Advancements in Hybrid Iterative Methods for Inverse Problems

Julianne Chung

Department of Mathematics, Virginia Tech
Computational Modeling and Data Analytics Division, Academy of Integrated Science

January 10, 2020 @ University of Potsdam

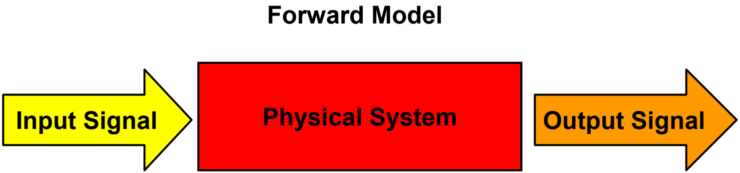
Unterstützt von / Supported by

DMS-1654175, CAREER
DMS-1723005, CDS&E-MSS

Overview

Outline

What is an inverse problem?



What is an inverse problem?

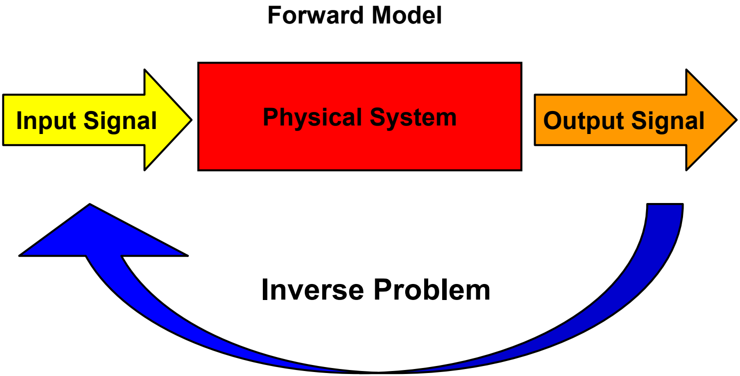
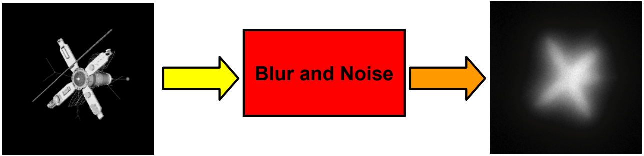


Image deblurring



- Given observed image, $\mathbf{d}$, and some information about the blur, $\mathbf{A}$
- Goal is to compute approximation of true image, $\mathbf{s}_{\text{true}}$

$$\mathbf{d} = \mathbf{A}\mathbf{s}_{\text{true}} + \boldsymbol{\epsilon}$$

Hansen, Nagy, and O'Leary, *Deblurring Images: Matrices, Spectra and Filtering*, SIAM (2006)

Gonzalez and Wintz, *Digital Image Processing*, Addison-Wesley (1977)

Time-lapse Photoacoustic Tomography (PAT)

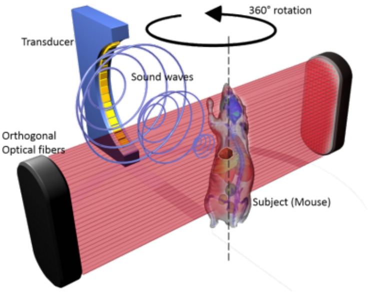
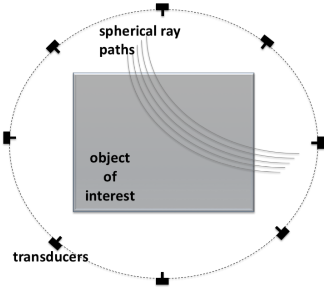


Image courtesy of Tomowave.

- non-invasive, non-ionizing imaging modality
- rich contrast of optical imaging with high resolution of ultrasound imaging

- Given: spherical projections
- Goal: estimate initial pressures



e.g., Oraevsky and Karabutov (2003), Katsevich (2010), Wang and Anastasio (2011), Wang and Wu (2012), Hahn (2014), Bal and Moradifam (2016), Lou et al (2016), Chung and Nguyen (2017), ...

