Generalized Hybrid Iterative Methods for Large-Scale Bayesian Inverse Problems

Julianne Chung[1] and Arvind Saibaba[2]

[1]Department of Mathematics, Virginia Tech
Computational Modeling and Data Analytics Division, Academy of Integrated Science
[2]Department of Mathematics, North Carolina State University

February 27, 2017 @ SIAM CSE

Overview

Bayesian inverse problems
- Modeling Gaussian priors
- Computing MAP estimates

Generalized Golub-Kahan hybrid approach
- Some theory and connections

Numerical results

Conclusions

Outline

Bayesian inverse problems
- Modeling Gaussian priors
- Computing MAP estimates

Generalized Golub-Kahan hybrid approach
- Some theory and connections

Numerical results

Conclusions

Linear inverse problem

$$\mathbf{d} = \mathbf{A}\mathbf{s} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

where

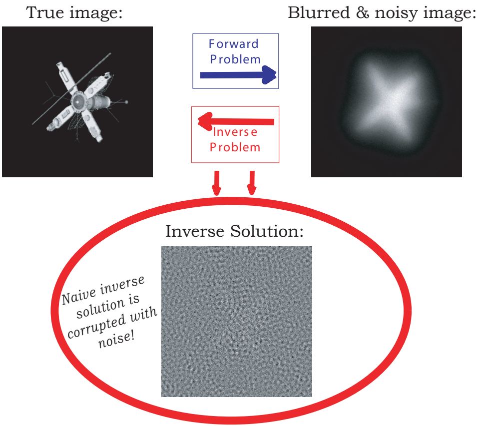
$\mathbf{d}$	:	observations or measurements
$\mathbf{s}$	:	desired parameters
$\mathbf{A}$	:	ill-conditioned matrix models forward process
$\boldsymbol{\epsilon}$	:	additive Gaussian noise, $\mathbf{R}$ diagonal

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

Ill-posed inverse problems

A problem is *ill-posed* if the solution

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



Bayesian approach

Assume

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

Using Bayes' rule, the posterior distribution

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

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

MAP Estimate

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

Overview of contributions

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

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

Matérn covariance family

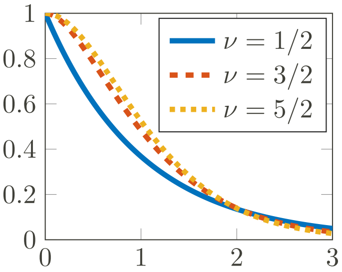
Gaussian prior: $\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, \lambda^{-2}\mathbf{Q})$

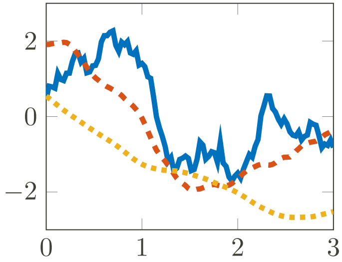
$$\mathbf{Q}_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j), \qquad \mathbf{x}_i \in \mathbb{R}^d$$

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 $\mathbf{Q}\mathbf{x}$

- FFT based methods
- Hierarchical Matrices, a.k.a. fast multipole method

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 estimates

Normal equations

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

Transformation to standard form / priorconditioning [2]

$$\mathbf{x} \leftarrow \mathbf{L}_{\mathbf{Q}}(\mathbf{s} - \boldsymbol{\mu}), \qquad \mathbf{b} \leftarrow \mathbf{d} - \mathbf{A}\boldsymbol{\mu}$$

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{L}_{\mathbf{R}}(\mathbf{A}\mathbf{L}_{\mathbf{Q}}^{-1}\mathbf{x} - \mathbf{b})\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{x}\|_2^2$$

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

Change of variables

We make a change of variables

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

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:

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

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

Outline

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

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

$$\Updownarrow$$

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

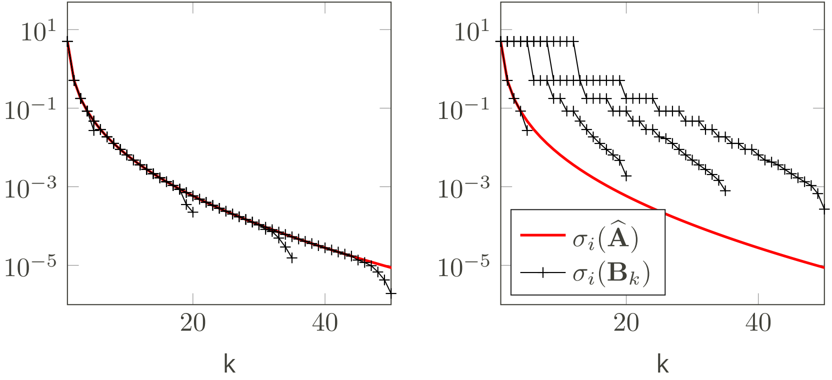
Observations:

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

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

Illustration

$\sigma_i(\widehat{\mathbf{A}})$ - singular values of $\widehat{\mathbf{A}} = \mathbf{L_R}\mathbf{A}\mathbf{L_Q}^{-1}$
$\sigma_i(\mathbf{B}_k)$ - singular values of $\mathbf{B}_k$



With reorthogonalization | Without reorthogonalization

Generalized hybrid (gen-HyBR) method

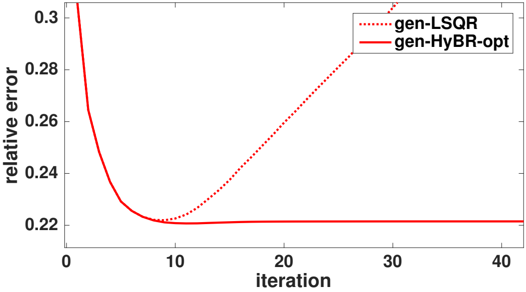
- Use gen-GK to project problem
- Choose λ with standard methods [4]

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

- 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.

Equivalence results

gen-LSQR iterates

- equivalent to k CG iterations with $\langle , \rangle_{\mathbf{Q}}$ inner products
- can be interpreted as filtered GSVD solutions

Theorem

Fix $\lambda \geq 0$ and let $\mathbf{z}_k = \arg\min_{\mathbf{z}} \|\mathbf{B}_k\mathbf{z} - \beta_1\mathbf{e}_1\|_2^2 + \lambda^2\|\mathbf{z}\|_2^2$. Then the k-th iterate of gen-HyBR,

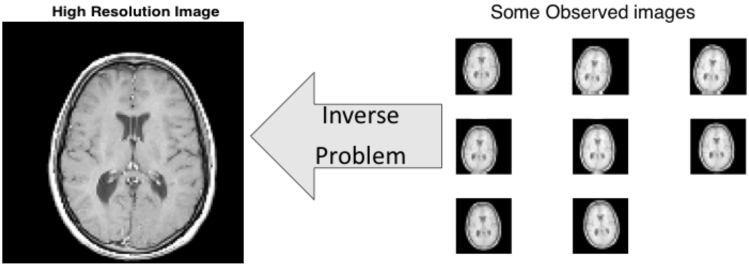
$$\boldsymbol{\mu} + \mathbf{Q}\mathbf{V}_k\mathbf{z}_k$$

is equivalent to $\boldsymbol{\mu} + \mathbf{L}_{\mathbf{Q}}^{-1}\mathbf{w}_k$, where $\mathbf{w}_k$ is the k-th iterate of LSQR on priorconditioned Tikhonov problem

$$\min_{\mathbf{w}} \left\| \begin{pmatrix} \mathbf{L}_{\mathbf{R}}\mathbf{A}\mathbf{L}_{\mathbf{Q}}^{-1} \\ \lambda\mathbf{I} \end{pmatrix} \mathbf{w} - \begin{pmatrix} \mathbf{L}_{\mathbf{R}}\mathbf{b} \\ \mathbf{0} \end{pmatrix} \right\|_2^2$$

Outline

Application: Super-resolution imaging

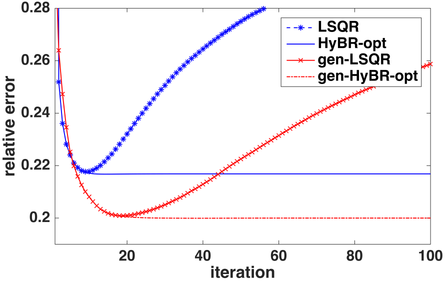


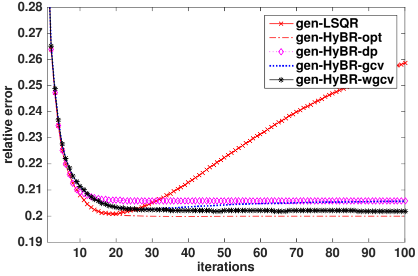
- Given low-resolution images, approximate high-resolution image

