# Efficient Krylov subspace methods for large-scale hierarchical Bayesian inverse problems

Elle Buser and Julianne Chung

May 28, 2024 @ SIAM IS24

Department of Mathematics
Emory University
ebuser@emory.edu

# Problem setup

Consider the inverse problem

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

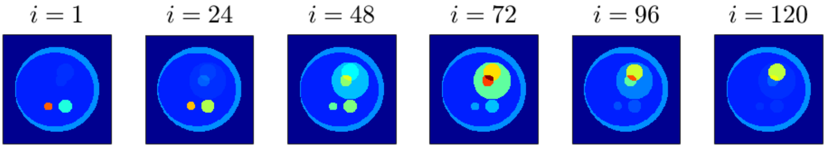
where

- $\mathbf{A} \in \mathbb{R}^{m \times n}$ represents the linear forward model
- $\mathbf{b} \in \mathbb{R}^{m}$ contains observations
- $\mathbf{x}_{\text{true}} \in \mathbb{R}^{n}$ is the true solution
- $\mathbf{e} \in \mathbb{R}^{m}$ is noise

**Complications:**

- underdetermined and ill-posed
- for large-scale problems, need efficient methods to
    - approximate solution
    - estimate uncertainties

# Dynamic tomography reconstruction




$$\underbrace{\begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_{n_t} \end{bmatrix}}_{\mathbf{b}} = \underbrace{\begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_{n_t} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{n_t} \end{bmatrix}}_{\mathbf{x}_{\text{true}}} + \underbrace{\begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_{n_t} \end{bmatrix}}_{\mathbf{e}}$$

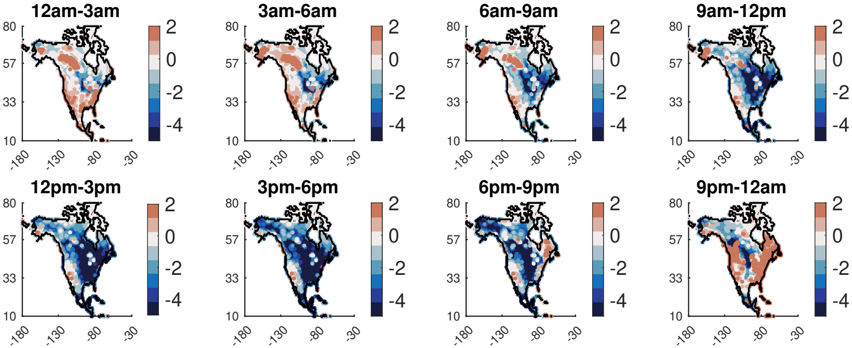
where

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

# Tracking carbon dioxide ($CO_2$)

$$\mathbf{b} = \underbrace{\begin{bmatrix} \mathbf{A}_1 & \cdots & \mathbf{A}_{2920} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{2920} \end{bmatrix}}_{\mathbf{x}_{\text{true}}} + \mathbf{e}$$

- $\mathbf{x}_i \in \mathbb{R}^{3,222} \quad \rightarrow 9.4 \times 10^6$ $CO_2$ fluxes
- $\mathbf{A}_i \in \mathbb{R}^{98,880 \times 3,222}$ represents atmospheric transport
- $\mathbf{b} \in \mathbb{R}^{98,880}$




Miller et. al. (2020), Miller and Michalak (2020), Crisp (2015)

# Bayesian approach

# Bayesian approach

**Idea:** Reformulate inverse problem as statistical inference where

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \mathbf{e}, \qquad \mathbf{e} \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{R}) \quad \text{and} \quad \mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \delta^{-1}\mathbf{Q})$$

| | | |
|---|---|---|
| $\boldsymbol{\mu} \in \mathbb{R}^n$ | : | prior mean |
| $\mathbf{R} \in \mathbb{R}^{m\times m}$, $\mathbf{Q} \in \mathbb{R}^{n\times n}$ | : | symmetric and positive definite (SPD) covariance matrices |
| $\lambda > 0$, $\delta > 0$ | : | hyper-parameters with gamma hyperpriors |

$$\pi(\lambda) \propto \lambda^{\alpha_\lambda - 1} \exp(-\beta_\lambda \lambda) \qquad \pi(\delta) \propto \delta^{\alpha_\delta - 1} \exp(-\beta_\delta \delta)$$

- Able to perform uncertainty quantification (UQ)
- Can assume $\mathbf{R}^{-1}$ is obtainable

---

Calvetti and Somersalo, Introduction to Bayesian Scientific Computing, Springer (2007)
Kaipio and Somersalo, Statistical and Computational Inverse Problems, Springer (2005)

# Bayesian approach

**Bayes' theorem**

Posterior distribution:

$$\pi(\mathbf{x}, \lambda, \delta \mid \mathbf{b}) = \frac{\pi(\mathbf{b} \mid \mathbf{x}, \lambda, \delta)\, \pi(\mathbf{x} \mid \delta)\, \pi(\lambda)\, \pi(\delta)}{\pi(\mathbf{b})} \propto \underbrace{\pi(\mathbf{b} \mid \mathbf{x}, \lambda, \delta)}_{\text{likelihood}}\, \underbrace{\pi(\mathbf{x} \mid \delta)}_{\text{prior}}\, \pi(\lambda)\, \pi(\delta)$$

- Solution to inverse problem
- Non-Gaussian
- Markov chain Monte Carlo (MCMC) for UQ
    - Variance estimates

