Lecture 21

Generalized Cross-Validation

The GCV function is defined to be
\[
G(s) = \frac{\| \hat{A} x_g - b \|_2^2}{\text{trace} \left( \mathbf{I}_m - \hat{A} A^H \right)^2} = \frac{\nu(s)}{\tau(s)}
\]

(Seek to minimize \( G(s) \))

To understand GCV, we introduce the concept of the resolution matrix:

If \( A^H \) is the regularized inverse of \( A \)
(i.e., \( x_{\text{reg}} = A^H b \)), then define the resolution matrix \( \Xi \) by
\[
\Xi = A^H A.
\]

Note: In special case of TSV, \( A^H = A_k^T = \sum_{i=1}^{k} u_i \sigma_i v_i^T \)

so \[
\Xi = (U_k \Sigma_k V_k^T)^{-1} (U \Sigma V) = U_k \Sigma_k V_k^T
\]

(see \( \Xi = \text{proj onto } \mathcal{R}(A_k^T) \))

If \( b = b_{\text{exact}} + e = A x_{\text{exact}} + e \) then
\[
x_{\text{reg}} = A^H b = \hat{A}^H (A x_{\text{exact}} + e) = \Xi x_{\text{exact}} + A^H e
\]

\( \Xi x_{\text{exact}} \) is a regularized version of \( x_{\text{exact}} \)
\( A^H e \) is the contribution to \( x_{\text{reg}} \) from the noise in \( b \).

So the pure reg error in \( x_{\text{reg}} \) is \( (I_n - \Xi) x_{\text{exact}} \).
So \( \Xi \) describes how well, in noise-free case, \( x_{\text{exact}} \) is approx. by \( A^H b_{\text{exact}} \). (Note \( \Xi \) is not diagonal)
In the statistical framework,
\[ E(x_{\text{reg}}) = \sum x_{\text{exact}} \]

If \( b = b_{\text{exact}} \) has els that are unbiased + uncorrelated with covariance \( m \times \sigma_0^2 I_n \), then the expected value of \( ||e||_2 \) satisfies
\[ E(||e||_2^2) = m \sigma_0^2 + E(1u_i^T e_1) = \sigma_i^2, \quad i = 1, \ldots, n \]

In this case, one can show
\[ ||Ax_5 - b||_2^2 = (||e||_2^2 - \sigma_0^2 \text{ trace } (AA^*))^{1/2} \]
\[ = \sigma_0 (m - \rho(\delta))^{1/2} \]
where \( \rho(\delta) \) is the sum of the filter factors:
\[ \rho(\delta) = \sum_{i=1}^{k} f_i \quad \text{if } L = I_n \]
\[ + \rho(\delta) = n - k + \sum_{i=k+1}^{K} f_i \quad \text{if } (L \neq I_n) \]

in this case \( f_i = 1 \) for \( i = k+1, \ldots, n \)

Side note: One can define the effective numerical rank \( r_{\text{eff}} \) of the ill-posed prob. by
\[ r_{\text{eff}} = \rho(S_{\text{opt}}) \]
where \( S_{\text{opt}} \) is the value of reg. par. \( S \) that minimizes
\[ ||x_{\text{exact}} - x_5||_2 \text{ or } ||Bx_{\text{exact}} - Lx_5||_2 \]

If \( r_{\text{eff}} \) is the closest integer to \( r_{\text{eff}} \), then
\[ r_{\text{eff}} \approx \# \text{ of svd components in } x_{\text{sopt}} \]
we can recover. So we define the effective resolution
\[ \text{limit } N_{\text{res}} \text{ by } N_{\text{res}} \approx \sqrt{\frac{r_{\text{eff}}}{r_{\text{eff}}}} ||x_{\text{exact}}||_1 \]
\[ \|Ax_0 - b\|_2 = \sigma_0 \left( m - n + k - \sum_{i=1}^{k} f_i \right)^{1/2} \]

\[ \text{trace } (I_m - AA^+) = m - n + k - \sum_{i=1}^{k} f_i = (Y(s))^2 \]

(Note \( T(s) \) is mon. incr.)

So in this case \( G(s) = \frac{\sigma_0}{(m - \rho(s))^{1/2}} \)

\( m - \rho(s) \) can be regarded as the "effective number of degrees of freedom" in the prob.

Graph of \( Y \) vs. \( s \) looks like a flat part at a point increasing with \( s \). The flat part is approx \( \sigma_0^2 \). So \( \|e\|_2^2 \approx m \) (value of \( Y \) at flat part)

\[ \min \text{ GCV func.} \]

prob: The unique min of the GCV func can be very flat. \( \Rightarrow \) numerical difficulties

Num. impl. concerns can be found in [Hansen]