Flexible Krylov Methods for $\ell_p$ Regularization

Silvia Gazzola
Joint work with Julianne Chung and Malena Sabaté Landman

Department of Mathematical Sciences

Modern Challenges in Imaging
Tufts University, August 6, 2019
What is this talk about?
What is this talk about?

Regularization of linear inverse problems

\[ Ax_{\text{true}} + \epsilon = b, \]

where

- \( b \in \mathbb{R}^M \) observations or measurements
- \( x_{\text{true}} \in \mathbb{R}^N \) desired parameters
- \( A \in \mathbb{R}^{M \times N} \) ill-conditioned matrix models forward process
- \( \epsilon \in \mathbb{R}^M \) additive Gaussian noise
What is this talk about?

Regularization of linear inverse problems

\[ \mathbf{Ax}_{\text{true}} + \epsilon = \mathbf{b}, \]

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
- \( \epsilon \in \mathbb{R}^M \): additive Gaussian noise

image deblurring and denoising

computed tomography
Outline

1. Introduction
   - $\ell^p$ variational regularization
   - Iteratively Re-weighted Norm (IRN) methods

2. Methods based on the Flexible Golub-Kahan (FGK) algorithm
   - Flexible Golub-Kahan (FGK) Algorithm
   - FLSQR and FLSMR
   - Hybrid FLSQR and Hybrid FLSMR

3. Sparsity under transform
   - Invertible transforms (wavelets)
   - Non-Invertible transforms (TV)

4. Numerical experiments

5. Conclusions
Applying variational regularization...

\[ \mathbf{x}^{\text{reg}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \| \mathbf{A} \mathbf{x} - \mathbf{b} \|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \quad \lambda > 0 \]
Applying variational regularization...

\[ \mathbf{x}^{\text{reg}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \| \mathbf{Ax} - \mathbf{b} \|^2_2 + \lambda \mathcal{R}(\mathbf{x}), \quad \lambda > 0 \]

\[ \mathcal{R}(\mathbf{x}) = \| \mathbf{Lx} \|^2_2 \]
Applying variational regularization...

\[ \mathbf{x}^{\text{reg}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \quad \lambda > 0 \]

\[ \mathcal{R}(\mathbf{x}) = \| \mathbf{L}\mathbf{x} \|_2^2 \]

Krylov methods are popular in this setting
Applying variational regularization...

\[
x^{\text{reg}} = \arg\min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \mathcal{R}(x), \quad \lambda > 0
\]

\[\mathcal{R}(x) = \|Lx\|_2^2\]

Krylov methods are popular in this setting

\[
x_k \in \mathcal{K}_k(C, d) = \text{span}\{d, Cd, \ldots, C^{k-1}d\}
\]

\[r_k = b - Ax_k \perp \mathcal{K}_k(C', d')\]
Applying variational regularization...

\[ \mathbf{x}^{\text{reg}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|^2_2 + \lambda \mathcal{R}(\mathbf{x}), \quad \lambda > 0 \]

\[ \mathcal{R}(\mathbf{x}) = \| \mathbf{L}\mathbf{x} \|^2_2 \]

Krylov methods are popular in this setting

\[ \mathbf{x}_k \in \mathcal{K}_k(\mathbf{C}, \mathbf{d}) = \text{span}\{ \mathbf{d}, \mathbf{C}\mathbf{d}, \ldots, \mathbf{C}^{k-1}\mathbf{d} \} \]

\[ \mathbf{r}_k = \mathbf{b} - \mathbf{A}\mathbf{x}_k \perp \mathcal{K}_k(\mathbf{C}', \mathbf{d}') \]

\[ \mathbf{A}\mathbf{Z}_k = \mathbf{W}_{k+1}\mathbf{G}_k \]

\[ \mathbf{x}_k = \mathbf{Z}_k\mathbf{y}_k, \quad \mathbf{y}_k = \arg \min_{\mathbf{y} \in \mathbb{R}^k} \| \mathbf{g}_k - \mathbf{G}_k\mathbf{y} \|^2_2 \]
Applying variational regularization...

\[ x^{\text{reg}} = \arg \min_{x \in \mathbb{R}^N} \|Ax - b\|^2_2 + \lambda R(x), \quad \lambda > 0 \]

- \( R(x) = \|Lx\|^2_2 \)

Krylov methods are popular in this setting

- \( x_k \in \mathcal{K}_k(C, d) = \text{span}\{d, C_d, \ldots, C^{k-1}d\} \)
- \( r_k = b - Ax_k \perp \mathcal{K}_k(C', d') \)

\[ AZ_k = W_{k+1}G_k \]

\[ x_k = Z_ky_k, \quad y_k = \arg \min_{y \in \mathbb{R}^k} \|g_k - G_ky\|^2_2 \]

Fast semi-convergence
Applying variational regularization...

\[
x^\text{reg} = \arg \min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \mathcal{R}(x), \quad \lambda > 0
\]

\[\mathcal{R}(x) = \|Lx\|_2^2\]

Krylov methods are popular in this setting

\[x_k \in \mathcal{K}_k(C, d) = \text{span}\{d, Cd, \ldots, C^{k-1}d\}\]

\[r_k = b - Ax_k \perp \mathcal{K}_k(C', d')\]

\[AZ_k = W_{k+1}G_k\]

\[x_k = Z_ky_k, \quad y_k = \arg \min_{y \in \mathbb{R}^k} \|g_k - G_ky\|_2^2 + \lambda_k \|L_ky\|_2^2\]

Fast semi-convergence
Applying variational regularization...

\[ x^{\text{reg}} = \arg \min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda R(x), \quad \lambda > 0 \]