Calvetti and Somersalo, Introduction to Bayesian Scientific Computing, Springer (2007)
Kaipio and Somersalo, Statistical and Computational Inverse Problems, Springer (2005)

# Challenges for large-scale UQ

⚠ Need efficient methods to explore the posterior

- Maximum a Posteriori estimate: $\max_{\mathbf{x},\lambda,\delta} \pi(\mathbf{x}, \lambda, \delta \mid \mathbf{b})$
- Estimating variances requires drawing samples, which can be expensive (non-Gaussian, MCMC)

⚠ $\mathbf{A}$ and $\mathbf{Q}$ may only be accessible via matrix-vector multiplications

⚠ Computing factorization $\mathbf{Q} = \mathbf{L}\mathbf{L}^\top$ is computationally infeasible

# Challenges for large-scale UQ

⚠ Need efficient methods to explore the posterior

- Maximum a Posteriori estimate: $\max_{\mathbf{x},\lambda,\delta} \pi(\mathbf{x}, \lambda, \delta \mid \mathbf{b})$
- Estimating variances requires drawing samples, which can be expensive (non-Gaussian, MCMC)

⚠ $\mathbf{A}$ and $\mathbf{Q}$ may only be accessible via matrix-vector multiplications

⚠ Computing factorization $\mathbf{Q} = \mathbf{L}\mathbf{L}^\top$ is computationally infeasible

**Proposed Hierarchical Bayesian Approach:**

- Use Gibbs sampler for $\pi(\mathbf{x}, \lambda, \delta \mid \mathbf{b})$
- Exploit generalized Golub-Kahan (genGK) methods

# Hierarchical Gibbs Sampling

Assume $\boldsymbol{\mu} = \mathbf{0}$

Sample from the conditional densities:

$$\lambda \mid \mathbf{b}, \mathbf{x} \sim \Gamma\big(m/2 + \gamma_\lambda,\ \frac{1}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{b}\right\|^2_{\mathbf{R}^{-1}} + \beta_\lambda\big)$$

$$\delta \mid \mathbf{b}, \mathbf{x} \sim \Gamma\big(n/2 + \gamma_\delta,\ \frac{1}{2}\left\|\mathbf{x}\right\|^2_{\mathbf{Q}^{-1}} + \beta_\delta\big)$$

$$\begin{aligned}\mathbf{x} \mid \mathbf{b}, \lambda, \delta &\sim \mathcal{N}\left(\underbrace{(\lambda\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A} + \delta\mathbf{Q}^{-1})^{-1}\lambda\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b}}_{\hat{\mathbf{x}}(\lambda,\delta)},\ \underbrace{(\lambda\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{A} + \delta\mathbf{Q}^{-1})^{-1}}_{\boldsymbol{\Gamma}(\lambda,\delta)}\right)\\ &= \mathcal{N}(\hat{\mathbf{x}}(\lambda,\delta), \boldsymbol{\Gamma}(\lambda,\delta))\end{aligned}$$

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

# Hierarchical Gibbs sampler

**Require:** Initialize $\mathbf{x}^0$

**for** $j = 1 : J$ **do**

  Sample $\lambda^j \sim \Gamma\left(m/2 + \gamma_\lambda, \frac{1}{2}\left\|\mathbf{A}\mathbf{x}^{j-1} - \mathbf{b}\right\|^2_{\mathbf{R}^{-1}} + \beta_\lambda\right)$

  Sample $\delta^j \sim \Gamma\left(n/2 + \gamma_\delta, \frac{1}{2}\left\|\mathbf{x}^{j-1}\right\|^2_{\mathbf{Q}^{-1}} + \beta_\delta\right)$

  Sample $\mathbf{x}^j \sim \mathcal{N}(\hat{\mathbf{x}}(\lambda^j, \delta^j), \boldsymbol{\Gamma}(\lambda^j, \delta^j))$

**end for**

- Generates a Markov chain $\{\mathbf{x}^i\}_{i=0}^J$ that converges to the target density $\pi(\mathbf{x}, \lambda, \delta \mid \mathbf{b})$

Bardsley (2018)

# Sampling from a Gaussian

Let $\lambda = \lambda^j,\ \delta = \delta^j$ be fixed, $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\lambda, \delta),\ \mathbf{\Gamma} = \mathbf{\Gamma}(\lambda, \delta)$

$$\mathbf{x}|\mathbf{b}, \delta, \lambda \sim \mathcal{N}(\hat{\mathbf{x}}, \mathbf{\Gamma})$$

**conditional $\mathcal{N}(\hat{\mathbf{x}}, \mathbf{\Gamma})$**

$$\mathbf{\Gamma} = (\lambda \mathbf{H} + \delta \mathbf{Q}^{-1})^{-1}, \quad \text{with} \quad \mathbf{H} = \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}$$

$$\hat{\mathbf{x}} = \lambda \mathbf{\Gamma} \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b}$$

Bardsley (2018)

# Sampling from a Gaussian

Let $\lambda = \lambda^j,\ \delta = \delta^j$ be fixed, $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\lambda, \delta),\ \boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\lambda, \delta)$

$$\mathbf{x}|\mathbf{b}, \delta, \lambda \sim \mathcal{N}(\hat{\mathbf{x}}, \boldsymbol{\Gamma})$$

**conditional $\mathcal{N}(\hat{\mathbf{x}}, \boldsymbol{\Gamma})$**

$$\boldsymbol{\Gamma} = (\lambda \mathbf{H} + \delta \mathbf{Q}^{-1})^{-1}, \quad \text{with} \quad \mathbf{H} = \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}$$

$$\hat{\mathbf{x}} = \lambda \boldsymbol{\Gamma} \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b}$$

