Flexible Krylov methods for ℓ_p-regularization

Julianne Chung[1] and Silvia Gazzola[2]

[1] Department of Mathematics, Virginia Tech
Computational Modeling and Data Analytics Division, Academy of Integrated Science
[2] Department of Mathematical Sciences, University of Bath

November 3, 2018 @ ICMDS

DMS-1654175 (CAREER Award)
DMS-1723005 (CDS&E-MSS)

Overview

Introduction and motivation

Background on iteratively re-weighted and flexible methods

Flexible Golub-Kahan based methods

Numerical results and conclusions

Outline

Introduction and motivation

Background on iteratively re-weighted and flexible methods

Flexible Golub-Kahan based methods

Numerical results and conclusions

Linear inverse problem

$$\mathbf{b} = \mathbf{A}\mathbf{x}_{\text{true}} + \boldsymbol{\epsilon}$$

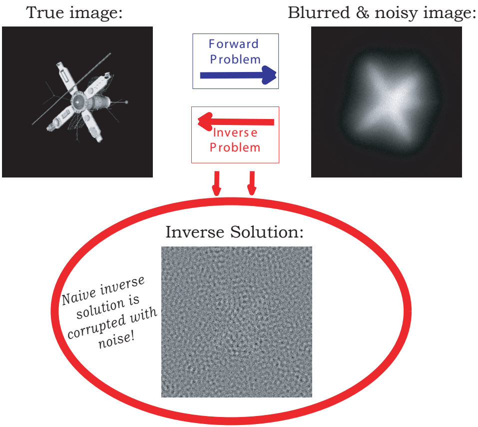
where

$\mathbf{b} \in \mathbb{R}^m$	:	observations or measurements
$\mathbf{x}_{\text{true}} \in \mathbb{R}^n$	:	desired parameters
$\mathbf{A} \in \mathbb{R}^{m \times n}$	:	ill-conditioned matrix models forward process
$\boldsymbol{\epsilon} \in \mathbb{R}^m$	:	additive Gaussian noise

- Goal: Given $\mathbf{b}$ and $\mathbf{A}$, compute approximation of $\mathbf{x}_{\text{true}}$

An example: image deblurring

- Given observed image, $\mathbf{b}$, and knowledge about the blur, $\mathbf{A}$
- Compute approximation of true image, $\mathbf{x}_{\text{true}}$



Hansen, Nagy, and O'Leary, *Deblurring Images: Matrices, Spectra and Filtering*, SIAM (2006), Hadamard (1902), Tikhonov and Arsenin (1977)

ℓ_p-regularization

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{\Psi}\mathbf{x}\|_p^p$$

where $\lambda > 0$ is a regularization parameter, $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$ invertible, $p \geq 1$

Tikhonov regularization ($p = 2$):

- Spectral filtering
- Bayesian maximum a posteriori (MAP) estimate:

$$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}\,,\, \mathbf{I}) \qquad \mathbf{x}_{\text{true}} \sim \mathcal{N}(\mathbf{0}, (\mathbf{\Psi}^\top \mathbf{\Psi})^{-\mathbf{1}})$$

ℓ_1–regularization ($p = 1$):

- Bayesian MAP estimate where the prior on $\mathbf{\Psi}\mathbf{x}$ is Laplacian
- Sparse reconstruction: $||\mathbf{\Psi}\mathbf{x}||_1 \approx ||\mathbf{\Psi}\mathbf{x}||_0$
- Preserve edges $||\mathbf{\Psi}\mathbf{x}||_1$: ℓ_1 is less sensitive to outliers than ℓ_2

A zoo of methods

1. Sub-gradient strategies
   - Fu (1998), Shevade and Keerthi (2003), Perkins (2003), Andrew and Gao (2007), ...
2. Constrained optimization
   - Kim et al (2007), Chen et al (1999), Bertsekas (2004), Gafni and Bertsekas (1984), ...
3. Iterative shinkage-thresholding algorithms (ISTA)
   - Bioucas-Dias and Figueiredo(2007), Giryes, Elad and Eldar (2011), Beck and Teboulle (2009), Goldstein and Osher (2009), Osher et al (2005) ....
4. Differentiable Approximations
   - Huber (1964), Lee et al (2006), Tibshirani (1996), Krishnapuram et al (2005), Figueiredo (2003) ...
5. Iteratively re-weighted norm
   - Rodriguez and Wohlberg (2008), Renaut et al (2017), ...
6. Generalized Krylov methods for $\ell_p - \ell_q$
   - Huang, Lanza, Morigi, Reichel, Sgallari (2017), Lanza et al (2015), ...
7. Flexible Arnoldi methods (for square problems)
   - Gazzola and Nagy (2014), Saibaba et al (2013), ...

