Today: Matching Pursuit for "solving" sparsity-constrained LS problems.

We have seen the LS problem:

$$\min_{\vec{x}} \|A\vec{x}-\vec{b}\|^2 \qquad (*)$$

which has elegant/concise solution in terms of normal equations

$\vec{x}$ is optimal in $(*)$

$\iff$ $\vec{x}$ solves $(A^TA)\vec{x} = A^T\vec{b}$.

In particular, if $N(A)=\{\vec{0}\}$,

then $\vec{x}^* = (A^TA)^{-1}A^T\vec{b}$.

Now, we are interested in solving:

$$\min_{\vec{x}} \|M\vec{x}-\vec{b}\|^2 \quad \text{subject to} \quad \overbrace{\|\vec{x}\|_0 \le k}^{\text{solution has } \le k \text{ nonzero entries}}.$$

"L0-norm" $\|\cdot\|_0$

$\|\vec{x}\|_0 = $ # of nonzero entries in $\vec{x}$.

Generally, this is a difficult problem (i.e., computationally intractable to find optimal solution)

For this reason, we resort to "heuristics"

<u>Assumption</u>: $M = [\vec{m}_1 \cdots \vec{m}_n]$,

columns are normalized $\|\vec{m}_i\| = 1$. (WLOG)

$$M\vec{x} = \sum_i \vec{m}_i x_i = \sum \underbrace{\frac{\vec{m}_i}{\|\vec{m}_i\|}}_{\text{unit vectors}} \underbrace{(\|\vec{m}_i\| x_i)}_{\text{rescaled coefficients}}$$

Often, but not always, the cols of M will be linearly dependent.

$\Rightarrow$ LS problem $\min_{\vec{x}} \|M\vec{x} - \vec{b}\|^2$ has infinitely many solutions, so we hope to identify a sparse one.

Last time: we motivated desirability of sparse solutions.

M is often called a "dictionary", so the job is to approximate $\vec{b}$ by linear comb. of "few" dictionary elements (cols of M).

Matching Pursuit is a "greedy" algorithm for approximately solving

$$\min_{\vec{x}} \|M\vec{x} - \vec{b}\|^2 \quad \text{s.t.} \quad \|\vec{x}\|_0 \leq k.$$

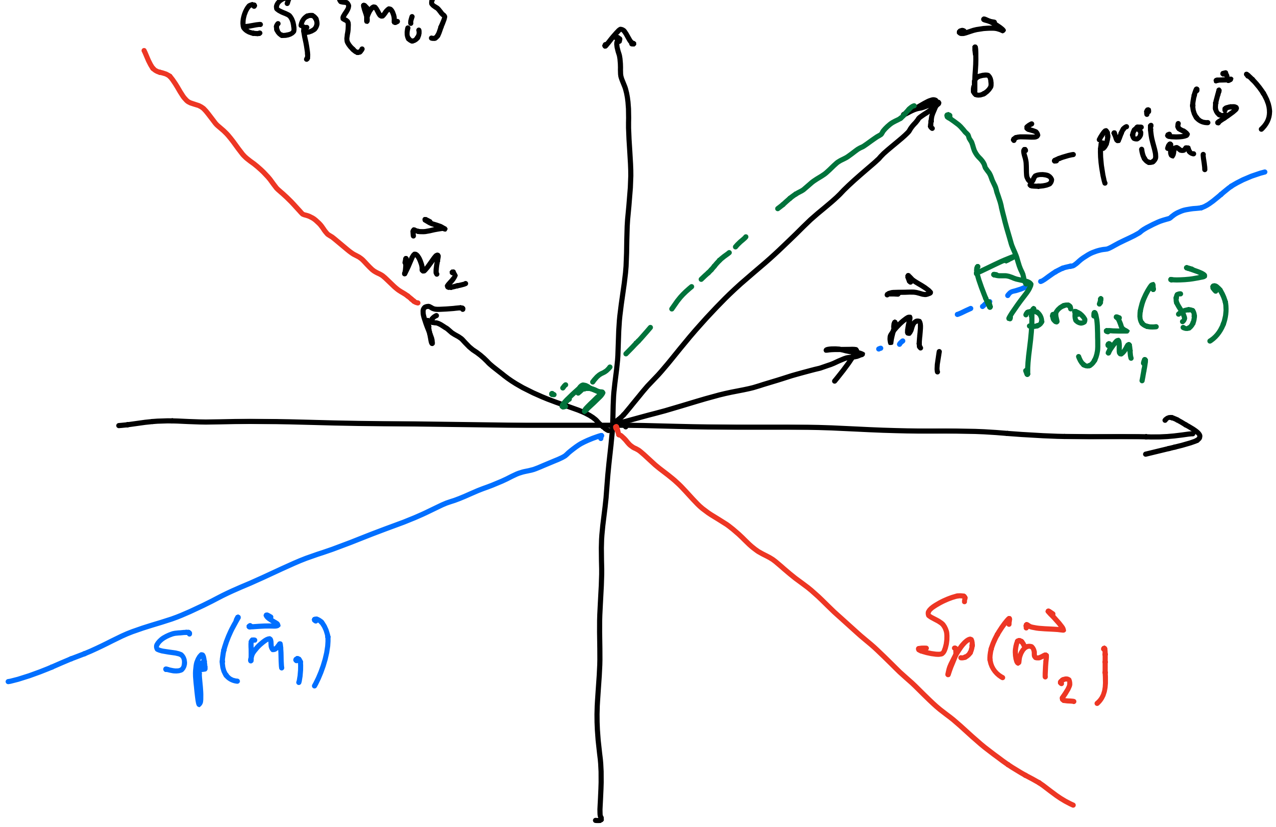
It is a heuristic (it works pretty well in practice, but no guarantees of optimality).