Compute sample:

$$\mathbf{x}^j = \hat{\mathbf{x}} + \boldsymbol{\Gamma}^{1/2} \boldsymbol{\xi} \quad \text{where} \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Bardsley (2018)

# Sampling from a Gaussian

Let $\lambda = \lambda^j,\ \delta = \delta^j$ be fixed, $\hat{\mathbf{x}} = \hat{\mathbf{x}}(\lambda, \delta),\ \mathbf{\Gamma} = \mathbf{\Gamma}(\lambda, \delta)$

$$\mathbf{x}|\mathbf{b}, \delta, \lambda \sim \mathcal{N}(\hat{\mathbf{x}}, \mathbf{\Gamma})$$

**conditional $\mathcal{N}(\hat{\mathbf{x}}, \mathbf{\Gamma})$**

$$\mathbf{\Gamma} = (\lambda \mathbf{H} + \delta \mathbf{Q}^{-1})^{-1}, \quad \text{with} \quad \mathbf{H} = \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}$$

$$\hat{\mathbf{x}} = \lambda \mathbf{\Gamma} \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{b}$$

Compute sample:

$$\mathbf{x}^j = \hat{\mathbf{x}} + \mathbf{\Gamma}^{1/2} \boldsymbol{\xi} \quad \text{where} \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Let $\mathbf{\Gamma} = \mathbf{S}\mathbf{S}^\top$, then a sample is given by

$$\mathbf{x}^j = \hat{\mathbf{x}} + \mathbf{S} \boldsymbol{\xi} \quad \text{where} \quad \boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Bardsley (2018)

# Sampling from a Gaussian

Use $K$ steps of the symmetric Lanczos process with $\boldsymbol{\Gamma}$ and $\boldsymbol{\xi}$

$$\mathbf{W}_K = \begin{bmatrix} \mathbf{w}_1 & \dots & \mathbf{w}_K \end{bmatrix} \in \mathbb{R}^{n \times K} \quad \text{and} \quad \mathbf{C}_K = \begin{bmatrix} \gamma_1 & \alpha_2 & & & \\ \alpha_2 & \gamma_2 & \alpha_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \alpha_{K-1} & \gamma_{K-1} & \alpha_K \\ & & & \alpha_K & \gamma_K \end{bmatrix} \in \mathbb{R}^{K \times K}$$

Then $\mathbf{x}^j = \hat{\mathbf{x}} + \mathbf{W}_K \mathbf{C}_K^{1/2} \alpha_1 \mathbf{e}_1$ is an approximate draw from $\mathcal{N}(\hat{\mathbf{x}}, \boldsymbol{\Gamma})$

# Sampling from an approximate conditional

⚠ Infeasible to explicitly compute $\mathbf{A}$, $\mathbf{Q}^{-1}$, $\hat{\mathbf{x}}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Gamma}^{1/2}$ or factorizations

Flath et al (2011), Bui-Thanh et al (2013), Spantini et al (2015), Saibaba and Kitanidis (2015), Brown et al (2018)

# Sampling from an approximate conditional

⚠ Infeasible to explicitly compute $\mathbf{A}$, $\mathbf{Q}^{-1}$, $\hat{\mathbf{x}}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Gamma}^{1/2}$ or factorizations

**approximate conditional $\mathcal{N}(\mathbf{x}_k, \widehat{\boldsymbol{\Gamma}})$**

$$\widehat{\boldsymbol{\Gamma}} \equiv (\lambda \widehat{\mathbf{H}} + \delta \mathbf{Q}^{-1})^{-1} \quad \text{with} \quad \widehat{\mathbf{H}} \approx \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}$$

$$\mathbf{x}_k \approx \hat{\mathbf{x}}$$

Flath et al (2011), Bui-Thanh et al (2013), Spantini et al (2015), Saibaba and Kitanidis (2015), Brown et al (2018)

# Sampling from an approximate conditional

⚠ Infeasible to explicitly compute $\mathbf{A}$, $\mathbf{Q}^{-1}$, $\hat{\mathbf{x}}$, $\mathbf{\Gamma}$, and $\mathbf{\Gamma}^{1/2}$ or factorizations

**approximate conditional $\mathcal{N}(\mathbf{x}_k, \widehat{\mathbf{\Gamma}})$**

$$\widehat{\mathbf{\Gamma}} \equiv (\lambda \widehat{\mathbf{H}} + \delta \mathbf{Q}^{-1})^{-1} \quad \text{with} \quad \widehat{\mathbf{H}} \approx \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}$$

$$\mathbf{x}_k \approx \hat{\mathbf{x}}$$

**Current Approaches for $\hat{\mathbf{H}}$:**

- Randomized SVD
  - ⚠ Requires lowrank $\mathbf{A}$

**Our Approach:**

- Find $\mathbf{x}_k$ with generalized Golub-Kahan based method
- Use for $\hat{\mathbf{H}}$ and $\hat{\mathbf{\Gamma}}^{1/2}$

Flath et al (2011), Bui-Thanh et al (2013), Spantini et al (2015), Saibaba and Kitanidis (2015), Brown et al (2018)

# Generalized Golub-Kahan (genGK) methods

# Estimate $\hat{\mathrm{x}}$ with genGK based method

$$\pi(\mathbf{x} \mid \mathbf{b}, \lambda, \delta) \propto \exp\left(-\frac{\lambda}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_{\mathbf{R}^{-1}}^{2}-\frac{\delta}{2}\|\mathbf{x}\|_{\mathbf{Q}^{-1}}^{2}\right)$$