Dynamic PAT

Simultaneous approach

$$\underbrace{\begin{bmatrix} \mathbf{d}_1 \\ \vdots \\ \mathbf{d}_{n_t} \end{bmatrix}}_{\mathbf{d}} = \underbrace{\begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_{n_t} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_{n_t} \end{bmatrix}}_{\mathbf{s}_{\text{true}}} + \underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_{n_t} \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

where

- $\mathbf{s}_i \in \mathbb{R}^{65,536}$ and $n_t = 120$ $\rightarrow 7,864,320$ unknowns
- $\mathbf{A}_i \in \mathbb{R}^{363 \times 65,536}$ represents discrete circular Radon transform
- $\mathbf{d}_i \in \mathbb{R}^{363} \rightarrow 43,560$ observations

Linear problem

$$\begin{aligned} \mathbf{d} &= \mathbf{A}\mathbf{s}_{\text{true}} + \boldsymbol{\epsilon} \\ &= \mathbf{d}_{\text{true}} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \end{aligned}$$

where

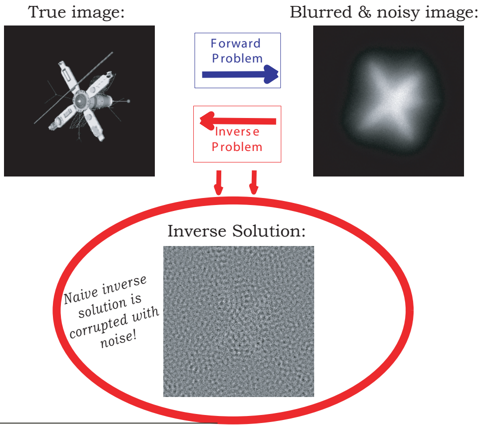
$\mathbf{d}$	:	observations or measurements
$\mathbf{s}_{\text{true}}$	:	desired parameters
$\mathbf{A}$	:	ill-conditioned matrix models forward process
$\boldsymbol{\epsilon}$	:	additive Gaussian noise, $\mathbf{R}$ diagonal

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

Ill-posed problem

A problem is *ill-posed* if the solution

- does not exist,
- is not unique, or
- does not depend continuously on the data.



Hadamard (1902), Tikhonov and Arsenin (1977)

Outline

Introduction

Hybrid projection methods for Tikhonov

Generalized hybrid methods with Arvind Saibaba (NCSU)

Flexible hybrid methods with Silvia Gazzola (Bath)

Conclusions

Tikhonov regularization

$$\min_{\mathbf{s}} \left\{ ||\mathbf{A}\mathbf{s} - \mathbf{d}||_2^2 + \lambda^2 ||\mathbf{s}||_2^2 \right\}$$

Let $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ be the singular value decomposition (SVD) of $\mathbf{A}$,

- $\boldsymbol{\Sigma} =$diag$(\sigma_1, \sigma_2, \ldots, \sigma_n)$, $\quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$
- $\mathbf{U}^\top\mathbf{U} = \mathbf{I}, \quad \mathbf{V}^\top\mathbf{V} = \mathbf{I}$

Tikhonov solution:

$$\mathbf{s}_\lambda = \sum_{i=1}^{n} \phi_i \frac{\mathbf{u}_i^\top \mathbf{d}}{\sigma_i} \mathbf{v}_i \equiv \mathbf{A}_\lambda^\dagger \mathbf{d}$$

with filter factors $\phi_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2}$

Choosing regularization parameter λ

- Discrepancy principle: $||(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger)\mathbf{d}||_2 < \delta$
- Generalized cross validation - Golub, Heath and Wahba (1979)

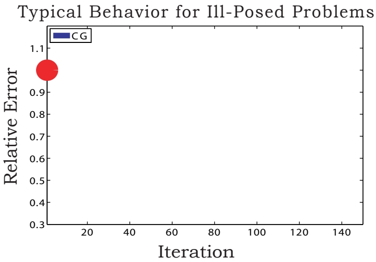
$$G_{\mathbf{A},\mathbf{d}}(\lambda) = \frac{n||(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger)\mathbf{d}||_2^2}{\left[\text{trace}(\mathbf{I} - \mathbf{A}\mathbf{A}_\lambda^\dagger)\right]^2}$$

- Unbiased predictive risk estimator (UPRE) - Mallow (1973), Giryes, Elad, Eldar (2011)

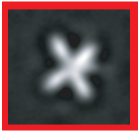
$$U_{\mathbf{A},\mathbf{d}}(\lambda) = \frac{1}{n}\|\mathbf{d} - \mathbf{A}\mathbf{s}_\lambda\|_2^2 + \frac{2\sigma^2}{n}\text{trace}(\mathbf{A}\mathbf{A}_\lambda^\dagger) - \sigma^2 .$$

