Regularization of large discrete inverse problems by iterative projection methods

I. Hnětynková

Faculty of Mathematics and Physics, Charles University in Prague

SNA 21 - January 2021

Outline

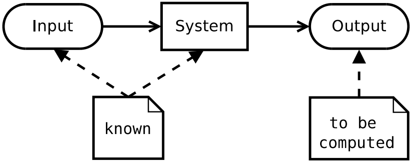
1. Inverse problems
2. Regularization by projection
3. Propagation of noise
4. Analysis of residuals
5. Hybrid methods
6. Conclusion

Outline

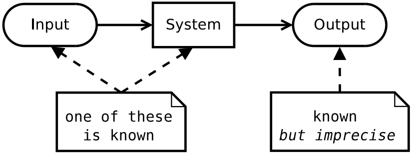
1. Inverse problems
2. Regularization by projection
3. Propagation of noise
4. Analysis of residuals
5. Hybrid methods
6. Conclusion

Basic illustration

Forward Problem



Inverse Problem



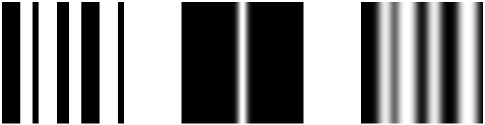
Fredholm integral equation

Given the continuous smooth kernel $K(s,t)$ and the (measured) data $g(s)$, the aim is to find the (source) function $f(t)$ such that

$$g(s) = \int_I K(s,t) f(t) \mathrm{d}t + e(s).$$

Fredholm integral has smoothing property, i.e. high frequency components in g are dampened compared to f.

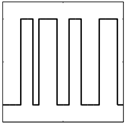
1D example: Barcode reading



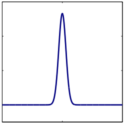
sharp barcode f(t) Gaussian blur measured data g(s)

Example: Fredholm integral equation - discretization

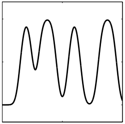
1D example: Barcode reading



sharp barcode f(t)



Gaussian blur



measured data g(s)

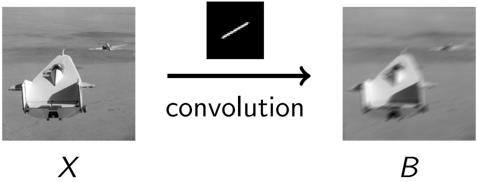
$$g(s) = \int_I K(s,t) f(t) \mathrm{d}t + e(s).$$

Using numerical quadrature formulas, we get a linearized model

$$b = Ax + e, \quad \text{with} \quad A \in \mathbb{R}^{N \times M}, \quad b, e \in \mathbb{R}^N, \quad x \in \mathbb{R}^M,$$

where A has the smoothing property.

2D Example: image deblurring problem



The data $B \in \mathbb{R}^{m \times n}$ are naturally discrete. Using the vectorization $x = vec(X), b = vec(B)$, we obtain a deconvolution problem

$$b = Ax + e, \quad \text{with} \quad A \in \mathbb{R}^{N \times N}, \quad N = mn.$$

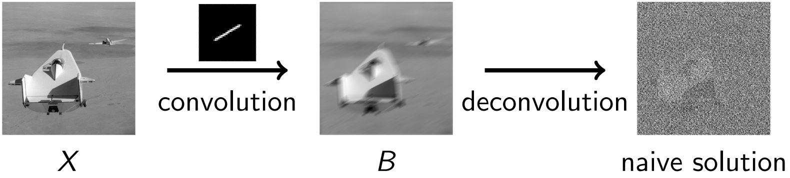
The model matrix is typically large, sparse and structured.

Naive solution

If A is square nonsingular, a naive approach is to solve directly

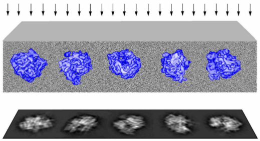
$$A\,x^{\mathrm{naive}} = b.$$

2D Example: image deblurring



3D Example: Electron microscopy

$$PSF_\omega * (P_\omega f + e^s_\omega) + e^b_\omega = g_\omega$$



- f: Unknown function representing the particle
- ω: Projection angle.
- PSF_ω: Point Spread Function.
- P_ω: X-Ray transform: $P_\omega f(s) := \int_{-\infty}^{\infty} f(t \cdot \omega + s)\mathrm{d}t, \quad s \in \omega^\perp$.
- e^s_ω, e^b_ω: Structure and background noise functions.
- g_ω: Measured data.
- $*$: Convolution operator.

Discrete model (one projection)

$$PSF_\omega * (P_\omega {\color{red}f} + e^s_\omega) + e^b_\omega = g_\omega \quad \text{Continuous model}$$

$$C_\omega \left(\bar{P}_\omega {\color{red}\bar{f}} + \bar{e}^s_\omega\right) + \bar{e}^b_\omega = \bar{g}_\omega \quad \text{Discrete model}$$

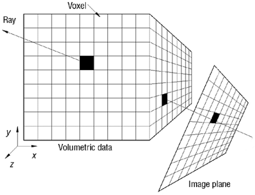


Figure: 3D grid discretization with unknown voxel values.