Point estimate:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} - \log\pi(\mathbf{x}|\mathbf{b},\lambda,\delta) = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min}\,\frac{\lambda}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_{\mathbf{R}^{-1}}^{2}+\frac{\delta}{2}\|\mathbf{x}\|_{\mathbf{Q}^{-1}}^{2}$$

Chung and Saibaba (2017), Calvetti and Somersalo (2005), Calvetti (2007), Arridge, Betcke, and Harhanen (2014)

# Estimate $\hat{\mathrm{x}}$ with genGK based method

$$\pi(\mathbf{x} \mid \mathbf{b}, \lambda, \delta) \propto \exp\left(-\frac{\lambda}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2_{\mathbf{R}^{-1}} - \frac{\delta}{2}\|\mathbf{x}\|^2_{\mathbf{Q}^{-1}}\right)$$

Point estimate:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} - \log \pi(\mathbf{x}|\mathbf{b}, \lambda, \delta) = \underset{\mathbf{x}\in\mathbb{R}^n}{\arg\min} \frac{\lambda}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|^2_{\mathbf{R}^{-1}} + \frac{\delta}{2}\|\mathbf{x}\|^2_{\mathbf{Q}^{-1}}$$

(1) Introduce a change in variables $\mathbf{s} = \mathbf{Q}^{-1}\mathbf{x}$

$$\min_{\mathbf{s}\in\mathbb{R}^n} \frac{\lambda}{2}\|\mathbf{A}\mathbf{Q}\mathbf{s}-\mathbf{b}\|^2_{\mathbf{R}^{-1}} + \frac{\delta}{2}\|\mathbf{s}\|^2_{\mathbf{Q}}$$

✔ Avoids computations with $\mathbf{Q}^{-1}$

Chung and Saibaba (2017), Calvetti and Somersalo (2005), Calvetti (2007), Arridge, Betcke, and Harhanen (2014)

# Estimate $\hat{\mathrm{x}}$ with genGK based method

(2) Project transformed problem onto a Krylov subspace of increasing dimension
$\Longrightarrow$ generalized Golub-Kahan bidiagonalization

Benbow (1999), Arioli (2013), Arioli & Orban (2013).

# Estimate $\hat{\mathrm{x}}$ with genGK based method

(2) Project transformed problem onto a Krylov subspace of increasing dimension
$\Longrightarrow$ generalized Golub-Kahan bidiagonalization

(a) Let $\gamma_1 = \|\mathbf{b}\|_{\mathbf{R}^{-1}}$ and $\mathbf{U}_{k+1}\gamma_1\mathbf{e}_1 = \mathbf{b}$

$$\mathbf{AQV}_k = \mathbf{U}_{k+1}\mathbf{B}_k, \quad \text{and} \quad \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{U}_{k+1} = \mathbf{V}_k\mathbf{B}_k^\top + \alpha_{k+1}\mathbf{v}_{k+1}\mathbf{e}_{k+1}^\top$$

with
$$\mathbf{U}_{k+1}^\top\mathbf{R}^{-1}\mathbf{U}_{k+1} = \mathbf{I}_{k+1} \quad \text{and} \quad \mathbf{V}_k^\top\mathbf{QV}_k = \mathbf{I}_k$$

Benbow (1999), Arioli (2013), Arioli & Orban (2013).

# Estimate $\hat{\mathrm{x}}$ with genGK based method

(2) Project transformed problem onto a Krylov subspace of increasing dimension
$\Longrightarrow$ generalized Golub-Kahan bidiagonalization

(a) Let $\gamma_1 = \|\mathbf{b}\|_{\mathbf{R}^{-1}}$ and $\mathbf{U}_{k+1}\gamma_1\mathbf{e}_1 = \mathbf{b}$

$$\mathbf{AQV}_k = \mathbf{U}_{k+1}\mathbf{B}_k, \quad \text{and} \quad \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{U}_{k+1} = \mathbf{V}_k\mathbf{B}_k^\top + \alpha_{k+1}\mathbf{v}_{k+1}\mathbf{e}_{k+1}^\top$$

with
$$\mathbf{U}_{k+1}^\top\mathbf{R}^{-1}\mathbf{U}_{k+1} = \mathbf{I}_{k+1} \quad \text{and} \quad \mathbf{V}_k^\top\mathbf{QV}_k = \mathbf{I}_k$$

**Krylov subspace**

$$\mathcal{R}(\mathbf{V}_k) = \mathcal{K}_k(\mathbf{A}^\top\mathbf{R}^{-1}\mathbf{AQ}, \mathbf{A}^\top\mathbf{R}^{-1}\mathbf{b})$$

where

$$\mathcal{K}_k(\mathbf{C}, \mathbf{g}) \equiv \text{span}\left\{\mathbf{g}, \mathbf{Cg}, \ldots\ldots, \mathbf{C}^{k-1}\mathbf{g}\right\}$$

Benbow (1999), Arioli (2013), Arioli & Orban (2013).

