Matrices, Moments, and Quadrature with Applications

Gene H. Golub

Stanford University

Outline

I. The Problem

II. Examples and Applications

III. Gauss-Radau Quadrature

IV. Solution of Problems

V. Other Extensions

I The Problem

Assumptions

$$A = A^T \qquad n \times n$$

A: positive definite,

$\mathbf{u}$ and $F(\cdot)$ given.

Determine U and L such that

$$L \le \mathbf{u}^T F(A)\mathbf{u} \le U,$$

or approximate

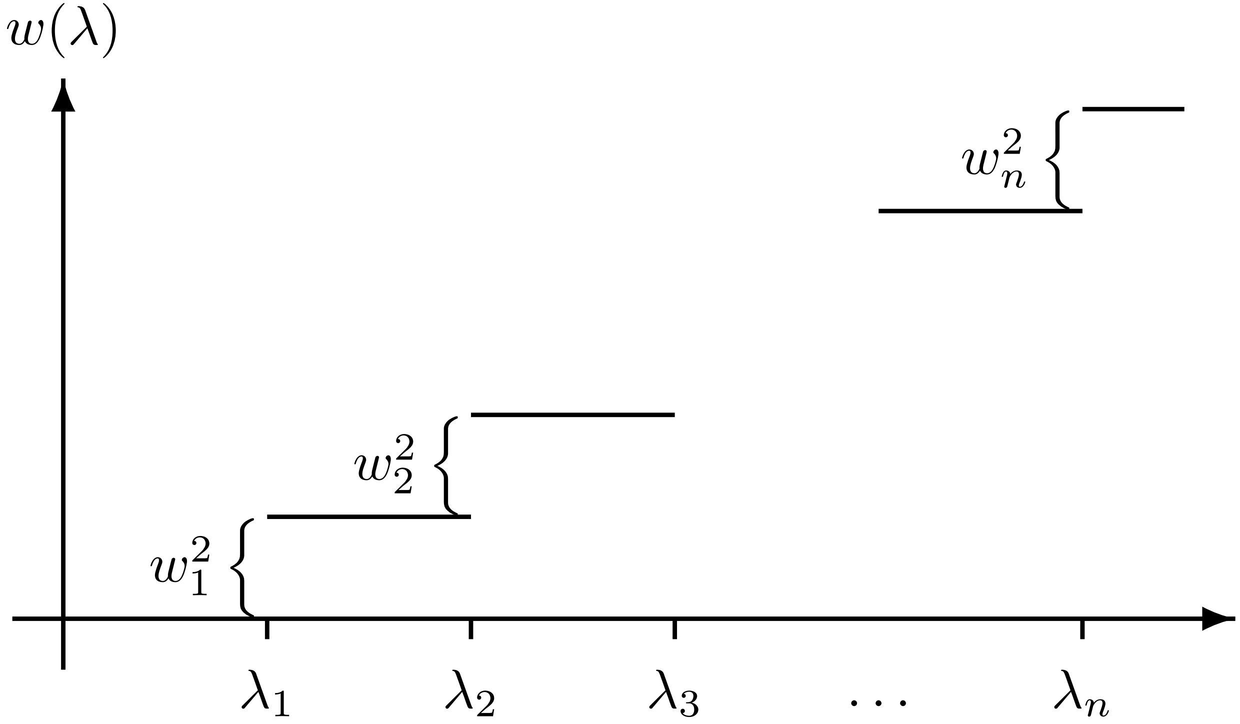
$$\mathbf{u}^T F(A)\mathbf{u}.$$

$$A = Q\Lambda Q$$

$$\begin{aligned}
\boldsymbol{u}^T F(A)\boldsymbol{u} &= \boldsymbol{u}^T F(Q\Lambda Q^T)\boldsymbol{u} \\
&= \boldsymbol{u}^T Q F(\Lambda) Q^T \boldsymbol{u} \\
&= \boldsymbol{w}^T F(\Lambda)\boldsymbol{w}
\end{aligned}$$

$$\boxed{\boldsymbol{w} = Q^T \boldsymbol{u}}$$

$$\begin{aligned}
\boldsymbol{u}^T F(A)\boldsymbol{u} &= \sum_{i=1}^{n} F(\lambda_i) w_i^2 \\
&= \int F(\lambda) dw(\lambda)
\end{aligned}$$



