MAX PLANCK INSTITUTE
FOR DYNAMICS OF COMPLEX
TECHNICAL SYSTEMS
MAGDEBURG

CSC
COMPUTATIONAL METHODS IN
SYSTEMS AND CONTROL THEORY

Low-Rank Methods for Bayesian Inverse Problems

Yue Qiu

joint work with Martin Stoll and Peter Benner

01-08-2017

Max Planck Institute for Dynamics of Complex Technical Systems

Outline

Inverse Problems

Many inverse problems governed by PDEs, we seek to infer unknown or uncertain varying parameter fields

- initial conditions
- coefficients
- boundary conditions ······

from limited and noisy observation.

Deterministic approach

- minimizes a regularized data misfit function
- does not incorporate uncertainties in the solution

Bayesian inference provides a systematic framework for incorporating uncertainties in

- observations
- forward models
- prior knowledge

to quantify uncertainties.

Bayesian Approach

In Bayesian approach, view all parameters as random variables and define the parameter-to-observable map $g : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^m$

$$Y = g(U, E).$$

Here, Y, U, E are all random variables.

- $u \in \mathbb{R}^n$: model parameters to be recovered;
- $e \in \mathbb{R}^k$: error vector;
- $y \in \mathbb{R}^m$: observable variables;

are realizations of U, E, and Y respectively.
Choose probability density functions (PDFs)

- $\pi_{\text{noise}} : \mathbb{R}^k \to \mathbb{R}$, modeling error and observation noise
- $\pi_{\text{prior}} : \mathbb{R}^n \to \mathbb{R}$, prior information of parameters u
- $\pi(y|u)$: describes relationship between observables y and parameters u

Bayesian Inference

To get posterior PDF $\pi_{\text{post}} : \mathbb{R}^m \to \mathbb{R}$, we apply Bayes' theorem,

$$\pi_{\text{post}} := \pi(u|y_{\text{obs}}) = \frac{\pi_{\text{prior}}(u)\pi(y_{\text{obs}}|u)}{\pi(y_{\text{obs}})} \propto \pi_{\text{prior}}(u)\pi(y_{\text{obs}}|u).$$

Here we assume additive noise, so

$$y = f(u) + e,$$

where $f : \mathbb{R}^n \to \mathbb{R}^m$ and e is additive noise that reflects both the modeling error and observation error.
Assume U and E are statistically independent, therefore,

$$\pi_{\text{post}} \propto \pi_{\text{prior}}(u)\pi_{\text{noise}}(y - f(u))$$

Linear Gaussian Case

Assuming that both probability density functions for u and e are Gaussian, we can rewrite the PDFs in the form

$$\pi_{\text{prior}}(u) \propto \exp\left(-\frac{1}{2}\left(u-\bar{u}_{\text{prior}}\right)^T \Gamma_{\text{prior}}^{-1}\left(u-\bar{u}_{\text{prior}}\right)\right),$$
$$\pi_{\text{noise}}(e) \propto \exp\left(-\frac{1}{2}\left(e-\bar{e}\right)^T \Gamma_{\text{noise}}^{-1}\left(e-\bar{e}\right)\right)$$

Bayes' theorem further gives

$$\pi_{\text{post}} \propto \exp\left(-\frac{1}{2}\left(u-\bar{u}_{\text{prior}}\right)^T \Gamma_{\text{prior}}^{-1}\left(u-\bar{u}_{\text{prior}}\right)-\frac{1}{2}\left(e-\bar{e}\right)^T \Gamma_{\text{noise}}^{-1}\left(e-\bar{e}\right)\right)$$
$$= \exp\left(-\frac{1}{2}\|u-\bar{u}_{\text{prior}}\|^2_{\Gamma_{\text{prior}}^{-1}}-\frac{1}{2}\|e-\bar{e}\|^2_{\Gamma_{\text{noise}}^{-1}}\right).$$