To describe MP, let's first try to solve a simpler problem.

$$\boxed{\min_{\vec{x}} \|M\vec{x}-\vec{b}\|^2 \quad \text{subject to} \quad \underbrace{\|\vec{x}\|_0 \le 1}_{\vec{x} \text{ has } \le 1 \text{ nonzero entry}}.}$$

Same as:

$$\min_{i,x} \|\underbrace{\vec{m}_i x}_{\in Sp\{\vec{m}_i\}} - \vec{b}\|^2 = \overbrace{\min_i}^{\text{min over columns}} \min_{\vec{v}\in Sp\{\vec{m}_i\}} \|\vec{v}-\vec{b}\|^2$$



In our simple problem, solution given by greedy search over columns to find subspace closest to $\vec{b}$.

Q: What is distance from $\vec{b}$ to $\mathrm{Span}(\vec{m}_1)$?

A: $\|\vec{b} - \mathrm{proj}_{\vec{m}_1}(\vec{b})\|^2 = \|\vec{b} - \dfrac{\langle \vec{b}, \vec{m}_1 \rangle}{\cancel{\langle \vec{m}_1, \vec{m}_1 \rangle} 1} \vec{m}_1 \|^2$

$= \|\vec{b}\|^2 + \underbrace{\|\vec{m}_1\|^2}_{1} - 2\langle \vec{b}, \vec{m}_1 \rangle \langle \vec{b}, \vec{m}_1 \rangle$

$= \|\vec{b}\|^2 + 1 - 2|\langle \vec{b}, \vec{m}_1 \rangle|^2$.

So, distance from $\vec{b}$ to $Sp(\vec{m}_i)$

$= \underbrace{\|\vec{b}\|^2 + 1}_{\text{constant}} - 2|\langle \vec{b}, \vec{m}_i \rangle|^2$.

Q: How to find which column of M has subspace $Sp(\vec{m}_i)$ closest to $\vec{b}$?