$w(\lambda)$
w_n^2
w_2^2
w_1^2
λ_1
λ_2
λ_3
$\dots$
λ_n

II Examples and Applications

1 Error Estimates

$$A\boldsymbol{x} = \boldsymbol{b}$$

$$\boldsymbol{x} = \boldsymbol{\xi} + \boldsymbol{e}$$

$\boldsymbol{\xi}$: given

$$\begin{aligned} \boldsymbol{r} &= \boldsymbol{b} - A\boldsymbol{\xi} \\ &= A\boldsymbol{e} \end{aligned}$$

$$\boldsymbol{e} = A^{-1}\boldsymbol{r}$$

$$||\boldsymbol{e}||_2^2 = \boldsymbol{r}^T A^{-2} \boldsymbol{r}$$

$$F(\lambda) = \lambda^{-2}.$$

2 Quadrature Constraint

$$\min \qquad \boldsymbol{x}^T A\boldsymbol{x} - 2\boldsymbol{b}^T\boldsymbol{x}$$

<u>constraint</u> $\qquad ||\boldsymbol{x}||_2 = \alpha$

$$\alpha < ||A^{-1}b||_2$$

Lagrangian

$$\phi\left(\boldsymbol{x};\mu\right) = \boldsymbol{x}^T A\boldsymbol{x} - 2\boldsymbol{b}^T\boldsymbol{x} + \mu\left(\boldsymbol{x}^T\boldsymbol{x} - \alpha^2\right)$$

"Regularization"

$$grad \;\; \phi = 0$$

$$\Rightarrow \left(A + \mu I\right)\boldsymbol{x} = \boldsymbol{b}$$

$$\boldsymbol{x} = \left(A + \mu I\right)^{-1}\boldsymbol{b}$$

$$||\boldsymbol{x}||^2 = \alpha^2$$

$$\boldsymbol{b}^T\left(A + \mu I\right)^{-2}\boldsymbol{b} = \alpha^2$$

$$\boldsymbol{u} = \boldsymbol{b}$$

$$F(A) = (A + \mu I)^{-2}$$

$$(\mu > 0)\,.$$

$$A = Q\Lambda Q^T$$

$$Q^T Q = I, \qquad \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

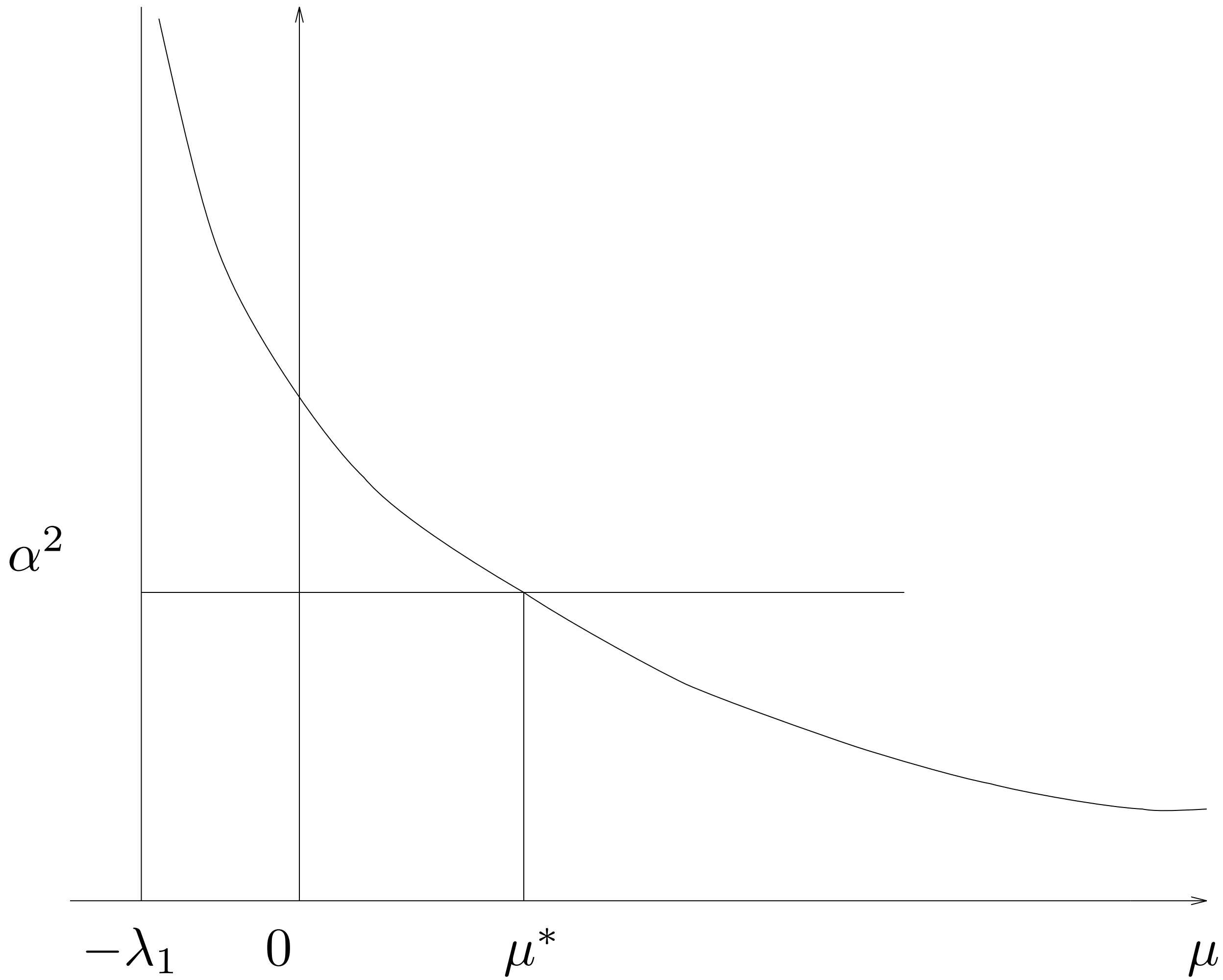
$$0 < \lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$$

