Recent progress in the solution of linear ill-posed problems

+ 3 advertisements

Michiel Hochstenbach

TU Eindhoven

www.win.tue.nl/∼hochsten

Joint work with **Lothar Reichel**

Advertisements with **Yvan Notay, David Singer, Paul Zachlin**

ILAS 2011

August 23, 2011

Talk

Advertisements:

(1) spectral inclusion regions based on the field of values
(joint with David Singer, Paul Zachlin)

(2) inner-outer control for Jacobi–Davidson
(joint with Yvan Notay)

(3) probabilistic bounds for eigenvalue problems

Main part: linear ill-posed problems

Advertisements have slogans

Slogans:

(1) spectral inclusion regions based on the field of values

"how to find a good inclusion region for all eigenvalues of a large matrix with just 5 MVs?"

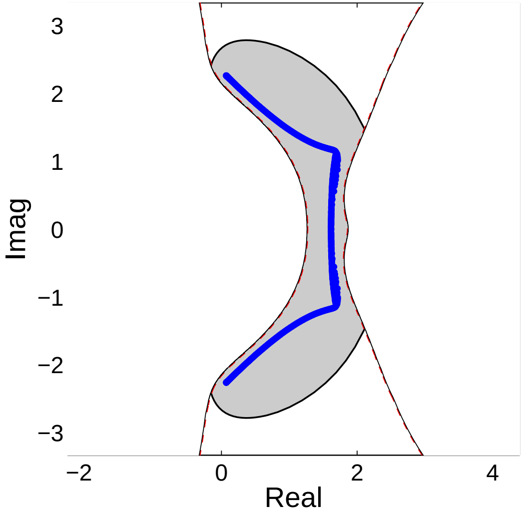
(2) inner-outer control for Jacobi–Davidson

"how to speed up Jacobi–Davidson by a factor 2 with just a few lines of code?"

(3) probabilistic bounds for eigenvalue problems

"how to find an interval $[\theta,\ 1.01\,\theta]$ that contains $\|A\|_2$ with probability 99.9% within 0.15 sec for a 23560×23560 matrix"

Advertisement 1: inclusion regions with the field of values



Inclusion region for 1000×1000 `grcar` matrix with just **5** MVs, how comes?

Advertisement 1: inclusion regions with the field of values

Wanted: (cheap, approximate) eigenvalue inclusion region

Field of values $\Lambda(A) \subset W(A) = \{\mathbf{x}^* A \mathbf{x} : \|\mathbf{x}\| = 1\}$

Advantages of $W(A)$ include:

- convex region $\Rightarrow$ easy to compute (discrete approximation)
- easy to approximate for large matrices
- "relatively parameter-free"

Alternatives include:

- based on matrix norm (e.g., disk with radius $\|A\|$)
- Gershgorin disks $\cup_i \{ |z - a_{ii}| \leq r_i \}$
- ovals of Cassini $\cup_{i \neq j} \{ |z - a_{ii}| \, |z - a_{jj}| \leq r_i \, r_j \}$
- residual bound (for 1 eigenvalue) $|\lambda - \widetilde{\lambda}| \leq \kappa \, \|\mathbf{r}\|$
- pseudospectra $\{ \sigma_{\min}(A - zI) < \varepsilon \}$

see, e.g., (Horn, Johnson 1985) (Beattie, Ipsen 2003) (Varga 2004)

Advertisement 1: inclusion regions with the field of values

Small A: compute $\mu^\alpha = \lambda_{\max}(\frac{1}{2}(e^{-i\alpha}A + e^{i\alpha}A^*))$
for angles $\alpha_1, \ldots, \alpha_m \in [0, 2\pi)$ (Johnson 1987)

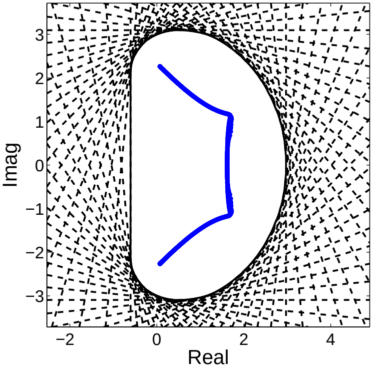
Compute and plot intersection $\bigcap \{ z : \mathrm{Re}(e^{i\alpha}z) \leq \mu^\alpha \}$

Large A, cheap approximation for $W(A)$ via Krylov space(s)

2 options:

- for angles $\alpha_1, \ldots, \alpha_m \in [0, \pi)$ (mind the π!)
  apply Lanczos to $\frac{1}{2}(e^{-i\alpha}A + e^{i\alpha}A^*)$ and compute
  $\lambda_{\max}(\frac{1}{2}(e^{-i\alpha}A + e^{i\alpha}A^*))$ and $\lambda_{\min}(\frac{1}{2}(e^{-i\alpha}A + e^{i\alpha}A^*))$
  this computes μ^α and $\mu^{\alpha+\pi}$ in 1 go
- "economy version": apply Arnoldi to A:
  $AV_k = V_k H_k + \beta_k \mathbf{v}_{k+1}\mathbf{e}_k^T$
  approximate $W(H_k) \approx W(A)$
  uses 1 Arnoldi process instead of $\frac{m}{2}$ Lanczos processes

Advertisement 1: inclusion regions with the field of values



However, approximation $W(H_k) \subset W(A)$ is often a crude eigenvalue inclusion region:

$\Lambda(A) \subset W(H_k)$ may be “unnecessarily wide”

Can we do better? yes, even from the same Krylov space!

(H., Singer, Zachlin 2006, 2008) (H., 2011)

Advertisement 1: inclusion regions with the field of values

$\Lambda(A) \subseteq W(A)$

$\Lambda(A) \subseteq W(A) \; \cap \; \dfrac{1}{W(A^{-1})}$ (Manteuffel, Starke 1996)

In fact (H., Singer, Zachlin 2008)

$$\Lambda(A) = \bigcap_{\tau \in \mathbb{C} \setminus \Lambda(A)} \frac{1}{W((A - \tau I)^{-1})} + \tau$$

$W(A)$ and $\dfrac{1}{W(A^{-1})}$ are special cases

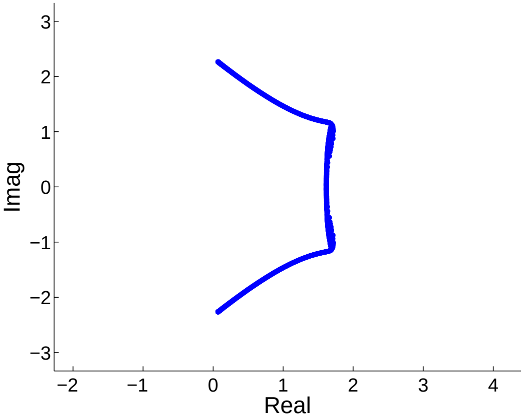
for $\tau = \infty$ and $\tau = 0$

Idea: for practical inclusion region: take finite intersection

For large matrices:

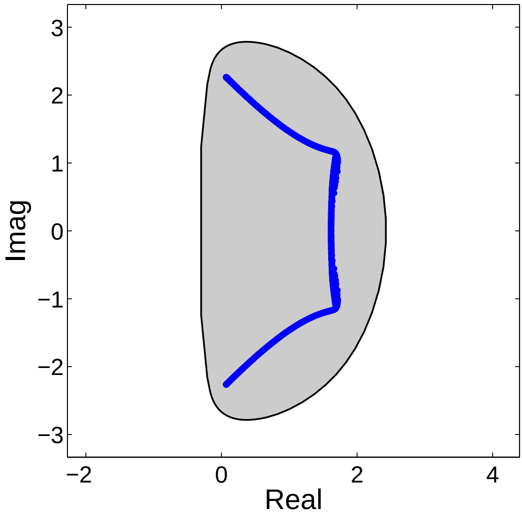
approximate $\dfrac{1}{W((A - \tau I)^{-1})}$ by projection onto same Krylov space

Advertisement 1: inclusion regions with the field of values



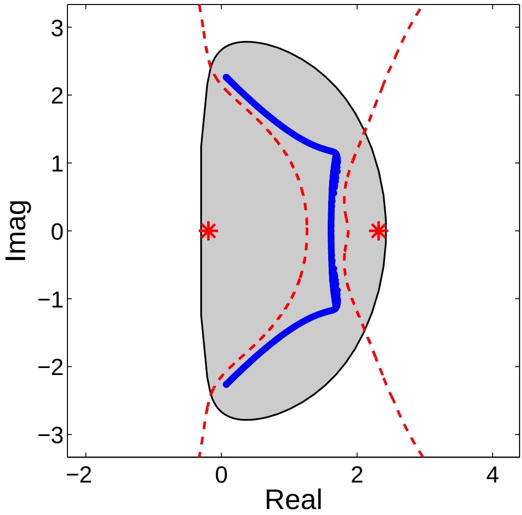
$A = 1000 \times 1000$ grcar matrix

Advertisement 1: inclusion regions with the field of values



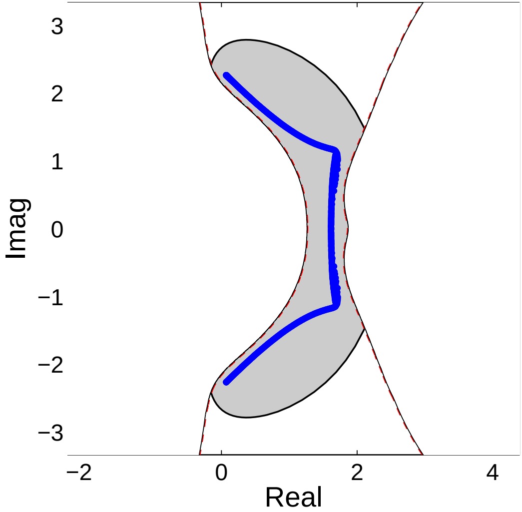
Approximation to field of values from $\mathcal{K}_5(A, \mathbf{v}_1)$

Advertisement 1: inclusion regions with the field of values



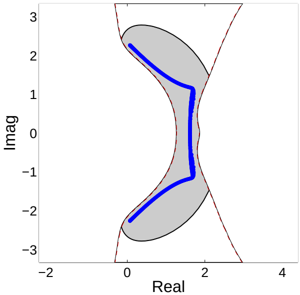
Intersection with two other regions determined with $\mathcal{K}_5(A, \mathbf{v}_1)$

Advertisement 1: inclusion regions with the field of values



Intersection with two other regions determined with $\mathcal{K}_5(A, \mathbf{v}_1)$ (gray area)

Advertisement 1: inclusion regions with the field of values



Surprisingly good eigenvalue inclusion region with only **5** MVs!
(gray area)

long before we have found any 1 eigenvalue accurately!