Linear model

Consider a linear ill-posed problem

$$b = Ax + e,$$

where the noise vector e

- is an unknown perturbation (rounding errors, errors of measurement, noise with physical sources, etc.),
- with the unknown noise level

$$\delta^{\text{noise}} \equiv \| e \| / \| b \| << 1$$

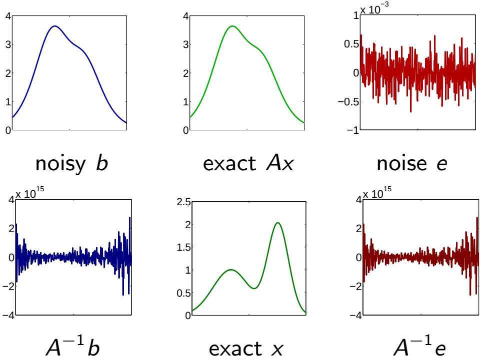
Properties of the problem:

- A dampens high frequencies (smoothing property),
- exact right-hand side is smooth, but noise is not,
- small changes in b cause large changes in the solution.

Naive solution - noise amplification

$$b = Ax + e, \quad \text{where} \quad \|Ax\| \gg \|e\| \qquad \textbf{BUT}$$
$$A^{-1}b = x + A^{-1}e, \quad \text{where} \quad \|x\| \ll \|A^{-1}e\|$$

1D Example: `shaw(400)`, $\delta^{\text{noise}} \approx 1e-4$, white noise



Naive solution - noise amplification

Singular value decomposition (SVD): $R = rank(A)$

$$A = U\Sigma V^T = \sum_{j=1}^{R} u_j^T \sigma_j v_j,$$

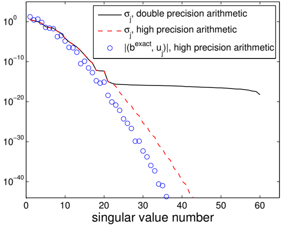
$$\Sigma = \text{diag}\{\sigma_1, \dots, \sigma_R, 0, \dots, 0\},$$

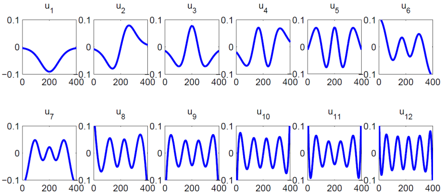
where $U = [u_1, \dots, u_N]$ and $V = [v_1, \dots, v_M]$ are unitary matrices. Then

$$x^{\text{naive}} \equiv \quad A^\dagger b = \underbrace{\sum\nolimits_{j=1}^{R} \frac{u_j^T b^{\text{exact}}}{\sigma_j} v_j}_{x^{\text{exact}}} + \underbrace{\sum\nolimits_{j=1}^{R} \frac{u_j^T e}{\sigma_j} v_j}_{\text{noise component}}.$$

Discrete Picard condition (DPC)

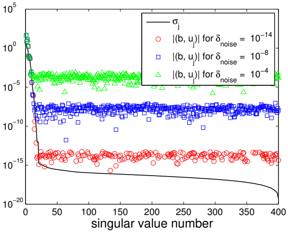
- singular values of A decay quickly without a noticeable gap;
- singular vectors u_i, v_j of A represent increasing frequencies;
- for the exact right-hand side, $|(b^{\mathrm{exact}}, u_j)|$ decay faster than the singular values σ_j of A (**DPC**)





Noise amplification

White noise: $|(e, u_j)|,\ j = 1, \ldots, N$ do not exhibit any trend



$$x^{\text{naive}} \equiv \quad A^{\dagger} b = \underbrace{\sum_{j=1}^{R} \frac{u_j^T b^{\text{exact}}}{\sigma_j} v_j}_{x^{\text{exact}}} + \underbrace{\sum_{j=1}^{R} \frac{u_j^T e}{\sigma_j} v_j}_{\text{amplified noise}}$$

Components corresponding to small σ_j's are dominated by e^{HF}.

2D imaging problem

For a blurred image B

$$x^{\text{naive}} = \sum_{j=1}^{R} \underbrace{\frac{u_j^T \operatorname{vec}(B)}{\sigma_j}}_{\text{scalar}} v_j, \qquad X = \operatorname{mtx}(x),$$

is a linear combination of right singular vectors v_j.

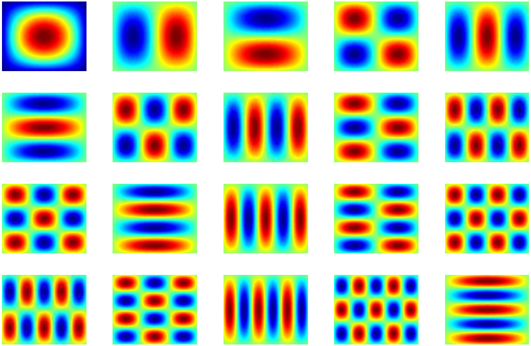
It can be further rewritten as

$$X^{\text{naive}} = \sum_{j=1}^{R} \frac{u_j^T \operatorname{vec}(B)}{\sigma_j} V_j, \qquad V_j = \operatorname{mtx}(v_j) \in \mathbb{R}^{m \times n}$$

using singular images V_j (the reshaped right singular vectors).

2D imaging problem: Singular images

Singular images $V_j \in \mathbb{R}^{m \times n}$ for 2D image deblurring model (Gaußian blur, zero BC, artificial colors).