⋮

A zoo of methods

1. Sub-gradient strategies
   - Fu (1998), Shevade and Keerthi (2003), Perkins (2003), Andrew and Gao (2007), ...
2. Constrained optimization
   - Kim et al (2007), Chen et al (1999), Bertsekas (2004), Gafni and Bertsekas (1984), ...
3. Iterative shinkage-thresholding algorithms (ISTA)
   - Bioucas-Dias and Figueiredo(2007), Giryes, Elad and Eldar (2011), Beck and Teboulle (2009), Goldstein and Osher (2009), Osher et al (2005) ....
4. Differentiable Approximations
   - Huber (1964), Lee et al (2006), Tibshirani (1996), Krishnapuram et al (2005), Figueiredo (2003) ...
5. **Iteratively re-weighted norm**
   - Rodriguez and Wohlberg (2008), Renaut et al (2017), ...
6. Generalized Krylov methods for $\ell_p - \ell_q$
   - Huang, Lanza, Morigi, Reichel, Sgallari (2017), Lanza et al (2015), ...
7. **Flexible Arnoldi methods (for square problems)**
   - Gazzola and Nagy (2014), Saibaba et al (2013), ...

⋮

Outline

Introduction and motivation

Background on iteratively re-weighted and flexible methods

Flexible Golub-Kahan based methods

Numerical results and conclusions

Iterative Re-weighted Norm (IRN)

Let $\mathbf{\Psi} = \mathbf{I}, p = 1$. Turn ℓ_1-problems into a sequence of ℓ_2-problems

$$\|\mathbf{x}\|_1 \approx \|\mathbf{L}(\mathbf{x})\mathbf{x}\|_2^2$$

where $\mathbf{L}(\mathbf{x}) = \mathrm{diag}\left(1/\sqrt{f_\tau(|\mathbf{x}|)}\right)$ and

$$f_\tau(|[\mathbf{x}]_i|) = \begin{cases} |[\mathbf{x}]_i| & \text{if } |[\mathbf{x}]_i| \geq \tau_1 \\ \tau_2 & \text{if } |[\mathbf{x}]_i| < \tau_1 \end{cases}$$

1: **for** $k = 0, 1, \ldots$ **do**
2: $\quad \mathbf{x}_{k+1} = \arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{L}(\mathbf{x}_k)\mathbf{x}\|_2^2$
3: **end for**

Let $\mathbf{L}_k = \mathbf{L}(\mathbf{x}_k)$, then $\mathbf{x}_{k+1} = \mathbf{L}_k^{-1}\mathbf{y}_{k+1}$ where

$$\mathbf{y}_{k+1} = \arg\min_{\mathbf{y}} \left\|\mathbf{A}\mathbf{L}_k^{-1}\mathbf{y} - \mathbf{b}\right\|_2^2 + \lambda \|\mathbf{y}\|_2^2$$

Rodriguez and Wohlberg (2008), Candes, Wakin and Boyd (2008), ...

