
We have seen the LS problem:

$$\min_{\tilde{x}} \| A \tilde{x} - \tilde{b} \|^2$$

(\*)

which has elegant/concise solution in terms of normal equations.

\(\tilde{x}\) is optimal in (\*):

(\implies) \(\tilde{x}\) solves \((A^T A) \tilde{x} = A^T \tilde{b}\).

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

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

Now, we are interested in solving:

$$\min_{\tilde{x}} \| M \tilde{x} - \tilde{b} \|^2 \text{ subject to } \| \tilde{x} \|_0 \leq k,$$

"\(0\)-norm" \(\| \cdot \|_0\)

\(\| \tilde{x} \|_0 = \# \text{ of nonzero entries in } \tilde{x}\).
Generally, this is a difficult problem (i.e., computationally intractable to find optimal solution).

For this reason, we resort to "heuristics."

Assumption: \( M = [\vec{m}_1, \ldots, \vec{m}_n] \), columns are normalized \( \| \vec{m}_i \| = 1 \), (WLOG)

\[
M \hat{x} = \sum \vec{m}_i x_i = \sum \frac{\vec{m}_i}{\| \vec{m}_i \|} (\| \vec{m}_i \| x_i)
\]

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

\[ \Rightarrow \] LS problem \( \min \| M \hat{x} - b \|^2 \) 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 $\frac{1}{b}$ by linear comb. of "few" dictionary elements (cols of $M$).

Matching Pursuit is a "greedy" algorithm for approximately solving

$$\min_{x} \| Mx - b \|_2 \text{ s.t. } \|x\|_0 \leq k.$$  

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.
\[
\min_{\vec{x}} \| M \vec{x} - \vec{b} \|^2 \quad \text{subject to} \quad \| \vec{x} \|_\infty \leq 1,
\]
\[\vec{x} \text{ has } \leq 1 \text{ nonzero entries}.\]

Same as:
\[
\min_{i, \vec{x}} \| \vec{m}_i \vec{x} - \vec{b} \|^2 = \min_{i, \vec{x}} \min_{\vec{v} \in \text{Sp} \{ \vec{m}_i \}} \| \vec{v} - \vec{b} \|^2
\]

In our simple problem, solution given by greedy search over columns to find
Subspace closest to \( \mathbf{b} \).

Q: What is distance from \( \mathbf{b} \) to \( \text{Span}(\mathbf{\hat{m}}_i) \)?

A: \[
\| \mathbf{b} - \text{proj}_{\mathbf{\hat{m}}_i}(\mathbf{b}) \|^2 = \| \mathbf{b} - \frac{\langle \mathbf{b}, \mathbf{\hat{m}}_i \rangle}{\langle \mathbf{\hat{m}}_i, \mathbf{\hat{m}}_i \rangle} \mathbf{\hat{m}}_i \|^2
= \| \mathbf{b} \|^2 + \| \mathbf{\hat{m}}_i \|^2 - 2 \langle \mathbf{b}, \mathbf{\hat{m}}_i \rangle \langle \mathbf{\hat{m}}_i, \mathbf{\hat{m}}_i \rangle
= \| \mathbf{b} \|^2 + 1 - 2 |\langle \mathbf{b}, \mathbf{\hat{m}}_i \rangle|^2.
\]

So, distance from \( \mathbf{b} \) to \( \text{Span}(\mathbf{\hat{m}}_i) \)

\[
= \sqrt{\| \mathbf{b} \|^2 + 1 - 2 |\langle \mathbf{b}, \mathbf{\hat{m}}_i \rangle|^2}.
\]

Q: How to find which column of \( \mathbf{M} \) has subspace \( \text{Span}(\mathbf{\hat{m}}_i) \) closest to \( \mathbf{b} \)?

A: Find the column \( \mathbf{\hat{m}}_i \) that maximizes inner product \( |\langle \mathbf{b}, \mathbf{\hat{m}}_i \rangle| \).
So: index $i^*$ that minimizes
\[
\min \min_i \| \hat{V} - \hat{b} \|^2
\]
is the same index that maximizes
\[
|\langle \hat{b}, \hat{m}_i \rangle |.
\]
\[ \hat{e} = \hat{b} - M \left[ \begin{array}{c} \langle \hat{b}, \hat{m}_1 \rangle \\ 0 \\ 0 \end{array} \right] \]

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

**Matching Pursuit**

Given matrix \( M \) with columns \( \hat{m}_1, ..., \hat{m}_n \), \( \| \hat{m}_i \|_2 = 1 \), want to solve:

\[
\min_{\hat{x}} \| M \hat{x} - \hat{b} \|_2^2 \quad \text{subject to} \quad \| \hat{x} \|_0 \leq k.
\]

Initialize: \( \hat{e} = \hat{b} \) and \( \hat{x} = 0 \).