- \( R(x) = \|Lx\|_2^2 \)

Krylov methods are popular in this setting

\[ x_k \in \mathcal{K}_k(C, d) = \text{span}\{d, Cd, \ldots, C^{k-1}d\} \]
\[ r_k = b - Ax_k \perp \mathcal{K}_k(C', d') \]
\[ AZ_k = W_{k+1}G_k \]
\[ x_k = Z_ky_k, \quad y_k = \arg \min_{y \in \mathbb{R}^k} \|g_k - G_ky\|_2^2 + \lambda_k \|L_ky\|_2^2 \]

Fast semi-convergence

Hanke (1995); Frommer and Maas (1999); O’Leary and Simmons (1981); Calvetti, Morigi, Reichel, Sgallari (2000); Kilmer, Hansen, Espanol (2007); Chung and Palmer (2015); G., Novati, Russo (2015) ...
Applying variational regularization...

$$x^{\text{reg}} = \arg \min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \mathcal{R}(x), \quad \lambda > 0$$
Applying variational regularization...

\[ \mathbf{x}^{\text{reg}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|_2^2 + \lambda \mathcal{R}(\mathbf{x}), \quad \lambda > 0 \]

- **\( \ell^p \) regularization**
  (we will consider \( \mathcal{R}(\mathbf{x}) = \| \mathbf{x} \|_p \), \( \mathcal{R}(\mathbf{x}) = \| \Psi \mathbf{x} \|_p \), \( \mathcal{R}(\mathbf{x}) = \text{TV}_p(\mathbf{x}) \); \( p \geq 1, p > 0 \)
Applying variational regularization...

\[ \mathbf{x}^{\text{reg}} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \| \mathbf{A}\mathbf{x} - \mathbf{b} \|^2_2 + \lambda \mathcal{R}(\mathbf{x}), \quad \lambda > 0 \]

- **\( \ell^p \) regularization**
  (we will consider \( \mathcal{R}(\mathbf{x}) = \| \mathbf{x} \|_p \), \( \mathcal{R}(\mathbf{x}) = \| \psi \mathbf{x} \|_p \), \( \mathcal{R}(\mathbf{x}) = \text{TV}_p(\mathbf{x}) \); \( p \geq 1, p > 0 \))

  - **Sub-gradient strategies**
    Shevade and Keerthi (2003), Perkins (2003), Andrew and Gao (2007), ...

  - **Constrained optimization**
    Chen et al (1999), Bertsekas (2004), Gafni and Bertsekas (1984), ...

  - **Iterative shrinkage-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) ....

  - **Iteratively re-weighted norm**
    Rodriguez and Wohlberg (2008), Renaut et al (2017), ...

  - **Generalized Krylov methods for \( \ell_p - \ell_q \)**
    Lanza et al (2015), Huang, Lanza, Morigi, Reichel, Sgallari (2017), Buccini and Reichel (2019), ...

  - **Flexible Arnoldi methods (for square problems)**
    Gazzola and Nagy (2014)
Applying variational regularization...

\[ x^{\text{reg}} = \arg\min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \mathcal{R}(x), \quad \lambda > 0 \]

- **\( \ell^p \) regularization**
  (we will consider \( \mathcal{R}(x) = \|x\|_p^p, \mathcal{R}(x) = \|\Psi x\|_p^p, \mathcal{R}(x) = \text{TV}_p(x); \ p \geq 1, \ p > 0 \))
  - Sub-gradient strategies
    - Shevade and Keerthi (2003), Perkins (2003), Andrew and Gao (2007), ...
  - Constrained optimization
    - Chen et al (1999), Bertsekas (2004), Gafni and Bertsekas (1984), ...
  - Iterative shrinkage-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) ...
  - **Iteratively re-weighted norm**
    - Rodriguez and Wohlberg (2008), Renaut et al (2017), ...
  - Generalized Krylov methods for \( \ell_p - \ell_q \)
    - Lanza et al (2015), Huang, Lanza, Morigi, Reichel, Sgallari (2017), Buccini and Reichel (2019), ...
  - **Flexible Arnoldi methods (for square problems)**
    - Gazzola and Nagy (2014)
A basic Iteratively Re-weighted Norm (IRN) strategy ...

Let $\Psi = I, \rho = 1$. 

A basic Iteratively Re-weighted Norm (IRN) strategy ...

Let $\Psi = I, p = 1$. Turn $\ell_1$-problems into a sequence of $\ell_2$-problems:
A basic Iteratively Re-weighted Norm (IRN) strategy ...

Let $\Psi = I, p = 1$. Turn $\ell_1$-problems into a sequence of $\ell_2$-problems:

$$\|x\|_1 \approx \|L(x)x\|_2^2$$
A basic Iteratively Re-weighted Norm (IRN) strategy ...

[Rodriguez and Wohlberg (2008)]

Let $\Psi = I, p = 1$. Turn $\ell_1$-problems into a sequence of $\ell_2$-problems:

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

**IRN algorithm**

Input: $A, b, x_0(= 0), L_0 = L(x_0)(= I)$

- For $k = 1, \ldots$, till a stopping criterion is satisfied

$$x_k = \arg \min_{x \in \mathbb{R}^N} \|b - Ax\|_2^2 + \lambda \|L_{k-1}x\|_2^2$$

- Update $L_k = \text{diag} \left(1/\sqrt{f_\tau(|x_k|)}\right)$, $f_\tau(||x_k||) = \begin{cases} |[x_k]_i| & \text{if } |[x_k]_i| \geq \tau_1 \\ \tau_2 & \text{if } |[x_k]_i| < \tau_1 \end{cases}$
A basic Iteratively Re-weighted Norm (IRN) strategy ...

