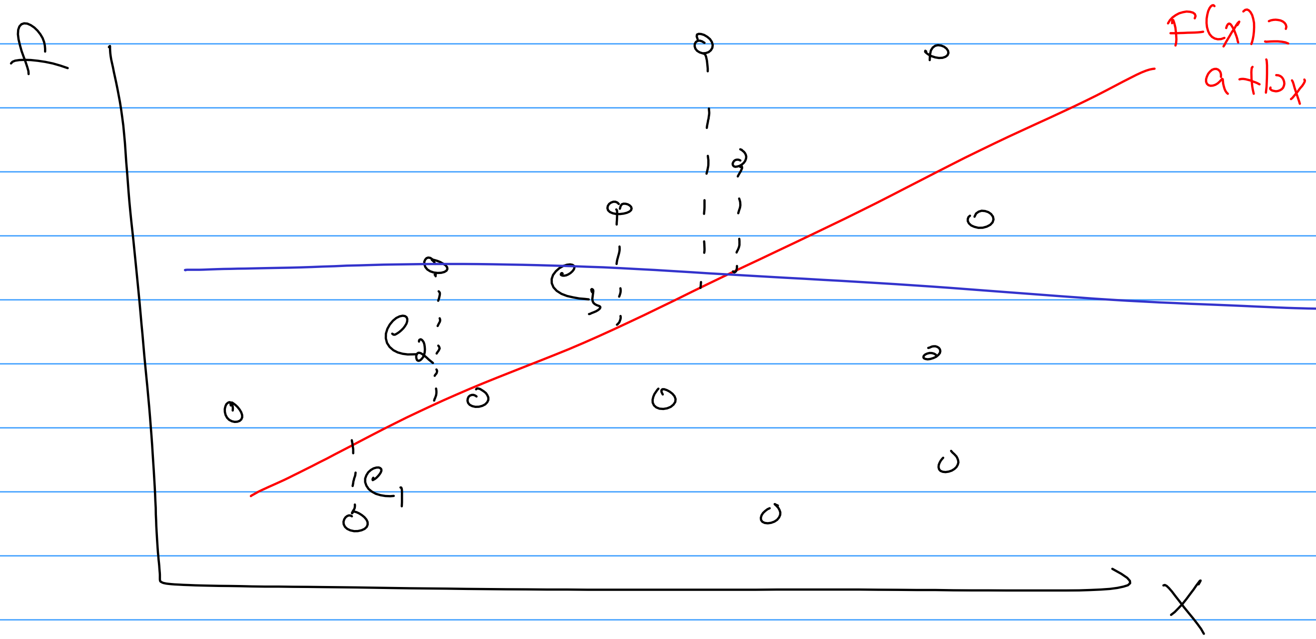
Later, we will talk about matrix decompositions that allow you to solve

$$\underline{A}^T\underline{A}\,\underline{x} = \underline{A}^T\underline{b} \quad \text{(or similar)}$$

without issue.

Example: Curve fitting

Say that you have many data points: (x_i, f_i)



You want to find an approximate relation $F(x) = a + bx$, that describes the data.

We want to minimize

$$e_1^2 + e_2^2 + e_3^2 + \dots + e_M^2$$

At each point, we wish that

$a + b x_i = f_i$ would hold.

or,

$a + b x_1 = f_1$

$a + b x_2 = f_2$

$\vdots$

$a + b x_m = f_m$

x_i & f_i are known,

a, b are unknown,

$$\Rightarrow \underbrace{\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}}_{\underline{A}} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$$

If at least one x_i does not equal 0, then the columns of $\underline{A}$ are independent.

If there are more than 2 data points, $\underline{A}^{-1}$ does not exist,

$\Rightarrow$ Called an over-constrained system,

but, $(\underline{A}^T \underline{A})^{-1}$ does exist,

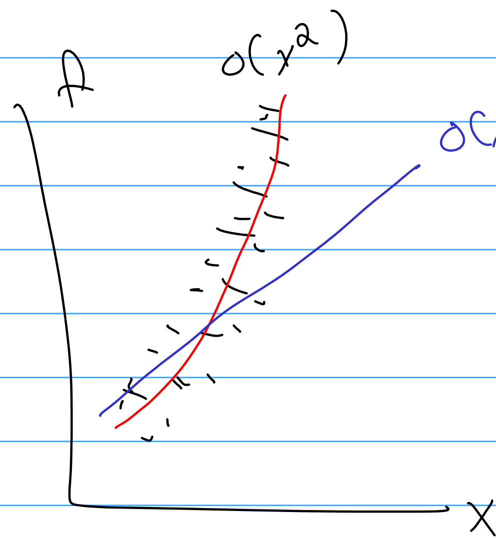
$\Rightarrow$ Solve

$$\underline{A}^T \underline{A} \begin{bmatrix} a \\ b \end{bmatrix} = \underline{A}^T \underline{f}$$

for a & b, to minimize $\|\underline{e}\|_2$.

Generalize to higher order:

$$F(x) = a + bx + cx^2$$



$$a + bx_1 + cx_1^2 = f_1$$
$$a + bx_2 + cx_2^2 = f_2$$
$$\vdots$$

$$\begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ & \vdots & \\ 1 & x_m & x_m^2 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$$

Multidimensional:

Fit $F(x,y) = a + bx + cy + dxy$ to the data

(x_i, y_i, f_i)

$$\begin{bmatrix} 1 & x_1 & y_1 & x_1 y_1 \\ 1 & x_2 & y_2 & x_2 y_2 \\ & & \vdots & \\ 1 & x_m & y_m & x_m y_m \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$$

Generic Functions:

Say that you have data you want to fit: (x_i, g_i)

Your interpolant is:

$$F(x) = a f_0(x) + b f_1(x) + c f_2(x)$$

$f_0(x)$, $f_1(x)$ etc are some functions (e.g. $f_0(x) = x$
$f_1(x) = 1 - x$
$f_2(x) = x^2 - x$
$\vdots$)

$$\Rightarrow \begin{bmatrix} f_0(x_1) & f_1(x_1) & f_2(x_1) \\ f_0(x_2) & f_1(x_2) & f_2(x_2) \\ & \vdots & \\ f_0(x_m) & f_1(x_m) & f_2(x_m) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{bmatrix}$$