Statistics 860 Lecture 9
See hw8.

1. The minimizer $\widehat{\lambda}$ of $V(\lambda)$ is a good estimator of the minimizer λ^* of $R(\lambda)$, the predicted mean square error.

2. Convergence rates of $R(\lambda^*)$.

$$y_i = L_i f + \varepsilon_i, \qquad i = 1, .., n, \qquad \varepsilon \sim N\left(0, \sigma^2 I\right)$$

$L_i f = f(t_i)$ or other bounded linear functional

$g = (L_1 f, ..., L_n f)'$, f_λ is the minimizer in $\mathcal{H}$ of

$$\frac{1}{n} \sum_{i=1}^{n} (y_i - L_i f)^2 + \lambda \|P_1 f\|^2$$

$$\begin{pmatrix} L_1 f_\lambda \\ ... \\ L_n f_\lambda \end{pmatrix} = A(\lambda) y$$

GCV estimate of λ is the minimizer of

$$
\begin{aligned}
V(\lambda) &= \frac{1}{n}\sum_{k=1}^{n}\frac{(y_k - L_k f_\lambda)^2}{\left(1-\frac{1}{n}\sum_{l=1}^{n} a_{ll}(\lambda)\right)^2} \\
&\equiv \frac{\frac{1}{n}\left|\left|(I - A(\lambda))\, y\right|\right|^2}{\left(\frac{1}{n} tr\,(I - A(\lambda))\right)^2} \\
&= n\frac{RSS(\lambda)}{(n - dfsig(\lambda))^2}
\end{aligned}
$$

$$
traceA(\lambda) = dfsignal(\lambda)
$$

$R(\lambda) =$ predictive mean square error when λ used

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(L_i f_\lambda - L_i f\right)^2 &= \frac{1}{n}\,||A(\lambda)\,y - g||^2 \\
&= \frac{1}{n}\,||A(\lambda)\,(g+\varepsilon) - g||^2 \\
&= \frac{1}{n}\{||(I - A(\lambda))\,g||^2 \\
&\quad -2g'A(\lambda)\,\varepsilon + \quad ||A(\lambda)\,\varepsilon||^2\}
\end{aligned}
$$

$$
\begin{aligned}
ER(\lambda) &= \frac{1}{n}\,||(I - A(\lambda))\,g||^2 + \frac{1}{n}\sigma^2 tr A^2(\lambda) \\
&\equiv \underbrace{b^2(\lambda)}_{squared\,bias} + \underbrace{\sigma^2 Var(\lambda)}_{+\ variance}
\end{aligned}
$$

$E\varepsilon' A^2(\lambda)\varepsilon$:

$$\begin{aligned} E\varepsilon' B\varepsilon &= E\sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_i\varepsilon_j b_{ij} \\ &= E\sum_{i=1}^{n}\varepsilon_i^2 b_{ii} \\ &= \sigma^2 trB \end{aligned}$$

Target Criteria: min

$$R(\lambda) = \frac{1}{n}\sum(L_i f_\lambda - L_i f)^2$$

Go after

$$ER(\lambda) = b^2(\lambda) + \sigma^2 Var(\lambda)$$

$$\begin{aligned} \frac{1}{n}trA^2(\lambda) &= \mu_2(\lambda) = \frac{1}{n}\sum_{v=1}^{n} eig_v^2(A) \\ \frac{1}{n}trA(\lambda) &= \mu_1(\lambda) = \frac{1}{n}\sum_{v=1}^{n} eig_v(A) \end{aligned}$$

Other target criteria: i.e.

$$\int_T (f_\lambda(t) - f(t))^2 \, dt$$

GCV may or may not be good for other target criteria, depending on $\mathcal{H}, f, \{L_i\}$ (See `insens.pdf`, ”When is the optimal regularization parameter insensitive to the choice of the loss function?”, GW and Wang, 1990)