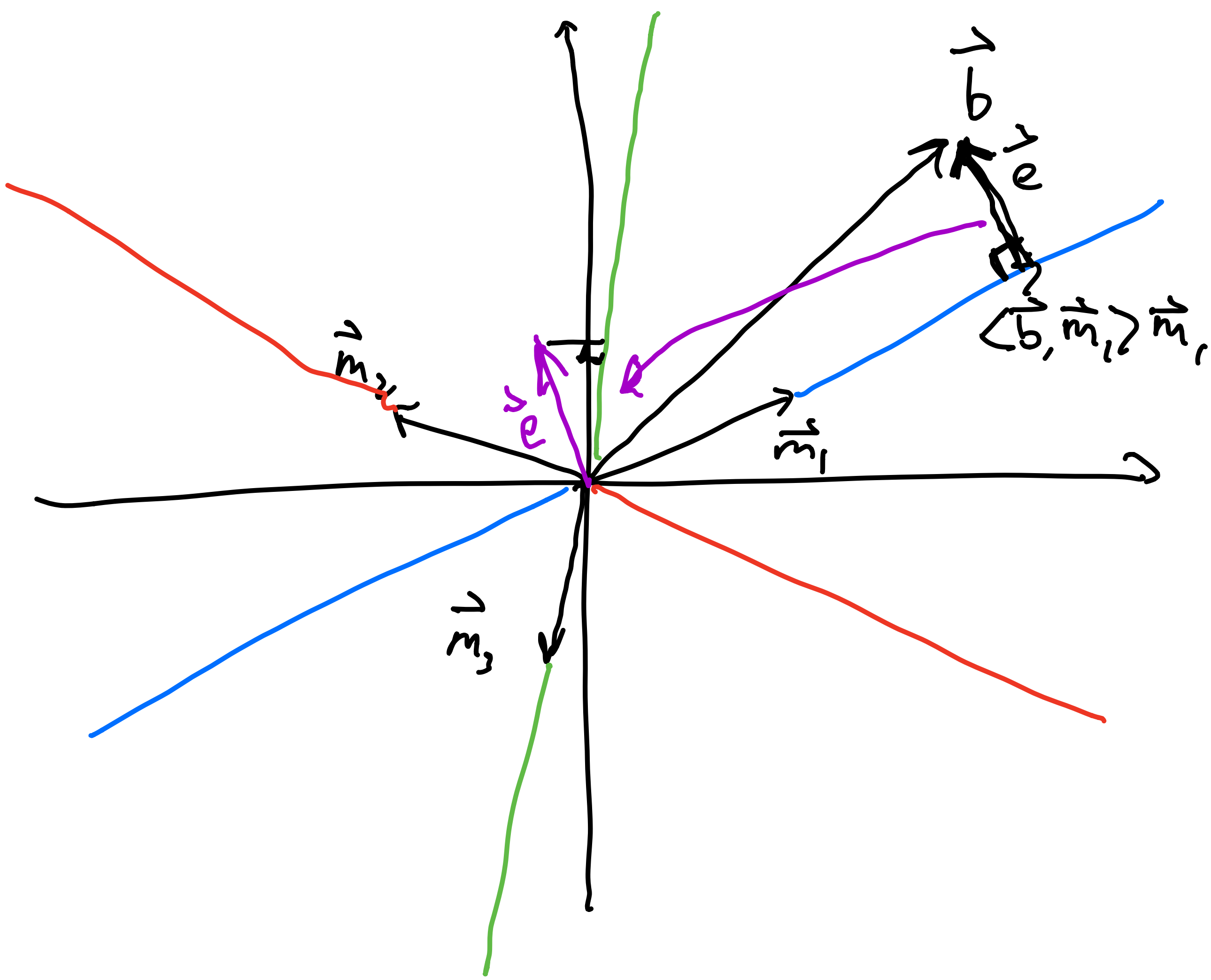
A: Find the column i that maximizes inner product $|\langle \vec{b}, \vec{m}_i \rangle|$

So: index i^* that minimizes

$$\min_i \min_{\vec{v} \in \mathrm{Span}\{\vec{m}_i\}} \|\vec{v} - \vec{b}\|^2$$

is the same index that maximizes

$$|\langle \vec{b}, \vec{m}_i \rangle|.$$



$$\vec{e} = \vec{b} - M \begin{bmatrix} \langle \vec{b}, \vec{m}_1 \rangle \\ 0 \\ 0 \end{bmatrix}$$

Now, we could repeat entire procedure, trying to approximate $\vec{e}$ by a multiple of a column of M.

Matching Pursuit

Given matrix M with columns $\vec{m}_1, \dots \vec{m}_m$, $\|\vec{m}_i\| = 1$, want to solve:

$$\min_{\vec{x}} \|M\vec{x} - \vec{b}\|^2 \quad \text{subject to } \|\vec{x}\|_0 \le k.$$

Initialize: $\vec{e} = \vec{b}$ and $\vec{x} = \vec{0}$.