Computations of Γ_{post}

For linear parameter-to-observable map, $y = f(u) = Au$, we have

$$\pi_{\text{post}} \propto \exp\left(-\frac{1}{2}\|u - \bar{u}_{\text{prior}}\|^2_{\Gamma^{-1}_{\text{prior}}} - \frac{1}{2}\|y_{\text{obs}} - Au - \bar{e}\|^2_{\Gamma^{-1}_{\text{noise}}}\right).$$

Then we get

$$\bar{u}_{\text{post}} = \operatorname{argmin}_u \underbrace{\left(\frac{1}{2}\|u - \bar{u}_{\text{prior}}\|^2_{\Gamma^{-1}_{\text{prior}}} + \frac{1}{2}\|y_{\text{obs}} - Au - \bar{e}\|^2_{\Gamma^{-1}_{\text{noise}}}\right)}_{\mathcal{J}(u)},$$

given by maximum a posterior (MAP) point. It is equivalent to solving a regularized deterministic inverse problem.

The posterior covariance matrix Γ_{post}

$$\Gamma_{\text{post}} = \left(A^T \Gamma^{-1}_{\text{noise}} A + \Gamma^{-1}_{\text{prior}}\right)^{-1}.$$

Since A comes from discretization of time-dependent PDE, A is large and dense. Direct computations of Γ_{post} is impossible.

Approximation of Γ_{post}

Note that

$$\begin{aligned}\Gamma_{\text{post}} &= \left(A^T \Gamma_{\text{noise}}^{-1} A + \Gamma_{\text{prior}}^{-1}\right)^{-1} \\ &= \Gamma_{\text{prior}}^{1/2} (\underbrace{\Gamma_{\text{prior}}^{1/2} A^T \Gamma_{\text{noise}}^{-1} A \Gamma_{\text{prior}}^{1/2}}_{\tilde{\mathcal{H}}_{\text{mis}}} + I)^{-1} \Gamma_{\text{prior}}^{1/2}\end{aligned}$$

It is shown that $\tilde{\mathcal{H}}_{\text{mis}}$ has low numerical rank [Flath, Wilkox, et.al. 2011], Lanczos method can be used, which gives

$$\tilde{\mathcal{H}}_{\text{mis}} \approx V \Lambda V^T.$$

With Sherman-Morrison-Woodbury formula,

$$\Gamma_{\text{post}} \approx \Gamma_{\text{prior}} - \Gamma_{\text{prior}}^{1/2} V \tilde{\Lambda} V^T \Gamma_{\text{prior}}^{1/2}$$

where $\tilde{\Lambda} = \text{diag}(\frac{\lambda_i}{1+\lambda_i})$

Computational Challenage

Applying Lanczos to $\tilde{\mathcal{H}}_{\text{mis}} = \Gamma_{\text{prior}}^{1/2} A^T \Gamma_{\text{noise}}^{-1} A \Gamma_{\text{prior}}^{1/2}$, at each iteration, we need to solve

- one forward PDE
- one adjoint PDE

For time-dependent PDE

$$\frac{\partial}{\partial t} y - \mathcal{L} y = f$$

computations of the solution $y = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_{n_t} \end{bmatrix}$ gives $\mathcal{O}(n_x n_t)$ complexity for both computations and storage.
Here

- n_x: number of variables in space
- n_t: number of time steps

curse of dimensionality

Low-Rank Lanczos

To beat the curse of dimensionality

- not directly computing $y = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_{n_t} \end{bmatrix}$,
- but computing a low-rank approximation of $y \approx u_x v_t^T$ with

$$\text{rank}(u_x) = \text{rank}(v_t) = r, \qquad r << \{n_x, n_t\}$$

low-rank in time approach, [Stoll and Breitan, 2015]

Low-rank in time approach gives

- $\mathcal{O}(n_x + n_t)$ computational complexity
- $\mathcal{O}(n_x + n_t)$ memory consumption

Applying the low-rank in time approach to Lanczos method gives low-rank Lanczos,

- reorthogonalization is necessary
- symmetry gets lost, but can be recovered with ‘good’ low-rank approximations