$R(\lambda) =$ predictive mean square error when λ used

$$ER(\lambda) = b^2(\lambda) + \sigma^2 Var(\lambda)$$

$$EV(\lambda) = \frac{\frac{1}{n} E \, ||(I - A(\lambda))(g + \varepsilon)||^2}{(1 - \mu_1(\lambda))^2}$$

$$\begin{aligned} \frac{1}{n} E \, ||(I - A(\lambda))(g + \varepsilon)||^2 &= \frac{1}{n} ||(I - A(\lambda)) g||^2 \\ &+ \frac{\sigma^2}{n} tr (I - A(\lambda))^2 \end{aligned}$$

$$EV(\lambda) = \frac{\frac{1}{n}||(I-A(\lambda))g||^2 + \frac{\sigma^2}{n}tr(I-A(\lambda))^2}{(1-\mu_1(\lambda))^2}$$

$$= \frac{b^2(\lambda) + \sigma^2(1-2\mu_1(\lambda)+\mu_2(\lambda))}{(1-\mu_1(\lambda))^2}$$

$$\frac{ER(\lambda) - \left[EV(\lambda) - \sigma^2\right]}{ER(\lambda)}$$

$$\equiv \frac{-\mu_1(2-\mu_1)}{(1-\mu_1)^2} + \frac{\sigma^2}{b^2+\sigma^2\mu_2}\frac{\mu_1^2}{(1-\mu_1)^2}$$

$$\leq h(\lambda)$$

where

$$h(\lambda) = \left[2\mu_1(\lambda) + \frac{\mu_1^2(\lambda)}{\mu_2(\lambda)}\right]\frac{1}{(1-\mu_1(\lambda))^2}$$

$$\mu_2(\lambda) \geq \mu_1^2(\lambda) \quad , \quad EV(\lambda) \geq \sigma^2 \Rightarrow$$

$$\frac{\sigma^2(1-2\mu_1(\lambda)+\mu_2(\lambda))}{\left(1-2\mu_1(\lambda)+\mu_1^2(\lambda)\right)} \geq \sigma^2$$

Get

$$ER(\lambda)(1-h(\lambda)) \leq EV(\lambda) - \sigma^2 \leq ER(\lambda)(1+h(\lambda))$$

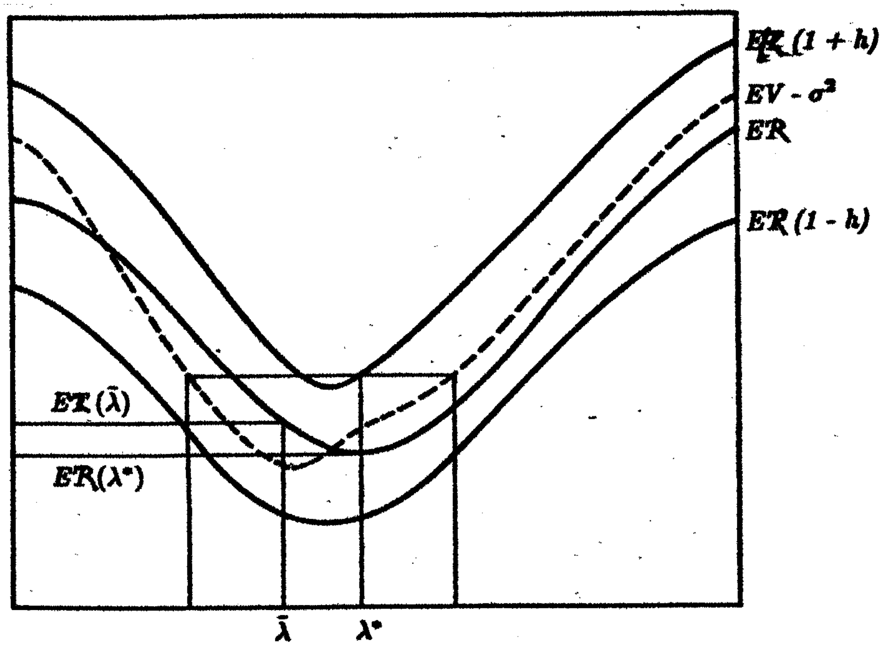


FIG. 4.9. *Graphical suggestion of the proof of the weak GCV theorem.*

$h(\lambda)$ will be small if $\mu_1(\lambda)$ and μ_1^2/μ_2 are small

$$\mu_1 = \frac{1}{n} tr A(\lambda), \qquad \mu_2(\lambda) = \frac{1}{n} tr A^2(\lambda)$$