Generalized Arnoldi Tikhonov (GAT)

1. For $\mathbf{A} \in \mathbb{R}^{n \times n}$, use *flexible* Arnoldi to generate basis vectors:

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{L}_1^{-1}\mathbf{v}_1 & \cdots & \mathbf{L}_k^{-1}\mathbf{v}_k \end{bmatrix} \in \mathbb{R}^{n \times k}$$

where

$$\mathbf{A}\mathbf{Z}_k = \mathbf{V}_{k+1}\mathbf{H}_k$$

- $\mathbf{V}_{k+1} = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_{k+1} \end{bmatrix}$ has orthonormal columns (ONC)
- $\mathbf{H}_k \in \mathbb{R}^{(k+1) \times k}$ is upper Hessenberg

2. Compute solution $\mathbf{x}_k = \mathbf{x}_0 + \mathbf{Z}_k \mathbf{y}_k$ where

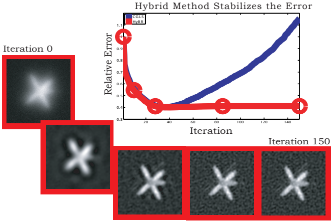
$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \frac{1}{2} \left\| \mathbf{H}_k \mathbf{y} - \|\mathbf{r}_0\|_2 \, \mathbf{e}_1 \right\|_2^2 + \lambda \|\mathbf{y}\|_2^2$$

Saad (1993, 2003), Gazzola and Nagy (2014)

Outline

Summary of benefits

- **Flexible Golub-Kahan**
  - ✔ Avoid inner-outer schemes: Current solution immediately incorporated in basis
  - ✔ Extensions to non-square problems
  - ✔ Optimality and equivalency results



- **Hybrid method**
  - ✔ Stabilize reconstruction errors
  - ✔ Automatic choice of λ and stopping criteria

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{\Psi}\mathbf{x}\|_p^p$$

- **Transformed problem**
  - ✔ Enforce sparsity in a transform
  - ✔ Connections to multi-parameter regularization

Flexible Golub-Kahan (GK) Process

Given $\mathbf{A}, \mathbf{b}$, initialize $\mathbf{u}_1 = \mathbf{b}/\beta_1$ where $\beta_1 = \|\mathbf{b}\|$.

After k iterations with changing preconditioners $\mathbf{L}_k$, we have

- $\mathbf{Z}_k = \begin{bmatrix} \mathbf{L}_1^{-1}\mathbf{v}_1 & \cdots & \mathbf{L}_k^{-1}\mathbf{v}_k \end{bmatrix} \in \mathbb{R}^{n \times k}$
- $\mathbf{M}_k \in \mathbb{R}^{(k+1)\times k}$ upper Hessenberg
- $\mathbf{T}_k \in \mathbb{R}^{k \times k}$ upper triangular
- $\mathbf{U}_{k+1} = \begin{bmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_{k+1} \end{bmatrix} \in \mathbb{R}^{m \times (k+1)}$ ONC
- $\mathbf{V}_k = \begin{bmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_k \end{bmatrix} \in \mathbb{R}^{n \times k}$ ONC

such that

$$\mathbf{A}\mathbf{Z}_k = \mathbf{U}_{k+1}\mathbf{M}_k \quad \text{and} \quad \mathbf{A}^\top \mathbf{U}_{k+1} = \mathbf{V}_{k+1}\mathbf{T}_{k+1}$$

Remarks:

- If $\mathbf{L}_k = \mathbf{L}$, get right-preconditioned GK bidiagonalization
- Additional orthogonalizations and storage

Related to inexact Krylov methods: Simoncini and Szyld (2007), Van Den Eshof and Sleijpen (2004)

Flexible LSQR and flexible LSMR

1. Use *flexible* GK to generate basis vectors:

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{L}_1^{-1}\mathbf{v}_1 & \cdots & \mathbf{L}_k^{-1}\mathbf{v}_k \end{bmatrix} \in \mathbb{R}^{n \times k}$$

$$\mathbf{A}\mathbf{Z}_k = \mathbf{U}_{k+1}\mathbf{M}_k \quad \text{and} \quad \mathbf{A}^\top \mathbf{U}_{k+1} = \mathbf{V}_{k+1}\mathbf{T}_{k+1}$$

2. Compute solution $\mathbf{x}_k = \mathbf{Z}_k \mathbf{y}_k$ where
    - Flexible LSQR (FLSQR)

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{M}_k \mathbf{y} - \beta_1 \mathbf{e}_1\|_2^2$$

    - Flexible LSMR (FLSMR)

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{T}_{k+1}\mathbf{M}_k \mathbf{y} - \beta_1 m_{11} \mathbf{e}_1\|_2^2$$