For \( j = 1, 2, ..., K \)
Finds 1-sparse solution to $M \hat{x} = \hat{b}$. (ie, our simplified problem)

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

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

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

\[ = \hat{e} - \langle \hat{m}_i, \hat{e} \rangle \hat{m}_i \]

Observations:

1. Since there are $K$ steps, we obtain $\hat{x}$ with $\| \hat{x} \|_0 \leq K$.
2. The error (residual) decreases in norm at each step.

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

$\Rightarrow \hat{e}$ is orthogonal to $\text{col}(M)$.

In particular, $\hat{x}$ is the LS solution to $\min_x \| M \hat{x} - \hat{b} \|_2^2$. 
"Sparse" means few nonzero entries.

"\( \hat{X} \) is \( k \)-sparse"\

\( \iff \) "\( \| \hat{X} \|_0 \leq k \)"

\( \iff \) "\( \hat{X} \) has at most \( k \) nonzero entries"
Step 1: Identify $Sp(m_1)$ as being closest to $b$. 

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

Step 2: Try to solve problem: 

$$\min_{\hat{x}} \| M \hat{x} - \hat{e} \| \text{ s.t. } \| \hat{x} \|_0 \leq 1.$$
we see that $Sp(\tilde{m}_3)$ is closest, new residual
\[
\tilde{e}' = \tilde{e} - M \begin{bmatrix} 0 \\ 0 \\ \langle \tilde{e}, \tilde{m}_3 \rangle \end{bmatrix}
\]
\[
= \tilde{e} - M \begin{bmatrix} \langle \tilde{e}, \tilde{b} \rangle \\ \langle \tilde{e}, \tilde{m}_1 \rangle \\ \langle \tilde{e}, \tilde{m}_3 \rangle \end{bmatrix}
\]

Remark: variations are possible, E.g., Fix threshold $\tau$, and run until $\|\tilde{e}\| \leq \tau$. (Must 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

\[
X_{i_1}, X_{i_2}, \ldots, X_{i_j}
\]

indices of the nonzero entries of \( x \).

This corresponds to following picture:

\[
M \hat{x} = \sum_{l=1}^{j} \hat{m}_{i_l} X_{sp, i_l}
\]

\[
M \hat{X} = \sum_{l=1}^{j} \hat{m}_{i_l} x_{i_l}
\]

\[
\text{Span}(\hat{m}_{i_1}, \ldots, \hat{m}_{i_j})
\]
\( \hat{e} \) is not necessarily orthogonal to
\( \text{Sp}(\hat{m}_i, ..., \hat{m}_j) \).
So, there is another vector \( \hat{x}_{sp} \),
which is just as sparse as \( \hat{x} \),
such that corresponding error \( \hat{e}' \) has
smaller norm.

Q: How do we find \( \hat{x}_{sp} \)?
A: Solve the corresponding LS
problem (i.e. projecting \( \hat{b} \) onto)
\( \text{Sp}(\hat{m}_i, ..., \hat{m}_j) \).

**OMP Algorithm:**

Initialize \( \hat{e} = \hat{b} \), \( A = [ ] \).

For \( j = 1, ... K \)

Find index \( i \) that maximizes \( |\langle \hat{e}, \hat{m}_i \rangle| \).
Update $A = [A \mid \vec{m}_i]$

Update $\hat{e} = \hat{b} - A(A^tA)^{-1}A^t\hat{b}$. (LS solution)

end

At end of procedure, we have

$A = [\vec{m}_i, \ldots, \vec{m}_{i_k}]$

Output $\vec{x}$ by defining

$$
\begin{bmatrix}
\vec{x}_{i_1} \\
\vdots \\
\vec{x}_{i_{k-1}}
\end{bmatrix} = (A^tA)^{-1}A^t\hat{b}.
$$