$$0 < a \le \lambda_1 \le \lambda_n \le b$$

$$\boldsymbol{\beta} = Q^T \boldsymbol{b}$$

$$\boldsymbol{\beta}^T (\Lambda + \mu I)^{-2} \boldsymbol{\beta} = \alpha^2$$

$$F(A;\mu) = \sum_{i=1}^{n} \frac{\beta_i^2}{(\lambda_i + \mu)^2}$$



How can we determine μ^* without computing the eigenvalues and eigenvectors of A ?

3 Bounds on elements of the inverse

$$\left\{A^{-1}\right\}_{jj}$$

$$\boldsymbol{u}^T = \boldsymbol{e}_j^T = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$$

$j^{\underline{th}}$ position.

$$F(A) = A^{-1}$$

4 Regularization

5 Backward Perturbations for Linear Least Squares Problems

(with Zheng Su) (NEW!)

$$\min_{\boldsymbol{x}} ||\boldsymbol{b} - A\boldsymbol{x}||_2, A : m \times n, \boldsymbol{b} : m \times 1.$$

$\boldsymbol{\xi}$: arbitrary vector, calculated.

$$\hat{\boldsymbol{x}} = \boldsymbol{\xi} + \boldsymbol{e}; \hat{\boldsymbol{x}} = A^{+}\boldsymbol{b}.$$

$$\boldsymbol{\xi} = (A + \delta A)^{+}\boldsymbol{b}.$$

$$\mu(\boldsymbol{\xi}) = \min_{\delta A} ||\delta A||_F.$$

$$\mu(\boldsymbol{\xi}) = \min\{\frac{||\boldsymbol{\rho}||_2}{||\boldsymbol{\xi}||_2}, \sigma_{\min}(A, B)\}$$

(Karlson, Waldén & Sun)

$$\boldsymbol{\rho} = \boldsymbol{b} - A\boldsymbol{\xi},$$

$$B = \frac{||\boldsymbol{\rho}||_2}{||\boldsymbol{\xi}||_2}(I - \frac{\boldsymbol{\rho}\boldsymbol{\rho}^T}{||\boldsymbol{\rho}||_2^2}).$$

$$\tilde{\mu}^2(\boldsymbol{\xi}) = \boldsymbol{\rho}^T A(\alpha A^T A + \beta I)^{-1} A^T \boldsymbol{\rho}.$$

$$\alpha = ||\boldsymbol{\xi}||_2^2, \quad \beta = ||\boldsymbol{\rho}||_2^2.$$

$$\boxed{\tilde{\mu}(\boldsymbol{\xi}) \sim \mu(\boldsymbol{\xi})}$$

$$\lim_{\boldsymbol{\xi} \longrightarrow \hat{\boldsymbol{x}}} \frac{\tilde{\mu}(\boldsymbol{\xi})}{\mu(\hat{\boldsymbol{x}})} = 1$$ (Grcar)

III Gauss-Radau Quadrature Rules

$$L \leq \int_a^b F(\lambda) dw(\lambda) \leq U$$

$$\mu_r = \int \lambda^r dw(\lambda) \qquad (r = 0, 1, \ldots, 2k + m - 1)$$

$$\int_a^b F(\lambda) dw(\lambda) = I[F] + R[F]$$

$$I[F] = \sum_{i=1}^{k} A_i F(t_i) + \sum_{j=1}^{m} B_j F(z_j)$$

$\{A_i, t_i\}_{i=1}^{k}$ unknown weights and nodes

$\{z_j\}_{j=1}^{m}$ prescribed

$\{B_j\}_{j=1}^{m}$ must be calculated

$$I[\lambda^r] = \mu^r$$

$$\mu_r = \sum_{i=1}^{k} A_i t_i^r + \sum_{j=1}^{m} B_j z_j^r$$

System of Non-linear Equations.