(H., Singer, Zachlin 2006, 2008; H. 2011)

∃ beautiful connections
with family of confocal boundary generating curves
(Kippenhahn 1951) (Zachlin, H. 2008) (H., Singer, Zachlin 2008)

Talk

Advertisements:

(1) spectral inclusion regions based on the field of values

(2) inner-outer control for Jacobi–Davidson

(3) probabilistic bounds for eigenvalue problems

Main part: linear ill-posed problems

Advertisement 2: Jacobi–Davidson inner-outer control

$Ax = \lambda x \qquad \lambda \approx \tau$

Pros Jacobi–Davidson include:

- accelerated inexact Newton method (inner-outer method)
- flexible, many parameters

Cons Jacobi–Davidson include:

- accelerated inexact Newton method (inner-outer method)
- flexible, many parameters

Advertisement 2: Jacobi–Davidson inner-outer control

Goal: find eigenpair $(\lambda, \mathbf{x})$ with $\lambda \approx \tau \in \mathbb{C}$

Starting vector $\mathbf{v}_1$

for $k = 1, 2, \ldots$ [outer iteration]

Rayleigh–Ritz on $\mathcal{V}_k$: $v = V_k\mathbf{c}$ with $(\theta, \mathbf{c})$ eigenpair of $V_k^* A V_k$

Solve approximately $\underbrace{(I - \mathbf{v}\mathbf{v}^*)(A - \tau I)\,\mathbf{t} = -(A - \theta I)\,\mathbf{v}}_{\text{inner iteration}}$

Use $\mathbf{t}$ to expand $\mathcal{V}_k$ to $\mathcal{V}_{k+1}$

end

Main question: when to stop inner iteration?
original method: fixed number of GMRES steps, but is not optimal

Key: cheap estimate for relation between inner and outer residual

Inner residual $\quad \mathbf{r}_{\text{in}} = -(A - \theta I)\,\mathbf{v} - (I - \mathbf{v}\mathbf{v}^*)(A - \tau I)\,\mathbf{t}$

Outer residual $\quad \mathbf{r}_{\text{out}} = \min_{\xi} \dfrac{\|(A - \xi I)(\mathbf{v} + \mathbf{t})\|}{\|\mathbf{v} + \mathbf{t}\|}$

Advertisement 2: Jacobi–Davidson inner-outer control

Hermitian case:

- JDCG: (Notay, 2002)

  Essence: $\underbrace{(I - \mathbf{v}\mathbf{v}^*)(A - \theta I)}_{\text{PD on } v^\perp \text{ if } \theta < \frac{1}{2}(\lambda_1 + \lambda_2)} \mathbf{t} = -(A - \theta I)\,\mathbf{v}$

- JDQMR: (Stathopoulos, 2008)

  Main advantages:
  - may use indefinite preconditioners
  - works for interior eigenvalues (indefinite systems)

Non-Hermitian case:

- JD + GMRES (H., Notay, 2009)

Results are very promising

also: one step towards a "parameter-free Jacobi–Davidson"

see also (JADAMILU: Bollhöfer, Notay 2008) for $A^* = A$

Advertisement 2: Jacobi–Davidson inner-outer control

One key result:

Correction equation $$(I - \mathbf{v}\mathbf{v}^*)(A - \tau I)\underbrace{\mathbf{t}}_{\text{solve}} = -(A - \theta I)\,\mathbf{v}$$

Inner residual $$\mathbf{r}_{\text{in}} = -(A - \theta I)\,\mathbf{v} - (I - \mathbf{v}\mathbf{v}^*)(A - \tau I)\,\mathbf{t}$$

Outer residual $$\mathbf{r}_{\text{out}} = \min_{\xi} \frac{\|(A - \xi I)(\mathbf{v} + \mathbf{t})\|}{\|\mathbf{v} + \mathbf{t}\|}$$

Thm $$\sqrt{\zeta} \le \mathbf{r}_{\text{out}} \le \sqrt{2}\,\sqrt{\zeta}$$

for $$\zeta := \underbrace{\frac{\|\mathbf{r}_{\text{in}}\|^2}{1 + \|\mathbf{t}\|^2}}_{\text{depends on } \|\mathbf{r}_{\text{in}}\|} + \underbrace{\left(\frac{\beta\,\|\mathbf{t}\|}{1 + \|\mathbf{t}\|^2}\right)^2}_{\text{m.o.l. independent}}$$ (for certain $\beta = \beta(\mathbf{t})$)

e.g., stop if first term $\approx$ second term

Talk

Advertisements:

(1) spectral inclusion regions based on the field of values

(2) inner-outer control for Jacobi–Davidson

(3) probabilistic bounds for eigenvalue problems

Main part: linear ill-posed problems

Advertisement 3: Probabilistic eigenvalue bounds

A large symmetric matrix, $\lambda_{\max}$ wanted

Random starting vector $\mathbf{v}_1$

Lanczos process $AV_k = V_k T_k + \beta_k \mathbf{v}_{k+1} \mathbf{e}^T$

Largest eigenvalue $\theta_{\max}^{(k)}$ of T_k is approximation to $\lambda_{\max}$ with

$$\theta_{\max}^{(k)} \leq \theta_{\max}^{(k+1)} \leq \cdots \leq \lambda_{\max}$$

Residual of Ritz pair $\quad \mathbf{r} = A\mathbf{v} - \theta_{\max}^{(k)} \mathbf{v}$

Bauer–Fike: there exists λ in interval $[\theta_{\max}^{(k)} - \|\mathbf{r}\|,\ \theta_{\max}^{(k)} + \|\mathbf{r}\|]$

BUT: do not know if this is $\lambda_{\max}$

If $\mathbf{v}_1$ is deficient in direction of eigenvector $\mathbf{x}_{\max}$
then $\lambda_{\max}$ may be arbitrarily larger

However, probability that this is the case is small

Advertisement 3: Probabilistic eigenvalue bounds

Probability that λ_{max} is much larger than θ_{max} is small

Vice versa: given a probability $1-\varepsilon$ ($0<\varepsilon\ll 1$: e.g., 99%)
we can derive an upper bound $\lambda_{\mathsf{max}} \leq \lambda^{\mathrm{up}}$
that holds with probability $1-\varepsilon$

Probabilistic upper bound, derived with Lanczos polynomial

$=$ polynomial that implicitly arises in Lanczos process

(see next slide)

Therefore: Lanczos

- not only gives approximation $\theta \leq \lambda_{\mathsf{max}}$
- but also (with very little extra work)
  probabilistic upper bound $\lambda_{\mathsf{max}} \leq \lambda^{\mathrm{up}}$ with probability $1-\varepsilon$

Advertisement 3: Probabilistic eigenvalue bounds

Derivation in brief:

Lanczos $\beta_k \mathbf{v}_{k+1} = A\mathbf{v}_k - \alpha_k \mathbf{v}_k - \beta_{k-1}\mathbf{v}_{k-1}$

Lanczos polynomial $\mathbf{v}_{k+1} = p_k(A)\,\mathbf{v}_1$

satisfies $\beta_k\, p_k(t) = t\, p_{k-1}(t) - \alpha_k p_{k-1}(t) - \beta_{k-1} p_{k-2}(t)$

in fact $p_k(t) = (\beta_1 \cdots \beta_k)^{-1} \det(tI - V_k^T A V_k)$

Decomposition of $\mathbf{v}_1$ in eigenvectors: $\mathbf{v}_1 = \sum_{j=1}^{n} \gamma_j\, \mathbf{x}_j$ then

$$1 = \|\mathbf{v}_{k+1}\|^2 = \|p_k(A)\,\mathbf{v}_1\|^2 = \sum_{j=1}^{n} \gamma_j^2\, p_k(\lambda_j)^2$$

$$\geq \ \gamma_n^2\, p_k(\lambda_n)^2 \ \geq_{\mathbf{P}} \ \delta^2\, p_k(\lambda_n)^2$$

so $\boxed{p_k(\lambda_n) \leq_{\mathbf{P}} \frac{1}{\delta}}$

where $\leq_{\mathbf{P}}$ means: holds with user-selected probability $1 - \varepsilon$

Advertisement 3: Probabilistic eigenvalue bounds

Technique has many applications:

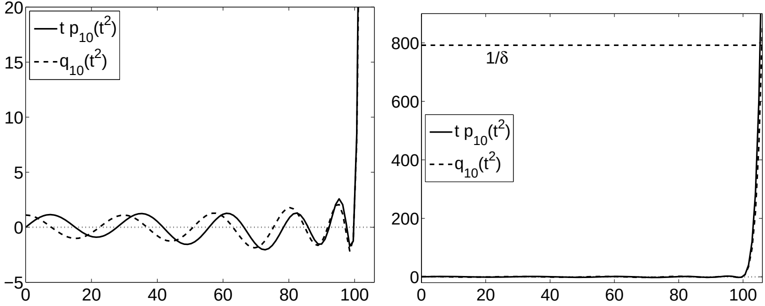
- Lanczos gives approximation $\theta \leq \lambda_{\max}$
  Probabilistic upper bound: $\lambda_{\max} \leq \lambda^{\mathrm{up}}$ with probability $1 - \varepsilon$
  (99%, 99.9%, ...) (van Dorsselaer, H., van der Vorst 2000)
- Lanczos bidiagonalization gives $\theta \leq \sigma_{\max} = \|A\|$
  Probabilistic upper bound: $\sigma_{\max} \leq \sigma^{\mathrm{up}}$ with probability $1 - \varepsilon$
  (H. 2011)
- Arnoldi gives approximations $\theta \approx \lambda$ to exterior eigenvalues
  Probabilistic inclusion region:
  eigenvalues $\subset \Omega$ with high probability (H. 2011)
- Lanczos gives "lower bound" $\mathcal{L} \subseteq W(A)$
  Probabilistic inclusion region ("upper bound"):
  $W(A) \subseteq \mathcal{U}$ with high probability (H. 2011)