Filtered solution

Unwanted components can be suppressed by

$$x^{\mathrm{filtered}} = \sum_{j=1}^{R} \phi_j \, \frac{u_j^T \, b}{\sigma_j} \, v_j, \qquad x^{\mathrm{filtered}} = V \Phi \Sigma^{-1} U^T b,$$

where $\Phi = \mathrm{diag}(\phi_1, \ldots, \phi_N)$. In image deblurring problem

$$X^{\mathrm{filtered}} = \sum_{j=1}^{R} \phi_j \, \frac{u_j^T \, \mathrm{vec}(B)}{\sigma_j} \, V_j.$$

The filter factors are given by some filter function

$$\phi_j = \phi(j, A, b, \ldots).$$

Classical regularization approaches

Spectral filtering (e.g., truncated SVD, Tikhonov): suitable for solving small ill-posed problems.

Projection on smooth subspaces: suitable for solving large ill-posed problems. The dimension of projection space represents a regularization parameter.

Hybrid techniques: combination of outer iterative regularization with a spectral filtering of the projected small problem.

... etc.

Large scale problems

- Direct filtering of SVD is too costly.
- The method should avoid work with full A.
- The method should take advantage of data properties (sparsity, structure, ...).
- The approximation must be dominated by low frequencies, high frequencies must be dumped.

We try to look for an approximation in some low dimensional subspace $\mathcal{W}_k$ dominated by low frequencies.

Outline

1. Inverse problems
2. Regularization by projection
3. Propagation of noise
4. Analysis of residuals
5. Hybrid methods
6. Conclusion

Projection methods

Consider a subspace

$$\mathcal{W}_k = \mathrm{span}(w_1, \ldots, w_k) \subset \mathbb{R}^N, \qquad W_k = [w_1, \ldots, w_k] \in \mathbb{R}^{N \times k},$$

such that $W_k^T W_k = I_k$ and w_j are dominated by low frequencies.

Then we solve the projected problem

$$\min_{x \in \mathcal{W}_k} \|b - Ax\| \quad \Leftrightarrow \quad \min_{y \in \mathbb{R}^k} \|b - (AW_k)y\|$$

$$\Leftrightarrow \quad W_k^T (A^T A) W_k y = W_k^T A^T b.$$

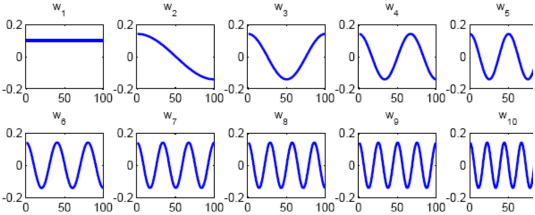
The question is, how to choose the basis w_j?

Projection using DCT basis

An example of a suitable basis is the DCT basis

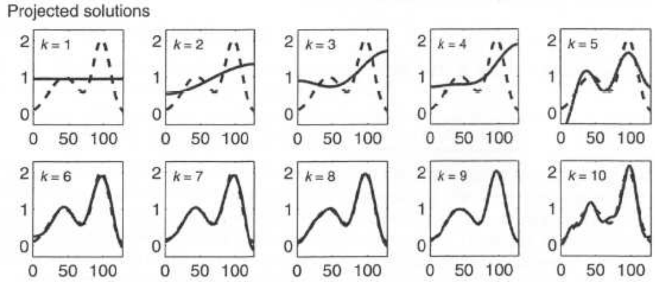
$$w_1 = \sqrt{\tfrac{1}{N}}\,(1, 1, \ldots, 1)^T,$$
$$w_j = \sqrt{\tfrac{2}{N}}\left(\cos\left(\tfrac{(j-1)\pi}{2N}\right), \cos\left(\tfrac{3(j-1)\pi}{2N}\right), \ldots \cos\left(\tfrac{(2N-1)(j-1)\pi}{2N}\right)\right)^T,$$

for $j > 1$.



Projection using DCT basis

Example: Solutions computed using the DCT basis $w_1, \ldots, w_k$, $k = 1, \ldots, 10$



A-priori known properties of the true solution (symmetry, periodicity, etc.) can be imposed by well-chosen basis.

Projection using DCT basis

Advantage:

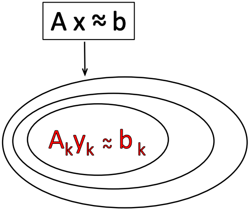
With a fixed set of basis Fourier-type vectors, computations can be performed efficiently, the basis is not stored.

Disadvantage:

The basis vectors are not always adapted to the particular problem.

Krylov subspace methods

When A is large/sparse/not given explicitly, approximation by projection onto a low dimensional Krylov subspace is advantageous.



$$\mathcal{K}_k(C, d) \equiv Span\{d,\ Cd,\ \dots,\ C^{k-1}d\}$$

$$\mathcal{K}_1(C, d) \subseteq \mathcal{K}_2(C, d) \subseteq \ \dots$$

For A square: $\mathcal{K}_k(A, b)$ ... GMRES, CG, MINRES

$\vec{\mathcal{K}}_k(A, b)$ ... RRGMRES, MINRES-II

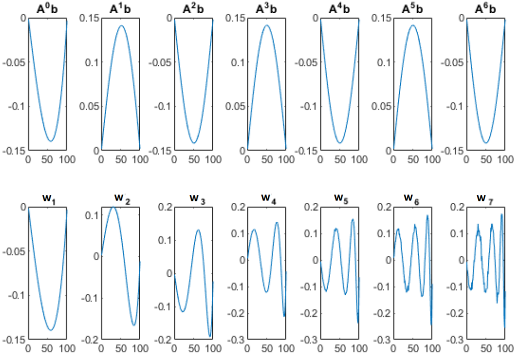
For A general: $\mathcal{K}_k(A^T A, A^T b)$ ... LSQR, LSMR, CGLS