Iterative methods: semi-convergence

$$\min_{\mathbf{s}} \|\mathbf{A}\mathbf{s} - \mathbf{d}\|_2$$

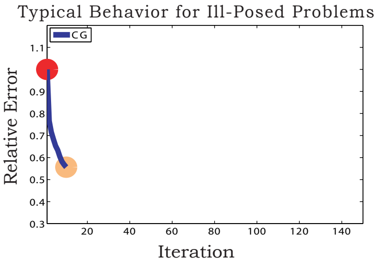


Iteration 0



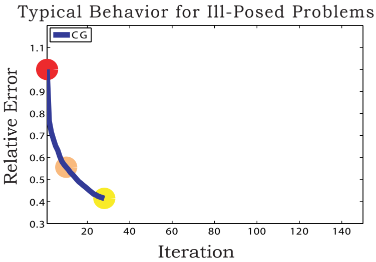
Iterative methods: semi-convergence

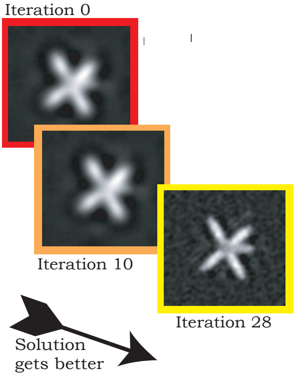
$$\min_{\mathbf{s}} \|\mathbf{A}\mathbf{s} - \mathbf{d}\|_2$$



Iterative methods: semi-convergence

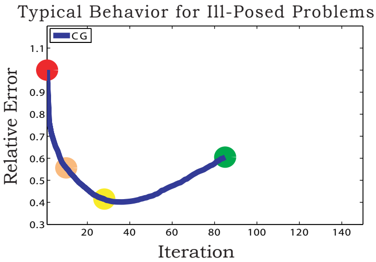
$$\min_{\mathbf{s}} \|\mathbf{A}\mathbf{s} - \mathbf{d}\|_2$$

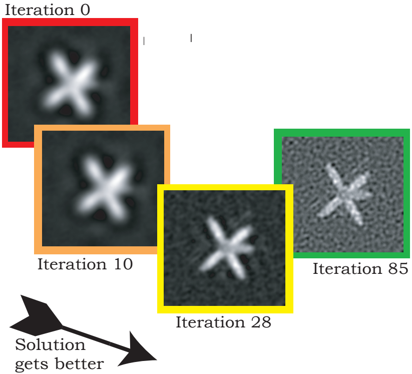




Iterative methods: semi-convergence

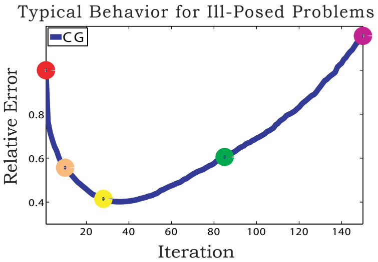
$$\min_{\mathbf{s}} \|\mathbf{A}\mathbf{s} - \mathbf{d}\|_2$$

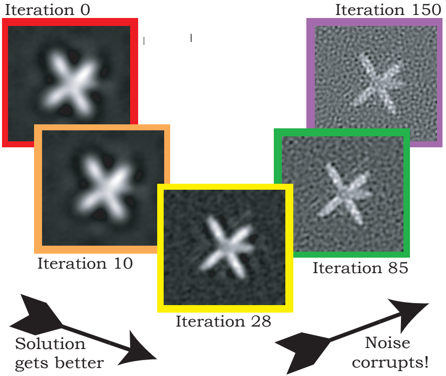




Iterative methods: semi-convergence

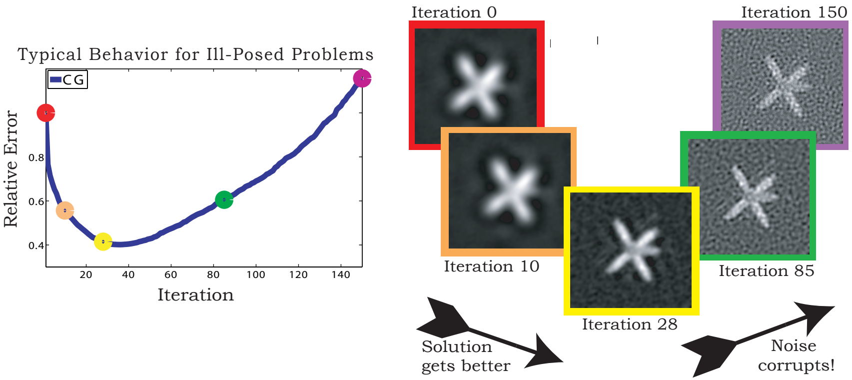
$$\min_{\mathbf{s}} \|\mathbf{A}\mathbf{s} - \mathbf{d}\|_2$$





Iterative methods: semi-convergence

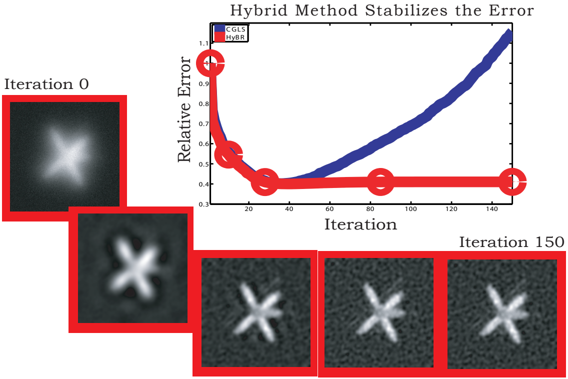
$$\min_{\mathbf{s}} \|\mathbf{A}\mathbf{s} - \mathbf{d}\|_2$$



Either find a good stopping criteria or ...