(See examples on next slides)

Advertisement 3: Probabilistic eigenvalue bounds

Lanczos bidiagonalization for $\|A\|$:

toy example `diag(1:100)`, $\varepsilon = 0.01$, 10 steps



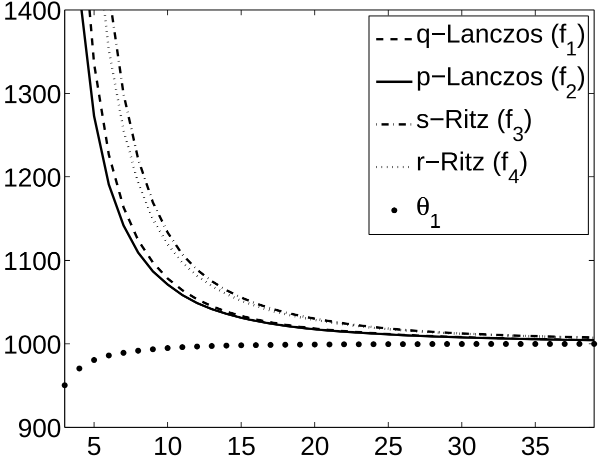
Largest zeros of polies give approximation to $\|A\| = 100$

$\varepsilon = 0.01 \Rightarrow$ intersection point with $1/\delta \approx 792$

gives probabilistic upper bound $\|A\| \lesssim 105$ with probability 99%

Advertisement 3: Probabilistic eigenvalue bounds

$\|A\|$ for `diag(1:1000)`, $\varepsilon = 0.01$



Ritz values and 4 types of probabilistic upper bounds for $\|A\|$ as a function of Krylov dimension k

Advertisement 3: Probabilistic eigenvalue bounds

More realistic example:

Lanczos bidiagonalization for $\|A\|$:

23560 $\times$ 23560 matrix `af23560`, 460598 nonzeros

10 steps of Lanczos bidiagonalization, 0.15 second:

$\theta \approx 645.7$ as approximation to $\|A\|$

Probabilistic upper bound

- $\varepsilon = 0.01$: $\sigma_{\mathrm{up}} \approx 646.8$
- $\varepsilon = 0.001$: $\sigma_{\mathrm{up}} \approx 652.0$

therefore: $\|A\| \in [\theta,\ 1.01\,\theta]$ with 99.9% probability

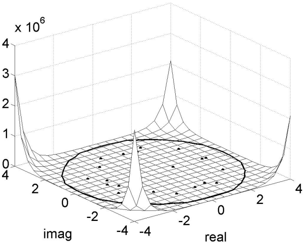
Results crush those of Matlab's `normest`

(based on power method on A^TA)

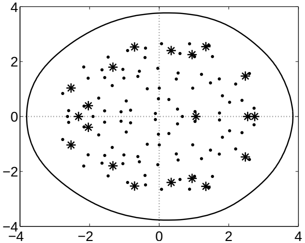
Advertisement 3: Probabilistic eigenvalue bounds

Also possible for non-symmetric problems
(although theoretical foundation weaker)

Arnoldi: toy problem `rand(100)-0.5`



Ritz values (·)

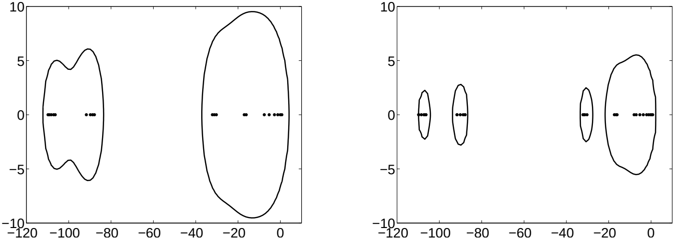


eigenvalues (·)
Ritz values (*)

Probabilistic inclusion region of level 99% (—)

Advertisement 3: Probabilistic eigenvalue bounds

Arnoldi: `dw8192`, 41746 nonzeros



20 Arnoldi steps 29 steps

Ritz values and probabilistic inclusion region of level 99%

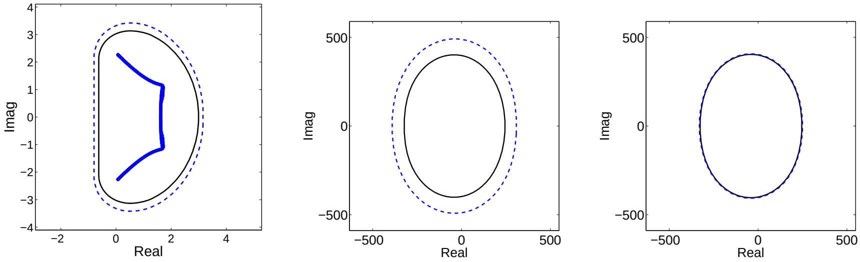
One advantage:

this (probabilistic) inclusion region is not necessarily convex!
(or even connected)

in contrast to, e.g., field of values $W(A)$

Advertisement 3: Probabilistic eigenvalue bounds

Also possible: probabilistic upper bound for field of values $W(A)$



`grcar(1000)`, $k = 20$ `af23560`, $k = 10$ `af23560`, $k = 20$

lower bound (—), probabilistic 99% upper bound (--) for $W(A)$

$\Rightarrow$ may also serve as excellent stopping criterion

Open problem: (prob.) upper bound for $\kappa(A) = \|A\| \, \|A^{-1}\|$

Difficulty: bound $\sigma_{\min}(A)$ from below

Advertisements have slogans

Slogans:

(1) spectral inclusion regions based on the field of values

"how to find a good inclusion region for all eigenvalues of a large matrix with just 5 MVs?"

(2) inner-outer control for Jacobi–Davidson

"how to speed up Jacobi–Davidson by a factor 2 with a few lines of code?"

(3) probabilistic bounds for eigenvalue problems

"how to find an interval $[\theta,\ 1.01\,\theta]$ that contains $\|A\|_2$ with probability 99.9% within 0.15 sec for a 23560×23560 matrix"

(Van Dorsselaer, H., Van der Vorst 2000) (H. 2011)
(H., Singer, Zachlin 2008) (H., Notay 2009)

Outline main part: linear ill-posed problems

- **Introduction** (see also Melina's talk)
- Tikhonov and other filter functions
  - new approach: fractional Tikhonov
- general regularization operator L
  - new approach: simple Krylov approach
- including apriori information
  - new approach: subspace restricted SVD
- combining different approaches
  - new approach: linear combination method
- Conclusions

Slogan: "how to wisely divide by 0 ?"

Overview: setting

Linear discrete ill-posed problem $A\mathbf{x} = \mathbf{b}$

A: $m \times n$ or $n \times m$; for convenience: $n \times n$

Noisy right-hand side $\mathbf{b} = \underbrace{\mathbf{b}_{\text{true}}}_{\text{error-free}} + \mathbf{e}$

in much we assume: error estimate $\|\mathbf{e}\| \approx \varepsilon$
(discrepancy principle)

Assume $A\mathbf{x} = \mathbf{b}_{\text{true}}$ is consistent,

look for solution $\mathbf{x}_{\text{true}}$ with minimal norm

Overview: introduction

$A\mathbf{x} = \mathbf{b}$, SVD $A = U\Sigma V^T$

Solution $\mathbf{x} = A^{-1}\mathbf{b} = \sum_{j=1}^{n} \sigma_j^{-1}(\mathbf{u}_j^T\mathbf{b})\,\mathbf{v}_j$

However $\mathbf{b} = \mathbf{b}_{\text{true}} + \mathbf{e}$

So unwanted component $\sum_{j=1}^{n} \sigma_j^{-1}(\mathbf{u}_j^T\mathbf{e})\,\mathbf{v}_j$

disastrous since many $\sigma_j \approx 0$

Essential difficulty with discrete ill-posed problems:

Noise is greatly magnified by ill-conditioned matrix

Regularization: replace system by system that is less sensitive to perturbations in right-hand side

(Groetsch 1984) (Hansen 1988) (Hanke 1995) (Engl, Hanke, Neubauer 1996)

Overview: Tikhonov

- $\min_{\mathbf{x}} \|\mathbf{x}\|^2$ given $\|A\mathbf{x} - \mathbf{b}\|^2 = \varepsilon^2$
- $\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2$ given $\|\mathbf{x}\|^2 = \Delta^2$

both lead to

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2 + \mu \|\mathbf{x}\|^2$$

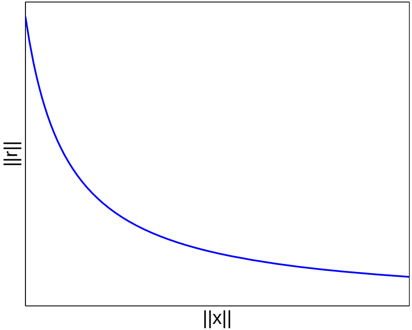
$\mu > 0$: regularization parameter; normal equations

$$(A^T A + \mu I)\,\mathbf{x}_\mu = A^T \mathbf{b}$$

albeit with different μ

Overview: L-curve

$\|\mathbf{r}\| = \|\mathbf{b} - A\mathbf{x}\|$ vs $\|\mathbf{x}\|$



If $\mathbf{x} = \mathbf{0}$ then $\|\mathbf{r}\| = \|\mathbf{b}\|$

If $\mathbf{r} \to \mathbf{0}$ then "$\|\mathbf{x}\| \to \infty$" (very large)

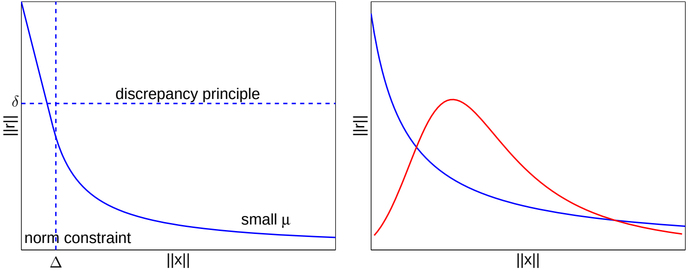
Seek balance between:

- $\mathbf{x} = A^{-1}\mathbf{b}$: small $\|\mathbf{r}\|_2 = \|\mathbf{b} - A\mathbf{x}\|$, but huge $\|\mathbf{x}\|$
- attractive $\|\mathbf{x}\| = \mathcal{O}(1)$, but what about $\|\mathbf{r}\|$?