Example

Uncertainty from initial condition

$$
\begin{aligned}
\frac{\partial}{\partial t} y - \Delta y &= 0, && \Omega \times (0, T) \\
y &= u, && \Omega \times \{t = 0\} \\
y &= 0, && \partial\Omega_D \times (0, T \\
\nabla y \cdot \mathbf{n} &= 0, && \partial\Omega_N \times (0, T)
\end{aligned}
$$

Objective function

$$
\min_u \left(\frac{\beta_{\text{noise}}}{2} \int_0^T \int_\Omega (y - y_{\text{obs}})^2 \, b(x,t) dx dt + \frac{\beta_{\text{prior}}}{2} \int_\Omega u^2 dx \right),
$$

Here y_{obs} can be only obtained at the vicinity of location of sensors and $b(x, t)$ is the observation operator with

$$
b(x, t) = \Sigma_j \delta(x - x_j).
$$

Settings for input uncertainty is similar.

Tensor Formulation

Discretizing using

- finite element in space
- implicit Euler in time

simultaneously gives $\mathcal{K}\mathbf{y} = \mathcal{C}u$ and

$$\min_u \left(\frac{1}{2} \left(\mathbf{y} - \mathbf{y}_{\text{obs}}\right)^T \mathcal{B}^T \Gamma_{\text{noise}}^{-1} \mathcal{B}(\mathbf{y} - \mathbf{y}_{\text{obs}}) + \frac{1}{2} u^T \Gamma_{\text{prior}}^{-1} u \right),$$

where, $\Gamma_{\text{noise}} = \frac{1}{\beta_{\text{noise}}} h^{-d} I_{n_x} \otimes I_{n_t}$, and $\Gamma_{\text{prior}} = \frac{1}{\beta_{\text{prior}}} h^{-d} I_{n_x}$.
Here $\mathcal{K} = C \otimes M + I_{n_t} \otimes (\tau * K)$, $\mathcal{C} = e_1 \otimes (\tau * M)$, M is the mass matrix, K is the stiffness matrix, e_1 is the canonical unit vector, $\tau = T/n_t$ and

$$C = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & \ddots \end{bmatrix}$$

Other settings of Γ_{prior} is also possible by defining appropriate inner product.

Preconditioning

Recall that

$$\tilde{\mathcal{H}}_{\text{mis}} = \Gamma_{\text{prior}}^{\frac{1}{2}} \mathcal{C}^T \mathcal{K}^{-T} \mathcal{B}^T \Gamma_{\text{noise}}^{-1} \mathcal{B} \mathcal{K}^{-1} \mathcal{C} \Gamma_{\text{prior}}^{\frac{1}{2}}.$$

Applying low-rank Lanczos to $\tilde{\mathcal{H}}_{\text{mis}}$, take $\mathcal{K}$ as an example.

$$(I_{n_t} \otimes L + C \otimes M)\, vec(X) = vec(F). \tag{1}$$

To solve (1), we use the alternative minimal energy (AMEn) approach and the tensor-train (TT) toolbox. At each AMEn iteration,

- left Galerkin projection

$$\left(\hat{I}_n \otimes L + \hat{C} \otimes M\right) x = \tilde{b}, \tag{2}$$

- or right Galerkin projection

$$\left(I_n \otimes \tilde{L} + C \otimes \tilde{M}\right) \tilde{x} = \hat{b}. \tag{3}$$

$P = \text{diag}(\hat{I}_n) \otimes L + \text{diag}(\hat{C}) \otimes M$ for (2) and direct method for (3)

Implementation Details

Consider the uncertainty in the initial condition case. Two problems are used

- Heat equation
- Convection-diffusion equation

For both cases, we have data observed from 9 sensors which are randomly located inside the domain.