$$R[F] = \frac{F^{(2k+m)}(\eta)}{(2k+m)!} \int_a^b \prod_{j=1}^{m} (\lambda - z_j) \left[\prod_{i=1}^{k} (\lambda - t_i) \right]^2 dw(\lambda)$$

$$a < \eta < b$$

$$\underline{m = 1}$$

$F^{(2k+1)}(\eta) \leq 0$ and $z_1 = a$, $R[F] \leq 0$

$$\Rightarrow I[F] = U$$

$$F^{(2k+1)}(\eta) \leq 0 \text{ and } z_1 = b,\ R[F] \geq 0$$

$$\Rightarrow I[F] = L$$

Problem 1

$$\boldsymbol{r}^T A^{-2} \boldsymbol{r} = \mu_{-2}$$

$$\mu_l = \boldsymbol{r}^T A^l \boldsymbol{r}, \qquad \boldsymbol{\rho} = Q^T \boldsymbol{r}$$
$$= \sum_{i=1}^{m} \lambda_i^l \rho_i^2$$

How can we determine quadrature rule in an efficient manner ?

Gauss Quadrature

$$\int p_r(\lambda)p_s(\lambda)d\alpha(\lambda) = 0, \qquad j \neq k$$

$$(r, s = 0, 1, \ldots, k)$$

$$p_{j+1}(\lambda) = (\lambda - \xi_{j+1})p_j(\lambda) - \eta_j^2 p_{j-1}(\lambda)$$

$$p_k(t_i) = 0, \qquad i = 1, 2, \ldots, k$$

$$J_k = \begin{pmatrix} \xi_1 & \eta_1 & & & \\ \eta_1 & \xi_2 & \eta_2 & & \\ & \eta_2 & \ddots & \ddots & \\ & & \ddots & \ddots & \eta_{k-1} \\ & & & \eta_{k-1} & \xi_k \end{pmatrix}$$

$$\boxed{\mu_0 = 1}$$

$$
\begin{aligned}
J_k \boldsymbol{v}_j &= t_j \boldsymbol{v}_j, \qquad j = 1, 2, \ldots, k \\
A_j &= v_{1j}^2, \qquad j = 1, 2, \ldots, k
\end{aligned}
$$

Gauss-Radau

(Inverse Eigenvalue Problem)

$$\bar{J}_{k+1} = \begin{pmatrix} & & & 0 \\ & J_k & & \vdots \\ & & & \eta_k \\ 0 & \dots & \eta_k & \bar{\xi}_{k+1} \end{pmatrix}$$

$$0 = p_{k+1}(t_0) = (t_0 - \bar{\xi}_{k+1})p_k(t_0) - \eta_k^2 p_{k-1}(t_0)$$

$$\bar{\xi}_{k+1} = t_0 - \eta_k^2 \frac{p_{k-1}(t_0)}{p_k(t_0)}$$

or

$$(J_k - t_0 I)\delta = \eta_k^2 \boldsymbol{e}_k$$

$$\bar{\xi}_{k+1} = t_0 + \delta_k$$

Evaluate

$$I[F] = \sum_{i=0}^{k} v_{1i}^2 F(t_i)$$

$$\bar{J}_{k+1} = VTV^T$$

$$V^T \boldsymbol{e}_1 = \langle \text{ first component of V } \rangle$$

$$\begin{aligned} I[F] &= \boldsymbol{e}_1^T V F(T) V^T \boldsymbol{e}_1 \\ &= \boldsymbol{e}_1^T F(VTV^T) \boldsymbol{e}_1 \\ &= \boldsymbol{e}_1^T F(\bar{J}_{k+1}) \boldsymbol{e}_1 \end{aligned}$$

How can we determine the orthogonal polynomials w.r.t. the measure $w(\lambda)$?