Regularization: avoiding large $\|\mathbf{x}\|$

Overview: parameter choice

$$(A^T A + \mu I)\,\mathbf{x}_\mu = A^T\mathbf{b}$$



- discrepancy principle $\|\mathbf{b} - A\mathbf{x}\| = \delta = \gamma\varepsilon$
- norm constraint $\|\mathbf{x}\| = \Delta$
- maximum curvature ("L-curve")
- generalized cross validation (e.g., leave-1-out)
  e.g. (Chung, Nagy, O'Leary 2008)
- quasi-optimal, . . . (Reginska 1996)

Overview: discrepancy principle (1)

In many applications, an estimate of the noise is available:

$A\mathbf{x} = \mathbf{b} = \mathbf{b}_{\text{true}} + \mathbf{e}, \quad \|\mathbf{e}\| = \varepsilon$

Aiming at $\mathbf{b} - A\mathbf{x} = \mathbf{0}$ is not necessary:

- greatly magnifies noise $\mathbf{e}$ (huge $\|\mathbf{x}\|$)
- $\|\mathbf{b} - A\mathbf{x}_{\text{true}}\| = \|\mathbf{b} - \mathbf{b}_{\text{true}}\| = \|\mathbf{e}\| = \varepsilon$

Sought solution $\mathbf{x}_{\text{true}}$ has residual norm ε

Idea: solve $A\mathbf{x} = \mathbf{b}$ such that residual norm $\|\mathbf{b} - A\mathbf{x}\| = \varepsilon$

To be more precise: $\|\mathbf{b} - A\mathbf{x}\| = \gamma\varepsilon$, for an $\gamma > 1$, then:

Thm: if $\varepsilon \to 0$ then $\mathbf{x} \to \mathbf{x}_{\text{true}}$

Overview: discrepancy principle (2)

What does $\|A\mathbf{x} - \mathbf{b}\| = \varepsilon$ mean geometrically?

Use SVD $A = U\Sigma V^T$:

$$\|U\Sigma V^T\mathbf{x} - \mathbf{b}\| = \varepsilon$$

$$\|\Sigma V^T\mathbf{x} - U^T\mathbf{b}\| = \varepsilon$$

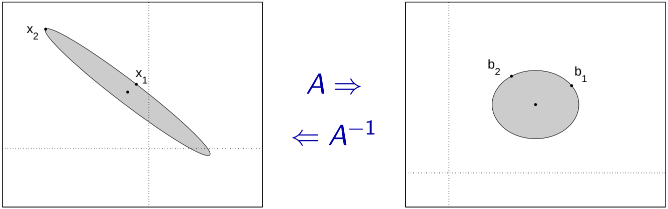
$$\|\Sigma\mathbf{y} - \mathbf{c}\|^2 = \varepsilon^2, \quad \mathbf{y} = V^T\mathbf{x}, \quad \mathbf{c} = U^T\mathbf{b}$$

$$\sum_{j=1}^{n}(\sigma_j\mathbf{y}_j - \mathbf{c}_j)^2 = \varepsilon^2 \quad \text{ellipsoid}$$

Overview: discrepancy principle (3)

$\|A\mathbf{x} - \mathbf{b}\| = \varepsilon$: ellipsoid

Which $\mathbf{x}$ on ellipsoid do we take ??



$A\mathbf{x}_1 = \mathbf{b}_1 \qquad A\mathbf{x}_2 = \mathbf{b}_2 \qquad \|\mathbf{b}_1 - \mathbf{b}_{\text{true}}\| = \|\mathbf{b}_2 - \mathbf{b}_{\text{true}}\|$

- $\|\mathbf{x}_2\| \gg \|\mathbf{x}_1\|$
- $\mathbf{x}_2$ more sensitive with respect to perturbations of $\mathbf{b}_2$

Therefore, generally $\mathbf{x}_1$ preferred, but still:

Discrete ill-posed problems are essentially underdetermined

⇒ use all available information!

⇒ use apriori information whenever possible!

Overview: large problems (1)

Iterative processes for $A\mathbf{x} = \mathbf{b}$, large A

- LSQR/Lanczos bidiagonalization on $A^TA\mathbf{x} = A^T\mathbf{b}$
  $\mathcal{K}_k(A^TA, A^T\mathbf{b}) = \text{span}(A^T\mathbf{b}, (A^TA)A^T\mathbf{b}, \dots, (A^TA)^{k-1}A^T\mathbf{b})$
- GMRES/MINRES on $A\mathbf{x} = \mathbf{b}$
  $\mathcal{K}_k(A, \mathbf{b}) = \text{span}(\mathbf{b}, A\mathbf{b}, \dots, A^{k-1}\mathbf{b})$
- Range-restricted GMRES on $A\mathbf{x} = \mathbf{b}$
  $\mathcal{K}_k(A, A\mathbf{b}) = \text{span}(A\mathbf{b}, A^2\mathbf{b}, \dots, A^k\mathbf{b})$

Key ideas:

- a few steps of iterative method works as regularization
  (not too many)
- $A\mathbf{b}$ and $A^T\mathbf{b}$ are smoother (contain less noise) than $\mathbf{b}$
- also enables "hybrid approaches": project onto Krylov space, then apply method "**X**" for small problems

Overview: large problems (2)

Lots of references . . .

- LSQR/Lanczos bidiagonalization on $A^T A\mathbf{x} = A^T\mathbf{b}$

  (O'Leary, Simmons 1981) (Björck 1988) (Golub, Von Matt 1997) (Calvetti, Golub, Reichel 1999) (Calvetti, Morigi, Reichel, Sgallari 2000) (Calvetti, Reichel 2003) (Hanke 2001) (Kilmer, O'Leary 2001) (Reichel, Sadok, Shyshkov 2007)

- GMRES/MINRES on $A\mathbf{x} = \mathbf{b}$

  (Hanke 1995) (Calvetti, Lewis, Reichel 2002)

- Range-restricted GMRES on $A\mathbf{x} = \mathbf{b}$

  (Calvetti, Lewis, Reichel 2000) (Neuman, Reichel, Sadok 2011)

Overview: extra constraints

Extra constraints

- nonnegativity constraints $(\mathbf{x}_i \geq 0)$
  (Calvetti, Lewis, Reichel, Sgallari 2004)
  (Nagy, Strakos 2000)
  (Hanke, Nagy, Vogel 2000)
- box constraints $(\ell_i \leq \mathbf{x}_i \leq u_i)$
  (Morigi, Reichel, Sgallari, Zama 2007)
  (Morigi, Plemmons, Reichel, Sgallari 2011)

Outline main part: linear ill-posed problems

- Introduction (see also Melina's talk)

⇒ **Tikhonov and other filter functions**
  - **new approach: fractional Tikhonov**
- general regularization operator L
  - new approach: simple Krylov approach
- including apriori information
  - new approach: subspace restricted SVD
- combining different approaches
  - new approach: linear combination method
- Conclusions

Overview: filter functions (1)

Recall: exact solution $\mathbf{x} = A^{-1}\mathbf{b} = \sum_{j=1}^{n} \sigma_j^{-1} (\mathbf{u}_j^T \mathbf{b})\, \mathbf{v}_j$

Many approaches of the form $\mathbf{x} = \sum_{j=1}^{n} \underbrace{\varphi(\sigma_j)}_{\text{filter function}} (\mathbf{u}_j^T \mathbf{b})\, \mathbf{v}_j$

TSVD $\varphi_{\mathrm{tsvd}}(\sigma) = \begin{cases} 1/\sigma & \text{if } \sigma \geq \tau \\ 0 & \text{otherwise} \end{cases}$

Tikhonov $\varphi_{\mathrm{tikh}}(\sigma) = \dfrac{\sigma}{\sigma^2 + \mu}$

See also (Klann, Ramlau 2008) (Calvetti, Reichel, Zhang 1999)

Overview: filter functions (2)

Recall: exact solution $\mathbf{x} = A^{-1}\mathbf{b} = \sum_{j=1}^{n} \sigma_j^{-1} (\mathbf{u}_j^T \mathbf{b}) \mathbf{v}_j$

Filter $\mathbf{x} = \sum_{j=1}^{n} \underbrace{\varphi(\sigma_j)}_{\text{filter function}} (\mathbf{u}_j^T \mathbf{b}) \mathbf{v}_j$

Desirable properties of filter functions:

- $\varphi(\sigma) = \sigma^{-1} + o(\sigma^{-1}) \qquad (\sigma \to \infty)$
  "true inversion for large singular values"
- $\varphi(\sigma) = o(1) \qquad (\sigma \to 0)$
  "small weights for small singular values"

Overview: filter functions (2)

Desirable properties of filter functions

- $\varphi(\sigma) = \sigma^{-1} + o(\sigma^{-1}) \qquad (\sigma \to \infty)$
- $\varphi(\sigma) = o(1) \qquad (\sigma \to 0)$

Tikhonov $\quad \varphi_{\mathrm{tikh}}(\sigma) = \dfrac{\sigma}{\sigma^2 + \mu}$

satisfies desired asymptotics:

$$\begin{aligned} \varphi_{\mathrm{tikh}}(\sigma) &= \sigma^{-1} + \mathcal{O}(\sigma^{-3}) \quad (\sigma \to \infty) \\ \varphi_{\mathrm{tikh}}(\sigma) &= \frac{\sigma}{\mu} + \mathcal{O}(\sigma^{3}) \qquad (\sigma \to 0) \end{aligned}$$