Hardware

- Intel i5 CPU, 3.60GHZ
- 8GB RAM

Software

- Matlab 2011b
- Tensor Train Toolbox (TT-Toolbox)
- tensor approximation tolerance: 10^{-8}
- PDE solver tolerance: 10^{-8}

Low-rank Lanczos vs 'eigs'

Heat equation, 32×32 mesh

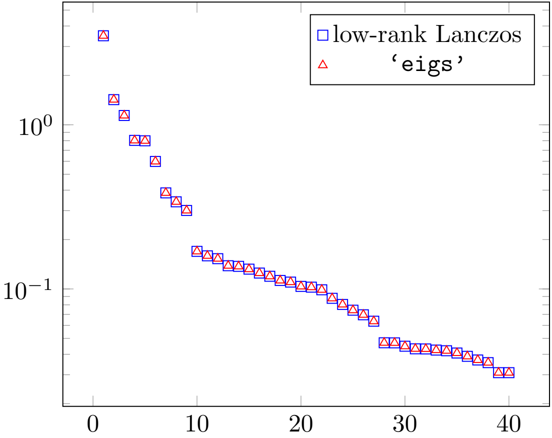


Figure: $\lambda(\tilde{\mathcal{H}}_{\text{mis}})$, $n_t = 60$

Maximum Rank vs n_t

64×64 mesh

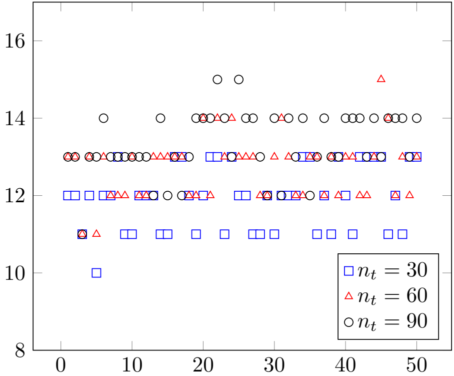


Figure: Heat equation

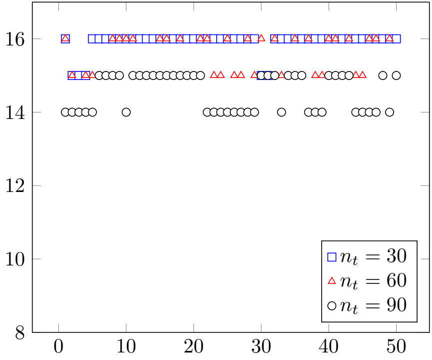


Figure: Convection-diffusion equation, $\nu = 10^{-2}$

Maximum Rank vs h and ν

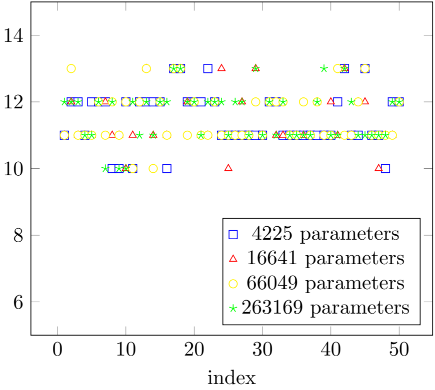


Figure: Heat equation, $n_t = 30$

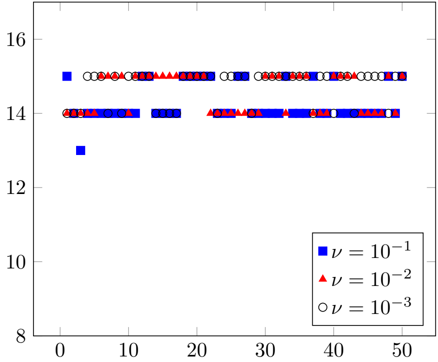
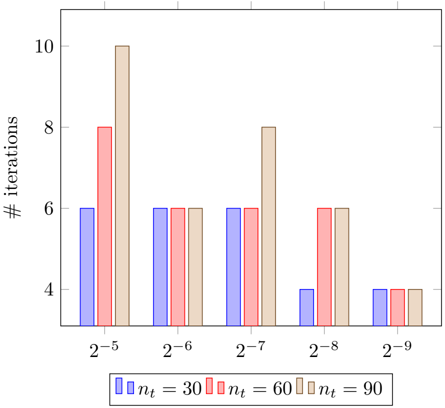
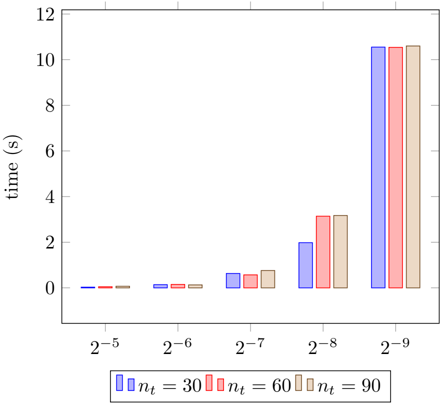