(Expectation) Inefficiency I is defined by

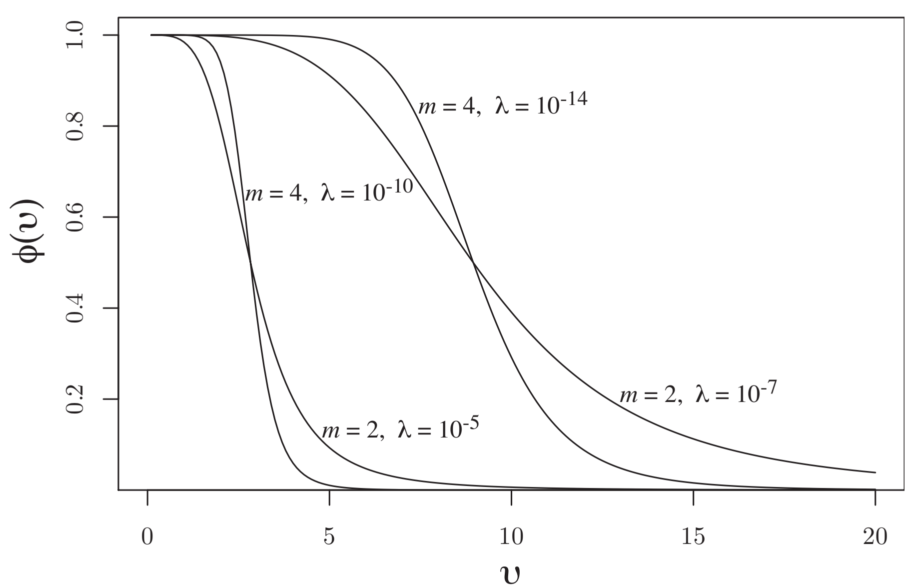
$$I = \frac{ER(\widehat{\lambda})}{ER(\lambda^*)}.$$

Then, if $h(\lambda)$ goes to 0, then $I \downarrow 1$ as $n \to \infty$.

Eigenvalues of $A(\lambda)$

$$\underbrace{1, 1, ..., 1}_{M} \qquad\qquad \frac{\lambda_{v_n}}{\lambda_{v_n} + n\lambda} = \frac{1}{1 + n\lambda/\lambda_{v_n}}$$

$\lambda_{v_n} \approx n/(2\pi v)^{2m}$ for equally spaced t_i, see periodic splines with $||P_1 f||^2 = \int \left(f^{(m)}(u)\right)^2 du$.



p

$\phi_n(v) = 1/(1 + n\lambda/\lambda_{v_n}) \approx 1/\left(1 + \lambda(2\pi v)^{2m}\right)$ for periodic splines with equally spaced observations.

"Low Pass Filter: Butterworth Filter"

$\mu_1(\lambda)$ small means average eigenvalue small

$$\mu_1^2(\lambda)/\mu_2(\lambda) \quad small \quad \frac{\left(\frac{1}{n}\sum\frac{1}{1+n\lambda/\lambda_{vn}}\right)^2}{\frac{1}{n}\sum\frac{1}{(1+n\lambda/\lambda_{vn})^2}}$$

This is satisfied if a small number of eigenvalues are near 1 the rest near 0.