# Estimate $\hat{x}$ with genGK based method

(b) Let $\mathbf{s} = \mathbf{V}_k \mathbf{y}$

$$\min_{\mathbf{s}\in\mathcal{R}(\mathbf{V}_k)} \frac{\lambda}{2}\left\|\mathbf{AQs} - \mathbf{b}\right\|^2_{\mathbf{R}^{-1}} + \frac{\delta}{2}\left\|\mathbf{s}\right\|^2_{\mathbf{Q}} \quad \Longleftrightarrow \quad \min_{\mathbf{y}\in\mathbb{R}^k} \frac{\lambda}{2}\left\|\mathbf{B}_k\mathbf{y} - \gamma_1\mathbf{e}_1\right\|^2_2 + \frac{\delta}{2}\left\|\mathbf{y}\right\|^2_2$$

$$\mathbf{y}_k = \min_{\mathbf{y}\in\mathbb{R}^k} \left\| \begin{bmatrix} \sqrt{\lambda}\mathbf{B}_k \\ \sqrt{\delta}\mathbf{I} \end{bmatrix} \mathbf{y} - \begin{bmatrix} \sqrt{\lambda}\gamma_1\mathbf{e}_1 \\ \mathbf{0} \end{bmatrix} \right\|^2_2$$

# Estimate $\hat{\mathrm{x}}$ with genGK based method

(b) Let $\mathbf{s} = \mathbf{V}_k \mathbf{y}$

$$\min_{\mathbf{s} \in \mathcal{R}(\mathbf{V}_k)} \frac{\lambda}{2} \left\| \mathbf{AQs} - \mathbf{b} \right\|_{\mathbf{R}^{-1}}^2 + \frac{\delta}{2} \left\| \mathbf{s} \right\|_{\mathbf{Q}}^2 \quad \Longleftrightarrow \quad \min_{\mathbf{y} \in \mathbb{R}^k} \frac{\lambda}{2} \left\| \mathbf{B}_k \mathbf{y} - \gamma_1 \mathbf{e}_1 \right\|_2^2 + \frac{\delta}{2} \left\| \mathbf{y} \right\|_2^2$$

$$\mathbf{y}_k = \min_{\mathbf{y} \in \mathbb{R}^k} \left\| \begin{bmatrix} \sqrt{\lambda} \mathbf{B}_k \\ \sqrt{\delta} \mathbf{I} \end{bmatrix} \mathbf{y} - \begin{bmatrix} \sqrt{\lambda} \gamma_1 \mathbf{e}_1 \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

(3) Project back: $\boxed{\mathbf{x}_k = \mathbf{Q}\mathbf{s}_k = \mathbf{Q}\mathbf{V}_k \mathbf{y}_k \approx \hat{\mathbf{x}}}$

# Recall: Sampling from approximate conditional

**approximate conditional $\mathcal{N}(\mathbf{x}_k, \widehat{\mathbf{\Gamma}})$**

$$\widehat{\mathbf{\Gamma}} \equiv (\sigma \widehat{\mathbf{H}} + \alpha \mathbf{Q}^{-1})^{-1} \quad \text{with} \quad \widehat{\mathbf{H}} \approx \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}$$

$$\mathbf{x}_k \approx \hat{\mathbf{x}}$$

$$\mathbf{x}^j = \mathbf{x}_k + \hat{\mathbf{\Gamma}}^{1/2}$$

Flath et al (2011), Bui-Thanh et al (2013), Spantini et al (2015), Saibaba and Kitanidis (2015), Saibaba et al (2020)

# Recall: Sampling from approximate conditional

**approximate conditional $\mathcal{N}(\mathbf{x}_k, \widehat{\boldsymbol{\Gamma}})$**

$$\widehat{\boldsymbol{\Gamma}} \equiv (\sigma \widehat{\mathbf{H}} + \alpha \mathbf{Q}^{-1})^{-1} \quad \text{with} \quad \widehat{\mathbf{H}} \approx \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}$$

$$\mathbf{x}_k \approx \hat{\mathbf{x}}$$

$$\mathbf{x}^j = \mathbf{x}_k + \hat{\boldsymbol{\Gamma}}^{1/2}$$

From genGK:

$$\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{U}_{k+1} \approx \mathbf{V}_k \mathbf{B}_k^\top \quad \Longrightarrow \quad \hat{\mathbf{H}} = \mathbf{V}_k \mathbf{B}_k^\top \mathbf{B}_k \mathbf{V}_k^\top$$

---

Flath et al (2011), Bui-Thanh et al (2013), Spantini et al (2015), Saibaba and Kitanidis (2015), Saibaba et al (2020)

# Estimate $\Gamma^{1/2}$ with genGK

Can assume computations with $\mathbf{Q}^{1/2}$ are accessible (using Lanczos method)

$$\begin{aligned}
\boldsymbol{\Gamma}^{1/2} &= (\delta \mathbf{Q}^{-1} + \lambda \mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A})^{-1/2} \\
&= \delta^{-1/2} \mathbf{Q}^{1/2} (\mathbf{I} + \frac{\lambda}{\delta} \mathbf{Q}^{1/2} \underbrace{\mathbf{A}^\top \mathbf{R}^{-1} \mathbf{A}}_{\approx \mathbf{V}_k \mathbf{B}_k^\top \mathbf{B}_k \mathbf{V}_k^\top} \mathbf{Q}^{1/2})^{-1/2} \\
&\approx \delta^{-1/2} \mathbf{Q}^{1/2} (\mathbf{I} + \frac{\lambda}{\delta} \underbrace{\mathbf{Q}^{1/2} \mathbf{V}_k \mathbf{B}_k^\top \mathbf{B}_k \mathbf{V}_k^\top \mathbf{Q}^{1/2}}_{\approx \mathbf{Z}_k \boldsymbol{\Theta}_k \mathbf{Z}_k^\top})^{-1/2} \\
&= \boxed{\delta^{-1/2} \mathbf{Q}^{1/2} (\mathbf{I} - \mathbf{Z}_k \mathbf{D}_k \mathbf{Z}_k^\top)} \equiv \hat{\boldsymbol{\Gamma}}^{1/2} \qquad \text{(Woodbury Identity)}
\end{aligned}$$

where $\mathbf{D}_k \equiv \mathbf{I}_k - (\mathbf{I}_k + \frac{\lambda}{\delta} \boldsymbol{\Theta}_k)^{-1/2}$

