Lecture 19

From last time:

The T-reg’d soln satisfies

\[
\min \{ \|Ax-m\|_2^2 + \delta \|Lx\|_2^2 \}
\]

We know from L-S theory that the minimizer of

\[
\|Ax-b\|_2 \quad \text{satisfies the normal eqns} \quad A^TAx = A^Tm.
\]

Regarding \((*)\) as two simultaneous minimization problems

\[
Ax = m \quad (A \in \mathbb{R}^{M \times n}, \ m \in \mathbb{R}^{M \times 1})
\]

and

\[
Lx = 0 \quad (L \in \mathbb{R}^{k \times n}, \ O \in \mathbb{R}^{k \times 1})
\]

We can write these in "stacked form" as

\[
\begin{bmatrix}
A \\
\sqrt{\delta} \ L
\end{bmatrix} x =
\begin{bmatrix}
m \\
0
\end{bmatrix}
\]

\((M+k) \times n \quad (M+k) \times 1\)

This is equivalent to the orig. min. prob. since,

defining \(\widetilde{A} = \begin{bmatrix} A \\ \sqrt{\delta} \ L \end{bmatrix}\) and \(\widetilde{m} = \begin{bmatrix} m \\ 0 \end{bmatrix}\),

\(\widetilde{A}^T \widetilde{A} = A^TA + \delta LTL\)

and

\(\widetilde{A}^T \widetilde{m} = A^Tm\)

So

\((A^TA + \delta LTL)x = A^Tm\)

ie,

\((A^TA + \delta LTL)x = -A^Tm + \sqrt{\delta} \ L \cdot 0\)

Note on the generalized T-reg:

For a mix \(L \in \mathbb{R}^{p \times n} + A \in \mathbb{R}^{m \times n}, m \geq n \geq p\),

where \(N(A) = N(L) \equiv \{0\} + L \) has full row rank,

the filter factors are given by
\[ s_i = \frac{x_i^2}{y_i^2 + s^2} \]

where the \( x_i \) are the generalized singular values of the matrix pair \((A, L)\). They are defined as the ratio \( y_i = \frac{s_i}{\mu_i} \), \( i = 1, \ldots, p \).

where the \( s_i \) are the singular values of \( A \) and the \( \mu_i \) of \( L \) in the GSVD:

\[
A = U [\Sigma 0] X^{-1}, \quad L = V [M, 0] X^{-1}
\]

The cols of \( U \in \mathbb{R}^{m \times m} \) and \( V \in \mathbb{R}^{p \times p} \) are orthonormal, \( U^T U = I_m, \quad V^T V = I_p \), \( X \in \mathbb{R}^{m \times n} \) is nonsingular, \( \text{diag}(\Sigma) = s_1 \geq \ldots \geq s_p \) (\( \Sigma \) is diag'd), \( \text{diag}(M) = \mu_1 \geq \ldots \geq \mu_p \) (\( M \) is diag'd), \( 0 \leq s_i \leq \ldots \leq s_p \leq 1 \)

\( 1 \geq \mu_1 \geq \mu_2 \geq \ldots \geq \mu_p > 0 \)

\( s_i^2, \mu_i^2 = 1 \), \( i = 1, \ldots, p \)

The choice of the regularization parameter

This is problem-dependent, often solved empirically. However, there are textbook mtds that can be effective.

Three popular mtds are

1. Morozov discrepancy principle
2. L-curve mtd
3. Generalized cross-validation
Morozov's discrepancy principle

Suppose we have an estimate of the error in the data:
\[ \| e \|_2 = \| b - b^{ex} \|_2 \]

Then: If an operator is compact + $\infty$-dim', then it is not possible to construct a convergent parameter choice rule that does not make explicit use of $\| e \|_2$. (Ref: Hanke, Engl, Neubauer)

Note: in practical apps' prob's are finite-dim'.

Idea: If $Ax^{ex} = b^{ex}$, choose $\delta$ s.t.
\[ \| Ax_\delta - b \|_2 = \epsilon_e, \quad \text{where} \quad \| e \|_2 < \epsilon_e. \]

Then: MDP gives a unique choice for $\delta > 0$ iff $\epsilon_e$ satisfies
\[ \| P b \|_2 < \epsilon_e < \| b \|_2 \]
where $P$ is the orthog. proj. onto $(\text{Im} A)^\perp$.

Pf:
We have $Ax_\delta = U \Sigma V^T V \Sigma_\delta^T U^T m$
\[ = U \Sigma \Sigma_\delta^T m. \]

So we have
\[ \| Ax_\delta - m \|_2 = \| P P_\delta^T m - m \|_2 \]
\[ = \sum_{j=1}^{\min(m,n)} \left( \frac{\delta_j^2}{\delta_j^2 + \delta} - 1 \right)^2 m_j^2 + \sum_{j=\min(m,n)+1}^{m} m_j^2 \]
\[ = \sum_{j=1}^{r} \left( \frac{\delta_j^2}{\delta_j^2 + \delta} \right)^2 m_j^2 + \sum_{j=\min(m,n)+1}^{m} m_j^2 \]

This shows that the mapping $\delta \mapsto \| Ax_\delta - m \|_2$ is mon. incr.
and
\[ \sum_{j=r+1}^m m_j' \leq \|Ax_0 - m\|^2 \leq \|Ax'_e - m\|^2 \]
\[ \leq \lim_{s \to \infty} \|Ax_s - m\| = \sum_{j=1}^m m_j'^2 \]

The claim follows from orthog. of \( U \).

Numerical implementation:
Find the unique zero of
\[ f(\delta) = \sum_{j=1}^r \left( \frac{\delta}{\delta_j^2 + \delta} \right)^2 m_j' + \sum_{j=r+1}^m m_j'^2 - \delta^2 \]

Discussed & defined L-curve mtd