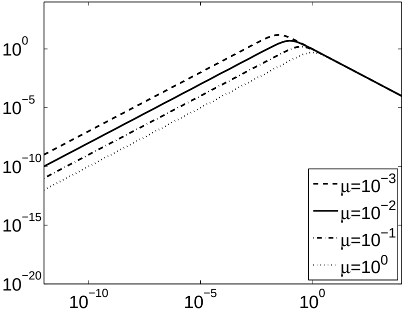
However, $\quad \dfrac{\sigma}{\mu} \quad (\sigma \to 0)$ may be "too fast"

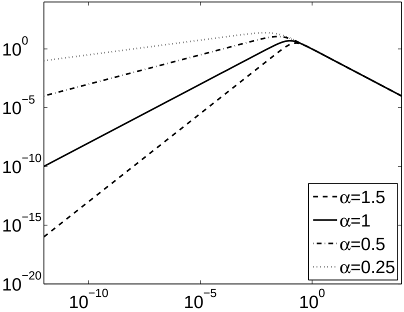
New idea: instead: fractional Tikhonov

$$\varphi_{\mathsf{frac}}(\sigma) = \frac{\sigma^{\alpha}}{\sigma^{\alpha+1} + \mu} \quad \text{for} \quad \alpha < 1 \qquad \text{(H., Reichel 2011)}$$

Filter functions: Tikhonov, fractional Tikhonov

As functions of σ: $(\mu = 10^{-2})$





$$\varphi_{\text{tikh}}(\sigma) = \frac{\sigma}{\sigma^2 + \mu} \qquad \varphi_{\text{frac}}(\sigma) = \frac{\sigma^\alpha}{\sigma^{\alpha+1} + \mu}$$

$$(\sigma \to \infty): \quad \sigma^{-1} + \mathcal{O}(\sigma^{-3}) \qquad (\sigma \to \infty): \quad \sigma^{-1} + \mathcal{O}(\sigma^{-(\alpha+2)})$$

$$(\sigma \to 0): \quad \frac{\sigma}{\mu} + \mathcal{O}(\sigma^3) \qquad (\sigma \to 0): \quad \frac{\sigma^\alpha}{\mu} + \mathcal{O}(\sigma^{2\alpha+1})$$

Fractional Tikhonov

Consider weighted minimization

$$\min_{\mathbf{x}\in\mathbb{R}^n} \|A\mathbf{x}-\mathbf{b}\|_W^2 + \mu \|\mathbf{x}\|^2$$

where $\|\mathbf{x}\|_W = (\mathbf{x}^T W \mathbf{x})^{1/2}$, $\quad W$ SPSD

Standard Tikhonov: $W = I$

Idea: take

$$W = (AA^T)^{(\alpha-1)/2}$$

Normal equations associated with Tikhonov minimization:

$$((A^TA)^{(\alpha+1)/2} + \mu I)\,\mathbf{x} = (A^TA)^{(\alpha-1)/2}A^T\mathbf{b}$$

Strengths of standard Tikhonov

Filter $\mathbf{x} = \sum_{j=1}^{n} \underbrace{\varphi(\sigma_j)}_{\text{filter function}} \; (\mathbf{u}_j^T \mathbf{b})\, \mathbf{v}_j$

Standard Tikhonov takes $\alpha = 1$ in $\dfrac{\sigma^{\alpha}}{\sigma^{\alpha+1} + \mu}$

Proposition: standard Tikhonov is optimal in two ways:

- If $\mathbf{x}_{\mu,\alpha}$ satisfies discrepancy principle $\|\mathbf{b} - A\mathbf{x}_{\mu,\alpha}\| = \delta$

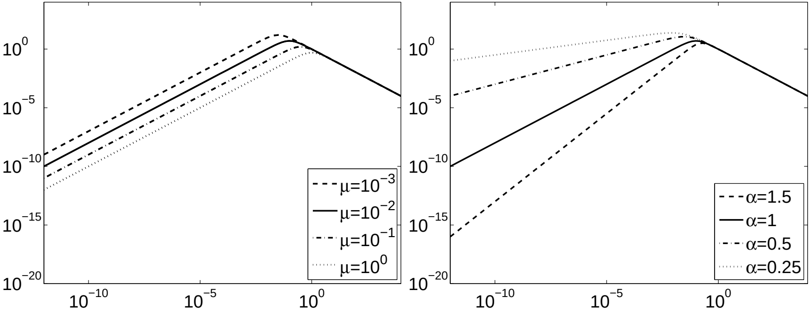
  $\|\mathbf{x}_{\mu,\alpha}\|$ "is minimal for $\alpha = 1$" (local minimum)

- If $\mathbf{x}_{\mu,\alpha}$ satisfies norm constraint $\|\mathbf{x}_{\mu,\alpha}\| = \Delta$

  $\|\mathbf{b} - A\mathbf{x}_{\mu,\alpha}\|$ "is minimal for $\alpha = 1$" (local minimum)

Weakness of standard Tikhonov

However, standard Tikhonov may provide **too much smoothing**



"solution too smooth"
"solution too short"

Idea: take $\alpha < 1$, for instance $\alpha = 0.5$

Also applicable to large-scale problems by first projecting onto Krylov space

Experiment fractional Tikhonov

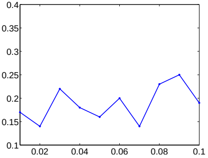
$n = 100$ examples for error-levels (1%, 5%, 10%) and $\alpha = 0.6$

quality of solutions: $\frac{\|\mathbf{x} - \mathbf{x}_{\text{true}}\|}{\|\mathbf{x}_{\text{true}}\|}$ for standard and fractional Tikhonov

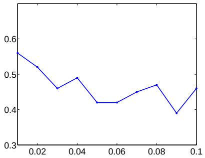
Error	1%			5%			10%
Problem	Tikh	Frac	Ratio	Tikh	Frac	Ratio	Ratio
baart	2.1e-1	1.9e-1	9.1e-1	3.3e-1	3.0e-1	9.0e-1	9.4e-1
deriv2-1	2.8e-1	2.7e-1	9.6e-1	3.8e-1	3.6e-1	9.3e-1	9.3e-1
deriv2-2	2.7e-1	2.5e-1	9.5e-1	3.8e-1	3.5e-1	9.2e-1	9.2e-1
deriv2-3	3.8e-2	4.6e-2	1.2e-0	7.0e-2	6.3e-2	8.9e-1	8.7e-1
foxgood	4.3e-2	4.0e-2	9.3e-1	1.4e-1	1.0e-1	7.3e-1	7.4e-1
gravity	3.7e-2	2.3e-2	6.4e-1	7.4e-2	5.6e-2	7.6e-1	8.2e-1
heat	1.5e-1	1.6e-1	1.1e-0	3.1e-1	3.1e-1	9.9e-1	9.5e-1
ilaplace	1.6e-1	1.5e-1	9.2e-1	2.0e-1	1.8e-1	9.2e-1	9.4e-1
phillips	2.8e-2	4.8e-2	1.7e-0	6.5e-2	7.3e-2	1.1e-0	9.8e-1
shaw	1.5e-1	1.3e-1	8.6e-1	1.8e-1	1.7e-1	9.5e-1	9.3e-1

Experiment fractional Tikhonov: optimal α

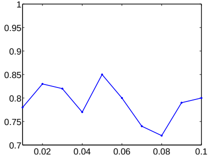
as function of noise level: often $\alpha \ll 1$ beneficial



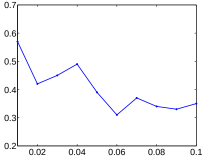
baart



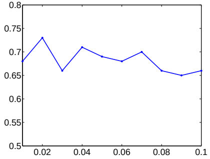
deriv2-1



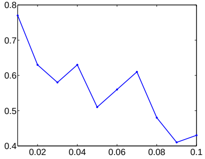
deriv2-3



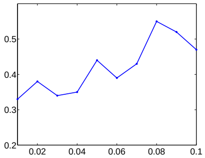
foxgood



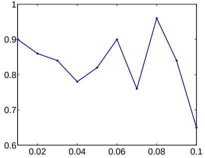
gravity



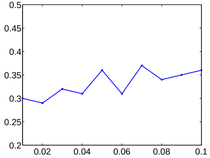
heat



ilaplace



phillips



shaw

Outline main part: linear ill-posed problems

- Introduction (see also Melina's talk)
- Tikhonov and other filter functions
  - new approach: fractional Tikhonov

$\Rightarrow$ **general regularization operator** L
  - **new approach: simple Krylov approach**
- including apriori information
  - new approach: subspace restricted SVD
- combining different approaches
  - new approach: linear combination method
- Conclusions

Overview: generalized Tikhonov with L (1)

L regularization operator

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|^2 + \mu\,\|L\mathbf{x}\|^2 \qquad L \in \mathbb{C}^{p\times n}$$

Normal equations

$$(A^TA + \mu\, L^T L)\,\mathbf{x} = A^T\mathbf{b}$$

Assumption:

$$\mathcal{N}(A) \cap \mathcal{N}(L) = \{\mathbf{0}\}$$

then there is a unique solution $\mathbf{x}_\mu$ for all $\mu > 0$

Overview: generalized Tikhonov with L (2)

For instance:

- $L = I$ no special regularization operator L
- $$L = L_1 = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}$$
  discretization of first-order derivative
- $$L = L_2 = \begin{bmatrix} -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & -1 & 2 & -1 \end{bmatrix} \in \mathbb{R}^{(n-2)\times n}$$
  discretization of second-order derivative
- $L = L_3$, etc
- $L = \begin{bmatrix} I \otimes L_1 \\ L_1 \otimes I \end{bmatrix}$ for 2D images
- ...