[Rodriguez and Wohlberg (2008)]

Let $\Psi = I$, $p = 1$. Turn $\ell_1$-problems into a sequence of $\ell_2$-problems:

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

**IRN algorithm**

**Input:** $A$, $b$, $x_0(=0)$, $L_0 = L(x_0)(=I)$

- For $k = 1, \ldots$, till a stopping criterion is satisfied
  - Till a stopping criterion is satisfied: run an (iterative) solver for
    $$x_k = \arg \min_{x \in \mathbb{R}^N} \|b - Ax\|_2^2 + \lambda \|L_{k-1}x\|_2^2$$

- Update $L_k = \text{diag} \left( 1/\sqrt{f_\tau(\|x_k\|)} \right)$, $f_\tau(\|x_k\|) = \begin{cases} \|x_k\| & \text{if } \|x_k\| \geq \tau_1 \\ \tau_2 & \text{if } \|x_k\| < \tau_1 \end{cases}$
A basic Iteratively Re-weighted Norm (IRN) strategy ...

[Rodriguez and Wohlberg (2008)]

Let $\Psi = I$, $p = 1$. Turn $\ell_1$-problems into a sequence of $\ell_2$-problems:

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

**IRN algorithm**

**Input:** $A$, $b$, $x_0(= 0)$, $L_0 = L(x_0)(= I)$

- For $k = 1, \ldots$, till a stopping criterion is satisfied
  - Till a stopping criterion is satisfied: run an (iterative) solver for

$$x_k = \arg\min_{x \in \mathbb{R}^N} \|b - Ax\|_2^2 + \lambda \|L_{k-1}x\|_2^2$$

- Update $L_k = \text{diag}\left(1/\sqrt{f_\tau(\|x_k\|)}\right)$, $f_\tau(\|x_k\|) = \begin{cases} \|x_k\|_i & \text{if } |x_k|_i \geq \tau_1 \\ \tau_2 & \text{if } |x_k|_i < \tau_1 \end{cases}$

Let $L_k = L(x_k)$, then $x_{k+1} = L_k^{-1}y_{k+1}$ where

$$y_{k+1} = \arg\min_{y} \left\|AL_k^{-1}y - b\right\|_2^2 + \lambda \|y\|_2^2$$
A basic Iteratively Re-weighted Norm (IRN) strategy ...

[Rodriguez and Wohlberg (2008)]

Let $\Psi = I, p = 1$. Turn $\ell_1$-problems into a sequence of $\ell_2$-problems:

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

**IRN algorithm**

Input: $A$, $b$, $x_0(= 0)$, $L_0 = L(x_0)(= I)$

- For $k = 1, \ldots$, till a stopping criterion is satisfied
  - Till a stopping criterion is satisfied: run an (iterative) solver for
    $$x_k = \arg \min_{x \in \mathbb{R}^N} \|b - Ax\|_2^2 + \lambda \|L_{k-1}x\|_2^2$$

- Update $L_k = \text{diag}\left(1/\sqrt{f_\tau(\|x_k\|)}\right)$, $f_\tau(\|x_k\|) = \begin{cases} \|x_k\|_i & \text{if } \|x_k\|_i \geq \tau_1 \\ \tau_2 & \text{if } \|x_k\|_i < \tau_1 \end{cases}$

Let $L_k = L(x_k)$, then $x_{k+1} = L_k^{-1}y_{k+1}$ where

$$y_{k+1} = \arg \min_y \|AL_k^{-1}y - b\|_2^2 + \lambda \|y\|_2^2$$
For $A \in \mathbb{R}^{N \times N}$, use flexible Arnoldi to generate basis vectors:

\[
Z_k = \begin{bmatrix}
L_1^{-1}v_1 & \cdots & L_k^{-1}v_k
\end{bmatrix} \in \mathbb{R}^{N \times k}
\]

where

\[
AZ_k = V_{k+1}H_k
\]

- $V_{k+1} = \begin{bmatrix} v_1 & \cdots & v_{k+1} \end{bmatrix} \in \mathbb{R}^{N \times (k+1)}$ has orthonormal columns (ONC)
- $H_k \in \mathbb{R}^{(k+1) \times k}$ is upper Hessenberg
... revisited within Flexible Krylov methods ...

[G. and Nagy (2014)]

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

   $Z_k = \left[ L_1^{-1}v_1 \quad \cdots \quad L_k^{-1}v_k \right] \in \mathbb{R}^{N \times k}$

   where

   $AZ_k = V_{k+1}H_k$

   - $V_{k+1} = \left[ v_1 \quad \cdots \quad v_{k+1} \right] \in \mathbb{R}^{N \times (k+1)}$ has orthonormal columns (ONC)
   - $H_k \in \mathbb{R}^{(k+1) \times k}$ is upper Hessenberg

2. Compute solution $x_k = x_0 + Z_k y_k$ where

   $y_k = \arg \min_{y} \frac{1}{2} \| H_k y - r_0 \|_2 \| e_1 \|_2^2 + \lambda \| y \|_2^2$
Flexible Golub-Kahan (FGK) Process

[Chung and G. (2018)]

A new flexible factorization
Flexible Golub-Kahan (FGK) Process

[Chung and G. (2018)]

Given \( A \in \mathbb{R}^{M \times N} \), \( b \in \mathbb{R}^{M} \), initialize \( u_1 = b / \beta_1 \) where \( \beta_1 = \| b \| \).