Will be satisfied if A is a “low pass” filter

To compute $V(\lambda)$ using eigenvalue – eigenvector decomposition

$$V(\lambda) = \frac{\frac{1}{n} \|(I - A(\lambda))\, y\|^2}{\left(\frac{1}{n} tr\,(I - A(\lambda))\right)^2}$$

$$\begin{aligned} I - A &= n\lambda Q_2 \underbrace{\left(Q_2' \Sigma Q_2 + n\lambda I_{n-m}\right)^{-1}}_{UDU' + n\lambda I_{n-M}} Q_2' \\ &= n\lambda Q_2 U' \left\{ \frac{n\lambda}{\lambda_{v_n} + n\lambda} \right\} U Q_2' \end{aligned}$$

Let $\Gamma_{n-m \times n} = UQ_2'$

$Z_{n-m} = \Gamma y$ (Note: projection of y onto columns of T disappears)

$$V(\lambda) = \frac{\frac{1}{n} \sum_{v=1}^{n-M} \left(\frac{n\lambda}{\lambda_{v_n} + n\lambda}\right)^2 z_v^2}{\left(\frac{1}{n} \sum \frac{n\lambda}{\lambda_{v_n} + n\lambda}\right)^2}$$

Note: cancel $n\lambda$ from top+bottom

$$
\begin{aligned}
V(\lambda) &= \frac{\frac{1}{n}\sum_{v=1}^{n-M}\left(\frac{1}{\lambda_{vn}+n\lambda}\right)^2 z_v^2}{\left(\frac{1}{n}\sum\frac{1}{\lambda_{vn}+n\lambda}\right)^2} \\
&= \frac{\frac{1}{n}\sum_{v=1}^{n-M}\frac{z_v^2}{\lambda_{v_n}^2}}{\left(\frac{1}{n}\sum_{v=1}^{n-M}\frac{1}{\lambda_{vn}}\right)^2} \qquad (\lambda \to 0) \\
&> 0
\end{aligned}
$$

(but may have unstable computation)

Let $n \to \infty$

$$
\begin{aligned}
V(\lambda) &= \frac{\frac{1}{n}\sum_{v=1}^{n-M} z_v^2}{\left(\frac{1}{n}\right)^2} \\
&= RSS \quad after \quad projection \quad onto \quad H_0
\end{aligned}
$$

$$R(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(L_i f_\lambda - L_i f)^2$$

$$\begin{aligned} ER(\lambda) &= \frac{1}{n}||(I - A(\lambda))g||^2 + \frac{1}{n}\sigma^2 tr A^2(\lambda) \\ &\equiv b^2(\lambda) + \sigma^2 Var(\lambda) \end{aligned}$$

where $g = (L_1 f, ..., L_n f)'$

Let $h = \Gamma g$, $\Gamma = UQ_2'$ as before

$$b^2(\lambda) = \frac{1}{n}\sum_{v=1}^{n-m}\left(\frac{n\lambda}{\lambda_{v_n} + n\lambda}\right)^2 h_{v_n}^2$$

$$var(\lambda) = \mu_2(\lambda) = \frac{\sigma^2}{n}\left[\sum_{v=1}^{n-m}\left(\frac{\lambda_{v_n}}{\lambda_{v_n} + n\lambda}\right)^2 + M\right]$$

Lemma: $b^2(\lambda) \leq \lambda\,||P_1 f||^2$

Proof of Lemma: $b^2(\lambda) \le \lambda \|P_1 f\|^2$

Find f_λ^* to min

$$\frac{1}{n}\sum_{i=1}^{n}(g_i - L_i f)^2 + \lambda\,||P_1 f||^2$$

y is replaced by g, so

$$\begin{pmatrix} L_1 f_\lambda^* \\ \ldots \\ L_n f_\lambda^* \end{pmatrix} = A\,(\lambda)\,g$$

so

$$\begin{aligned} & \frac{1}{n}\sum (g_i - L_i f_\lambda^*)^2 + \lambda\,||P_1 f_\lambda^*||^2 \\ \le \; & \frac{1}{n}\sum (g_i - L_i f)^2 + \lambda\,||P_1 f|| \qquad\qquad any \quad f \end{aligned}$$