$$x_k \longrightarrow x^{\text{naive}}$$

Key role of orthonormal basis

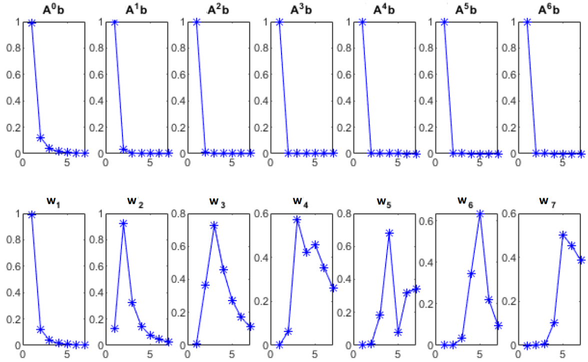
Generating Krylov vectors are smooth. In order to approximate less smooth features, it is necessary to use orthonormal basis.

Example: Generating vectors and orthonormal basis vectors w_i (computed by Arnoldi process) for $\mathcal{K}_5(A, b)$



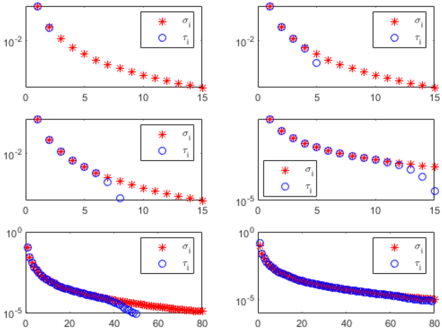
Key role of orthonormal basis

Example: Generating vectors and orthonormal basis vectors w_i in frequency basis U (left singular vectors of A)



Inheritance of DPC

Example: Singular values σ_i of A and singular values τ_i of H_k from the Arnoldi process for $k = 2, 5, 8, 5, 50, 80$



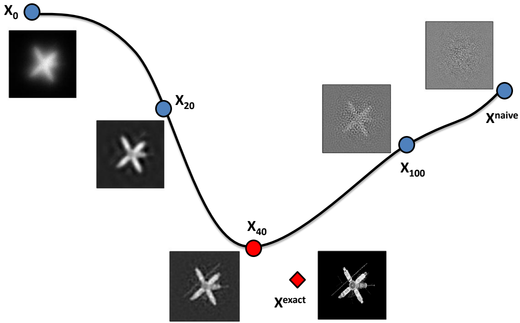
The projected problem $A_k y_k \approx b_k$ then subsequently inherits DPC properties of the original problem.

Semiconvergence of Krylov subspace methods

With growing k:

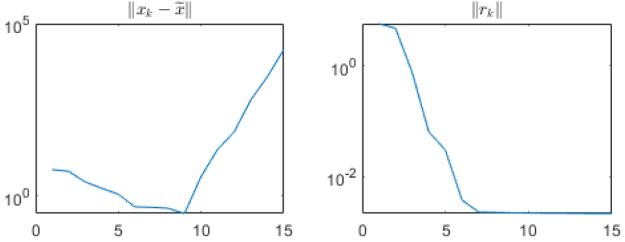
- we include HF features to the solution,
- noise e propagates to the projection.

small k = over-smoothed solution **large** k = noisy solution



Semiconvergence of Krylov subspace methods

Example: True errors and residual norms of LSQR approximations x_k for the problem `shaw(400)` contaminated by white noise e



Number of iterations = regularization parameter

Stopping criteria

Since $b - Ax^{exact} = e$, a reasonable requirement could be

$$r_k \equiv b - Ax_k \quad \approx \quad e.$$

Stopping criteria: this idea can be used if a priori information is available, e.g., $\|e\|$ in DP, spectral properties of e (white) in NCP.

However, e is often not known.

Understanding noise propagation:

- consider $\mathcal{K}_k(A^T A, A^T b)$ for a general A,
- study how e propagates to the projections,
- study the relation between e and $r_1, r_2, \ldots$.

Outline

Golub-Kahan iterative bidiagonalization (GK)

Given $w_0 = 0$, $s_1 = b / \beta_1$, $\beta_1 = \|b\|$, for $j = 1, 2, \ldots$

$$\begin{aligned} \alpha_j w_j &= A^T s_j - \beta_j w_{j-1}, \qquad \|w_j\| = 1, \\ \beta_{j+1} s_{j+1} &= A w_j - \alpha_j s_j, \qquad \|s_{j+1}\| = 1. \end{aligned}$$

Output:

- $S_k = [s_1, \ldots, s_k]$ - orthonormal bases of $\mathcal{K}(AA^T, b)$,
- $W_k = [w_1, \ldots, w_k]$ - orthonormal bases of $\mathcal{K}(A^T A, A^T b)$,
- bidiagonal matrices of the normalization coefficients

$$L_k = \begin{bmatrix} \alpha_1 & & & \\ \beta_2 & \alpha_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \alpha_k \end{bmatrix}, \quad L_{k+} = \begin{bmatrix} L_k \\ e_k^T \beta_{k+1} \end{bmatrix}.$$