After \( k \) iterations with changing preconditioners \( L_k \), we have

Related to inexact Krylov methods [Simoncini and Szyld (2007)]

- \( Z_k = \begin{bmatrix} L_1^{-1}v_1 & \cdots & L_k^{-1}v_k \end{bmatrix} \in \mathbb{R}^{N \times k} \)
- \( M_k \in \mathbb{R}^{(k+1) \times k} \) upper Hessenberg
- \( T_k \in \mathbb{R}^{k \times k} \) upper triangular
- \( U_{k+1} = \begin{bmatrix} u_1 & \cdots & u_{k+1} \end{bmatrix} \in \mathbb{R}^{M \times (k+1)} \) ONC
- \( V_k = \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} \in \mathbb{R}^{N \times k} \) ONC

such that

\[
AZ_k = U_{k+1}M_k \quad \text{and} \quad A^T U_{k+1} = V_{k+1}T_{k+1}
\]
Flexible Golub-Kahan (FGK) Process

[Chung and G. (2018)]

Given \( A \in \mathbb{R}^{M \times N}, b \in \mathbb{R}^M \), initialize \( u_1 = b / \beta_1 \) where \( \beta_1 = \|b\| \).

After \( k \) iterations with changing preconditioners \( L_k \), we have

Related to inexact Krylov methods [Simoncini and Szyld (2007)]

- \( Z_k = \begin{bmatrix} L_1^{-1}v_1 & \cdots & L_k^{-1}v_k \end{bmatrix} \in \mathbb{R}^{N \times k} \)
- \( M_k \in \mathbb{R}^{(k+1) \times k} \) upper Hessenberg
- \( T_k \in \mathbb{R}^{k \times k} \) upper triangular
- \( U_{k+1} = \begin{bmatrix} u_1 & \cdots & u_{k+1} \end{bmatrix} \in \mathbb{R}^{M \times (k+1)} \) ONC
- \( V_k = \begin{bmatrix} v_1 & \cdots & v_k \end{bmatrix} \in \mathbb{R}^{N \times k} \) ONC

such that

\[ AZ_k = U_{k+1}M_k \quad \text{and} \quad A^T U_{k+1} = V_{k+1}T_{k+1} \]

Remarks:
- If \( L_k = L \), get right-preconditioned GK bidiagonalization
- Additional orthogonalizations and storage
Flexible LSQR and flexible LSMR

New flexible solvers
Flexible LSQR and flexible LSMR

1 Use *flexible* GK to generate basis vectors:

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

\[ \mathbf{A} \mathbf{Z}_k = \mathbf{U}_{k+1} \mathbf{M}_k \quad \text{and} \quad \mathbf{A}^\top \mathbf{U}_{k+1} = \mathbf{V}_{k+1} \mathbf{T}_{k+1} \]
Flexible LSQR and flexible LSMR

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

\[
Z_k = \begin{bmatrix}
L_1^{-1}v_1 & \cdots & L_k^{-1}v_k
\end{bmatrix} \in \mathbb{R}^{n \times k}
\]

\[
AZ_k = U_{k+1}M_k \quad \text{and} \quad A^T U_{k+1} = V_{k+1}T_{k+1}
\]

2. Compute solution \( x_k = Z_k y_k \) where

- **Flexible LSQR (FLSQR)**

\[
y_k = \arg \min_{y \in \mathbb{R}^k} \| M_k y - \beta_1 e_1 \|_2^2
\]

- **Flexible LSMR (FLSMR)**

\[
y_k = \arg \min_{y \in \mathbb{R}^k} \| T_{k+1} M_k y - \beta_1 m_{11} e_1 \|_2^2
\]
Flexible LSQR and flexible LSMR

1. Use flexible GK to generate basis vectors:

\[ Z_k = \begin{bmatrix} L_1^{-1} v_1 & \cdots & L_k^{-1} v_k \end{bmatrix} \in \mathbb{R}^{n \times k} \]

\[ AZ_k = U_{k+1} M_k \quad \text{and} \quad A^T U_{k+1} = V_{k+1} T_{k+1} \]

2. Compute solution \( x_k = Z_k y_k \) where

- **Flexible LSQR (FLSQR)**
  \[ y_k = \arg\min_{y \in \mathbb{R}^k} \| M_k y - \beta_1 e_1 \|_2^2 \]

  **Optimality property:**
  \( x_k \) minimizes \( \| A x_k - b \|_2 \) over \( x_0 + \text{span}\{Z_k\} \).

- **Flexible LSMR (FLSMR)**
  \[ y_k = \arg\min_{y \in \mathbb{R}^k} \| T_{k+1} M_k y - \beta_1 m_{11} e_1 \|_2^2 \]
Flexible LSQR and flexible LSMR

1. Use flexible GK to generate basis vectors:

\[ Z_k = \begin{bmatrix} L_1^{-1}v_1 & \cdots & L_k^{-1}v_k \end{bmatrix} \in \mathbb{R}^{n \times k} \]

\[ AZ_k = U_{k+1}M_k \quad \text{and} \quad A^\top U_{k+1} = V_{k+1}T_{k+1} \]

2. Compute solution \( x_k = Z_k y_k \) where

- Flexible LSQR (FLSQR)

\[ y_k = \arg \min_{y \in \mathbb{R}^k} \| M_k y - \beta_1 e_1 \|_2^2 \]

Optimality property:
\( x_k \) minimizes \( \| Ax_k - b \|_2 \) over \( x_0 + \text{span}\{Z_k\} \).

- Flexible LSMR (FLSMR)

\[ y_k = \arg \min_{y \in \mathbb{R}^k} \| T_{k+1}M_k y - \beta_1 m_{11} e_1 \|_2^2 \]

Optimality property:
\( x_k \) minimizes \( \| A^\top (Ax_k - b) \|_2 \) over \( x_0 + \text{span}\{Z_k\} \).