Overview: generalized Tikhonov with L (3)

Small case

- TGSVD: truncated GSVD of (A, L) (Hansen 1989)
- Tikhonov: $(A^T A + \mu L^T L)\,\mathbf{x} = A^T \mathbf{b}$

Large case, transformation to standard form with $\mathbf{y} = L\mathbf{x}$

- if L square and invertible: $\min_{\mathbf{y}} \|AL^{-1}\mathbf{y} - \mathbf{b}\|^2 + \mu \|\mathbf{y}\|^2$
- otherwise: $\min_{\mathbf{y}} \|AL_A^{+}\mathbf{y} - \bar{\mathbf{b}}\|^2 + \mu \|\mathbf{y}\|^2$

  $L_A^{+} = (I - (A(I - L^{+}L))^{+}A)\,L^{+} \in \mathbb{R}^{n\times p}$
  the A-weighted generalized inverse of L (Elden 1982)

  $\bar{\mathbf{b}} = \mathbf{b} - A\bar{\mathbf{x}}, \qquad \bar{\mathbf{x}} = (A(I - L^{+}L))^{+}\mathbf{b}$

  Solution of original problem $\mathbf{x}_\mu = L_A^{+}\mathbf{y}_\mu + \bar{\mathbf{x}}$

only suitable for some problems!

Overview: generalized Tikhonov with L (4)

$$L_A^+ = \left(I - (A(I - L^+L))^+ A\right) L^+ \in \mathbb{R}^{n \times p}$$

However, matrix-vector products required:

- with L_A^+, AL_A^+
- possibly also with $(L_A^+)^T$ and $(AL_A^+)^T$

Only realistic for certain classes of L such as

- small bandwidth
- circulant matrices
- orthogonal projections
- sparse nonsingular matrices with fast LU factorization

Moreover, efficient evaluation of $(A(I - L^+L))^+$ requires $\mathcal{N}(L)$ to be known and low-dimensional

$\Rightarrow$ only suitable for some problems!

Generalized Tikhonov: new approach

Large case, transformation to standard form not feasible:

- joint bidiagonalization (Kilmer, Hansen, Espanol 2007)
- generalized Krylov method (Reichel, Sgallari, Ye 2011)

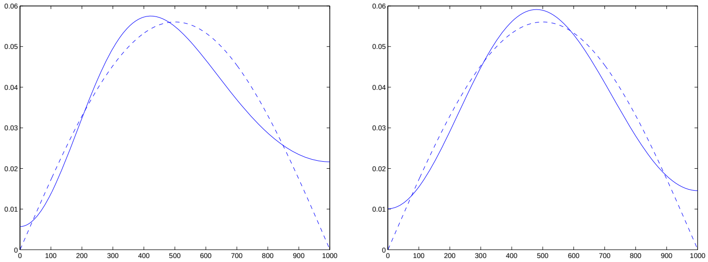
New approach: simple Krylov (H., Reichel 2010)

- just take simple Krylov subspace
  $\mathcal{K}_k(A^TA,\, A^T\mathbf{b}) = \text{span}\{A^T\mathbf{b}, (A^TA)A^T\mathbf{b}, \ldots, (A^TA)^{k-1}A^T\mathbf{b}\}$
  as search space

  Observation: this space is often of good quality
- use L only to extract solution from this space:
  $$\min_{\mathbf{x}\in\mathcal{K}_k(A^TA,\, A^T\mathbf{b})} \|A\mathbf{x}-\mathbf{b}\|^2 + \mu\,\|L\mathbf{x}\|^2$$

efficient computation with incremental QR-decompositions

Generalized Tikhonov: new approach



$L = I$ $\qquad$ $L = L_2$

Approximation — (5 Krylov steps), exact solution (--)

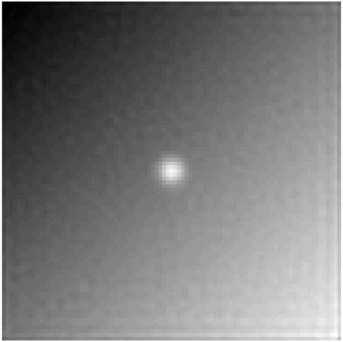
Generalized Tikhonov: new approach

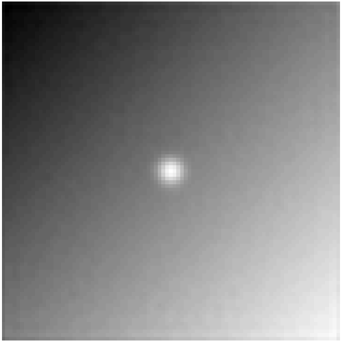
Blurred

restored with $L = \begin{bmatrix} I \otimes L_1 \\ L_1 \otimes I \end{bmatrix}$

size 16380×8281

Generalized Tikhonov: new approach





restored with $L = \begin{bmatrix} I \otimes L_1 \\ L_1 \otimes I \end{bmatrix}$

size 16380×8281

idem, plus splitting

(see later)

Outline main part: linear ill-posed problems

- Introduction (see also Melina's talk)
- Tikhonov and other filter functions
  - new approach: fractional Tikhonov
- general regularization operator L
  - new approach: simple Krylov approach

$\Rightarrow$ **including apriori information**
  - **new approach: subspace restricted SVD**
- combining different approaches
  - new approach: linear combination method
- Conclusions

Overview: using apriori information (1)

Discrete ill-posed problems are essentially underdetermined

$\Rightarrow$ use apriori information wherever possible!

Examples of "sensible vectors" to include:

- polynomial vectors:

$$W = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad W = \begin{bmatrix} 1 & 1 \\ \vdots & \vdots \\ 1 & n \end{bmatrix}, \quad W = \begin{bmatrix} 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{bmatrix}$$

  not assuming much knowledge about problem
- approximate solution provided by user
- approximate solutions obtained with other methods
- ...

Assumption (w.l.o.g.): columns of W are orthonormal

Overview: using apriori information (2)

Exploiting apriori information W

- LSQR with extra vectors (Calvetti, Reichel, Shuibi 2003)

Other option:

$P_W = WW^T, \ \ P_W^\perp = I - WW^T$

$AW = QR$

$P_Q = QQ^T, \ \ P_W^\perp = I - QQ^T$

Orthogonal projection, split $A\mathbf{x} = \mathbf{b}$ into $\mathbf{y} = P_W\mathbf{x}, \ \ \mathbf{z} = P_W^\perp\mathbf{x}$

$$\begin{aligned} P_Q\,A\,\mathbf{y} + P_Q\,A\,\mathbf{z} &= P_Q\,\mathbf{b} \\ P_Q^\perp\,A\,\mathbf{z} &= P_Q^\perp\,\mathbf{b} \end{aligned}$$

- projection + TSVD (Morigi, Reichel, Sgallari 2006)
- projection + Tikhonov (Morigi, Reichel, Sgallari 2007)
- projection + LSQR/GMRES (Baglama, Reichel 2007)

Overview: using apriori information (3)

Exploiting apriori information W

- GSVD of $(A, I - WW^T)$
  that is, take $L = I - WW^T$ (Morigi, Reichel, Sgallari 2007)
- often better:

New approach: SRSVD: subspace-restricted SVD

  exploits SVD of $A(I - WW^T)$ (H., Reichel 2010)

Motivation

TSVD: $\mathbf{x}_k = \sum_{j=1}^{k} \sigma_j^{-1} (\mathbf{u}_j^T \mathbf{b})\, \mathbf{v}_j$

One option: choose k such that discrepancy principle is satisfied:

$\min_k$ such that $\|\mathbf{b} - A\mathbf{x}_k\| \leq \gamma\varepsilon$ $\qquad (\gamma > 1)$

For some problems, TSVD is disappointing

Hansen et al (1992); Calvetti et al (2002); Eldén (2004); Morigi et al (2006)

often, issue is often **not** bad choice of k

instead, span$(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ does not contain good approximation

Problem may be seen as underdetermined

$\Rightarrow$ try to use (user-supplied) extra information W

Idea: TSVD where user can prescribe some right/left singular vectors

Subspace restricted SVD

Select orthonormal vectors $W = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_p]$

$$A(I - WW^T) = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^T$$

$$\widetilde{U} \in \mathbb{R}^{n \times n}, \ \ \widetilde{\Sigma} \in \mathbb{R}^{n \times (n-p)}, \ \ \widetilde{V} \in \mathbb{R}^{n \times (n-p)}$$

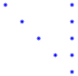
$[\widetilde{V} \ \ W]$: orthonormal system

$$A[\widetilde{V} \ W] = [\widetilde{U}\widetilde{\Sigma} \ \ AW]$$

$$\widetilde{U}^T A[\widetilde{V} \ W] = [\widetilde{\Sigma} \ \ \widetilde{U}^T AW]$$

$$A = \widetilde{U} \ \underbrace{[\widetilde{\Sigma} \ \ \widetilde{U}^T AW]}_{\widetilde{S}} \ [\widetilde{V} \ W]^T$$

spy($\widetilde{S}$) of type :



Thm $\sigma_j \geq \widetilde{\sigma}_j \geq \sigma_{j+p}$

Subspace restricted SVD: left vectors

Similar technique possible "from the left"

Select orthonormal vectors $\quad Z = [\mathbf{z}_1\ \mathbf{z}_2\ \ldots\ \mathbf{z}_q]$

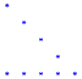
$$(I - ZZ^T)\,A = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^T$$

$\widetilde{U} \in \mathbb{R}^{n\times(n-q)}$, $\widetilde{\Sigma} \in \mathbb{R}^{(n-q)\times n}$, $\widetilde{V} \in \mathbb{R}^{n\times n}$

$[\widetilde{U}\ \ Z]$: orthonormal system

$$[\widetilde{U}\ Z]^T A = \begin{bmatrix} \widetilde{\Sigma}\widetilde{V}^T \\ Z^T A \end{bmatrix}$$

$$A = [\widetilde{U}\ Z]\ \underbrace{\begin{bmatrix} \widetilde{\Sigma} \\ Z^T A \widetilde{V} \end{bmatrix}}_{\widetilde{S}}\ \widetilde{V}^T$$

spy($\widetilde{S}$) of type : 