Some theoretical results

Optimality properties:

- The FLSQR solution $\mathbf{x}_k$ obtained at the kth step minimizes the residual norm $\|\mathbf{A}\mathbf{x}_k - \mathbf{b}\|_2$ over $\mathbf{x}_0 + \text{span}\{\mathbf{Z}_k\}$.
- The FLSMR solution $\mathbf{x}_k$ obtained at the kth step minimizes $\left\|\mathbf{A}^\top(\mathbf{A}\mathbf{x}_k - \mathbf{b})\right\|_2$ over $\mathbf{x}_0 + \text{span}\{\mathbf{Z}_k\}$.

Equivalency results:

- FLSQR is equivalent to flexible conjugate gradients[1] on the normal equations

$$\mathbf{A}^\top\mathbf{A}\mathbf{x} = \mathbf{A}^\top\mathbf{b}$$

- FLSMR is equivalent to FGMRES applied to the normal equations

[1] Notay (2000)

Flexible GK *hybrid* methods

1. Use *flexible* GK to generate basis vectors:

$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{L}_1^{-1}\mathbf{v}_1 & \cdots & \mathbf{L}_k^{-1}\mathbf{v}_k \end{bmatrix} \in \mathbb{R}^{n \times k}$$

$$\mathbf{A}\mathbf{Z}_k = \mathbf{U}_{k+1}\mathbf{M}_k \quad \text{and} \quad \mathbf{A}^\top \mathbf{U}_{k+1} = \mathbf{V}_{k+1}\mathbf{T}_{k+1}$$

2. Compute solution $\mathbf{x}_k = \mathbf{Z}_k \mathbf{y}_k$ where
    - Flexible GK Tikhonov - R (FLSQR-R)

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{M}_k \mathbf{y} - \beta_1 \mathbf{e}_1\|_2^2 + \lambda \|\mathbf{R}_k \mathbf{y}\|_2^2 \,, \quad \mathbf{Z}_k = \mathbf{Q}\mathbf{R}_k$$

    - Flexible GK Tikhonov - I (FLSQR-I)[2]

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{M}_k \mathbf{y} - \beta_1 \mathbf{e}_1\|_2^2 + \lambda \|\mathbf{y}\|_2^2$$

[2] Based on Gazzola & Nagy (2014)

FLSQR-R: Approximate singular values of $\mathbf{A}$

$$\mathbf{R}_k^{-\top}\mathbf{M}_k^\top\mathbf{M}_k\mathbf{R}_k^{-1} = \mathbf{R}_k^{-\top}\mathbf{M}_k^\top\mathbf{U}_{k+1}^\top\mathbf{U}_{k+1}\mathbf{M}_k\mathbf{R}_k^{-1}$$
$$= \mathbf{Q}_k^\top\mathbf{A}^\top\mathbf{A}\mathbf{Q}_k$$

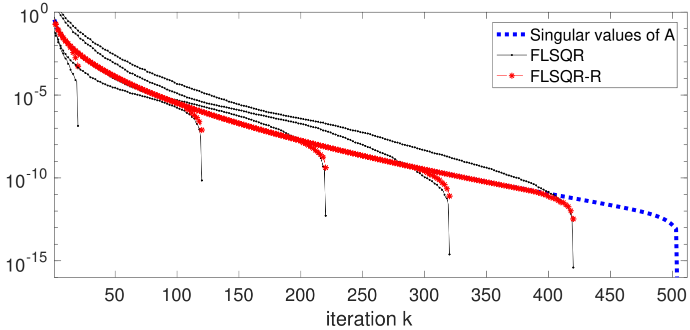


Figure: This plot compares the singular values of $\mathbf{A}$ to the singular values of $\mathbf{M}_k$ from FLSQR and of $\mathbf{M}_k\mathbf{R}_k^{-1}$ from FLSQR-R, for iterations k between 20 and 420 in increments of 100.