Now

$$
\begin{aligned}
p_{j+1}(\lambda) &= (\lambda - \xi_{j+1}) p_j(\lambda) - \eta_j^2 p_{j-1}(\lambda) \\
p_{j+1}(A) &= (A - \xi_{j+1} I) p_j(A) - \eta_j^2 p_{j-1}(A) \\
p_{j+1}(A)\boldsymbol{u} &= (A - \xi_{j+1} I) p_j(A)\boldsymbol{u} - \eta_j^2 p_{j-1}(A)\boldsymbol{u}
\end{aligned}
$$

Set $\boldsymbol{w}_j = p_j(A)\boldsymbol{u}$

We define ξ_{j+1} and η_j^2 so that

$$
\begin{aligned}
\boldsymbol{w}_{j+1}^T \boldsymbol{w}_j &= 0 \\
\boldsymbol{w}_{j+1}^T \boldsymbol{w}_{j-1} &= 0
\end{aligned}
$$

$$\Rightarrow \boldsymbol{w}_{j+1}^T \boldsymbol{w}_r = 0 \qquad \text{for } r < j-1$$

$$\xi_{j+1} = \frac{(\boldsymbol{w}_j, A\boldsymbol{w}_j)}{(\boldsymbol{w}_j, \boldsymbol{w}_j)}$$

$$\eta_j^2 = \frac{(\boldsymbol{w}_j, \boldsymbol{w}_j)}{(\boldsymbol{w}_{j-1}, \boldsymbol{w}_{j-1})}$$

<u>LANCZOS METHOD</u>

$$\begin{aligned} (\boldsymbol{w}_j, \boldsymbol{w}_k) &= 0 \\ &= (p_j(A)\boldsymbol{u}, p_k(A)\boldsymbol{u}) \\ &= \int p_j(\lambda)p_k(\lambda)dw(\lambda) \end{aligned}$$

IV Solution of Problems

1. **Find upper and lower bounds on**

$$\mu_{-2} = \boldsymbol{r}^T A^{-2} \boldsymbol{r}.$$

Use Lanczos process with initial vector $\boldsymbol{r}$.

Compute

$$\bar{J}_{k+1} = \begin{pmatrix} \xi_1 & \eta_1 & & 0 \\ \eta_1 & \ddots & \ddots & \\ & \ddots & \xi_k & \eta_k \\ 0 & & \eta_k & \bar{\xi}_{k+1} \end{pmatrix}.$$

Then

$$\boldsymbol{e}_1^T \bar{J}_{k+1}^{-2} \boldsymbol{e}_1$$

will yield upper and lower bounds

on μ_{-2} depending on $\bar{\xi}_{k+1}$.