Motivation for hybrid methods

... avoid semi-convergence behavior altogether!



Hybrid projection method for standard Tikhonov

Basic Idea: Combine iterative method with variational regularization

Input: $\mathbf{A}$, $\mathbf{d}$, $k = 1$

1: **while** stopping criteria not satisfied **do**
2: Expand the projection subspace $\mathbf{V}_k$ $\longleftarrow$ *projection method*
3: Select λ_k $\longleftarrow$ *parameter selection method*
4: Solve the regularized, projected problem

$$\mathbf{s}_k(\lambda_k) = \operatorname*{arg\,min}_{\mathbf{s} \in \mathcal{R}(\mathbf{V}_k)} \|\mathbf{d} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda_k^2 \|\mathbf{s}\|_2^2$$

5: $k = k + 1$
6: **end while**

Historical overview

1981-1990

- seminal publication by O'Leary and Simmons (1981)
- independently by Björck (1988)

1991-2000

- regularizing properties of Krylov methods, stopping criteria, reorthogonalization, nonlinear inversion, extensions to GMRES and Arnoldi, regularization parameter selection
- Nemirovskii, Hanke, Hansen, Bjorck, Grimme, Van Dooren, Calvetti, Golub, Reichel, Haber, Von Matt, Frommer, Maass, ...

2001-2010

- computational software, regularization parameter selection, general regularization terms, noise estimation, HPC, new applications
- Hanke, Kilmer, O'Leary, Nagy, Chung, Mead, Renaut, Borges, Hostenbach, Español, Hnětynková, Plešinger, Strakos ...

2010-2020

- Arnoldi-Tikhonov, general regularization terms and constraints, flexible methods, uncertainty quantification,...

Outline

Introduction

Hybrid projection methods for Tikhonov

Generalized hybrid methods with Arvind Saibaba (NCSU)

Flexible hybrid methods with Silvia Gazzola (Bath)

Conclusions

Bayesian approach

Assume

$$\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, \lambda^{-2}\mathbf{Q})$$

Using Bayes' rule, the posterior distribution

$$\begin{aligned} \pi(\mathbf{s}|\mathbf{d}) &\propto \pi(\mathbf{d}|\mathbf{s})\pi(\mathbf{s}) \\ &\propto \exp\left(-\frac{1}{2}\|\mathbf{A}\mathbf{s} - \mathbf{d}\|^2_{\mathbf{R}^{-1}} - \frac{\lambda^2}{2}\|\mathbf{s} - \boldsymbol{\mu}\|^2_{\mathbf{Q}^{-1}}\right) \end{aligned}$$

where $\|\mathbf{x}\|_{\mathbf{M}} = \sqrt{\mathbf{x}^\top \mathbf{M}\mathbf{x}}$

MAP Estimate

$$\begin{aligned} \mathbf{s}_\lambda &= \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{A}\mathbf{s} - \mathbf{d}\|^2_{\mathbf{R}^{-1}} + \frac{\lambda^2}{2}\|\mathbf{s} - \boldsymbol{\mu}\|^2_{\mathbf{Q}^{-1}} \\ &= \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{L}_\mathbf{R}(\mathbf{A}\mathbf{s} - \mathbf{d})\|^2_2 + \frac{\lambda^2}{2}\|\mathbf{L}_\mathbf{Q}(\mathbf{s} - \boldsymbol{\mu})\|^2_2 \end{aligned}$$

where $\mathbf{Q}^{-1} = \mathbf{L}_\mathbf{Q}^\top \mathbf{L}_\mathbf{Q}$ and $\mathbf{R}^{-1} = \mathbf{L}_\mathbf{R}^\top \mathbf{L}_\mathbf{R}$

Overview

Efficient generalized Golub-Kahan methods for computing

- **MAP estimate:**

$$\mathbf{s}_\lambda = \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{A}\mathbf{s} - \mathbf{d}\|^2_{\mathbf{R}^{-1}} + \frac{\lambda^2}{2}\|\mathbf{s} - \boldsymbol{\mu}\|^2_{\mathbf{Q}^{-1}}$$