Figure: Convection-diffusion equation, 64×64 mesh, $n_t = 30$

Preconditioner Performance

Heat equation, 1st Lanczos iteration, 2nd AMEn (outer) iteration, GMRES for inner iteration





Preconditioner Performance

Convection-diffusion problem, 1st Lanczos iteration, 2nd AMEn (outer) iteration, GMRES for inner iteration

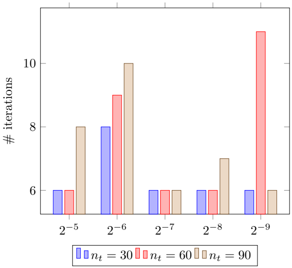


Figure: $\nu = 10^{-1}$

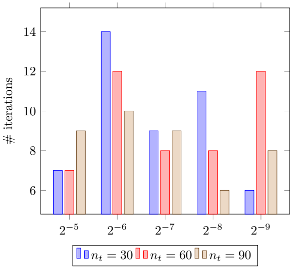


Figure: $\nu = 10^{-2}$

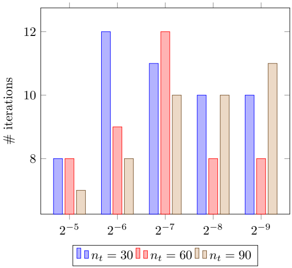


Figure: $\nu = 10^{-3}$

Preconditioner Performance

Convection-diffusion problem, 1st Lanczos iteration, 2nd AMEn (outer) iteration, GMRES for inner iteration

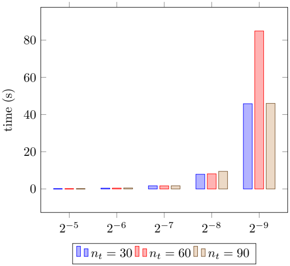


Figure: $\nu = 10^{-1}$

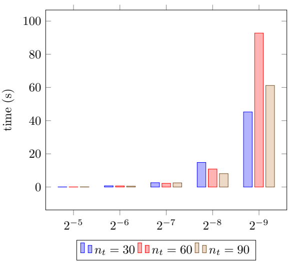


Figure: $\nu = 10^{-2}$

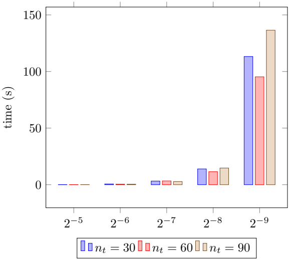


Figure: $\nu = 10^{-3}$

Conclusions & Remarks

- Low-rank method enables to solve even larger problems;
- Maximum rank remain bounded with the increase of n_x and n_t;
- Block diagonal preconditioner + AMEn solver give satisfactory performance;
- Extending to nonlinear inverse problem is challenging (non Gaussian).

Bibliographies

- H. P. Flath, L. C. Wilcox, V. Akelik, J. Hill, B. van Bloemen Waanders, and O. Ghattas, Fast algorithms for bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations, SIAM Journal on Scientific Computing 33 (2011), no. 1, 407 - 432.
- M. Stoll and T. Breiten, A low-rank in time approach to PDE-constrained optimization, SIAM Journal on Scientific Computing 37 (2015), no. 1, B1 - B29.
- P. Benner, Y. Qiu, and M. Stoll. Low-rank computations of posterior covariance matrices in Bayesian inverse problems, arXiv:1703.05638.
- TT-Toolbox, `https://github.com/oseledets/TT-Toolbox`