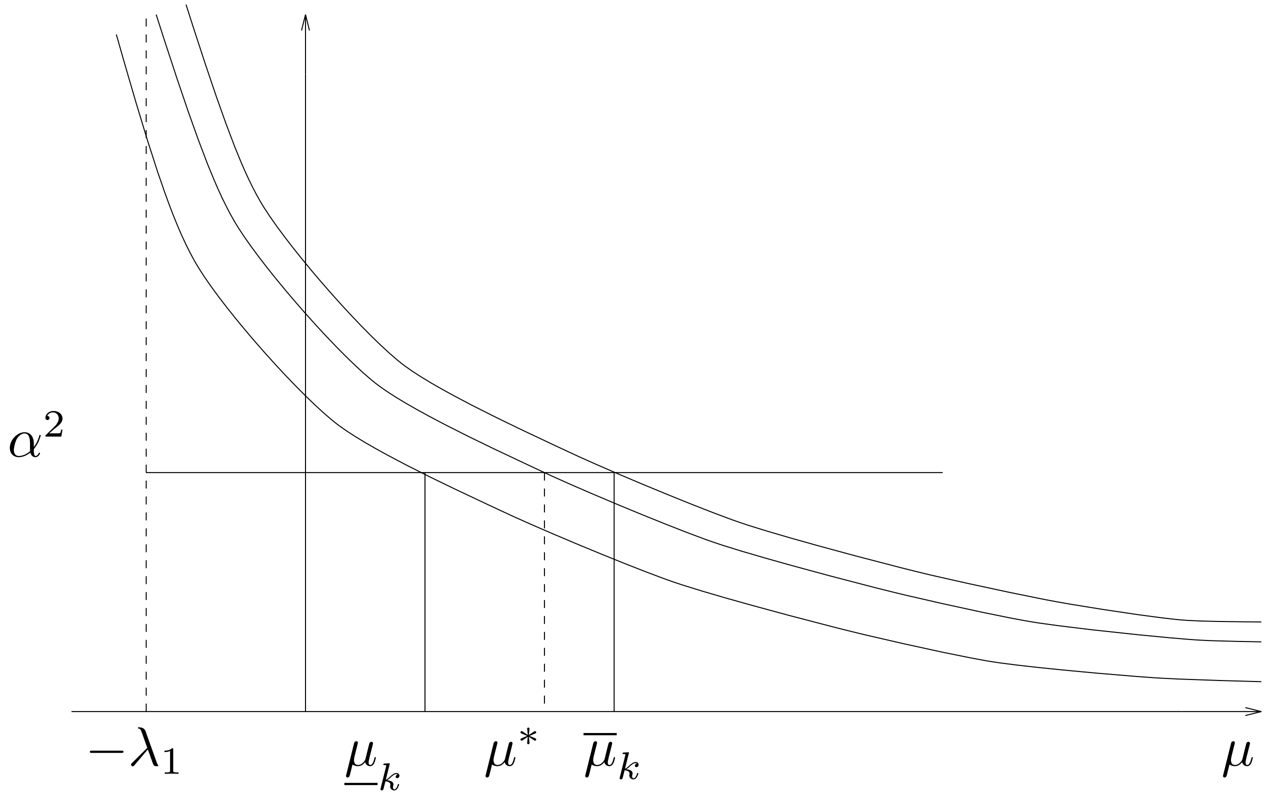
2. **We need to solve**

$$\boldsymbol{b}^T(A+\mu I)^{-2}\boldsymbol{b} = \alpha^2$$

a) Begin Lanczos process with $\boldsymbol{u} = \boldsymbol{b}$.

b) Construct $\bar{J}_{k+1}$.

c) Solve $\boldsymbol{e}_1^T(\bar{J}_{k+1} + \mu I)^{-2}\boldsymbol{e}_1 = \alpha^2$



$$\underline{\mu}_1 \leq \underline{\mu}_2 \leq \dots \mu^* \leq \dots \leq \overline{\mu}_2 \leq \overline{\mu}_1$$

Solving the equation

$$\boldsymbol{c}_1^T \left(\bar{J}_{k+1} + \mu I\right)^{-2} \boldsymbol{e}_1 = \alpha^2$$

Can be accomplished in various ways.

For instance, one can use Newton's method, or better

$$\boldsymbol{e}_1^T \left(\bar{J}_{k+1} + \mu I\right)^{-2} \boldsymbol{e}_1 \sim \left(\frac{a}{b + c\mu}\right)^2$$

We can also use the direction generated

by Lanzos process for updating solution.

3. **Find upper and lower bounds on**

$$\boldsymbol{e}_j^T A^{-1} \boldsymbol{e}_j.$$

Apply Lanczos with $\boldsymbol{u} = \boldsymbol{e}_j$.

For $k = 2$,

$$\frac{a_{ii} - b + \frac{s_i^2}{b}}{a_{ii} - a_{ii}b + s_i^2} \le (A^{-1})_{ii} \le \frac{a_{ii} - a + \frac{s_i^2}{a}}{a_{ii}^2 - a_{ii}a + s_i^2}$$

$$s_i = \sum_{j \neq i} a_{ij}^2$$

optimal

We can generalize these results for

$$\left\{A^{-1}\right\}_{j,k}$$

4. **Tikhonov Regularization**

Ill-posed problem:

$$||A\boldsymbol{x} - \boldsymbol{b}||_2 = \min .$$

$$||A\boldsymbol{x} - \boldsymbol{b}||_2^2 + \alpha||\boldsymbol{x}||_2^2 = \min .$$

$$(\alpha > 0)$$

$$(A^T A + \alpha I)\boldsymbol{x}_\alpha = A^T \boldsymbol{b}$$

$$\left\| \begin{pmatrix} A \\ \sqrt{\alpha} I \end{pmatrix} \boldsymbol{x} - \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix} \right\|_2 = \min .$$

Choice of α?

5. **Generalized Cross-Validation**

Omit observation i:

$$A_i := \begin{pmatrix} \boldsymbol{a}_1^T \\ \vdots \\ \boldsymbol{a}_{i-1}^T \\ \boldsymbol{a}_{i+1}^T \\ \vdots \\ \boldsymbol{a}_m^T \end{pmatrix} \quad , \quad b_i := \begin{pmatrix} b_1^T \\ \vdots \\ b_{i-1}^T \\ b_{i+1}^T \\ \vdots \\ b_m^T \end{pmatrix}$$

Regularization solution:

$$\boldsymbol{x}_i(\alpha) = (A_i^T A_i + \alpha I)^{-1} A_i^T \boldsymbol{b}_i.$$

Residual for observation i:

$$r_i(\alpha) = b_i - \boldsymbol{a}_i^T \boldsymbol{x}_i(\alpha).$$

Cross-validation: choose α such that

$$||\boldsymbol{r}(\alpha)||_2 = \min .$$

Regularization solution:

$$\boldsymbol{x}_i(\alpha) = (A_i^T A_i + \alpha I)^{-1} A_i^T \boldsymbol{b}_i.$$

