Approximate Solution of Large-scale Linear Inverse Problems with Monte Carlo Simulation

Nick Polydorides\textsuperscript{1}  D.P. Bertsekas\textsuperscript{2}  M. Wang\textsuperscript{2}

\textsuperscript{1}The Cyprus Institute,

\textsuperscript{2}Massachusetts Institute of Technology.

August 15, 2010
Introduction & Motivation
  Inverse problems

The New Framework for Large-Scale Problems
  Approximation – Simulation – Regularization
  Everything but the simulation

Simulated Matrix Algebra
  Design of sampling distributions

Results & Conclusions
Ill-posed Inverse Problems

Input $\rightarrow$ [Model(p)] $\rightarrow$ ‘Observation’

- From indirect observations infer model parameters.
- Models are linear/nonlinear in differential/integral form.
- Impact of noise on solution existence, uniqueness, stability.
- Applied in geosciences, biomedical imaging, industrial NDT.
- Bayesian inference: Find the posterior density of the unknown conditioned on the observations. MC Simulation/Integration
Integral equations of the first kind


- In continuum,

\[ b(t) = \int_0^1 ds \ A(s, t)x(s) + \epsilon, \]

with integral operator compact.

- In discrete, approximated on a uniform 1d grid of resolution \( n^{-1} \to 0 \)

\[ b = Ax + \epsilon. \]

\( A \in \mathbb{R}^{n \times n} \) dense, ill-conditioned, of smooth structure.
Main phases of methodology

- Initial hd problem

\[ x^* = \arg \min_x \{ \|Ax - b\|^2 \} \]

- Approximate unknown in ld subspace

\[ x \approx \Phi r \]

- Use simulation to estimate \( G = \Phi' A' A \Phi \) and \( c = \Phi' A' b \) in

\[ \hat{c} = \hat{G} \ r + \text{simul. error} + \text{approx. error} + \epsilon \]

- Final ld regularization problem (MAP – Gaussian model assumption)

\[ r^* = \arg \min_r \{ \| \hat{G} r - \hat{c} \|_{\Sigma^{-1}}^2 + \| r - \bar{r} \|_{\Sigma_r^{-1}}^2 \} \]
What to expect: Probing the solution error

Figure: The approximation ($e_1$), simulation ($e_2$) and numerical ($e_3$) errors affecting the solution. Π is the projection mapping from $\mathbb{R}^n$ to $S$, and $r^*(G, c)$ is the calculated and $\hat{r}(\hat{G}, \hat{c})$ the simulation-based ld solution.
The target of simulation

- Notice that the elements of the symmetric $G$ and the vector $c$ are 3d sums

$$G_{k,w} = \phi_k' A' A \phi_w = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} A_{i,j} \Phi_{j,k} \right) \left( \sum_{\tilde{j}=1}^{n} A_{i,\tilde{j}} \Phi_{\tilde{j},w} \right),$$

$$c_k = \phi_k' A' b = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} A_{i,j} \Phi_{j,k} \right) b_i$$

needing $n^3$ and $n^2$ additions respectively. If $n \sim O(10^9)$ ???

- Proposed simulation-based algorithm has numerical complexity independent of $n!$
Simulation instead of Calculation

- Suppose $\hat{G}$ and $\hat{c}$ are estimators of $G$ and $c$ respectively, simulated element-by-element independently,

- Let $\nu_{G_{kw}} = \text{var}(\hat{G}_{kw})$ and $\nu_{c_k} = \text{var}(\hat{c}_k)$ sample-based, then

  \[ \Sigma(r) = \text{diag}(\nu_G r^2) + \text{diag}(\nu_c) \]

- For large sample numbers, CLT implies the error $\hat{G}r^* - \hat{c}$ approaches zero mean Gaussian with covariance $\Sigma(r) > 0$.

- Case is suitable for Bayesian inference under Gaussian model and data uncertainty.
Ill-posed integral eqs. have smooth kernels

Implication: Matrix $A$ has smooth structure.

Figure: The kernels of some classical Fredholm integral eqs. of the first kind: heat, gravity, 2nd derivative and the Fox-Goodwin equation, discretized on a grid of dimension 1000.
Sampling with Monte Carlo

- Instead of computing

\[ G_{k,w} = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} A_{i,j} \Phi_{j,k} \right) \left( \sum_{\bar{j}=1}^{n} A_{i,\bar{j}} \Phi_{\bar{j},w} \right), \]

estimate

\[ \hat{G}_{k,w} = \frac{1}{T} \sum_{t=1}^{T} \frac{A_{i_t,j_t}\Phi_{j_t,k}\Phi_{\bar{j}_t,w}}{n^{-1}}, \quad k = 1, \ldots, s \quad w = k, \ldots, s \]

and the variance statistic \( \nu_{G_{kw}} \), where \((i_t,j_t,\bar{j}_t) \in \mathbb{N}^3\) are uniformly sampled indices from \([1, \ldots, n]^3\).