Solving the transformed problem

Equivalent problems (for $\widetilde{\boldsymbol{\Psi}}$ orthogonal):

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\boldsymbol{\Psi}\mathbf{x}\|_p^p$$

$$\Updownarrow$$

$$\min_{\mathbf{x}} \Bigg\| \underbrace{\widetilde{\boldsymbol{\Psi}}\mathbf{A}\boldsymbol{\Psi}^{-1}}_{\mathbf{H}} \underbrace{\boldsymbol{\Psi}\mathbf{x}}_{\mathbf{s}} - \underbrace{\widetilde{\boldsymbol{\Psi}}\mathbf{b}}_{\mathbf{d}} \Bigg\|_2^2 + \lambda \Bigg\| \underbrace{\boldsymbol{\Psi}\mathbf{x}}_{\mathbf{s}} \Bigg\|_p^p$$

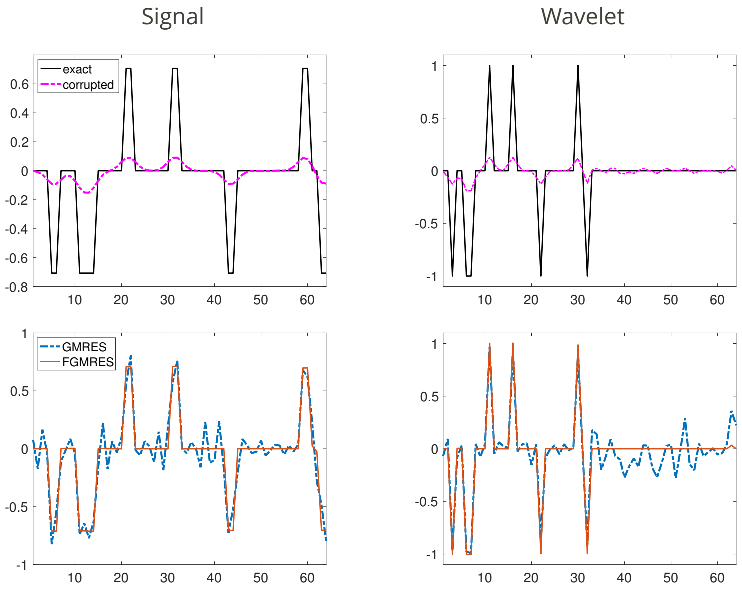
Solution subspace for flexible Arnoldi:

$$\mathrm{span}\{\mathbf{L}_1^{-1}\mathbf{d}, \mathbf{L}_2^{-1}\mathbf{H}\mathbf{L}_1^{-1}\mathbf{d}, \dots, \mathbf{L}_k^{-1}\mathbf{H}\cdots\mathbf{L}_2^{-1}\mathbf{H}\mathbf{L}_1^{-1}\mathbf{d}\}$$