Equivalency result:
FLSMR is equivalent to FGMRES applied to the normal equations.
Flexible GK (FGK) *hybrid* methods

New flexible solvers used in a hybrid framework
Use flexible GK to generate basis vectors:

\[ Z_k = \begin{bmatrix} L_1^{-1}v_1 & \cdots & L_k^{-1}v_k \end{bmatrix} \in \mathbb{R}^{N \times k} \]

\[ AZ_k = U_{k+1}M_k \quad \text{and} \quad A^\top U_{k+1} = V_{k+1}T_{k+1} \]
Flexible GK (FGK) *hybrid* methods

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

   \[ Z_k = \begin{bmatrix} L_1^{-1}v_1 & \cdots & L_k^{-1}v_k \end{bmatrix} \in \mathbb{R}^{N \times k} \]

   \[ AZ_k = U_{k+1}M_k \quad \text{and} \quad A^\top U_{k+1} = V_{k+1}T_{k+1} \]

2. Compute solution \( x_k = Z_k y_k \)
Flexible GK (FGK) *hybrid* methods

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

   \[ Z_k = \begin{bmatrix} L_1^{-1}v_1 & \cdots & L_k^{-1}v_k \end{bmatrix} \in \mathbb{R}^{N \times k} \]

   \[ AZ_k = U_{k+1}M_k \quad \text{and} \quad A^\top U_{k+1} = V_{k+1}T_{k+1} \]

2. Compute solution \( x_k = Z_ky_k \), where

   - Flexible GK Tikhonov - R (FLSQR-R)

   \[ y_k = \arg \min_{y \in \mathbb{R}^k} \| M_k y - \beta_1e_1 \|_2^2 + \lambda_k \| R_k y \|_2^2, \quad Z_k = Q_kR_k \]
Flexible GK (FGK) *hybrid* methods

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

   \[
   Z_k = \begin{bmatrix}
   L_1^{-1}v_1 \\
   \vdots \\
   L_k^{-1}v_k
   \end{bmatrix} \in \mathbb{R}^{N \times k}
   \]

   \[
   AZ_k = U_{k+1}M_k \quad \text{and} \quad A^\top U_{k+1} = V_{k+1}T_{k+1}
   \]

2. Compute solution \( x_k = Z_ky_k \), where
   - Flexible GK Tikhonov - R (FLSQR-R)
     \[
     y_k = \arg\min_{y \in \mathbb{R}^k} \|M_ky - \beta_1e_1\|_2^2 + \lambda_k \|R_ky\|_2^2, \quad Z_k = Q_kR_k
     \]
   - Flexible GK Tikhonov - I (FLSQR-I)
     \[
     y_k = \arg\min_{y \in \mathbb{R}^k} \|M_ky - \beta_1e_1\|_2^2 + \lambda_k \|y\|_2^2
     \]
FLSQR-R: Approximate singular values of $\mathbf{A}$

$$
R_k^{-\top} M_k^\top M_k R_k^{-1} = R_k^{-\top} M_k^\top U_{k+1}^\top U_{k+1} M_k R_k^{-1} = Q_k^\top A^\top A Q_k
$$

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

Let \( \Psi \neq I \) (invertible), \( p = 1 \) (e.g., \( \Psi \) : image domain \( \rightarrow \) wavelet domain)
**Solving the transformed problem**

Let $\Psi \neq I$ (invertible), $p = 1$ (e.g., $\Psi$ : image domain $\rightarrow$ wavelet domain)

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

$\text{[Belge, Kilmer, Miller (2000)]}$

$$
\begin{align*}
\min_{x \in \mathbb{R}^N} \| Ax - b \|_2^2 + \lambda \| \Psi x \|_1 & \iff \\
\min_{x \in \mathbb{R}^N} \| \tilde{\Psi} A \Psi^{-1} \Psi x - \tilde{\Psi} b \|_2^2 + \lambda \| \Psi x \|_1
\end{align*}
$$

Solution subspace for flexible Arnoldi:

$\hat{s}_k \in \text{span}\{ L_1^{-1}\hat{v}_1, L_2^{-1}\hat{v}_2, \ldots, L_k^{-1}\hat{v}_k \}$, where

$$
\begin{align*}
\hat{v}_1 &= d / \| d \|_2 \\
\hat{v}_2 &= \text{ONC}(HL_1^{-1}\hat{v}_1) \\
\hat{v}_3 &= \text{ONC}(HL_2^{-1}\hat{v}_2) \\
&\vdots
\end{align*}
$$
Solving the transformed problem

Let $\Psi \neq I$ (invertible), $p = 1$ (e.g., $\Psi : \text{image domain} \rightarrow \text{wavelet domain}$)

Equivalent problems (for $\tilde{\Psi}$ orthogonal):
[Belge, Kilmer, Miller (2000)]

\[
\min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \|\Psi x\|_1 \iff \min_{x \in \mathbb{R}^N} \|\tilde{\Psi} A \tilde{\Psi}^{-1} H s - \tilde{\Psi} b\|_2^2 + \lambda \|\tilde{\Psi} x\|_1
\]

Solution subspace for flexible Arnoldi:

$s_k \in \text{span}\{L_1^{-1}\hat{v}_1, L_2^{-1}\hat{v}_2, \ldots, L_k^{-1}\hat{v}_k\}, \quad \text{where}

x_k = \Psi^{-1} s_k$

\[
\begin{align*}
\hat{v}_1 &= d/\|d\|_2 \\
\hat{v}_2 &= \text{ONC}(HL_1^{-1}\hat{v}_1) \\
\hat{v}_3 &= \text{ONC}(HL_2^{-1}\hat{v}_2) \\
\ldots
\end{align*}
\]
Solving the transformed problem