For $j = 1, 2, \dots k$

Finds 1-sparse solution to $M\vec{x} \approx \vec{e}$. (ie, our simplified problem)

Find index i that maximizes $|\langle \vec{m}_i, \vec{e} \rangle|$.

Update $x_i \leftarrow x_i + \langle \vec{m}_i, \vec{e} \rangle$

Update $\vec{e} \leftarrow \vec{b} - M\vec{x}$

$$\left(= \vec{e} - \langle \vec{m}_i, \vec{e} \rangle \vec{m}_i \right)$$

end

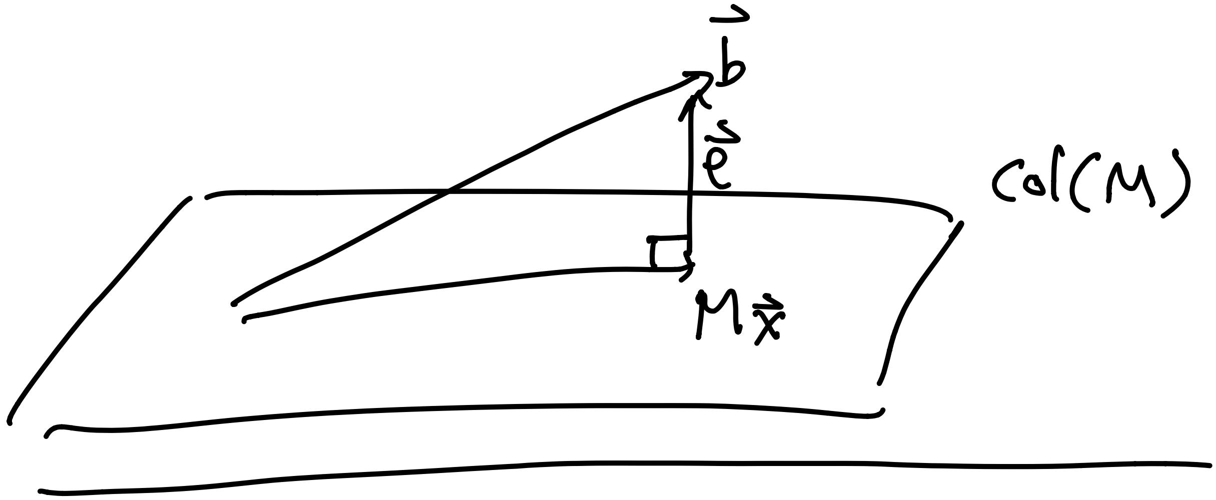
Observations: • Since there are k steps, will obtain $\vec{x}$ with $\|\vec{x}\|_0 \leq k$.

• The error (residual) decreases in norm at each step.

In particular, if $k = \infty$, we would only stop once $\langle \vec{e}, \vec{m}_i \rangle = 0 \ \forall i$.

$\Rightarrow$ $\vec{e}$ is orthogonal to $\mathrm{col}(M)$

in particular, $\vec{x}$ is the LS solution to $\min_{\vec{x}} \|M\vec{x} - \vec{b}\|^2$.