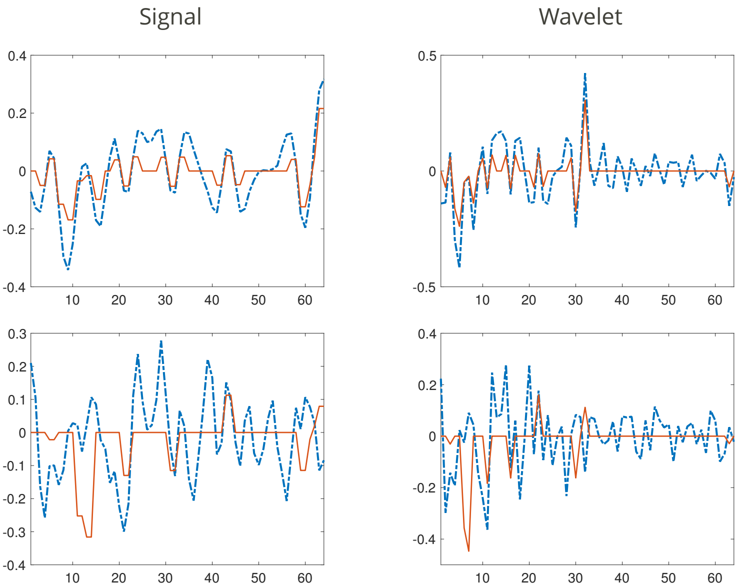
$$\Updownarrow$$

$$\boldsymbol{\Psi}^\top \mathrm{span}\{\mathbf{L}_1^{-1}\boldsymbol{\Psi}\mathbf{b}, \mathbf{L}_2^{-1}\boldsymbol{\Psi}\mathbf{A}\mathbf{b}, \dots, \mathbf{L}_k^{-1}\boldsymbol{\Psi}\mathbf{A}\boldsymbol{\Psi}^\top\cdots\mathbf{L}_2^{-1}\boldsymbol{\Psi}\mathbf{A}\boldsymbol{\Psi}^\top\mathbf{L}_1^{-1}\boldsymbol{\Psi}\mathbf{b}\}$$

An illustration: Sparsity in a wavelet domain



2^{nd} and 4^{th} basis vectors

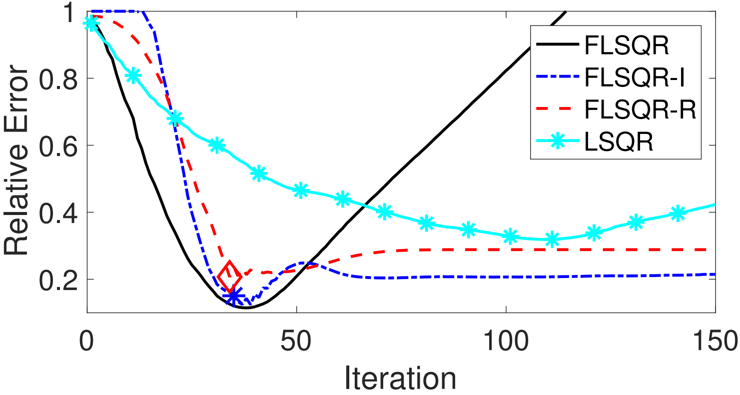


Outline

Image deblurring example with $\mathbf{\Psi} = \mathbf{I}$

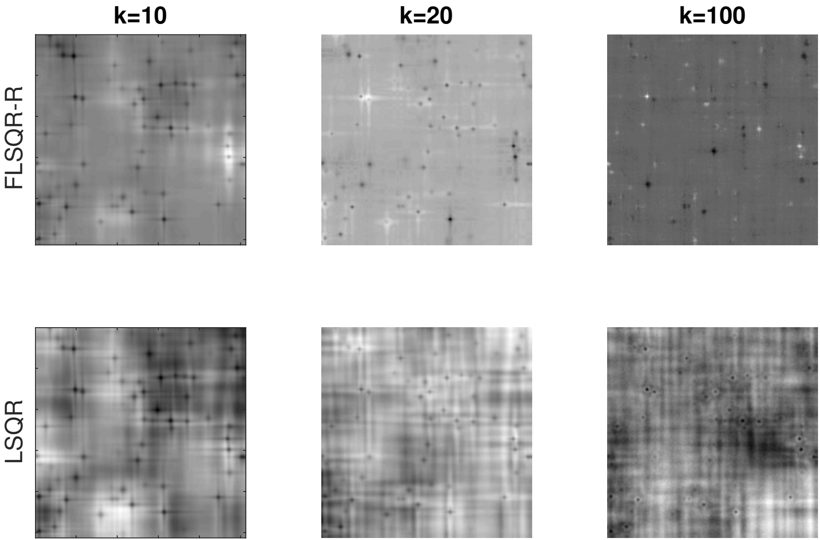
- Image is 256×256
- Noise level is 5×10^{-2}
- Reflexive boundary conditions