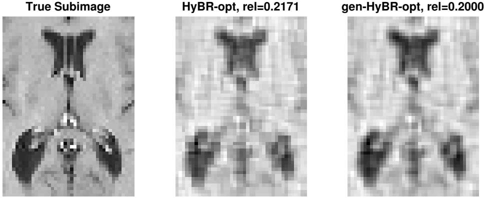
$$\underbrace{\begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_K \end{bmatrix}}_{\mathbf{b}} = \underbrace{\begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_K \end{bmatrix}}_{\mathbf{A}} \mathbf{s} + \underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_K \end{bmatrix}}_{\boldsymbol{\epsilon}}$$

- $\mathbf{R} = \mathbf{I}$, $\boldsymbol{\mu} = \mathbf{0}$, 2% noise
- LSQR and HyBR-opt use $\mathbf{Q} = \mathbf{I}$
- gen-LSQR and gen-HyBR-opt use $\mathbf{Q}$ with $\nu = 1/2$, $\alpha = 7 \times 10^{-3}$

Numerical Results

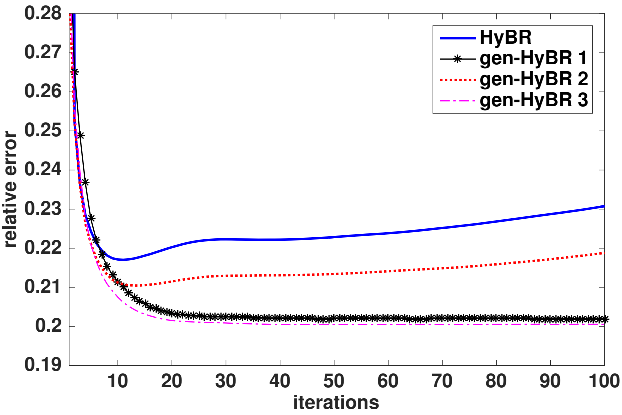


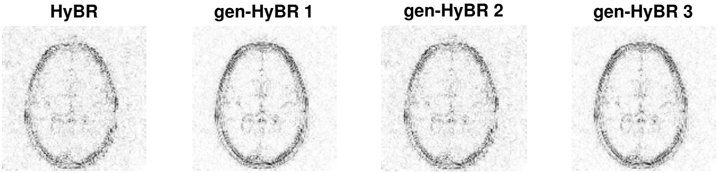




Different covariance kernels

- gen-HyBR-1:
  $\nu = 1/2$
  $\alpha = 7 \times 10^{-3}$
- gen-HyBr-2:
  $\nu = 1/2$
  $\alpha = 3 \times 10^{-3}$
- gen-HyBr-3:
  $\nu = \infty$
  $\alpha = 7 \times 10^{-3}$

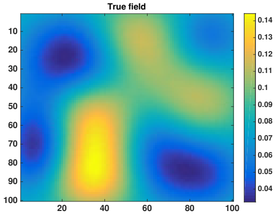




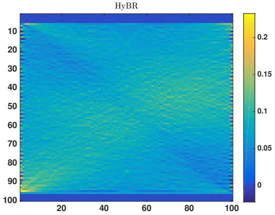
Application: Seismic tomography

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

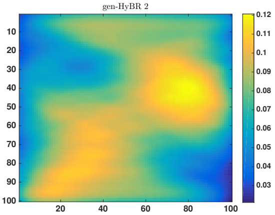
- Given observed travel times, $\mathbf{d}$
- Approximate slowness in the medium, $\mathbf{s}$



True image



HyBR, optimal $\lambda^2 = 4.71 \times 10^{-1}$



gen-HyBR ($\nu = 3/2$), optimal $\lambda^2 = 2.32 \times 10^{-7}$

Outline

Conclusions and outlook

- Developed hybrid method for large Bayesian inverse problems[5]
  - **Flexible**: Matérn covariance family offers a rich class of priors, black box use
  - **Efficient**: Avoid $\mathbf{Q}^{-1}$ and $\mathbf{L}_\mathbf{Q}$
  - **Automatic**: Hybrid methods allow automatic regularization parameter selection and stopping criteria
  - **Robust**: Equivalence results

- Ongoing work
  - Use gen-GK bidiagonalization to compute uncertainty estimates (e.g., posterior variances and samples)
  - Exploit separability in covariance matrix $\mathbf{Q}$

Thank you!

[5] Chung and Saibaba. "Generalized Hybrid Iterative Methods for Large-Scale Bayesian Inverse Problems." To appear in SISC, 2017.