($\widetilde{S}$ has extra 0-block if $m > n$)

Thm $\quad \sigma_j \geq \widetilde{\sigma}_j \geq \sigma_{j+q}$

Subspace restricted SVD: left and right vectors

Similar technique possible from left and right simultaneously

Select orthonormal vectors

$$W = [\mathbf{w}_1 \; \mathbf{w}_2 \; \ldots \; \mathbf{w}_p], \quad Z = [\mathbf{z}_1 \; \mathbf{z}_2 \; \ldots \; \mathbf{z}_q]$$

$$(I - ZZ^T)A(I - WW^T) = \widetilde{U}\widetilde{\Sigma}\widetilde{V}^T$$

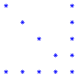
$\widetilde{U} \in \mathbb{R}^{n\times(n-q)}$, $\widetilde{\Sigma} \in \mathbb{R}^{(n-q)\times(n-p)}$, $\widetilde{V} \in \mathbb{R}^{n\times(n-p)}$

$[\widetilde{V} \;\; W]$ and $[\widetilde{U} \;\; Z]$: orthonormal systems

$$[\widetilde{U} \; Z]^T A [\widetilde{V} \; W] = \begin{bmatrix} \widetilde{\Sigma} & \widetilde{U}^T AW \\ Z^T A\widetilde{V} & Z^T AW \end{bmatrix}$$

$$A = [\widetilde{U} \; Z] \underbrace{\begin{bmatrix} \widetilde{\Sigma} & \widetilde{U}^T AW \\ Z^T A\widetilde{V} & Z^T AW \end{bmatrix}}_{\widetilde{S}} [\widetilde{V} \; W]^T$$

spy($\widetilde{S}$) :



Thm $\quad \sigma_j \geq \widetilde{\sigma}_j \geq \sigma_{j+p+q}$

TSRSVD

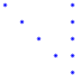
SVD: $A = U\Sigma V^T \approx U_k \Sigma_k V_k^T$

Recall: TSVD solves

$$\min \|U_k \Sigma_k V_k^T \mathbf{x} - \mathbf{b}\| \qquad \mathbf{x} \in \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$$

$$\min \|\Sigma_k \mathbf{y} - U_k^T \mathbf{b}\| \qquad \mathbf{x} = V_k \mathbf{y}$$

SRSVD: $A = \widetilde{U}\; [\widetilde{\Sigma}\;\; \widetilde{U}^T AW]\; [\widetilde{V}\;\; W]^T$

Idea: replace most (small) diagonal elements by 0 in 

TSRSVD = truncated SRSVD: solves

$$\min \|[\widetilde{\Sigma}_{k-p}\;\; \widetilde{U}^T AW]\; [\widetilde{V}_{k-p}\;\; W]^T \mathbf{x} - \widetilde{U}^T \mathbf{b}\|$$

$$\mathbf{x} \in \text{span}(\widetilde{\mathbf{v}}_1, \ldots, \widetilde{\mathbf{v}}_{k-p}, \mathbf{w}_1, \ldots, \mathbf{w}_p)$$

SRSVD + Tikhonov

Recall: standard Tikhonov

$$\begin{aligned} (A^TA + \mu I)\,\mathbf{x} &= A^T\mathbf{b}, \qquad\qquad A = U\Sigma V^T \\ (\Sigma^T\Sigma + \mu I)\,\mathbf{y} &= \Sigma^T U^T\mathbf{b}, \qquad \mathbf{x} = V\mathbf{y} \end{aligned}$$

SRSVD: $A = \widetilde{U}\ \underbrace{[\widetilde{\Sigma}\quad \widetilde{U}^TAW]}_{\widetilde{S}}\ [\widetilde{V}\ \ W]^T$

Tikhonov type SRSVD method:

$$(\widetilde{S}^T\widetilde{S} + \mu I)\,\mathbf{y} = \widetilde{S}^T\widetilde{U}^T\mathbf{b}, \qquad \mathbf{y} = [\widetilde{V}\ \ W]^T\mathbf{x}$$

SRSVD: relations with other SVDs

Thm If $W = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_p]$ (right) singular vectors, then SRSVD = permuted SVD

Thm If $W = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_p]$ (right) singular vectors, then

- TGSVD(A, $I - WW^T$) and
- TSRSVD of A with respect to W

give same solution to $A\mathbf{x} = \mathbf{b}$

So:

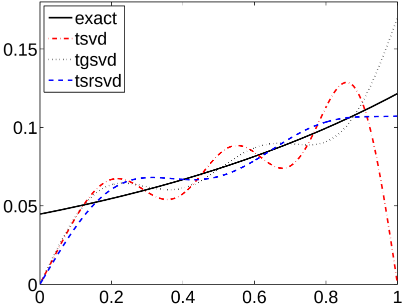
- if W contains singular vectors, then SRSVD(A, W) = TGSVD(A, $I - WW^T$)
- if W contains approximate singular vectors, then SRSVD(A, W) $\approx$ TGSVD(A, $I - WW^T$)

However, SRSVD often seems (somewhat to much) better

Numerical experiment I: deriv2

$n = 500$, 1% noise, discrepancy principle $\|\mathbf{r}\| = 1.1 \cdot 0.01 \cdot \|\mathbf{b}\|$

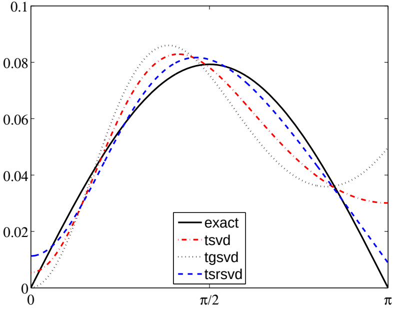
Method \ W	1	t^2
TSVD	$2.94 \cdot 10^{-1}$	
TGSVD(A, $I - WW^T$)	$1.38 \cdot 10^{-1}$	$1.64 \cdot 10^{-1}$
TSRSVD(A, W)	$1.36 \cdot 10^{-1}$	$1.34 \cdot 10^{-1}$



Numerical experiment II: baart

$n = 500$, 1% noise, discrepancy principle $\|\mathbf{r}\| = 1.1 \cdot 0.01 \cdot \|\mathbf{b}\|$

Method \ W	$[1,\ t]$	t^2	$\sin(t)$
TSVD	$1.67 \cdot 10^{-1}$		
TGSVD	$1.74 \cdot 10^{-1}$	$2.73 \cdot 10^{-1}$	$3.43 \cdot 10^{-3}$
TSRSVD	$1.41 \cdot 10^{-1}$	$6.56 \cdot 10^{-2}$	$3.43 \cdot 10^{-3}$



Numerical experiment III: eye

Image eye, $71 \times 71 \Rightarrow \mathbf{b} \in \mathbb{R}^{5041}$

$$A = (2\pi\sigma^2)^{-1} T \otimes T$$

T: 71×71 symmetric banded Toeplitz matrix; first row

```
[exp(-((0:band-1).^2)/(2*sigma^2)); zeros(1,n-band)]
```

Underlying Gaussian point spread function

$$h(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

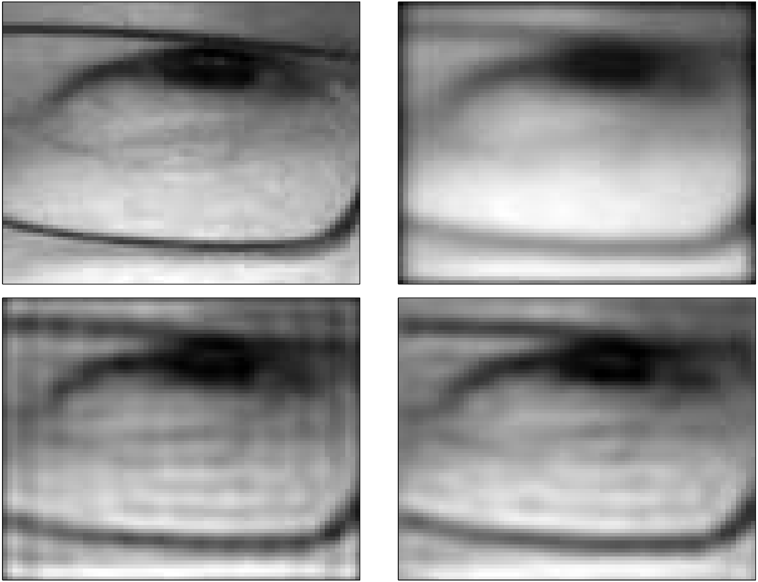
(see, e.g., Hansen, Nagy, O'Leary (2006))

Here `band` $= 16$, $\sigma = 1.5$

	iter	rel. error
TSVD	392	$6.81 \cdot 10^{-2}$
TSRSVD	358	$5.23 \cdot 10^{-2}$

Numerical experiment III: eye

Original, blur+noise, TSVD, TSRSVD "$W = [1\ t\ t^2]$"



Outline main part: linear ill-posed problems

- Introduction (see also Melina's talk)
- Tikhonov and other filter functions
  - new approach: fractional Tikhonov
- general regularization operator L
  - new approach: simple Krylov approach
- including apriori information
  - new approach: subspace restricted SVD

$\Rightarrow$ **combining different approaches**
  - **new approach: linear combination method**
- Conclusions

Linear combination approach

Many competitive approaches:

- TSVD
- TGSVD(A, L) (Hansen 1989, 1990)
- Tikhonov (discrepancy principle)
- Tikhonov (quasi-optimal)
- Tikhonov (L-curve based)
- Modified TSVD (Hansen, Sekii, Shibahashi 1992)
- TPSVD (Morigi, Reichel, Sgallari 2006)
- TSRSVD (this talk)
- TTLS (Van Huffel, Vandewalle 1991)
- (etc, etc) ...

Idea: take "linear combination" to improve

Linear combination approach

Assume have approximate solutions $\mathbf{v}_1, \ldots, \mathbf{v}_k$

Interested in $\mathbf{v} \in \mathcal{V} = \text{span}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$

More precisely:

$$\mathbf{v} = \underset{\mathbf{x} \in \mathcal{V}}{\text{argmin}} \|\mathbf{b} - A\mathbf{x}\| \qquad \text{under} \quad \|\mathbf{v}\| = M$$

Choice of M ?