Regularization based on GK

$x_k = W_k y_k$, where the columns of W_k span $\mathcal{K}_k(A^T A, A^T b)$

LSQR method: minimize the residual

$$\min_{x \in \mathcal{K}_k(A^T A, A^T b)} \|Ax - b\| = \min_{y \in \mathbb{R}^k} \|L_{k+} y - \beta_1 e_1\|$$

CRAIG method: minimize the error

$$\min_{x \in \mathcal{K}_k(A^T A, A^T b)} \|x^* - x\| = \min_{y \in \mathbb{R}^k} \|L_k y - \beta_1 e_1\|$$

LSMR method: minimize $A^T r_k$

$$\min_{x \in \mathcal{K}_k(A^T A, A^T b)} \|A^T (Ax - b)\| = \min_{y \in \mathbb{R}^k} \|L_{k+1}^T L_{k+} y - \beta_1 \alpha_1 e_1\|$$

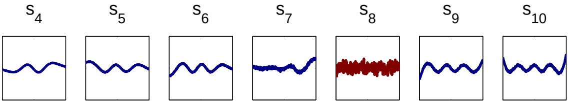
Noise propagation in GK

Recall that we are interested in the relation between

$$\tilde{r} \equiv b - A\tilde{x} \qquad \longleftrightarrow \qquad e.$$

Since $x_k = W_k y_k \in \mathcal{K}_k(A^T A, A^T b)$, then

$$r_k \equiv b - AW_k y_k = \beta_1 s_1 - S_{k+1} L_{k+} y_k = S_{k+1} p_k \in \mathcal{K}_k(AA^T, b).$$



Analyzed in [H., Plešinger, Strakoš - 09], [H., Plešinger, Kubínová - 17].

Exact and noise component in s_k

- $s_1 = b/||b|| = Ax/||b|| + e/||b||$
- for $k = 2, 3, \ldots$

$$\beta_{k+1} s_{k+1} = A w_k - \alpha_k s_k$$

Thus

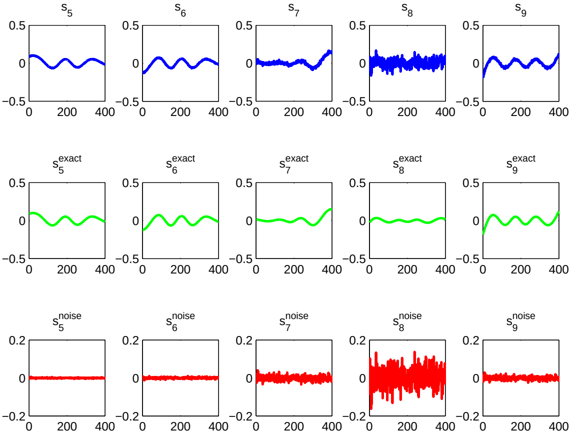
$$s_k = \ (\cdot) + \gamma_k e^{HF}, \quad \text{where} \quad \gamma_k \equiv \varphi_{k-1}(0) = (-1)^{k-1} \frac{1}{\beta_k} \prod_{j=1}^{k-1} \frac{\alpha_j}{\beta_j}.$$

Here $(\cdot)$ is smooth and the amplification factor γ_k of e^{HF} is the absolute term of the Lanczos polynomial,

$$s_{k+1} = \varphi_k(AA^T)b, \qquad \varphi_k \in \mathcal{P}_k.$$

Exact and noise component in s_k

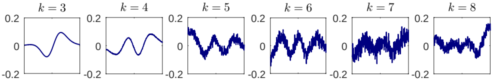
$$s_k = s_k^{exact} + s_k^{noise}$$



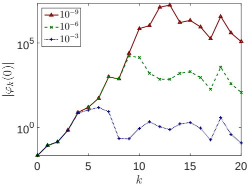
Noise propagation in GK - behavior

The size of γ_k (on average) rapidly grows until it reaches the noise revealing iteration k_{rev}. Then it decreases.

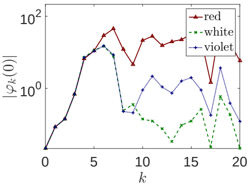
Example: `shaw(400)`, reortogonalization in GK



s_k, $\delta_{\text{noise}} = 10^{-3}$



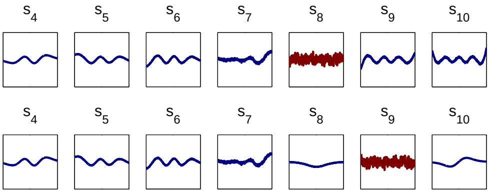
white noise

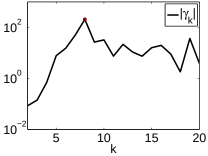


colored noise

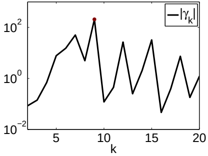
Influence of the loss of orthogonality

Comparison GK with and without reorthogonalization:





with ReOG

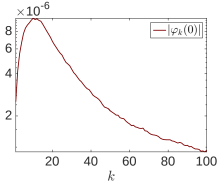


without ReOG

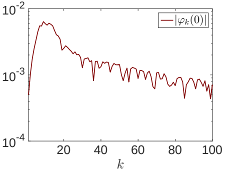
Aggregation may be necessary [Gergelits, H., Kubínová - 18].