Rank-one modifications:

$$A_i^T A_i = A^T A - \boldsymbol{a}_i \boldsymbol{a}_i^T$$

$$A_i^T \boldsymbol{b}_i = A^T \boldsymbol{b} - b_i \boldsymbol{a}_i.$$

Use Sherman-Morrison formula:

$$\begin{aligned}(A_i^T A_i + \alpha I)^{-1} &= (A^T A + \alpha I)^{-1} \\ &+ \frac{(A^T A + \alpha I)^{-1} \boldsymbol{a}_i \boldsymbol{a}_i^T (A^T A + \alpha I)^{-1}}{1 - \boldsymbol{a}_i^T (A^T A + \alpha I)^{-1} \boldsymbol{a}_i}\end{aligned}$$

$$r_i(\alpha) = \frac{b_i - \boldsymbol{a}_i^T (A^T A + \alpha I)^{-1} A^T \boldsymbol{b}}{1 - \boldsymbol{a}_i^T (A^T A + \alpha I)^{-1} \boldsymbol{a}_i}.$$

Define

$$D(\alpha) := diag(I - A(A^TA + \alpha I)^{-1}A^T).$$

Residual:

$$\begin{aligned} ||\boldsymbol{r}||_2^2 &= ||D^{-1}(I - A(A^TA + \alpha I)^{-1}A^T)\boldsymbol{b}||_2^2 \\ &\geq \frac{||(I - A(A^TA + \alpha I)^{-1}A^T)\boldsymbol{b}||_2^2}{||D||_F^2} \\ &= \frac{||(I - A(A^TA + \alpha I)^{-1}A^T)\boldsymbol{b}||_2^2}{(trace(I - A(A^TA + \alpha I)^{-1}A^T))^2} \\ &=: \phi_{GCV}^2(\alpha) \end{aligned}$$

$$\phi_{GCV}(\alpha) := \frac{||(A^TA+\alpha I)^{-1}\boldsymbol{b}||_2}{trace((AA^T+\alpha I)^{-1})}$$

$$P\{U_i = +1\} = \frac{1}{2}$$

$$P\{U_i = -1\} = \frac{1}{2}$$

$\boldsymbol{u}$ is a random vector., entries are independent and from U.

$$\tilde{t}(\alpha) = \boldsymbol{u}^T(AA^T+\alpha I)^{-1}\boldsymbol{u}$$

$$E\{\tilde{t}(\alpha)\} = t(\alpha) = trace((AA^T+\alpha I)^{-1})$$

$$\tilde{\phi}_{GCV}(\alpha) := \frac{||(A^TA+\alpha I)^{-1}\boldsymbol{b}||_2}{\boldsymbol{u}^T(AA^T+\alpha I)^{-1}\boldsymbol{u}}$$

6. **Backward Perturbations**

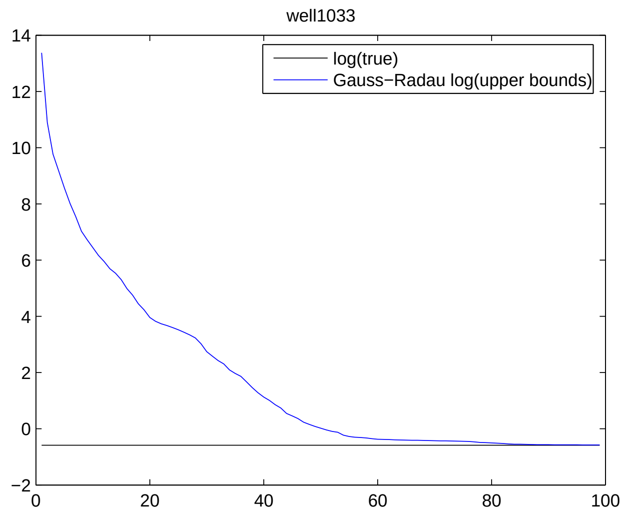
$$\tilde{\mu}^2(\boldsymbol{\xi}) = \boldsymbol{\rho}^T A(\alpha A^T A + \beta I)^{-1} A^T \boldsymbol{\rho}.$$