Flath et al (2011), Bui-Thanh et al (2013), Spantini et al (2015), Saibaba and Kitanidis (2015), Saibaba et al (2020)

# Draw sample from approximate conditional

Sample:

$$\mathbf{x}^j = \underbrace{\mathbf{Q}\mathbf{V}_k\mathbf{y}_k}_{\mathbf{x}_k} + \underbrace{\delta^{-1/2}\mathbf{Q}^{1/2}(\mathbf{I} - \mathbf{Z}_k\mathbf{D}_k\mathbf{Z}_k^\top)}_{\hat{\mathbf{\Gamma}}^{1/2}}\boldsymbol{\xi}$$

# Draw sample from approximate conditional

Sample:

$$\mathbf{x}^j = \underbrace{\mathbf{Q}\mathbf{V}_k\mathbf{y}_k}_{\mathbf{x}_k} + \underbrace{\delta^{-1/2}\mathbf{Q}^{1/2}(\mathbf{I} - \mathbf{Z}_k\mathbf{D}_k\mathbf{Z}_k^\top)}_{\hat{\boldsymbol{\Gamma}}^{1/2}}\boldsymbol{\xi}$$

**Note:**

- To sample $\delta$, need

$$(\mathbf{x}^j)^\top\mathbf{Q}^{-1}(\mathbf{x}^j) = \left(\mathbf{V}_k\mathbf{y}^j\right)^\top \mathbf{x}^j + \boldsymbol{\xi}^\top\delta^{-1}\left(\mathbf{I} - \mathbf{Z}_k\left((\mathbf{D}_k^j)^2 - 2\mathbf{D}_k^j\right)\mathbf{Z}_k^\top\right)\boldsymbol{\xi} + 2\left(\mathbf{V}_k\mathbf{y}^j\right)^\top\hat{\boldsymbol{\Gamma}}^{1/2}$$

- The genGK process does not depend on $\lambda$ or $\delta$

| | |
|---|---|
| precompute: | $\mathbf{B}_k,\quad \mathbf{V}_k,\quad \gamma_1,\quad \mathbf{Z}_k,\quad \boldsymbol{\Theta}_k$ |
| update: | $\mathbf{D}_k$ |

# Hierarchical Gibbs sampler with genGK

**Require:** Initialize $(\lambda^0, \delta^0)$

Use genGK to get $\mathbf{x}^0 = (\lambda^0 \mathbf{A}^\top \mathbf{R}^{-1}\mathbf{A} + \delta^0 \mathbf{Q}^{-1})^{-1} \lambda^0 \mathbf{A}^\top \mathbf{R}^{-1}\mathbf{b}$

Save bidiagonal matrix $\mathbf{B}_k$, basis vectors $\mathbf{V}_k$

Compute and save $\mathbf{Z}_k$ and $\boldsymbol{\Theta}_k$

1: **for** $j = 1 : J$ **do**

2: Compute $\lambda^j \sim \Gamma\left(m/2 + \gamma_\lambda, \frac{1}{2}\left\|\mathbf{A}\mathbf{x}^{j-1} - \mathbf{b}\right\|^2_{\mathbf{R}^{-1}} + \beta_\lambda\right)$

3: Compute $\delta^j \sim \Gamma\left(n/2 + \gamma_\delta, \frac{1}{2}\left\|\mathbf{x}^{j-1}\right\|^2_{\mathbf{Q}^{-1}} + \beta_\delta\right)$

4: Compute $\mathbf{x}_k(\lambda^j, \delta^j)$ using genGK relations

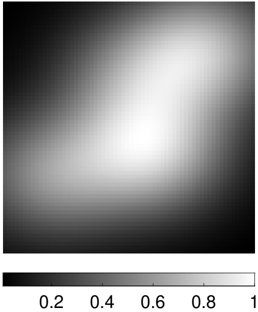
5: Update $\mathbf{D}_k$ and draw sample from approximate conditional $\mathbf{x}^j \sim \mathcal{N}(\mathbf{x}_k, \widehat{\boldsymbol{\Gamma}})$

