Probabilistic Linear Solvers

Jon Cockayne, Chris Oates, Ilse Ipsen, Mark Girolami

May 29, 2019

Solving Linear Systems

The Problem

Goal: find $\boldsymbol{x}^*$ in

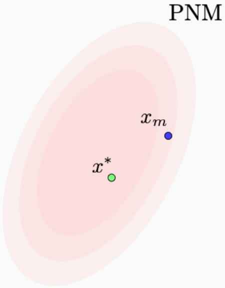
$$A\boldsymbol{x}^* = \boldsymbol{b}$$

$A \in \mathbb{R}^{d \times d}$ invertible (not necessarily SPD).

$\boldsymbol{x}^*, \boldsymbol{b} \in \mathbb{R}^d$.

Probabilistic Numerical Methods

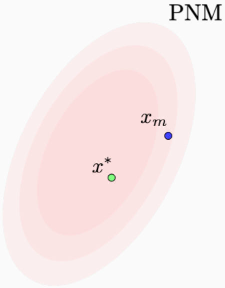
Numerical methods that return probability measures.



Probabilistic Numerical Methods

Numerical methods that return probability measures.

Those measures are designed to describe where the truth might lie given the computational effort expended.



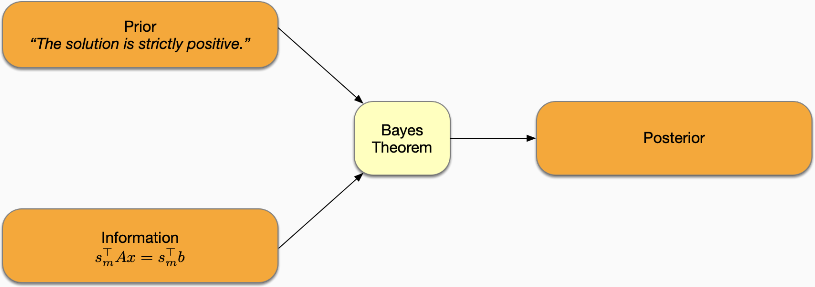
Why Use PNM?

- Contemporary numerical problems involve composition of many base numerical methods into pipelines.

Why Use PNM?

- Contemporary numerical problems involve composition of many base numerical methods into pipelines.
- PNM can be straightforwardly composed under mild conditions.

All of Bayesian Statistics in One Slide



BayesCG

Probabilistic Linear Solvers

- Start with a Gaussian prior

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}_0, \Sigma_0)$$

Probabilistic Linear Solvers

- Start with a Gaussian prior

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}_0, \Sigma_0)$$

- Use the data provided by some set of search directions to construct the posterior:

$$\boldsymbol{s}_m^\top A \boldsymbol{x}^* = \boldsymbol{s}_m^\top \boldsymbol{b} := y_m$$

Probabilistic Linear Solvers

- Start with a Gaussian prior

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{x}_0, \Sigma_0)$$

- Use the data provided by some set of search directions to construct the posterior:

$$\boldsymbol{s}_m^\top A \boldsymbol{x}^* = \boldsymbol{s}_m^\top \boldsymbol{b} := y_m$$

Let

$$S_m = \begin{pmatrix} \boldsymbol{s}_1 & \cdots & \boldsymbol{s}_m \end{pmatrix}$$

$$
\begin{aligned}
\boldsymbol{x} | \boldsymbol{y}_m &\sim \mathcal{N}(\boldsymbol{x}_m, \Sigma_m) \\
\boldsymbol{x}_m &= \boldsymbol{x}_0 + \Sigma_0 A^\top S_m \Lambda_m^{-1} S_m^\top (\boldsymbol{b} - A\boldsymbol{x}_0) \\
\Sigma_m &= \Sigma_0 - \Sigma_0 A^\top S_m \Lambda_m^{-1} S_m^\top A \Sigma_0
\end{aligned}
$$

$$
\Lambda_m = S_m^\top A \Sigma_0 A^\top S_m
$$

A Problem

To compute the posterior we must invert

$$\Lambda_m = S_m^\top A \Sigma_0 A^\top S_m$$

A Problem

To compute the posterior we must invert

$$\Lambda_m = S_m^\top A \Sigma_0 A^\top S_m$$

However

$$(\Lambda_m)_{ij} = \langle \boldsymbol{s}_i, \boldsymbol{s}_j \rangle_{A\Sigma_0 A^\top}$$

Choosing $A\Sigma_0 A^\top$-orthonormal search directions makes this more practical.