Set $f = ''true'' \quad f$ for which $g_i = L_i f$, get

$$= \lambda\,||P_1 f||^2$$

$$
\begin{aligned}
\mu_2(\lambda) &\approx \frac{1}{n}\sum_{v=1}^{n}\left(\frac{1}{\left(1+\lambda(2\pi v)^{2m}\right)}\right)^2 \\
&\approx \int_0^\infty \frac{dx}{\left(1+\lambda(2\pi x)^{2m}\right)^2}
\end{aligned}
$$

Let $y = \lambda(2\pi x)^{2m}, x = \frac{1}{2\pi}\left(\frac{y}{\lambda}\right)^{1/2m}$

$$
\mu_2(\lambda) \approx \frac{c_2}{n\lambda^{1/2m}}, \qquad b^2(\lambda) \leq c_1\lambda
$$

minimize

$$
c_1\lambda + \frac{c_2}{n}\frac{1}{\lambda^{1/2m}}
$$

$$
\lambda_{opt} = \frac{c_3}{n^{2m/2m+1}}
$$

$$ER(\lambda_{opt}):$$

$$\begin{aligned}
b^2\left(\lambda_{opt}\right) + \sigma^2 Var\left(\lambda_{opt}\right) &\approx \frac{c_4}{n^{2m/2m+1}} \\
&= \frac{c}{n^{4/5}} \qquad for \quad m = 2 \\
&= \frac{c}{n^{.8}}
\end{aligned}$$

$$\log E(R(\lambda_{opt})) \approx -0.8 \log n + c$$

If you assume more on g than just $f \in \mathcal{H}$

$$b^2(\lambda) = \frac{1}{n} \sum_{v=1}^{n-m} \left(\frac{n\lambda}{\lambda_{v_n} + n\lambda} \right)^2 h_{v_n}^2$$

$$\begin{aligned} h &= \Gamma g \qquad g = \begin{pmatrix} L_1 f \\ \dots \\ L_n f \end{pmatrix} \\ &\equiv \lambda^p \frac{1}{n} \sum_{v=1}^{n} \left(\frac{n\lambda}{\lambda_{v_n} + n\lambda} \right)^{2-p} \frac{h_{v_n}^2 / n}{(\lambda_{v_n}/n + \lambda)^p} \\ &\leq \lambda^p \sum_{v=1}^{n} \frac{h_{v_n}^2 / n}{(\lambda_{v_n}/n)^p} \end{aligned}$$

for any $p \in [1, 2]$.

$$b^2(\lambda) \leq \lambda^p \sum_{v=1}^{n} \frac{h_{v_n}^2/n}{(\lambda_{v_n}/n)^p} \qquad (*)$$

Note: for $p = 1$ the sum is $\sum \frac{h_{v_n}^2}{\lambda_{v_n}} = ||P_1 f||_{\mathcal{H}}^2$

For $1 < p \leq 2$ this is, roughly $||f||_{\mathcal{H}^p}^2$

where $\mathcal{H}^p$ is the RKHS with

$$R^p(s,t) = \sum \lambda_v^p \phi_v(s)\,\phi_v(t) \qquad with \quad \{\lambda_v, \phi_v\}$$

the eigenvalues and eigenfunctions of R

If $f \in \mathcal{H}^p$, then the sum in $(*)$ is uniformly bounded

$$b^2(\lambda) \leq c_1 \lambda^p \qquad Var(\lambda) \approx \frac{c_2}{n^{2mp/2mp+1}}$$

For $||P_1 f||^2 = \int (f''(u))^2 \, du, \quad m = 2$

$$p = 1 \qquad R\left(\lambda_{opt}\right) \approx \frac{const}{n^{4/5}} \approx cn^{-.8}$$

$$p = 2 \qquad R\left(\lambda_{opt}\right) \approx \frac{const}{n^{8/9}} \approx cn^{-.89}$$