- **Uncertainty estimates:**

$$\mathbf{s}|\mathbf{d} \sim \mathcal{N}(\boldsymbol{\Gamma}_{\text{post}}\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b}, \boldsymbol{\Gamma}_{\text{post}}), \quad \boldsymbol{\Gamma}_{\text{post}} = (\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A} + \lambda^2\mathbf{Q}^{-1})^{-1}$$

Benefits:

- *Flexible*: Matérn class of covariance kernels
- *Efficient*: avoid $\mathbf{Q}^{-1}, \mathbf{L}_\mathbf{Q}$, and $\mathbf{L}_\mathbf{Q}^{-1}$
- *Automatic*: choice of λ and stopping criteria
- *Equivalent*: “project-then-regularize” vs. “regularize-then-project” on prior-conditioned problem

Matérn covariance family

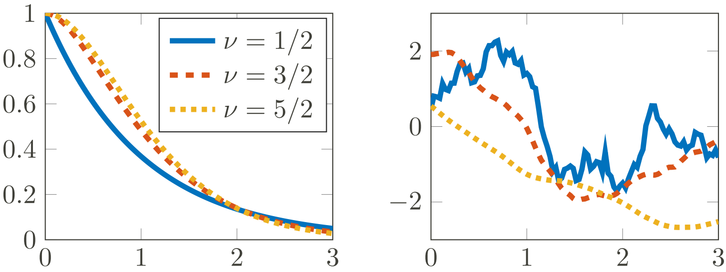
Gaussian prior: $\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, \lambda^{-2}\mathbf{Q})$

$$\mathbf{Q}_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j), \qquad \mathbf{x}_i \in \mathbb{R}^d$$

Matérn class of covariance kernels (isotropic):

$$\kappa(r) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\sqrt{2\nu}\alpha r\right)^{\nu} K_\nu\left(\sqrt{2\nu}\alpha r\right), \qquad r = \|\mathbf{x}_i - \mathbf{x}_j\|_2$$

Examples: Exponential kernel ($\nu = 1/2$), Gaussian kernel ($\nu = \infty$)



Fast covariance evaluations[1]

$$\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, \lambda^{-2}\mathbf{Q})$$

- Covariance matrices are dense - expensive to store and compute
- e.g., a dense $10^6 \times 10^6$ matrix requires 7.45 TB in storage

Available approaches for evaluating $\mathbf{Qx}$

- FFT based methods
- Hierarchical Matrices

Compared to the naive $\mathcal{O}(n^2)$

Storage cost: $\mathcal{O}(n \log n)$ Matvec cost: $\mathcal{O}(n \log n)$

[1] Saibaba et al. (2012)., Ambikasaran et al. (2013), Nowak et al (2003)

Computing MAP estimate

Normal equations

$$(\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{A} + \lambda^2 \mathbf{Q}^{-1})\mathbf{s} = \mathbf{A}^\top \mathbf{R}^{-1}\mathbf{d} + \lambda^2 \mathbf{Q}^{-1}\boldsymbol{\mu}$$

We make a change of variables[2]

$$\begin{aligned} \mathbf{x} &\leftarrow \mathbf{Q}^{-1}(\mathbf{s} - \boldsymbol{\mu}) \\ \mathbf{b} &\leftarrow \mathbf{d} - \mathbf{A}\boldsymbol{\mu} \end{aligned}$$

and get

$$(\mathbf{A}^\top \mathbf{R}^{-1}\mathbf{A}\mathbf{Q} + \lambda^2 \mathbf{I})\mathbf{x} = \mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b}$$

Equivalent transformed problem

MAP estimate:

$$\mathbf{s}_\lambda = \boldsymbol{\mu} + \mathbf{Q}\mathbf{x}_\lambda \quad \text{where}$$

$$\mathbf{x}_\lambda = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{Q}\mathbf{x} - \mathbf{b}\|^2_{\mathbf{R}^{-1}} + \frac{\lambda^2}{2}\|\mathbf{x}\|^2_{\mathbf{Q}} \tag{1}$$

[2] Similar ideas in Calvetti and Somersalo (2005), Calvetti (2007), Arridge, Betcke, and Harhanen (2014).