Theorem (BayesCG)

Let

$$\tilde{\boldsymbol{s}}_m = \boldsymbol{r}_{m-1} - \langle \boldsymbol{s}_{m-1}, \boldsymbol{r}_{m-1} \rangle_{A\Sigma_0 A^\top} \cdot \boldsymbol{s}_{m-1}$$

Then after normalisation the directions $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$ *are* $A\Sigma_0 A^\top$*-orthonormal.*

Theorem (BayesCG)

Let

$$\tilde{\boldsymbol{s}}_m = \boldsymbol{r}_{m-1} - \langle \boldsymbol{s}_{m-1}, \boldsymbol{r}_{m-1} \rangle_{A\Sigma_0 A^\top} \cdot \boldsymbol{s}_{m-1}$$

Then after normalisation the directions $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_m$ *are* $A\Sigma_0 A^\top$*-orthonormal.*

Furthermore we have

$$\boldsymbol{x}_m = \boldsymbol{x}_{m-1} + \Sigma_0 A^\top \boldsymbol{s}_m (\boldsymbol{s}_m^\top \boldsymbol{r}_{m-1})$$
$$\Sigma_m = \Sigma_{m-1} - \Sigma_0 A^\top \boldsymbol{s}_m \boldsymbol{s}_m^\top A \Sigma_0$$

Cost

- $\mathcal{O}(md^2)$ computation. (2-3 matrix-vector multiplications per-iter).
- $\mathcal{O}(md)$ storage (need to store search directions).

Cost

- $\mathcal{O}(md^2)$ computation. (2-3 matrix-vector multiplications per-iter).
- $\mathcal{O}(md)$ storage (need to store search directions).

More costly than CG, but comes with UQ.

Theoretical Results I

Theorem (Krylov Subspace Method)

Let

$$K_m^* = \boldsymbol{x}_0 + \Sigma_0 A^\top K_m(A\Sigma_0 A^\top, \boldsymbol{r}_0)$$

Theoretical Results I

Theorem (Krylov Subspace Method)

Let

$$K_m^* = \boldsymbol{x}_0 + \Sigma_0 A^\top K_m(A\Sigma_0 A^\top, \boldsymbol{r}_0)$$

Then the BayesCG posterior mean satisfies

$$\boldsymbol{x}_m = \underset{\boldsymbol{x} \in K_m^*}{\arg\min} \left\| \boldsymbol{x} - \boldsymbol{x}^* \right\|_{\Sigma_0^{-1}}$$

Theoretical Results I

Theorem (Krylov Subspace Method)

Let

$$K_m^* = \boldsymbol{x}_0 + \Sigma_0 A^\top K_m(A\Sigma_0 A^\top, \boldsymbol{r}_0)$$

Then the BayesCG posterior mean satisfies

$$\boldsymbol{x}_m = \arg\min_{\boldsymbol{x} \in K_m^*} \|\boldsymbol{x} - \boldsymbol{x}^*\|_{\Sigma_0^{-1}}$$

Note that $\Sigma_0 = A^{-1}$ replicates CG!

Theoretical Results II

Theorem (Convergence Rate)

$$\frac{\|\boldsymbol{x}_m - \boldsymbol{x}^*\|_{\Sigma_0^{-1}}}{\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_{\Sigma_0^{-1}}} \leq 2\left(\frac{\sqrt{\kappa(\Sigma_0 A^\top A)} - 1}{\sqrt{\kappa(\Sigma_0 A^\top A)} + 1}\right)^m$$

Theoretical Results II

Theorem (Convergence Rate)

$$\frac{\|\boldsymbol{x}_m - \boldsymbol{x}^*\|_{\Sigma_0^{-1}}}{\|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_{\Sigma_0^{-1}}} \leq 2\left(\frac{\sqrt{\kappa(\Sigma_0 A^\top A)} - 1}{\sqrt{\kappa(\Sigma_0 A^\top A)} + 1}\right)^m$$

Fastest convergence achieved when $\kappa(\Sigma_0 A^\top A) \approx 1$.