Several approximations are guaranteed to satisfy $\|\mathbf{v}_j\| < \|\mathbf{x}_{\text{true}}\|$

(and/or $\|\mathbf{v}_j\|/\|\mathbf{x}_{\text{true}}\|$ very close to 1)

Choose: $$M = \max_j \|\mathbf{v}_j\| \qquad \left(\lesssim \|\mathbf{x}_{\text{true}}\|\right)$$

Efficient computation with QR-decomposition of AV (H., Reichel 2011)

Linear combination approach: numerical results

Let $\mathbf{v}_j$ have relative error $\quad q_j = \dfrac{\|\mathbf{x} - \mathbf{x}_{\text{true}}\|}{\|\mathbf{x}_{\text{true}}\|}$

Indicator of success of linear combination approach:

$$\rho = \frac{q - q_{\text{best}}}{q_{\text{worst}} - q_{\text{best}}}$$

Meaning:

- $\rho < 0$: better than best of the $\mathbf{v}_j$
- $\rho = 0$: as good as best of the $\mathbf{v}_j$
- $\rho = 1$: as bad as worst of the $\mathbf{v}_j$
- $\rho > 1$: worse than worst of the $\mathbf{v}_j$

Linear combination approach: numerical results

$n = 100$, **0.1%** noise, $\|\mathbf{b} - A\mathbf{x}\| = \mathbf{1.2}\,\varepsilon$

$$\rho = \frac{q - q_{\text{best}}}{q_{\text{worst}} - q_{\text{best}}}$$

$\rho < 0$ means better than the best

Problem	Tikh (di.pr.)	TSVD	Quasi	Lin.comb.	ρ
`baart`	$1.62 \cdot 10^{-1}$	$1.66 \cdot 10^{-1}$	$1.57 \cdot 10^{-1}$	$1.38 \cdot 10^{-1}$	-2.1
`deriv2-1`	$1.94 \cdot 10^{-1}$	$2.12 \cdot 10^{-1}$	$1.99 \cdot 10^{-1}$	$1.79 \cdot 10^{-1}$	-0.87
`deriv2-2`	$1.84 \cdot 10^{-1}$	$2.08 \cdot 10^{-1}$	$1.90 \cdot 10^{-1}$	$1.68 \cdot 10^{-1}$	-0.64
`deriv2-3`	$1.91 \cdot 10^{-2}$	$2.68 \cdot 10^{-2}$	$1.70 \cdot 10^{-2}$	$1.67 \cdot 10^{-2}$	-0.036
`foxgood`	$2.58 \cdot 10^{-2}$	$3.11 \cdot 10^{-2}$	$2.46 \cdot 10^{-2}$	$1.44 \cdot 10^{-2}$	-1.6
`gravity`	$1.87 \cdot 10^{-2}$	$2.81 \cdot 10^{-2}$	$2.59 \cdot 10^{-2}$	$2.59 \cdot 10^{-2}$	0.76
`heat`	$5.60 \cdot 10^{-2}$	$7.41 \cdot 10^{-2}$	$4.50 \cdot 10^{-2}$	$4.38 \cdot 10^{-2}$	-0.041
`ilaplace`	$1.25 \cdot 10^{-1}$	$1.27 \cdot 10^{-1}$	$1.08 \cdot 10^{-1}$	$1.08 \cdot 10^{-1}$	0
`shaw`	$7.26 \cdot 10^{-2}$	$4.76 \cdot 10^{-2}$	$5.50 \cdot 10^{-2}$	$6.09 \cdot 10^{-2}$	0.53

Linear combination approach: numerical results

$n = 100$, **0.1%** noise, $\|\mathbf{b} - A\mathbf{x}\| = \mathbf{1.5}\,\varepsilon$

$$\rho = \frac{q - q_{\text{best}}}{q_{\text{worst}} - q_{\text{best}}}$$

$\rho < 0$ means better than the best

Problem	Tikh (di.pr.)	TSVD	Quasi	Lin.comb.	ρ
`baart`	$1.66 \cdot 10^{-1}$	$1.66 \cdot 10^{-1}$	$1.57 \cdot 10^{-1}$	$1.38 \cdot 10^{-1}$	-2
`deriv2-1`	$2.10 \cdot 10^{-1}$	$2.29 \cdot 10^{-1}$	$1.99 \cdot 10^{-1}$	$1.84 \cdot 10^{-1}$	-0.49
`deriv2-2`	$2.00 \cdot 10^{-1}$	$2.11 \cdot 10^{-1}$	$1.90 \cdot 10^{-1}$	$1.71 \cdot 10^{-1}$	-0.91
`deriv2-3`	$2.32 \cdot 10^{-2}$	$2.68 \cdot 10^{-2}$	$1.70 \cdot 10^{-2}$	$1.64 \cdot 10^{-2}$	-0.061
`foxgood`	$2.86 \cdot 10^{-2}$	$3.11 \cdot 10^{-2}$	$2.46 \cdot 10^{-2}$	$1.44 \cdot 10^{-2}$	-1.6
`gravity`	$2.37 \cdot 10^{-2}$	$2.81 \cdot 10^{-2}$	$2.59 \cdot 10^{-2}$	$2.59 \cdot 10^{-2}$	0.49
`heat`	$6.35 \cdot 10^{-2}$	$9.62 \cdot 10^{-2}$	$4.50 \cdot 10^{-2}$	$4.50 \cdot 10^{-2}$	0
`ilaplace`	$1.33 \cdot 10^{-1}$	$1.45 \cdot 10^{-1}$	$1.08 \cdot 10^{-1}$	$1.08 \cdot 10^{-1}$	0
`shaw`	$9.45 \cdot 10^{-2}$	$1.10 \cdot 10^{-1}$	$5.50 \cdot 10^{-2}$	$5.50 \cdot 10^{-2}$	0

Linear combination approach: numerical results

$n = 100$, **0.1%** noise, $\|\mathbf{b} - A\mathbf{x}\| = \mathbf{2}\,\varepsilon$

$$\rho = \frac{q - q_{\text{best}}}{q_{\text{worst}} - q_{\text{best}}}$$

$\rho < 0$ means better than the best

Problem	Tikh (di.pr.)	TSVD	Quasi	Lin.comb.	ρ
baart	$1.72 \cdot 10^{-1}$	$1.66 \cdot 10^{-1}$	$1.57 \cdot 10^{-1}$	$1.39 \cdot 10^{-1}$	-1.2
deriv2-1	$2.27 \cdot 10^{-1}$	$2.52 \cdot 10^{-1}$	$1.99 \cdot 10^{-1}$	$1.97 \cdot 10^{-1}$	-0.041
deriv2-2	$2.18 \cdot 10^{-1}$	$2.32 \cdot 10^{-1}$	$1.90 \cdot 10^{-1}$	$1.83 \cdot 10^{-1}$	-0.18
deriv2-3	$2.81 \cdot 10^{-2}$	$4.79 \cdot 10^{-2}$	$1.70 \cdot 10^{-2}$	$1.70 \cdot 10^{-2}$	0
foxgood	$3.12 \cdot 10^{-2}$	$3.11 \cdot 10^{-2}$	$2.46 \cdot 10^{-2}$	$1.43 \cdot 10^{-2}$	-1.5
gravity	$2.82 \cdot 10^{-2}$	$4.00 \cdot 10^{-2}$	$2.59 \cdot 10^{-2}$	$2.59 \cdot 10^{-2}$	0
heat	$7.36 \cdot 10^{-2}$	$9.62 \cdot 10^{-2}$	$4.50 \cdot 10^{-2}$	$4.50 \cdot 10^{-2}$	0
ilaplace	$1.39 \cdot 10^{-1}$	$1.45 \cdot 10^{-1}$	$1.08 \cdot 10^{-1}$	$1.08 \cdot 10^{-1}$	0
shaw	$1.18 \cdot 10^{-1}$	$1.10 \cdot 10^{-1}$	$5.50 \cdot 10^{-2}$	$5.50 \cdot 10^{-2}$	0

Work in progress: noise estimation

- (Hnetynkova, Plesinger, Strakos 2009 + in progress) (**THU 18:40!**)
- (H., Reichel in progress): method "X"

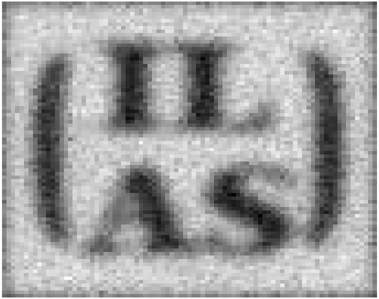
New method "X": noise estimate qualities $\frac{\text{estimated noise level}}{\text{true noise level}}$

for $n = 100$ examples for different noise levels:

Problem \ noise	0.1%	1%	5%	10%
baart	1.03	0.98	1.01	1.00
deriv2-1	0.991	0.962	0.971	1.14
deriv2-2	0.962	1.05	1.36	1.05
deriv2-3	0.939	1.00	1.06	1.00
foxgood	0.975	0.985	0.997	0.974
gravity	0.984	0.99	0.985	0.931
heat	0.891	0.89	0.932	0.934
ilaplace	0.984	1.03	1.01	0.985
phillips	0.997	0.925	0.972	0.984
shaw	0.966	0.995	0.999	0.959

Surprisingly good estimates; theoretically not yet fully understood . . .

Finally, to impress you, if we combine ALL techniques . . .



Conclusions

3 advertisements:

- inclusion regions based on the field of values
- when to stop inner iterations in Jacobi–Davidson
- probabilistic eigenvalue bounds (for $\lambda_{\max}$, $\|A\|$, $\Lambda(A)$, $W(A)$)

Linear ill-posed problems:

- fractional Tikhonov to find "longer, less smooth" solutions
- simple Krylov approach for general L
- subspace-restricted SVD to include apriori information W
- possible to combine different solutions