Generalized Golub-Kahan (gen-GK)[3]

Given $\mathbf{A}, \mathbf{b}, \mathbf{R}, \mathbf{Q}$, initialize $\beta_1 = \|\mathbf{b}\|_{\mathbf{R}^{-1}}$, then at the kth iteration,

$$\begin{aligned} \mathbf{U}_{k+1}\beta_1\mathbf{e}_1 &= \mathbf{b} \\ \mathbf{A}\mathbf{Q}\mathbf{V}_k &= \mathbf{U}_{k+1}\mathbf{B}_k \\ \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{U}_{k+1} &= \mathbf{V}_k\mathbf{B}_k^\top + \alpha_{k+1}\mathbf{v}_{k+1}\mathbf{e}_{k+1}^\top \end{aligned}$$

with

$$\mathbf{U}_{k+1}^\top\mathbf{R}^{-1}\mathbf{U}_{k+1} = \mathbf{I}_{k+1} \qquad \mathbf{V}_k^\top\mathbf{Q}\mathbf{V}_k = \mathbf{I}_k$$

Krylov subspace

$$\mathcal{S}_k \equiv \text{Span}\{\mathbf{V}_k\} = \mathcal{K}_k(\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A}\mathbf{Q}, \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b}),$$

where

$$\mathcal{K}_k(\mathbf{C}, \mathbf{g}) \equiv \text{Span}\{\mathbf{g}, \mathbf{C}\mathbf{g}, \ldots, \ldots, \mathbf{C}^{k-1}\mathbf{g}\}.$$

[3] Similar to Benbow (1999), Arioli (2013), Arioli & Orban (2013).

Generalized LSQR to solve (1)

Basic idea: Search for solutions $\mathbf{x}_k = \mathbf{V}_k \mathbf{z}_k \in \mathcal{S}_k$

gen-LSQR subproblem

$$\min_{\mathbf{x}_k \in \mathcal{S}_k} \frac{1}{2} \|\mathbf{A}\mathbf{Q}\mathbf{x}_k - \mathbf{b}\|^2_{\mathbf{R}^{-1}} + \frac{\lambda^2}{2} \|\mathbf{x}_k\|^2_{\mathbf{Q}}$$

$$\Updownarrow$$

$$\min_{\mathbf{z}_k \in \mathbb{R}^k} \frac{1}{2} \|\mathbf{B}_k \mathbf{z}_k - \beta_1 \mathbf{e}_1\|_2^2 + \frac{\lambda^2}{2} \|\mathbf{z}_k\|_2^2$$

Observations:

- Efficient regularization parameter selection for projected problem
- Singular values of $\mathbf{B}_k$ approximate singular values of

$$\widehat{\mathbf{A}} \equiv \mathbf{L}_{\mathbf{R}} \mathbf{A} \mathbf{L}_{\mathbf{Q}}^{-1}$$

Generalized hybrid (gen-HyBR) method

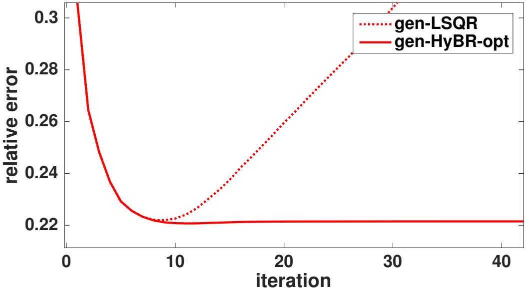
- Use gen-GK to project problem
- Choose λ with standard methods [4]

$$\min_{\mathbf{z}_k \in \mathbb{R}^k} \frac{1}{2}\|\mathbf{B}_k\mathbf{z}_k - \beta_1\mathbf{e}_1\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{z}_k\|_2^2$$

- Undo change of variables

$$\hat{\mathbf{s}}_k = \boldsymbol{\mu} + \mathbf{Q}\mathbf{x}_k = \boldsymbol{\mu} + \mathbf{Q}\mathbf{V}_k\mathbf{z}_k$$

- Overcome semi-convergence and automatic stopping criteria



[4] Discrepancy principle (DP), Generalized cross-validation (GCV), Unbiased predictive-risk estimate (UPRE), etc.

Space-time covariance kernels

Assume Gaussian prior: $\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, \lambda^{-2}\mathbf{Q})$

$$\mathbf{Q}_{ij} = C(\mathbf{p}_1 - \mathbf{p}_2, t_1 - t_2), \qquad \forall(\mathbf{p}, t) \in \mathbb{R}^d \times \mathbb{R}$$

- Separable covariance function

$$C(\mathbf{p}, t) = C_S(\|\mathbf{p}\|)C_T(|t|) \qquad \Longrightarrow \qquad \mathbf{Q} = \mathbf{Q}_t \otimes \mathbf{Q}_s$$

e.g., random walk forecast model[5] or differential operator [6]:

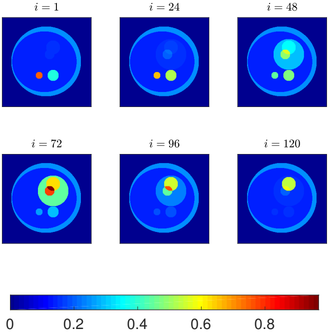
$$\mathbf{Q} = \underbrace{\left(\mathbf{I} + \frac{\lambda_t^2}{\lambda_s^2}\mathbf{L}_t^\top \mathbf{L}_t\right)^{-1}}_{\mathbf{Q}_t} \otimes \underbrace{\lambda_s^{-2}\mathbf{I}}_{\mathbf{Q}_s}$$

- Non-separable: $C(\mathbf{p}, t) = \varphi(\sqrt{c_1\|\mathbf{p}\|^2 + c_2|t|^2})$

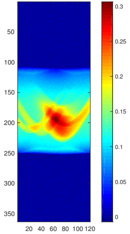
[5] Vauhkonen et al. (1998), Kim et al. (2001), Soleimani et al (2007), Neena (2011)

[6] Schmitt and Louis (2002), Schmitt, Louis, Wolters, and Vauhkonen (2002)

Dynamic PAT

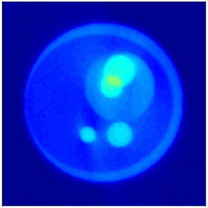


Sample true images



Sinogram

$(\mathbf{R} = 0.0082^2\mathbf{I})$

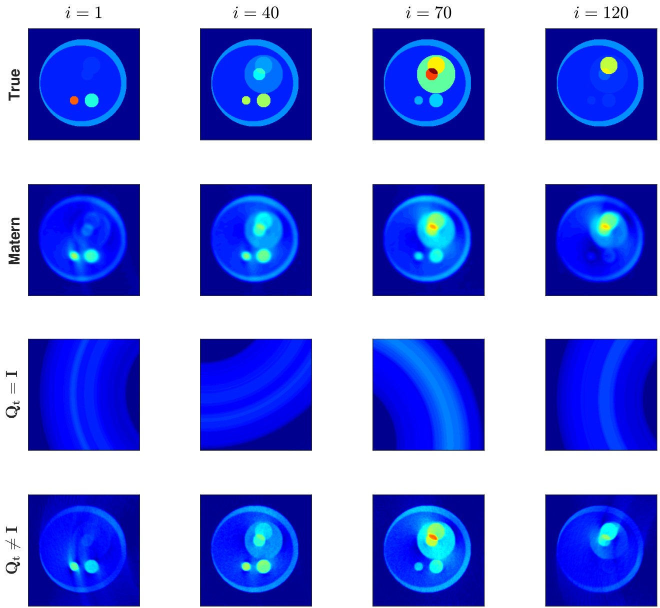


Static reconstruction

genHyBR results where

1. $\mathbf{Q}$ is generated from a Matérn kernel: $C_{1,.01}(\sqrt{\|\mathbf{p}\|^2 + 0.0025|t|^2})$
2. $\mathbf{Q} = \mathbf{I} \otimes \mathbf{Q}_s$ where $\mathbf{Q}_s$ corresponds to $C_S(\cdot) = C_{1,.01}(\cdot)$
3. $\mathbf{Q} = \mathbf{Q}_t \otimes \mathbf{Q}_s$ where $C_T(\cdot) = C_{\infty,.01}(\cdot)$ and $C_S(\cdot) = C_{1,.01}(\cdot)$

Dynamic PAT


$i = 1$
$i = 40$
$i = 70$
$i = 120$
True
Matern
$\mathbf{Q}_t = \mathbf{I}$
$\mathbf{Q}_t \neq \mathbf{I}$

Outline

ℓ_p-regularized problem

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

where $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{\Psi} \in \mathbb{R}^{n \times n}$ invertible, $p \geq 1$

Interpretations (for $p = 1$):

- Bayesian perspective: maximum a posteriori 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

There exists a zoo of methods:

sub-gradient strategies, constrained optimization, iterative shrinkage-threshholding algorithms, differential approximations, ...

Iterative Reweighted 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 $\mathbf{L}(\mathbf{x}) = \mathrm{diag}\left(1/\sqrt{|[\mathbf{x}]_i|}\right)$ and $[\mathbf{x}]_i$ is the ith element of $\mathbf{x}$

Input: $\mathbf{x}_0, \lambda$

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**

Algorithm 1: IRN for ℓ_1-regularization

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

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

Flexible Arnoldi Tikhonov (Flexi-AT)[7]

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
- $\mathbf{H}_k \in \mathbb{R}^{(k+1) \times k}$ is upper Hessenberg

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

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

Benefits of flexible hybrid approach:

- *Automatic*: choice of λ and stopping criteria
- *Efficient*: current solution immediately incorporated

