\[ \ell^1 \text{-norm minimization} \]

Consider

\[
(P_1) \quad \min_x \| x \|_1, \\
\text{s.t. } A x = b
\]

How many solutions are there? 
(or is the solution for $P_1$ unique?)

Note that the norm $\| x \|_1$ is convex but not strictly convex. Therefore, if $\hat{x}_1$ and $\hat{x}_2$ are both solutions of $(P_1)$, and suppose $\text{sign}(\hat{x}_1) = \text{sign}(\hat{x}_2)$, then, any linear combination

\[ \hat{x} = \alpha \hat{x}_1 + (1-\alpha) \hat{x}_2, \quad \alpha \in [0,1] \]

will have the same sign as $\hat{x}_1$ and $\hat{x}_2$, and

\[
\| \hat{x} \|_1 = \| \alpha \hat{x}_1 + (1-\alpha) \hat{x}_2 \|_1 \\
= \alpha \| \hat{x}_1 \|_1 + (1-\alpha) \| \hat{x}_2 \|_1 \\
= \alpha \| \hat{x}_1 \|_1 + (1-\alpha) \| \hat{x}_1 \|_1, \quad \text{both are minimizers}
\]
Therefore, the solution of \((P_i)\) is **not**

unique.

What if \(\hat{x}_1\) and \(\hat{x}_2\) do not have same

sign ?

If \(\hat{x}_1\) and \(\hat{x}_2\) do not have same sign,

we can show that

\[
\|\hat{x}\|_1 = \|\alpha \hat{x}_1 + (1-\alpha) \hat{x}_2\|_1
\]

\[
\leq \alpha \|\hat{x}_1\|_1 + (1-\alpha) \|\hat{x}_2\|_1,
\]

by convexity of \(\|\cdot\|_1\).

\[
= \alpha \|\hat{x}_1\|_1 + (1-\alpha) \|\hat{x}_1\|_1
\]

\[
= \|\hat{x}_1\|_1.
\]

But since \(\hat{x}_1\) is already a minimizer,

it is impossible to have \(\hat{x}\) having

\[
\|\hat{x}\|_1 < \|\hat{x}_1\|_1,
\]

or otherwise \(\hat{x}\) is a better minimizer.

Therefore, we must have \(\|\hat{x}\|_1 = \|\hat{x}_1\|_1\).

Moreover, if we let

\[
\mathcal{R} = \{\hat{x} \mid \hat{x} = \text{argmin}_x \|x\|_1 \text{ s.t. } Ax = b\},
\]

then \(\mathcal{R}\) is convex.
Among all the solutions in \( \mathcal{S} \), is there anyone which has at most \( n \) non-zeros?

Suppose that \( \mathbf{x} \) is a solution and \( \mathbf{x} \) has \( k > n \) non-zeros.

Then, the \( k \) columns must be linearly dependent. Consequently, \( \exists \mathbf{h} \neq 0 \) s.t.

\[
A \mathbf{h} = 0 \quad \text{(support}( \mathbf{h}) \subset \text{support}(\mathbf{x}))
\]

Let \( \mathbf{x} = \mathbf{x}^* + \varepsilon \mathbf{h} \), such that

\[
\text{Sign}(\mathbf{x}) = \text{Sign}(\mathbf{x}).
\]

**How to make sure this?**

\[
\text{let } |\varepsilon| < \min_i \left| \frac{\hat{x}_i}{|h_i|} \right| \tag{\ref{c}}
\]

Then two things will happen:

1. \( A\mathbf{x} = \mathbf{b}, \)

   because \( A(\hat{\mathbf{x}} + \varepsilon \mathbf{h}) = A\hat{\mathbf{x}} + \varepsilon A\mathbf{h} = \mathbf{b} \cdot \varepsilon = 0 \)

2. \( \forall |\varepsilon| < \min_i \left| \frac{\hat{x}_i}{|h_i|} \right|, \)

\[
\| \mathbf{x} \|_1 = \| \mathbf{x}^* + \varepsilon \mathbf{h} \|_1 \geq \| \mathbf{x}^* \|_1,
\]

because \( \hat{\mathbf{x}} \) is optimal.
Interestingly, the above inequality is an equality if we want the inequality to hold for all \( \varepsilon \) satisfy (6).

\[ \| \mathbf{x} + \varepsilon \mathbf{h} \|_1 \geq \| \mathbf{x} \|_1 \text{ for all } \varepsilon \in (-\infty, 0) \]

is to have \( h = 0 \), and

\[ \| \mathbf{x} \|_1 = \| \mathbf{x} + \varepsilon \mathbf{h} \|_1. \]

If we move along this slope line then \( l_1 \) norm is preserved, regardless of the \( \varepsilon \) or \(-\varepsilon\).

If we move along this direction, then \( +\varepsilon \) \( \varepsilon \) gives \( \| \mathbf{x} \|_1 \geq \| \mathbf{x} \|_1 \), but \(-\varepsilon \) \( \varepsilon \) gives \( \| \mathbf{x} \|_1 \leq \| \mathbf{x} \|_1 \).

So impossible.
Therefore, we must have
\[ \| \hat{X} + \varepsilon h \|_1 = \| \hat{X} \|_1. \]

Consequently, we have
\[
\| \hat{X} \|_1 = \| \hat{X} + \varepsilon h \|_1 = \sum_{i=1}^{m} (\hat{X}_i + \varepsilon h_i) \text{sgn}(\hat{X}_i)
= \sum_{i=1}^{m} \hat{X}_i \text{sgn}(\hat{X}_i) + \varepsilon \sum_{i=1}^{m} h_i \text{sgn}(\hat{X}_i)
\]
\[
= \| \hat{X} \|_1.
\]

So
\[
\sum_{i=1}^{m} h_i \text{sgn}(\hat{X}_i) = 0 \quad (\iff h^T(\text{sgn}(X)) = 0).
\]

Pictorially, it means \( h \) has to align with the -1 slope line (if \( X \) is all five).

Now, our next step is to null one entry of \( \hat{X} \) and maintain the \( \| \cdot \|_1 \).

Choose \( \varepsilon = -\frac{\hat{X}_{i^*}}{h_{i^*}} \) where \( i^* = \arg \min_i \frac{|\hat{X}_i|}{|h_i|} \)

then, \( (\hat{X} + \varepsilon h)_{i^*} = 0 \) after this choice of \( \varepsilon \). But since the sign of \( X \) is preserved, the norm is also maintained:
\[ \| \hat{X} + \varepsilon h \|_1 = \| \hat{X} \|_1. \]
So, now we have at most \((k-1)\) non-zeros.

Then, repeat, until we get \(k=n\).

\[
\text{find } h \text{ s.t. } Ah=0 \text{, where the support of } h \text{ now becomes smaller.}
\]

Bottom line message:

(1) \(P_i\) does promote sparsity.

(2) But our theory is weak:

   we can only show that there is one solution whose \# non-zero is at most \(n\). \(n\) is very dense.

(3) if \# non-zero of \(X\) is actually < \(n\),

    then this theory does not help.

   we need something stronger.