Experimental Results

Priors Considered

- $\Sigma_0 = A^{-1}$: Replicates CG.

Priors Considered

- $\Sigma_0 = A^{-1}$: Replicates CG.
- $\Sigma_0 = I$: "Uninformative".

Priors Considered

- $\Sigma_0 = A^{-1}$: Replicates CG.
- $\Sigma_0 = I$: "Uninformative".
- A-Priori Optimal Directions: Essentially random.

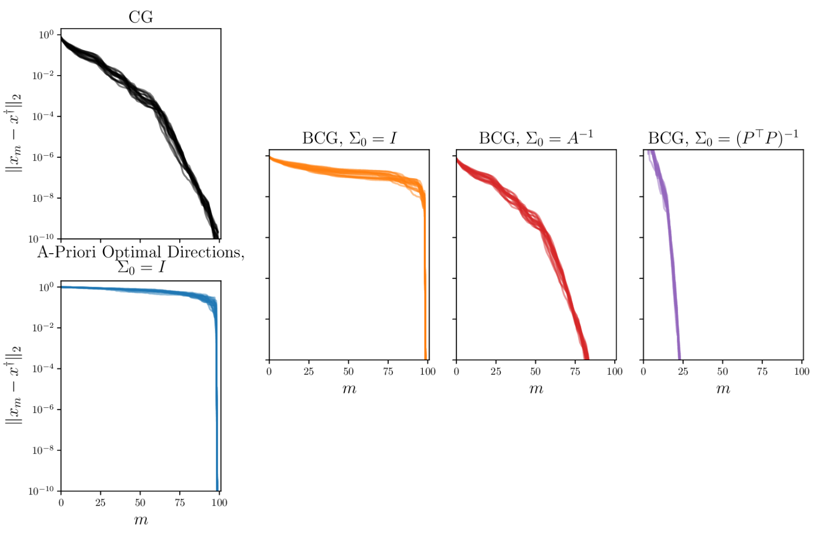
Priors Considered

- $\Sigma_0 = A^{-1}$: Replicates CG.
- $\Sigma_0 = I$: "Uninformative".
- A-Priori Optimal Directions: Essentially random.
- Preconditioner Prior: Given a preconditioner P for A, set $\Sigma_0 = (P^\top P)^{-1}$.

Experimental Setup

- A a random sparse matrix (drawn using the matlab function `sprandsym`).
- $d = 100$.
- Many test problems $\boldsymbol{x}^*$ are drawn from $\mathcal{N}(\mathbf{0}, I)$.
- BayesCG applied to $m = 100$.

Convergence of Posterior Mean



Assessment of Posterior UQ

To assess the UQ we make the ansatz that if the posterior is "well-calibrated" then $\boldsymbol{x}^*$ should look like a draw from the posterior.

Assessment of Posterior UQ

To assess the UQ we make the ansatz that if the posterior is "well-calibrated" then $\boldsymbol{x}^*$ should look like a draw from the posterior.

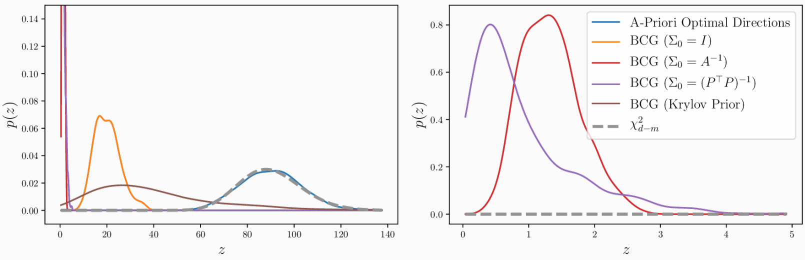
Then for the Z-statistic:

$$Z(\boldsymbol{x}^*) := \|\boldsymbol{x}^* - \boldsymbol{x}_m\|^2_{\Sigma_m^\dagger}$$

we can prove that under the ansatz:

$$Z(\boldsymbol{x}^*) \sim \chi^2_{d-m}$$

Assessment of Posterior UQ



Conclusions

Current Developments

- Stability properties in finite-precision.

Current Developments

- Stability properties in finite-precision.
- Accelerating convergence while obtaining better UQ:
    - Further work on the Krylov prior.
    - "Pushforward" methods.

Questions?