Let $\Psi \neq I$ (invertible), $p = 1$ (e.g., $\Psi$ : image domain $\rightarrow$ wavelet domain)

Equivalent problems (for $\tilde{\Psi}$ orthogonal):
[Belge, Kilmer, Miller (2000)]

\[
\min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \|\Psi x\|_1 \iff \min_{x \in \mathbb{R}^N} \|\tilde{\Psi} A \tilde{\Psi}^{-1} \tilde{\Psi} x - \tilde{\Psi} b\|_2^2 + \lambda \|\tilde{\Psi} x\|_1
\]

Solution subspace for flexible Arnoldi:

$s_k \in \text{span}\{L_1^{-1}\hat{v}_1, L_2^{-1}\hat{v}_2, \ldots, L_k^{-1}\hat{v}_k\}$, where

\[
\hat{v}_1 = d/\|d\|_2
\]
\[
\hat{v}_2 = \text{ONC}(H L_1^{-1}\hat{v}_1)
\]
\[
\hat{v}_3 = \text{ONC}(H L_2^{-1}\hat{v}_2)
\]
\[
\vdots
\]

$x_k = \Psi^{-1}s_k$

Analogously for flexible Golub-Kahan (possibly without $\tilde{\Psi}$).
An illustration: sparsity in a wavelet domain

Signal

Wavelet

exact

corrupted

GMRES

FGMRES

S. Gazzola (UoB) Regularization by Flexible Krylov Methods August 6, 2019
An illustration: 2\textsuperscript{nd} and 4\textsuperscript{th} basis vectors
TV penalization

Let $\mathcal{R}(x) = TV(x)$.

- **1d case:**
  \[
  TV(x) = \|D_{1d}x\|_1 \approx \|W_{1d}Dx\|_2^2, \text{ where}
  \]
  \[
  D_{1d} = \begin{bmatrix}
  1 & -1 \\
  \vdots & \vdots \\
  1 & -1
  \end{bmatrix} \in \mathbb{R}^{(N-1) \times N}, \quad W = \text{diag}\left(|D_{1d}x|^{-1/2}\right)
  \]

- **2d case:**
  \[\text{[Wohlberg and Rodriguez. An iteratively reweighted norm algorithm for TV. IEEE, 2007]}\]
  \[
  TV(x) = \|\left((D^h x)^2 + (D^v x)^2\right)^{1/2}\|_1 \approx \|WD_{2d}x\|_2^2, \text{ where}
  \]
  \[
  D_{2d} = \begin{bmatrix}
  D^h \\
  D^v
  \end{bmatrix} = \begin{bmatrix}
  D_{1d} \otimes I \\
  I \otimes D_{1d}
  \end{bmatrix}, \quad \hat{W} = \text{diag}\left(\left((D^h x)^2 + (D^v x)^2\right)^{-1/4}\right), \quad W = \begin{bmatrix}
  \hat{W} & 0 \\
  0 & \hat{W}
  \end{bmatrix}
  \]
Smoothing Norm, $\mathbf{A} \in \mathbb{R}^{N \times N}$

Standard form transformation:

$$\tilde{y}_L = \arg \min_{\tilde{y}} \| \tilde{A} \tilde{y} - \tilde{b} \|^2_2 + \lambda \| \tilde{y} \|^2_2,$$

where

$$\tilde{A} = \mathbf{A} \mathbf{L}^*_\mathbf{A} = \mathbf{A} \left[ I - (\mathbf{A}(I - \mathbf{L}^* \mathbf{L}))^* \mathbf{A} \right]$$

$$\tilde{b} = \mathbf{b} - \mathbf{A}x_0$$

$$x_L = \mathbf{L}^*_\mathbf{A} \tilde{y}_L + x_0 = \bar{x}_L + x_0$$
Smoothing Norm, \( \mathbf{A} \in \mathbb{R}^{N \times N} \)

Standard form transformation:

\[
\tilde{\mathbf{y}}_L = \arg \min_{\tilde{\mathbf{y}}} \| \tilde{\mathbf{A}} \tilde{\mathbf{y}} - \tilde{\mathbf{b}} \|_2^2 + \lambda \| \tilde{\mathbf{y}} \|_2^2,
\]

where

\[
\begin{align*}
\tilde{\mathbf{A}} &= \mathbf{A} \mathbf{L}_A^\dagger = \mathbf{A} \left[ \mathbf{I} - \left( \mathbf{A} (\mathbf{I} - \mathbf{L}_L^\dagger \mathbf{L}_L) \right)^\dagger \mathbf{A} \right] \\
\tilde{\mathbf{b}} &= \mathbf{b} - \mathbf{A} \mathbf{x}_0 \\
\mathbf{x}_L &= \mathbf{L}_A^\dagger \tilde{\mathbf{y}}_L + \mathbf{x}_0 = \tilde{\mathbf{x}}_L + \mathbf{x}_0.
\end{align*}
\]

[Hansen and Jensen. Smoothing-Norm Preconditioning for Reg. Min.-Res. SIMAX, 2007]

Write:

\[
\mathbf{x}_L = \tilde{\mathbf{x}}_L + \mathbf{x}_0 = \mathbf{L}_A^\dagger \tilde{\mathbf{y}}_L + \mathbf{x}_0 = \mathbf{L}_A^\dagger \tilde{\mathbf{y}}_L + \mathbf{K} \mathbf{t}_0,
\]

where \( \mathcal{R}(\mathbf{K}) = \mathcal{N}(\mathbf{L}) \), \( \mathbf{L}_A^\dagger \) rectangular.