[7] Gazzola and Nagy (2014)

Flexible Golub-Kahan Process [8]

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

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

such that

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

Remarks:

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

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

Flexible GK Tikhonov

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

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

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

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

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{T}_k \mathbf{y} - \beta_1 \mathbf{e}_1\|_2 \qquad (\lambda = 0)$$

   - Flexible GK Tikhonov - R (FLSQR-R)

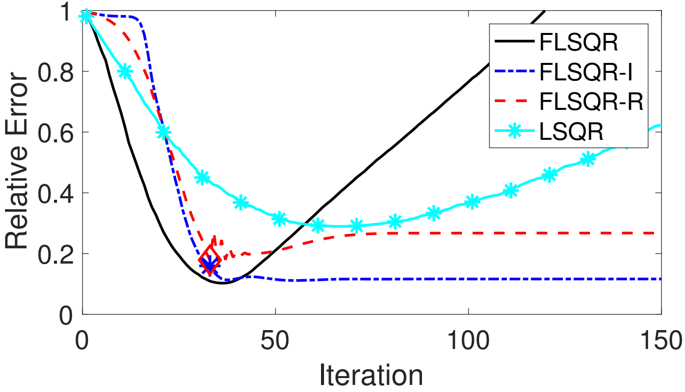
$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{T}_k \mathbf{y} - \beta_1 \mathbf{e}_1\|_2 + \lambda \|\mathbf{R}_k \mathbf{y}\|_2^2 \,, \quad \mathbf{Z}_k = \mathbf{Q}\mathbf{R}_k$$

   - Flexible GK Tikhonov - I (FLSQR-I)

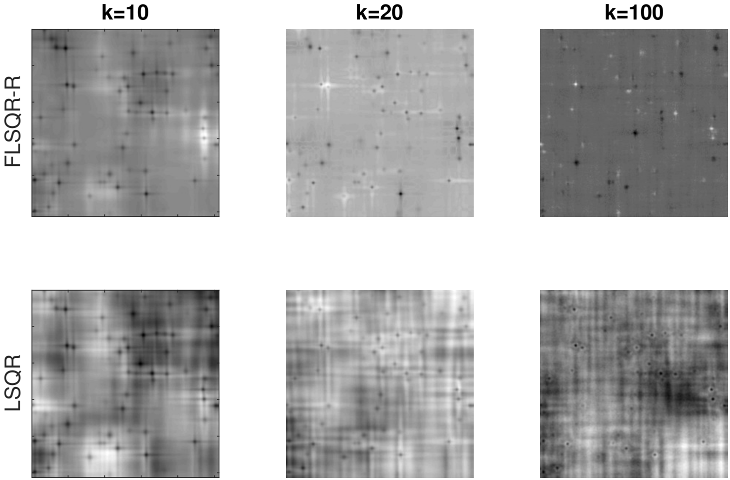
$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{T}_k \mathbf{y} - \beta_1 \mathbf{e}_1\|_2 + \lambda \|\mathbf{y}\|_2^2$$

Image deblurring example

- Noise level: $5e-2$
- Relative error: $\|\mathbf{x}_k - \mathbf{x}_{\text{true}}\|_2 \,/\, \|\mathbf{x}_{\text{true}}\|_2$



Basis Images



Outline

Conclusions

- Hybrid methods are very nice...
  - **Automatic**: Hybrid methods allow automatic regularization parameter selection and stopping criteria
  - **Robust**: Equivalence results

- **Generalized GK methods**
  - **Flexible**: Matérn covariance family offers a rich class of priors, black box use
  - **Efficient**: Avoid $\mathbf{Q}^{-1}$ and $\mathbf{L}_{\mathbf{Q}}$

- **Flexible GK methods**
  - **Efficient**: Solve ℓ_p-regularized problems
  - **Extensions**: Solve non-square problems, can incorporate multilevel decompositions

References

- Chung and Saibaba. "Generalized Hybrid Iterative Methods for Large-Scale Bayesian Inverse Problems." SISC 39(5), 2017.
- Chung, Saibaba, Brown, and Westman. "Efficient generalized Golub-Kahan based methods for dynamic inverse problems." Inverse Problems, 2018.
- Saibaba, Chung, and Petroske. "Efficient Krylov subspace methods for uncertainty quantification in large Bayesian linear inverse problems." arXiv, 2019.

- Gazzola and Nagy. "Generalized Arnoldi-Tikhonov Method for Sparse Reconstruction." SISC 36(2), 2014.
- Chung and Gazzola. "Flexible Krylov Methods for ℓ_p Regularization." SISC 41(5), 2019.

Thank you for your attention!