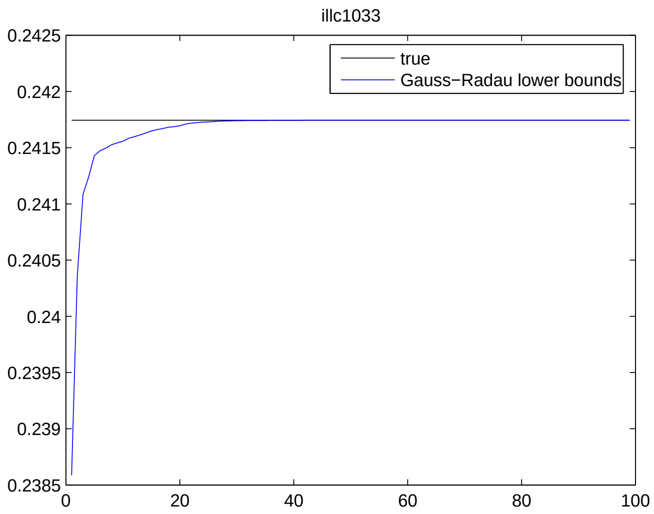
Use Lanczos on $A^T A$ (or G/Kahan bi-diagonalization) with initial vector $\boldsymbol{u} = A^T \boldsymbol{\rho}$.

Test problems $\min ||Ax - b||_2$: They originated in the least-squares analysis of gravity-meter observations.

matrix	rows m	columns n	$\kappa_2(A)$
(a) `well1033`	1033	320	$1.7e{+}2$
(b) `illc1033`	1033	320	$1.9e{+}4$

$\boldsymbol{\xi}$: QR solution of $\min ||Ax - b||_2$ for well1033 (left) and illc1033 (right).


well1033
log(true)
Gauss−Radau log(upper bounds)
14
12
10
8
6
4
2
0
−2
0
20
40
60
80
100


illc1033
true
Gauss−Radau lower bounds
0.2425
0.242
0.2415
0.241
0.2405
0.24
0.2395
0.239
0.2385
0
20
40
60
80
100

V Other Extensions

$$1 \qquad W: \quad n \times p.$$

Estimate : $W^T A W$

Block Lanczos Process

$$J_k = \begin{pmatrix} \Omega_1 & \Gamma_1^T & & 0 \\ \Gamma_1^T & \ddots & \ddots & \\ & \ddots & \Omega_{k-1} & \Gamma_{k-1}^T \\ 0 & & \Gamma_{k-1} & \Omega_k \end{pmatrix}.$$

$$\Omega_j: \quad p \times p.$$

Non-symmetric variant of Lanczos Process

2 Modified Moments

$$\mu_k = \int \lambda^k dw(\lambda).$$

Sometimes, we are given

$$\nu_k = \int \pi_k(\lambda) dw(\lambda) \qquad \text{"modified moments"}$$

where

$$\pi_{k+1}(\lambda) = (\lambda - a_{k+1})\,\pi_k(\lambda) - b_k^2 \pi_{k-1}(\lambda).$$

Can we determine $\{p_k(\lambda)\}$ so that

$$\int p_k(\lambda) p_l(\lambda) dw(\lambda) = 0 \qquad k \neq l$$

$$p_{k+1}(\lambda) = (\lambda - \xi_{k+1})\,p_k(\lambda) - \eta_k^2 p_{k-1}(\lambda) \qquad ?$$

Answer. YES! (Gautschi, Wheeler, Sack).

Chebyshev Method for Solving Linear Equation

$$A\boldsymbol{x} = \boldsymbol{b}$$

$$\left(\boldsymbol{x}^k - \boldsymbol{x}\right) = P_k(A)\left(\boldsymbol{x}^0 - \boldsymbol{x}\right)$$

$$\boldsymbol{r}^k = P_k(A)\boldsymbol{r}^0$$

$$G_{rs} = \left(\boldsymbol{r}^k, \boldsymbol{r}^s\right)$$

We can construct $\left(\xi_{k+1}, \eta_k^2\right)$ and estimate eigenvalues.

3 Updating/Downdating

We can also determine the coeffitients from the Grammian.

$$G_{rs} = \int \Pi_r(\lambda)\Pi_s(\lambda)dw(\lambda)$$

Updating orthogonal polynomials

$$\begin{aligned} Y^{(N)} &= \{x_i, w_i\}_{i=1}^{N} \qquad \left\{p_k^{(N)}(\lambda)\right\}_k \\ Y^{(N+1)} &= \{x_i, w_i\}_{i=1}^{N+1} \end{aligned}$$

$$\begin{aligned} G_{rs} &= \int p_r^{(N)} p_s^{(N)} dw^{(N+1)}(\lambda) \\ &= \sum_{i=1}^{N+1} p_r^{(N)}(x_i) P_s^{(N)}(x_i) w_i \\ &= \delta_{rs} + p_r^{(N)}(x_{N+1}) p_s^{(N)}(x_{N+1}) w_{N+1} \end{aligned}$$

$$G = I + \boldsymbol{u}\boldsymbol{u}^T.$$

Applies to downdating, too.