Equivalently:

\[
\mathbf{A} \begin{bmatrix} \mathbf{L}_A^\dagger, \mathbf{K} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{y}}_L \\ \mathbf{t}_0 \end{bmatrix} = \mathbf{b},
\]

and, further:

\[
\begin{bmatrix} (\mathbf{L}_A^\dagger)^T \mathbf{A} \mathbf{L}_A^\dagger & (\mathbf{L}_A^\dagger)^T \mathbf{A} \mathbf{K} \\ \mathbf{K}^T \mathbf{A} \mathbf{L}_A^\dagger & \mathbf{K}^T \mathbf{A} \mathbf{K} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{y}}_L \\ \mathbf{t}_0 \end{bmatrix} = \begin{bmatrix} (\mathbf{L}_A^\dagger)^T \mathbf{b} \\ \mathbf{K}^T \mathbf{b} \end{bmatrix}.
\]
Smoothing Norm, \( A \in \mathbb{R}^{N \times N} \)

Standard form transformation:

\[
\bar{y}_L = \arg \min_{\bar{y}} \| \bar{A} \bar{y} - \bar{b} \|_2^2 + \lambda \| \bar{y} \|_2^2 , \quad \text{where} \quad \bar{A} = AL_A^\dagger = A[I - (A(I - L_L^\dagger L))L^\dagger] \quad \bar{b} = b - Ax_0 \quad x_L = L_A^\dagger \bar{y}_L + x_0 = \bar{x}_L + x_0
\]

[Hansen and Jensen. Smoothing-Norm Preconditioning for Reg. Min.-Res. SIMAX, 2007]

Write:

\[
x_L = \bar{x}_L + x_0 = L_A^\dagger \bar{y}_L + x_0 = L_A^\dagger \bar{y}_L + Kt_0 , \quad \text{where} \quad \mathcal{R}(K) = \mathcal{N}(L) , \quad L_A^\dagger \text{ rectangular} .
\]

Equivalently:

\[
A \begin{bmatrix} L_A^\dagger & K \end{bmatrix} \begin{bmatrix} \bar{y}_L \\ t_0 \end{bmatrix} = b ,
\]

and, further:

\[
\begin{bmatrix} (L_A^\dagger)^T AL_A^\dagger & (L_A^\dagger)^T AK \\ K^T AL_A^\dagger & K^T AK \end{bmatrix} \begin{bmatrix} \bar{y}_L \\ t_0 \end{bmatrix} = \begin{bmatrix} (L_A^\dagger)^T b \\ K^T b \end{bmatrix} .
\]

Schur complement system:

\[
(L_A^\dagger)^T PAL_A^\dagger \bar{y} = (L_A^\dagger)^T Pb , \quad \text{where} \quad P = I - AK(K^T AK)^{-1}K^T \in \mathbb{R}^{N \times N} .
\]
TV regularization, $\mathbf{A} \in \mathbb{R}^{N \times N}$

[G. and Sabaté Landman (2019)]

Similar idea, with reweighting...

$$(\mathbf{D}^\dagger)^T \mathbf{P} \mathbf{A} (\mathbf{W} \mathbf{D})^\dagger_A \tilde{\mathbf{y}} = (\mathbf{D}^\dagger)^T \mathbf{P} \mathbf{b}$$

Building a better approximation subspace for the solution!
TV regularization, $\mathbf{A} \in \mathbb{R}^{N \times N}$

[G. and Sabaté Landman (2019)]

Similar idea, with reweighting...

$$(\mathbf{D}^\dagger)^T \mathbf{P} \mathbf{A} (\mathbf{W} \mathbf{D})^\dagger \mathbf{A} \bar{\mathbf{y}} = (\mathbf{D}^\dagger)^T \mathbf{P} \mathbf{b}$$

Building a better approximation subspace for the solution!

- $\mathbf{L} = \mathbf{W} \mathbf{D}$ (with $\mathbf{W} = \mathbf{W}(x_k)$):
  flexible GMRES (instead of restarted GMRES);
TV regularization, $A \in \mathbb{R}^{N \times N}$

[G. and Sabaté Landman (2019)]

Similar idea, with reweighting...

$$(D^\dagger)^T PA(WD)^\dagger A\bar{y} = (D^\dagger)^T Pb$$

Building a better approximation subspace for the solution!

- $L = WD$ (with $W = W(x_k)$):
  flexible GMRES (instead of restarted GMRES);
- large-scale computations:
  - approximating $L^\dagger$
    (exploiting structure, and running preconditioned LSQR or LSMR)
  - thresholding the weights
A simple 1D example...
A simple 1D example...
A simple 1D example...
Image deblurring example with $\Psi = I$

[G., Hansen, Nagy. IR Tools (2018)]
https://github.com/silviagazzola/IRtools
http://www2.compute.dtu.dk/pcha/IRtools/

- Image is $256 \times 256$ pixels
- Noise level is $5 \times 10^{-2}$
- Reflexive boundary conditions
Reconstruction errors computed as \( \frac{\|x_k - x_{\text{true}}\|_2}{\|x_{\text{true}}\|_2} \)

- \( \lambda \) for FLSQR-I and FLSQR-R use discrepancy principle
Basis images

k=10

k=20

k=100

FLSQR-R

LSQR
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)
Tomography example with $\Phi \neq I$
Tomography example with $\Phi \neq I$

[G., Hansen, Nagy. *IR Tools* (2018)]

$n = 256; \text{optn} = \text{PRtomo('defaults')}$;
$\text{optn} = \text{PRset(optn,'angles',0:2:179,'p',\text{round}(\sqrt{2} \times n),'d',\sqrt{2} \times n);$\
$[A, b, x, \text{ProbInfo}] = \text{PRtomo}(n, \text{optn});$