6: **end for**

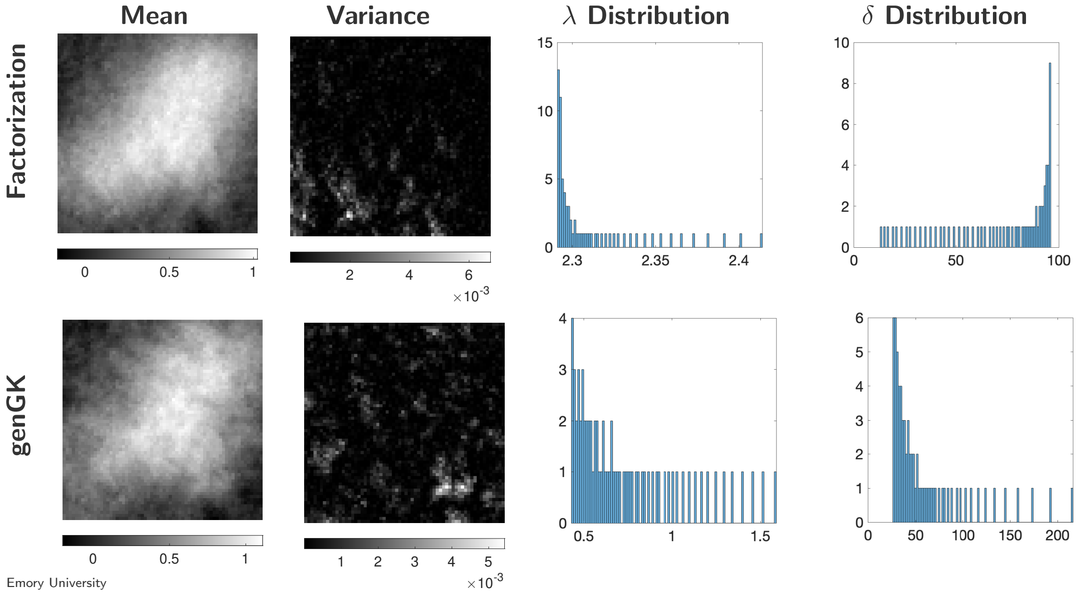
# Numerical Examples

# Seismic Tomography

- $\mathbf{x}_{\text{true}}$ represents a vectorized $64 \times 64$ image of the Earth's interior
- observations $\mathbf{b} \in \mathbb{R}^{3200}$ contains simulated measurements with 2% noise
- IRTools 'PRseismic'
- $\mathbf{Q}$: Matérn Kernel with smoothness $\nu = 0.5$ and correlation $\ell = 0.25$



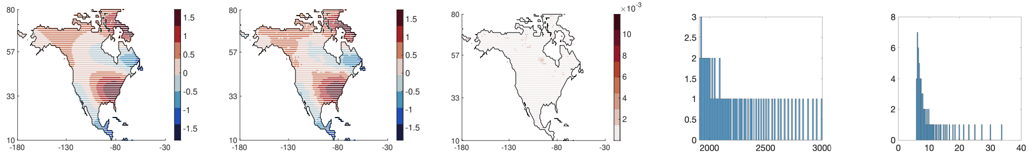
# Seismic Tomography



# Atmospheric Inverse Modeling

- Simplified atmospheric problem that is not dynamic
- $\mathbf{A} \in \mathbb{R}^{98,880 \times 11,900}$ represents the forward atmospheric transport model
- observations $\mathbf{b} \in \mathbb{R}^{98,880}$ contains satellite data with $4\%$ noise
- Solution $\mathbf{x} \in \mathbb{R}^{11,900}$ is average $CO_2$ emissions over North America
- $\mathbf{Q}$ : Matérn Kernel with smoothness $\nu = 2.5$ and correlation $\ell = 0.05$

**True** **Mean** **Variance** **$\lambda$ Distribution** **$\delta$ Distribution**



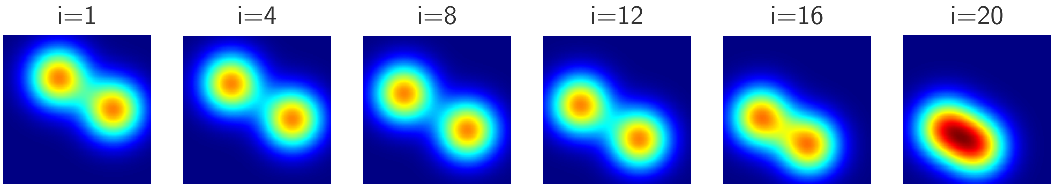

Miller et. al. (2020), Miller and Michalak (2020), Crisp (2015)

# Dynamic Photoacoustic Tomography

- $\mathbf{x}_i \in \mathbb{R}^{16,384}$ is a vectorized $128 \times 128$ image
- forward model $\mathbf{A} \in \mathbb{R}^{65,160 \times 327,680}$ is spherical projection with 18 angles
- observations $\mathbf{b} \in \mathbb{R}^{65,160}$ contain projection data with $2\%$ noise

$$\mathbf{x}_{\text{true}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{20} \end{bmatrix}, \mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_{20} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_{20} \end{bmatrix}, \mathbf{Q} = \mathbf{Q}_t \otimes \mathbf{Q}_s$$

- $\mathbf{Q}_s, \mathbf{Q}_t$ Matérn Kernel with smoothness $\nu = 0.5, \nu = 2.5$ and correlations $\ell = 0.25, \ell = 0.1$