Reconstruction errors

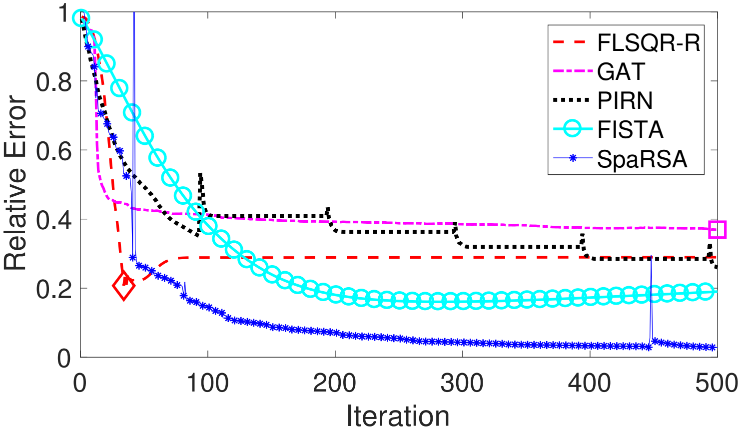


- Reconstruction errors computed as $\frac{\|\mathbf{x}_k - \mathbf{x}_{\text{true}}\|_2}{\|\mathbf{x}_{\text{true}}\|_2}$
- λ for FLSQR-I and FLSQR-R use discrepancy principle

Basis images



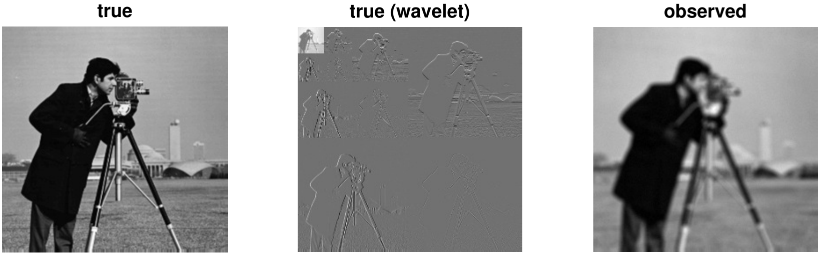
Comparison to other methods



- GAT = Generalized Arnoldi-Tikhonov
- PIRN† = Preconditioned iteratively re-weighted norm
- FISTA† = Fast iterative-shrinkage-thresholding algorithm
- SpaRSA† = Sparse Reconstruction by Separable Approximation

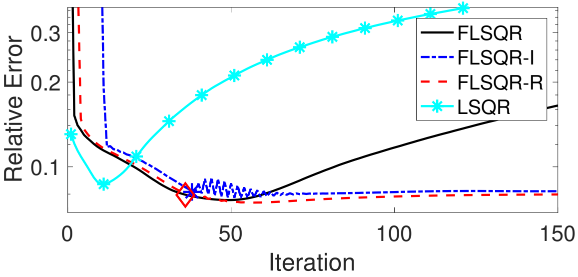
† uses λ from FLSQR-R

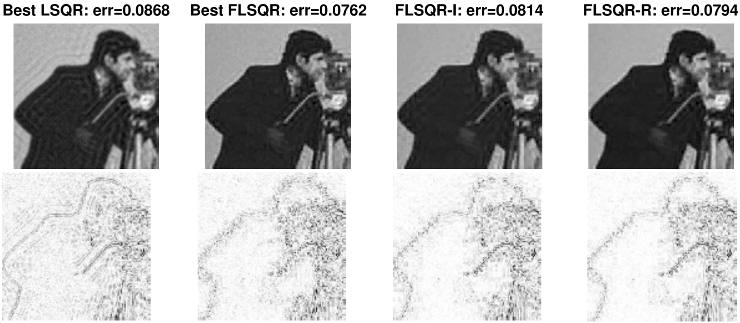
Image deblurring with $\Phi \neq \mathrm{I}$



- Image is 256×256
- Φ is a 2D Haar wavelet decomposition with 3 levels
- Blur is out-of-focus blur with radius 4
- Noise level 0.01

Golub-Kahan based reconstructions





Conclusions and references

- Introduced *flexible* and *hybrid* Golub-Kahan methods

- Easy-to-use methods for ℓ_p-regularization with 2 main benefits:
  1. Immediate incorporation of the current solution

     - thanks to the flexible approach
  2. Automatic regularization parameter selection

     - thanks to the hybrid framework

- Extensions to non-square and transformed problems

- Theoretical results show optimality and equivalences

Thank you!

Gazzola and Nagy. "Generalized Arnoldi-Tikhonov Method for Sparse Reconstruction." SISC 36(2), 2014.

Chung and Gazzola. "Flexible Krylov Methods for ℓ_p-Regularization." arXiv: 1806.06502