Noise propagation in GK - large 2D problems

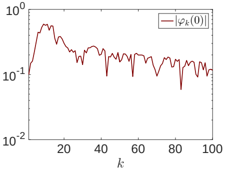
Example: $\delta_{\text{noise}} \approx 10^{-2}$, various physical noise, without ReOG



vargaussianblur, $N = 99856$



seismictomo, $N = 20000$

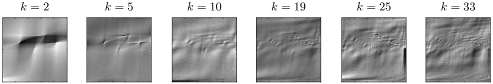


paralleltomo, $N = 65160$

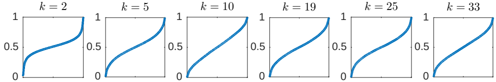
There is no particular noise revealing iteration k, but rather a noise revealing phase represented by a group of subsequent iterations k, see [H., Plešinger, Kubínová - 17].

Noise propagation in GK - large 2D problems

Example: `seismictomo`, $\delta_{\text{noise}} \approx 10^{-2}$, without ReOG



reshaped s_k



cumulative periodograms of s_k

Cumulative periodogram (examining distribution of frequencies) of s_{10} is flatter, thus s_{10} belong to the noise revealing phase.

Application in regularization process

- Stopping criterion - before noise propagates seriously to s_k.
- If k_{rev} can be identified, we can estimate the high frequency part of e:

$$s_{k_{\text{rev}}} \equiv (\cdot) + \gamma_{k_{\text{rev}}} e^{HF} \approx \gamma_{k_{\text{rev}}} e^{HF}$$

gives the estimate by scaled left bidiagonalization vector

$$\tilde{e} \equiv \gamma_{k_{\text{rev}}}^{-1} s_{k_{\text{rev}}}.$$

- We can obtain a cheap estimate of the unknown noise level $\| e \| / \| b \|$, see [H., Kubínová, Plešinger - 16] for application in image deblurring.

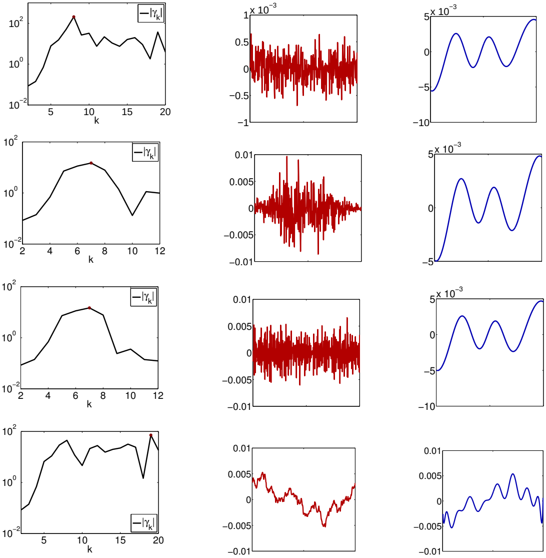
Noise estimate for shaw(400)

White:

Data correlated:

High-frequency:

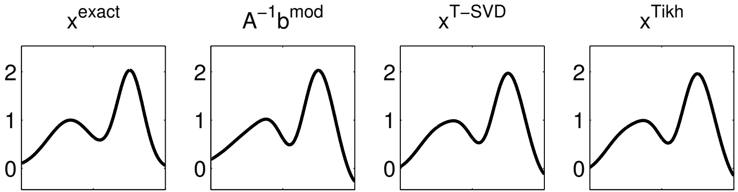
Low-frequency:



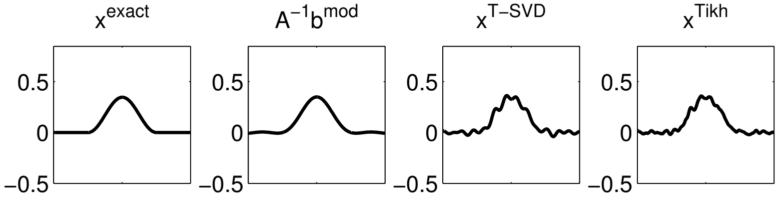
e $\qquad$ $e - \tilde{e}$

Comparison of noise reduction to spectral filtering

shaw(400), white noise



phillips(400), white noise



Outline

1. Inverse problems
2. Regularization by projection
3. Propagation of noise
4. Analysis of residuals
5. Hybrid methods
6. Conclusion

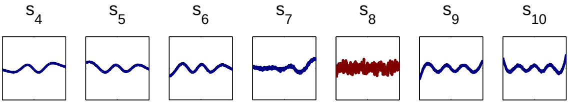
Regularization based on GK

Recall that we are interested in the relation between

$$\tilde{r} \equiv b - A\tilde{x} \qquad \longleftrightarrow \qquad e.$$

For GK based methods with $x_k = W_k y_k \in \mathcal{K}_k(A^T A, A^T b)$, we have

$$r_k = S_{k+1} p_k.$$



Based on noise propagation in S_k, we can analyze CRAIG, LSQR, LSMR by studing p_k, see [H., Kubínová, Plešinger - 17].

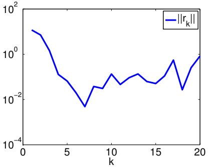
Residual of CRAIG method

$$\min_{x \in \mathcal{K}_k(A^TA, A^Tb)} \|x^* - x\| = \min_{y \in \mathbb{R}^k} \|L_k y - \beta_1 e_1\|, \quad x_k = W_k y_k$$

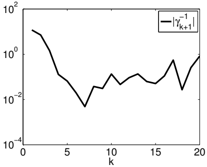
Theorem: x_k^{CRAIG} is the exact solution to the consistent system

$$A x_k^{\mathrm{CRAIG}} = b - \varphi_k(0)^{-1} s_{k+1}.$$

Consequently, $\|r_k^{\mathrm{CRAIG}}\| = |\varphi_k(0)^{-1}| \equiv |\gamma_{k+1}|^{-1}$ reaches its minimum in the noise revealing iteration.