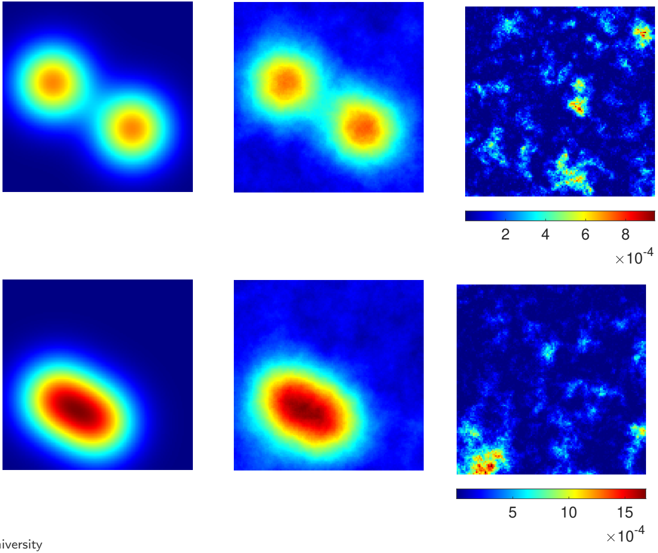


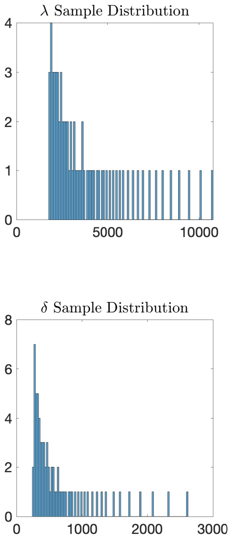
# Dynamic Photoacoustic Tomography

True | Mean | Variance

$i = 10$

$i = 20$



# Future Directions

- Extend to separable non-linear problems of the form

$$\mathbf{b} = \mathbf{A}(\mathbf{y}_{\mathrm{true}})\mathbf{x}_{\mathrm{true}} + \mathbf{e}$$

  where $\mathbf{A}(\cdot)$ maps model parameters $\mathbf{y}$ to large, ill-conditioned matrix
  - Dimension $\mathbf{y} << $ dimension $\mathbf{x}$
  - Linear in $\mathbf{x}$
  - Non-linear $\implies$ may be nonGaussian (need MCMC)
  - For fixed $\mathbf{y}$, use genGK methods to sample conditional of $\mathbf{x}$
- Use for large-scale dynamic atmospheric example
- Use for MRI reconstructions

# Take-home messages

1. UQ for large-scale inverse problems can be challenging
2. For Gaussian posterior distributions, generalized Krylov methods can be used to efficiently
   - compute point estimates
   - perform UQ (e.g., draw samples)
3. For non-Gaussian posterior distributions, hierarchical Bayesian approaches are needed
   - Sampling can be computationally infeasible and slow
   - Generalized Krylov methods can be exploited within hierarchical Bayesian approaches

## References

[1] J.M. Bardsley. *Computational Uncertainty Quantification for Inverse Problems*. Society for Industrial and Applied Mathematics. 2018.

[2] J. Chung, A.K. Saibaba, M. Brown, E. Westman. "Efficient generalized Golub-Kahan based methods for dynamic inverse problems". *Inverse Problems* 34.2 (2018).

**Thank you!**

**Support**

- NSF DMS-2026841
- NSF DMS-1654175

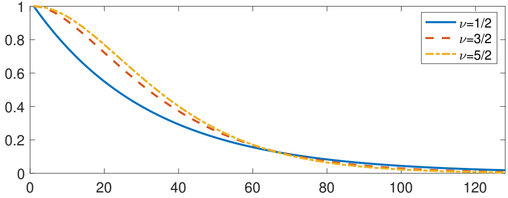
# Gaussian priors: Matérn covariance family

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

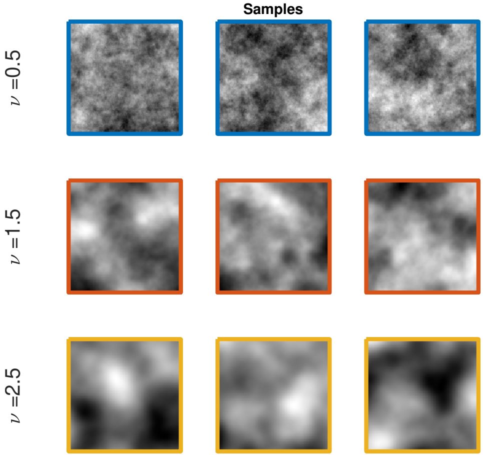
Matérn kernel:

$$\mathbf{Q}_{ij} = \kappa(r) = \underbrace{\frac{2^{1-\nu}}{\Gamma(\nu)}}_{Gamma\ function} \left(\frac{\sqrt{2\nu r}}{\ell}\right)^{\nu} \underbrace{K_\nu\left(\frac{\sqrt{2\nu r}}{\ell}\right)}_{Modified\ Bessel\ function}$$

where $r = \|\mathbf{p}_i - \mathbf{p}_j\|_2$




$\mathbf{Q}$ dense but fast matvecs



# Fast covariance evaluations

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \lambda^{-1}\mathbf{Q})$$

- 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 matvec with $\mathbf{Q}$

- FFT based methods
- Hierarchical Matrices

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

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

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

# How good is the approximation?

**Theorem**

*At the end of $k$ iterations, the Kullback-Leibler (KL) divergence from the approximate to the true posterior measure satisfies*

$$0 \leq D_{KL}(\widehat{\rho}_{\text{post}} \| \rho_{\text{post}}) \leq \frac{\lambda^{-1}}{2} \left[ \theta_k + \frac{\omega_k^2}{\lambda^2 + \omega_k} \delta_1^2 \beta_1^2 \right].$$

*where*

- $\omega_k = \|\mathbf{H}_\mathbf{Q} - \widehat{\mathbf{H}}_\mathbf{Q}\|_F$ *(monotonically decreasing)*
- $\theta_k = \text{trace}(\mathbf{H}_\mathbf{Q} - \widehat{\mathbf{H}}_\mathbf{Q})$ *(monotonically decreasing)*
- *with*

$$\mathbf{H}_\mathbf{Q} = \mathbf{Q}^{1/2}\mathbf{H}\mathbf{Q}^{1/2} \quad \textit{and} \quad \widehat{\mathbf{H}}_\mathbf{Q} = \mathbf{Q}^{1/2}\widehat{\mathbf{H}}\mathbf{Q}^{1/2}$$