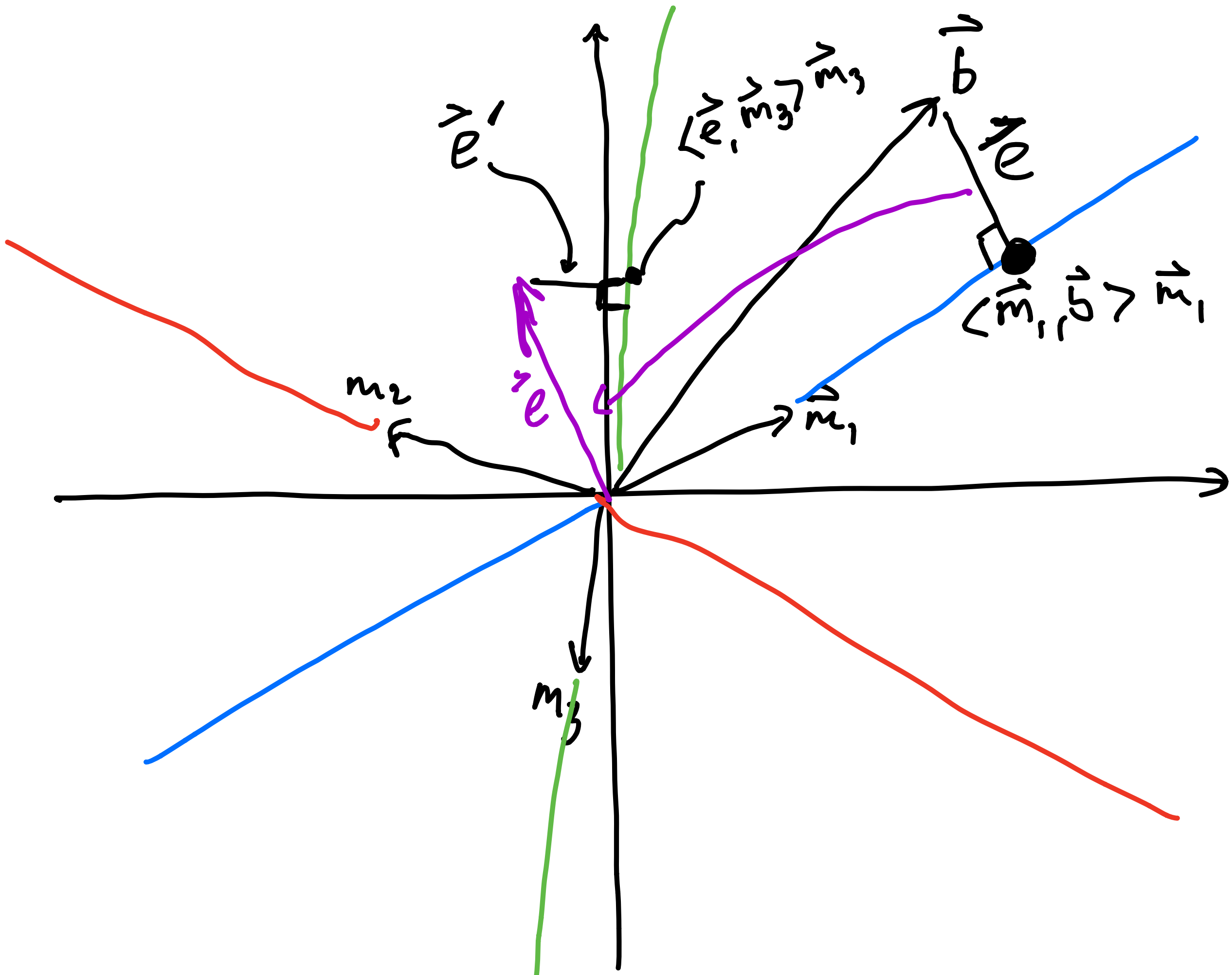


"Sparse" means few nonzero entries.

"$\vec{x}$ is k-sparse"

$\iff$ "$\|\vec{x}\|_0 \le k$"

$\iff$ "$\vec{x}$ has at most k nonzero entries".



<u>Step 1:</u> identify $Sp(\vec{m}_1)$ as being closest to $\vec{b}$.

residual error $\vec{e} = \underset{\sim}{b} - M \begin{bmatrix} \langle \vec{m}_1, \vec{b} \rangle \\ 0 \\ 0 \end{bmatrix}$

<u>Step 2:</u> Try to solve problem:

$$\min_{\vec{x}} \| M \vec{x} - \underset{\sim}{\vec{e}} \| \quad \text{s.t} \quad \| \vec{x} \|_0 \leq 1 .$$

we see that $Sp(\vec{m}_3)$ is closest,

new residual

$$\vec{e}\,' = \vec{e} - M \begin{bmatrix} 0 \\ 0 \\ \langle \vec{e}, \vec{m}_3 \rangle \end{bmatrix}$$

$$= \vec{b} - M \begin{bmatrix} \langle \vec{b}, \vec{m}_1 \rangle \\ 0 \\ \langle \vec{e}, \vec{m}_3 \rangle \end{bmatrix}.$$

<u>Remark</u>: variations are possible. E.g.,

Fix threshold τ, and run until

$\|\vec{e}\| \leq \tau$. (won't guarantee certain level of sparsity, but still yields sparse solution in practice).