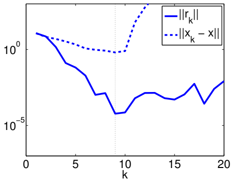
CRAIG residual



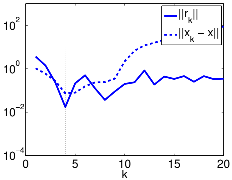
inverted ampl. factor

Comparison of the error and the residual

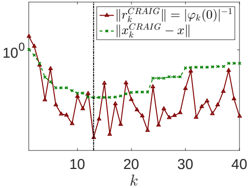
Measuring the size of the residual seems to be a valid stopping criterion for CRAIG. The minimal error is reached approximately at the iteration with the minimal residual.



shaw(400)



phillips(1000)



phillips, no ReOG

Residual of LSQR method

$$\min_{x \in \mathcal{K}_k(A^T A, A^T b)} \|Ax - b\| = \min_{y \in \mathbb{R}^k} \|L_{k+} y - \beta_1 e_1\|, \quad x_k = W_k y_k$$

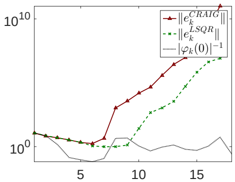
Theorem: The residual corresponding to x_k^{LSQR} satisfies

$$r_k^{\text{LSQR}} = \frac{1}{\sum_{l=0}^{k} \varphi_l(0)^2} \sum_{l=0}^{k} \varphi_l(0) s_{l+1}.$$

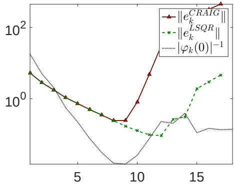
Consequently, the size of the component of r_k in the direction of s_j is proportional to the amount of propagated noise e^{HF} in s_j.

Comparison of CRAIG and LSQR

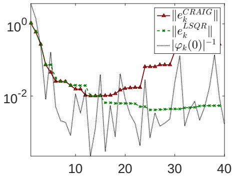
Typically, LSQR can reach better approximation than CRAIG.



`shaw(400)`, white



`gravity(400)`, Poisson

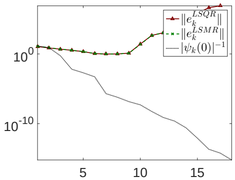


`phillips(400)`, white, no ReOG

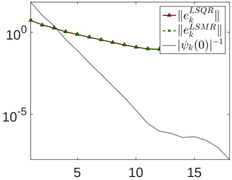
Residual of LSMR method

$$\min_{x \in \mathcal{K}_k(A^TA, A^Tb)} \|A^T(Ax - b)\| = \min_{y \in \mathbb{R}^k} \|L_{k+1}^T L_{k+} y - \beta_1 \alpha_1 e_1\|$$

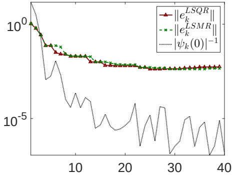
Components of r_k in LSMR behave similarly as in LSQR. The errors resemble as long as $|\psi_k(0)|$ (the absolute term of the Lanczos polynomial for GK vectors w_k) grows rapidly.



`shaw(400)`, white

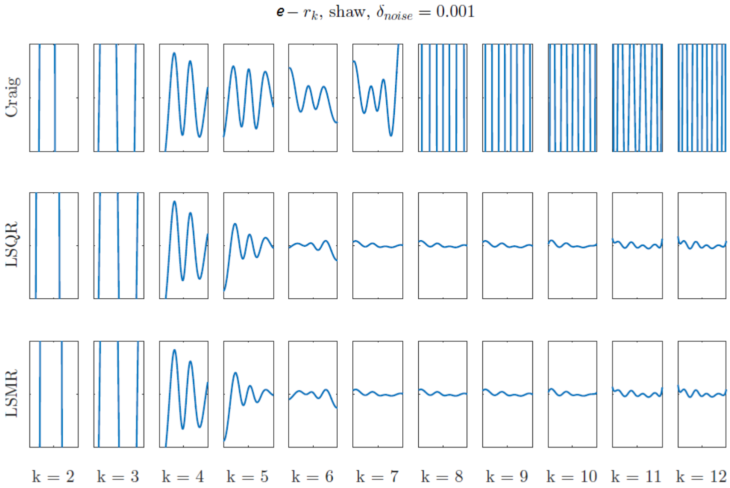


`gravity(400)`, Poisson



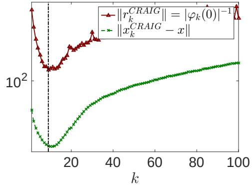
`phillips(400)`, white, no ReOG

Comparison of noise and residuals

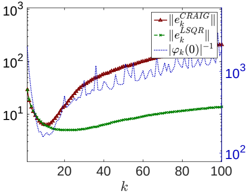


Comparison of the methods - large 2D problems

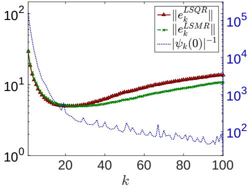
Example: `seismictomo(100,100,200)`, white noise, $\delta_{\text{noise}} = 0.01$, $A \in \mathbb{R}^{20000 \times 10000}$, no ReOG