- Repeat as appropriate for \( \hat{c}_k, k = 1, \ldots, s \).
Variance Reduction with Importance Sampling

- Instead design an optimal importance sampling distribution customized for $G_{k,w}$, or $c_k$,

- The optimal $\xi^* : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$ is

$$
\xi^*_{G_{k,w}}(i,j,\bar{j}) \propto (\Phi_{j,k} \| A_j \|_1)(\Phi_{\bar{j},w} \| A_{\bar{j}} \|_1) \frac{A_{i,j} A_{i,\bar{j}}}{\| A_j \|_1 \| A_{\bar{j}} \|_1}
$$

where $A_j$ is the j’th column of $A$ and $\| A_j \|_1 = \sum_{i=1}^n |A_{i,j}|$. 
Variance Reduction with Importance Sampling

- How to sample a 3D distribution:

\[ \tilde{\xi}^*(i, j, \bar{j}) = \xi(\bar{j}|i, j)\xi(i, j) = \xi(\bar{j}|i, j)\xi(j|i)\xi(i) \propto G_{w,k}(i, j, \bar{j}) \]

where \( \xi(i, j) = \sum_{j=1}^{n} \xi(i, j, \bar{j}) \), and \( \xi(i) = \sum_{j=1}^{n} \xi(i, j) \).

- Evaluate \( G_{w,k}(i, j, \bar{j}) \) on a coarse grid in \([1, \ldots, n]^3\),
- Approximate \( G_{w,k} \) over \( \ell d \) polynomial bases,
- Compute approximate sums analytically,
- Scale to make sampling distributions.
IS distribution approximation in pictures.

Figure: Top row an approximation of $\xi^*$ in dimension 8, below at 20.
Figure: Estimators of $G = \Phi' A' A \Phi$ with $n = 10^6$, $s = 50$, and $\Phi$ piecewise constant basis functions. Top row results with IS and below MC. $A$ is derived from the second derivative kernel.
MC Vs IS scheme comparison: estimator variances

10 samples 100 samples 1000 samples

Figure: Variances of the elements of $G = \Phi' A' A \Phi$ with $n = 10^6$, $s = 50$, and $\Phi$ piecewise constant basis functions. Top row results with IS and below MC. $A$ is derived from the second derivative kernel.
One test example: Inverse heat conduction

Starting from the familiar elliptic pde, for \( u(y, t) \) the temperature at point \( y \) at time \( t \).

\[
\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial y^2}, \quad y \geq 0, \ t \geq 0
\]

\[
u(y, 0) = 0, \quad u(0, t) = x(t)
\]

using the Green’s function method

\[
b(t) = \int_0^T d\tau A(\tau, t)x(\tau),
\]

measured at point \( y_m \) away from the source \( y = 0 \), where

\[
A(\tau, t) = \begin{cases} 
\frac{y_m/\alpha}{\sqrt{4\pi(\tau-t)^3}} \exp\left(-\frac{(y_m/\alpha)^2}{4(\tau-t)}\right) & \text{if } 0 \leq t < \tau \leq T, \\
0 & \text{otherwise}.
\end{cases}
\]
One test example: Solution plot

Figure: Results with 50 and 100 piecewise constant basis function. Large problem dimension is $10^9$. The optimal $\xi^*$ was approximated on a linear basis. Results with $5 \times 10^4$ samples per simulated entry, and zero additive noise!
One test example: Inverse heat conduction

Figure: Simulation error metric: Trace of the covariance of $\hat{G}$ and $\hat{c}$. Results with different number of samples, matrix partitions and polynomial approximation of $\xi^*$. Tests with inverse heat problem at $n = 10^9$ and $s = 50$ piecewise constant $\Phi$. MC, $\hat{\xi}$ in pwc basis, and $\hat{\xi}$ in pwI.
Conclusion

- Simulation timings: 50-80 $\mu s$ per sample for pwc - pwq $\hat{\xi}$.
- Method robust for ‘applied’ ill-posed inverse problems
- Simulation scheme is suitable for multi-thread or parallel processing
- Can be utilized in the context of model reduction
- More analysis, error bounds and results at web.mit.edu/dimitrib/www/publ.html
- Thank you. - Questions?