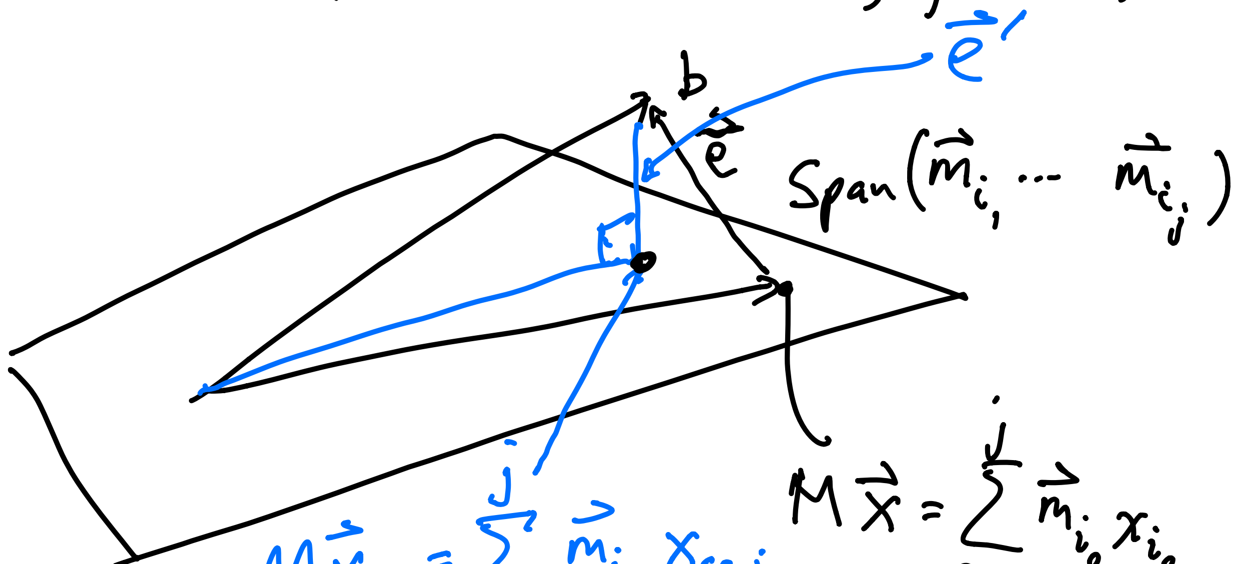
Orthogonal Matching Pursuit (OMP)

At each step of MP, we have a set of x_i's that are nonzero: Call them

$$x_{i_1}, x_{i_2}, \dots x_{i_j}$$

indices of the nonzero entries of $\vec{x}$

This corresponds to following picture:



$$M\vec{x}_{sp} = \sum_{\ell=1}^{j} \vec{m}_{i_\ell} x_{sp,i_\ell}$$

$$M\vec{x} = \sum_{\ell=1}^{j} \vec{m}_{i_\ell} x_{i_\ell}$$

$\vec{e}$ is not necessarily orthogonal to $Sp(\vec{m}_{i_1}, \dots \vec{m}_{i_j})$.

So, there is another vector $\vec{X}_{sp}$, which is just as sparse as $\vec{X}$, such that corresponding error $\vec{e}'$ has smaller norm.

<u>Q</u>: How do we find $\vec{X}_{sp}$?

<u>A</u>: Solve the corresponding LS problem (i.e. projecting $\vec{b}$ onto $Sp(\vec{m}_{i_1}, \dots \vec{m}_{i_j})$).

<u>OMP Algorithm</u>:

<u>Initialize</u> $\vec{e} = \vec{b}$, $A = [\ ]$.

<u>For</u> $j = 1, \dots K$

Find index i that maximizes $|\langle \vec{e}, \vec{m}_i \rangle|$

Update $A = [A \mid \vec{m}_i]$

Update $\vec{e} = \vec{b} - A\underbrace{(A^TA)^{-1}A^T\vec{b}}_{\text{LS solution}}$.

<u>end</u>

At end of procedure, we have

$$A = [\vec{m}_{i_1} \cdots \vec{m}_{i_k}]$$

Output $\vec{x}$ by defining

$$\begin{bmatrix} x_{i_1} \\ \vdots \\ x_{i_k} \end{bmatrix} = (A^TA)^{-1}A^T\vec{b}.$$