- phantom is $256 \times 256$ pixels
- $A$ has size $32580 \times 65536$ (approx. 50% undersampling)
Tomography example with $\Phi \neq I$

[G., Hansen, Nagy. *IR Tools* (2018)]

\[ n = 256; \text{optn} = \text{PRtomo}(\text{‘defaults’}); \]
\[ \text{optn} = \text{PRset}(	ext{optn}, \text{‘angles’}, 0:2:179, \text{‘p’}, \text{round}(\text{sqrt}(2)*n), \text{‘d’}, \text{sqrt}(2)*n); \]
\[ [A, b, x, \text{ProbInfo}] = \text{PRtomo}(n, \text{optn}); \]
\[ \text{figure, PRshowx(x, ProbInfo)} \]

- phantom is $256 \times 256$ pixels
- $A$ has size $32580 \times 65536$ (approx. 50% undersampling)
Tomography example with $\Phi \neq I$

[G., Hansen, Nagy. *IR Tools* (2018)]

$$n = 256; \text{optn} = \text{PRtomo(‘defaults’);}$$
$$\text{optn} = \text{PRset(optn, ‘angles’, 0:2:179, ‘p’, round(sqrt(2)*n), ‘d’, sqrt(2)*n);}$$
$$[A, b, x, \text{ProbInfo}] = \text{PRtomo(n, optn);}$$
$$\text{figure, PRshowx(x, ProbInfo);}$$
$$b_n = \text{PRnoise(b, 1e-2);}$$

- phantom is $256 \times 256$ pixels
- $A$ has size $32580 \times 65536$ (approx. 50% undersampling)
Tomography example with $\Phi \neq I$

[G., Hansen, Nagy. *IR Tools* (2018)]

```matlab
n = 256; optn = PRtomo('defaults');
optn=PRset(optn,'angles',0:2:179,'p',round(sqrt(2)*n),'d',sqrt(2)*n);
[A, b, x, ProbInfo] = PRtomo(n, optn);
figure, PRshowx(x, ProbInfo)
bn = PRnoise(b, 1e-2);
```

- phantom is $256 \times 256$ pixels
- $A$ has size $32580 \times 65536$ (approx. 50% undersampling)
- $\Psi$ is a 4-level 2D Haar wavelet transform
Reconstructed phantoms

**exact**

**FLSQR-I dp**
(0.1626, # 28)

**FISTA**
(0.1722, # 150)

**SpaRSA**
(0.8829, # 150)

**IRN**
(0.2200, # 60)

**PIRN**
(0.1155, # 150)
Image deblurring example

[G., Hansen, Nagy. *IR Tools* (2018)]

Cameraman example: 256 × 256 pixels.

**blurred & noisy**

**SN-GMRES**

**TV-FGMRES**

**fast gradient-based method**
Image deblurring example

Relative Error History

Total Variation History
Tomography example with flexible TV regularization

small PRtomo example: 32 × 32 pixels; \( A \in \mathbb{R}^{2025 \times 1024} \)

... ongoing work

Exact phantom
noisy (Gaussian white noise, \( \| e \| / \| b^{\text{true}} \| = 10^{-2} \)) image
Tomography example with flexible TV regularization

small PRtomo example: $32 \times 32$ pixels; $\mathbf{A} \in \mathbb{R}^{2025 \times 1024}$

... ongoing work

LSQR
Tomography example with flexible TV regularization

small PRtomo example: $32 \times 32$ pixels; $A \in \mathbb{R}^{2025 \times 1024}$

... ongoing work

LSQR (D)
Tomography example with flexible TV regularization

small PRtomo example: $32 \times 32$ pixels; $A \in \mathbb{R}^{2025 \times 1024}$

... ongoing work

TV-LSQR
Tomography example with flexible TV regularization

small PRtomo example: $32 \times 32$ pixels; $\mathbf{A} \in \mathbb{R}^{2025 \times 1024}$

... ongoing work

TV-LSQR “0 norm”
Summary of benefits ...

- **Flexible Krylov methods**
  - ✓ Avoid inner-outer schemes (current solution immediately incorporated in basis)
  - ✓ Both square (flexible Arnoldi) non-square problems (flexible Golub-Kahan)
  - ✓ Optimality and equivalency results

- **Hybrid method**
  - ✓ Stabilize reconstruction errors
  - ✓ Automatic choice of $\lambda$ and stopping criteria

- **Transformed problem**
  - ✓ Enforce sparsity in a transform
  - ✓ Connections to multi-parameter regularization
... and (hopefully) (much) more work to do ...
... and (hopefully) (much) more work to do ...

- deeper theoretical analysis (convergence, recovery guarantees);
... and (hopefully) (much) more work to do ...

- deeper theoretical analysis (convergence, recovery guarantees);
- extension to yet other regularization terms or constraints;
... and (hopefully) (much) more work to do ...

- deeper theoretical analysis (convergence, recovery guarantees);
- extension to yet other regularization terms or constraints;
- parameter choice for nonlinear problems and solvers;
... and (hopefully) (much) more work to do ...

- deeper theoretical analysis (convergence, recovery guarantees);
- extension to yet other regularization terms or constraints;
- parameter choice for nonlinear problems and solvers;

References

- J. Chung and S. G. Flexible Krylov methods for $\ell^p$ regularization. SISC, 2019.
- S. G. and M. Sabaté Landman. Flexible GMRES for Total Variation regularization. BIT, 2019.
- S. G., P. C. Hansen, and J. Nagy. IR Tools: A MATLAB Package of Iterative Regularization Methods and Large-Scale Test Problems. Numer. Algorithms, 2018. https://github.com/silviagazzola/IRtools