CRAIG error and residual



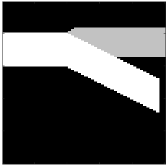
CRAIG vs LSQR



LSQR vs LSMR

Comparison of reconstructions

Reconstructions for `seismictomo(100,100,200)`. Iteration is selected as $k = \operatorname{argmax}_{k=1,2,\dots} |\varphi_k(0)|$.



exact solution

noisy projections

LSQR

LSMR

Outline

1. Inverse problems
2. Regularization by projection
3. Propagation of noise
4. Analysis of residuals
5. Hybrid methods
6. Conclusion

Basic idea

Two stage inner (Krylov projection) - outer (direct) regularization.

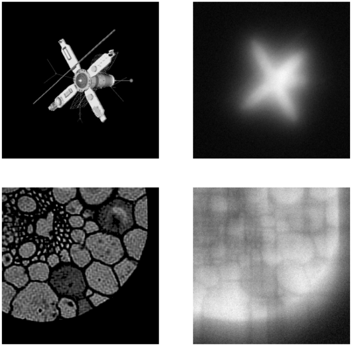
Algorithm: Hybrid LSQR

- Golub-Kahan iterative bidiagonalization
  - $L_{k+} y_k \approx \beta_1 e_1$
- Tikhonov regularization of the projected problem
  - $y_k^\lambda = \arg\min_y \{\|L_{k+} y - \beta_1 e_1\|_2^2 + \lambda^2 \|y\|_2^2\}$
  - Parameter selection approach.
- Back projection $x_k^\lambda = W_k y_k^\lambda$
- Stopping criterion.

See [Calvetti, Reichel - 03], [Chung, Nagy, O'Leary - 08], [Kilmer, Hansen, Español - 07], [Renaut, H., Mead - 10], [Chung, Palmer - 15], ...

2D image deblurring

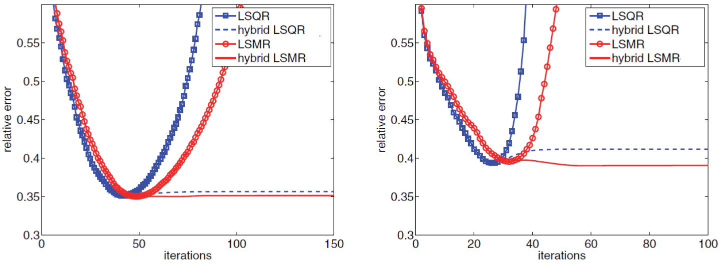
Examples: Satelite and grain test image, Gaussian blur, white noise with $\delta_{\text{noise}} = 0.05$.



True Image Blurred & 5% noise

2D image deblurring

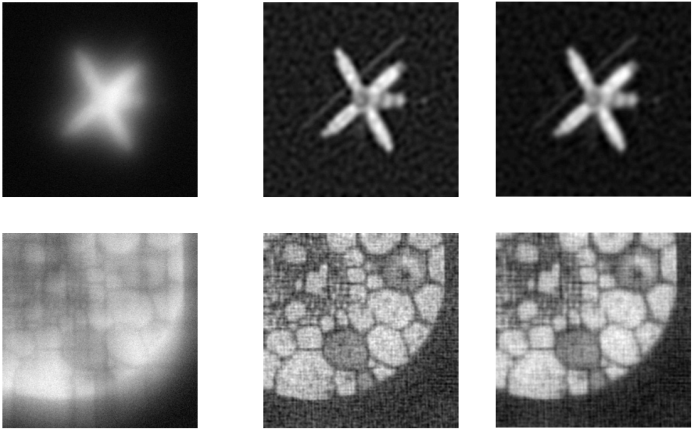
Example: LSQR and LSMR with inner Tikhonov regularization



Satelite image **Grain image**

- overcomes the semiconvergence phenomenon,
- two regularization parameters (outer - number of iterations, inner - direct regularizer) must be tuned.

2D image deblurring - reconstructions



Blurred & 5% noise | LSMR | hybrid LSMR

Outline

Conclusion

- Iterative projective regularization is a powerfull tool to solve large problems.
- Noise propagates subsequentially to the projections, early stopping is necessary.
- Combinations of projection and direct regularization is advantageous.
- Constraints (e.g. nonnegativity or sparsity of the solution) can be incorporated.

Selected references

Software:

- S. Gazzola, P. C. Hansen, and J. G. Nagy: IR Tools Version 1.0, 2019.
- P. C. Hansen, and J. S. Jrgensen: AIR Tools II Version 1.0, 2018.
- P. C. Hansen: Regularization Tools Version 4.0, 2007.

Overview books:

- R C. Gonzalez, R. E. Woods: Digital Image Processing, Pearson, 4th Edition 2018.
- M. Hanke: A Taste of Inverse Problems: Basic Theory and Examples, SIAM, 2017.
- P. C. Hansen: Discrete Inverse Problems: Insight and Algorithms, SIAM, 2010.
- P. C. Hansen, J. G. Nagy, and D. P. O'Leary: Deblurring Images: Matrices, Spectra, and Filtering, SIAM, 2006.

Thank you for your attention!