Dimensionality Reduction and Uncertainty Quantification for Inverse Problems

Tristan van Leeuwen - Utrecht University, the Netherlands

SIAM CSE - MS35 -March 14, 2015

Inverse problems

Estimate parameters from noisy measurements

$$\mathbf{d}_i = F(\mathbf{m})\mathbf{q}_i + \mathbf{n}_i,$$

with

$\mathbf{d}_i$ - *observations*
F - *forward operator, typically involves a PDE solve*
$\mathbf{m}$ - *parameters*
$\mathbf{q}_i$ - *input/source*
$\mathbf{n}_i \sim \mathcal{N}(0, C)$

Non-linear data-fitting

Maximum likelihood estimation can be formulated as

$$\min_{\mathbf{m}} \sum_{i=1}^{n} \|F(\mathbf{m})\mathbf{q}_i - \mathbf{d}_i\|_{C^{-1}}^2,$$

which can be solved using a Newton-like method.

- Noise covariance may not be know a-priori
- Evaluation of the misfit and gradient requires $2n$ PDE solves.

Non-linear data-fitting

Agenda:

Estimation of the noise covariance matrix

- May give different estimate of the parameters $\mathbf{m}$,
- Important for uncertainty quantification

Non-linear data-fitting

Agenda:

Estimation of the noise covariance matrix

- May give different estimate of the parameters $\mathbf{m}$,
- Important for uncertainty quantification

Reduction of the effective number of simulations

- Random subsampling
- Exploit structure of Covariance matrix

Extended least-squares

Formulate an extended LS problem:

$$\min_{\mathbf{m},C} \log(|C|) + \sum_{i=1}^{n} \|F(\mathbf{m})\mathbf{q}_i - \mathbf{d}_i\|_{C^{-1}}^2.$$

For fixed **m** we have a closed-form solution

$$C(\mathbf{m}) = \sum_{i=1}^{n} \mathbf{r}_i(\mathbf{m})\mathbf{r}_i(\mathbf{m})^T,$$

where

$$\mathbf{r}_i(\mathbf{m}) = F(\mathbf{m})\mathbf{q}_i - \mathbf{d}_i.$$

[Aravkin et. al. 12]

Intermezzo: Variational projection

Given a twice differentiable function $g(x, y)$, define $\overline{y}(x) = \min_y g(x, y)$ and define a *reduced* function

$$f(x) = g(x, \overline{y}(x)),$$

then

$$\nabla f(x) = \nabla_x g(x, \overline{y}(x)),$$
$$\nabla^2 f(x) = \nabla_x^2 g(x, \overline{y}(x)) - \nabla_{x,y}^2 g^T \left(\nabla_y^2 g\right)^{-1} \nabla_{x,y}^2 g.$$

[Bell & Burke 08; Aravkin et. al. 12]

Extended least-squares

Define a *reduced* objective

$$f(\mathbf{m}) = \log(|C(\mathbf{m})|) + \sum_{i=1}^{n} \|F(\mathbf{m})\mathbf{q}_i - \mathbf{d}_i\|^2_{C(\mathbf{m})^{-1}},$$

with gradient

$$\nabla f(\mathbf{m}) = \sum_{i=1}^{n} DF(\mathbf{m}, \mathbf{q}_i)^T C(\mathbf{m})^{-1} \mathbf{r}_i(\mathbf{m}).$$

where

$$C(\mathbf{m}) = \sum_{i=1}^{n} \mathbf{r}_i(\mathbf{m}) \mathbf{r}_i(\mathbf{m})^T.$$

Extended least-squares

Main tasks:

1. Compute all residuals (n PDE-solves),

$$\mathbf{r}_i = F(\mathbf{m})\mathbf{q}_i - \mathbf{d}_i,$$

2. Estimate the covariance matrix,

$$C = \sum_{i=1}^{n} \mathbf{r}_i \mathbf{r}_i^T,$$

3. Compute gradient ($2n$ PDE solves)

$$\nabla f(\mathbf{m}) = \sum_{i=1}^{n} DF(\mathbf{q}_i)^T C^{-1} \mathbf{r}_i$$

Estimating the covariance

Organizing the residuals in an $m \times n$ matrix R, we have

$$C = RR^T.$$

Expanding $R = U_k \Sigma_k V_k^T$, we have $C = U_k \Sigma_k^2 U_k^T$, and

$$f = n + 2 \sum_{i=1}^{k} \log(\sigma_i),$$

$$\nabla f = \sum_{i=1}^{k} \sigma_i^{-1} DF(\widetilde{\mathbf{q}}_i)^T \mathbf{v}_i,$$

with $\widetilde{Q} = QV_k$.

Estimating the covariance

Observation:

If the covariance matrix has rank k, we need only $2k$ PDE-solves to evaluate the gradient

Question:

How do we efficiently obtain the (truncated) SVD of R ?

Intermezzo: Randomized trace estimation

Given a matrix A, we can estimate the trace

$$\mathrm{tr}\left(A^T A\right) \approx \sum_{i=1}^{k} \mathbf{w}_i A^T A \mathbf{w}_i,$$

where $\mathbf{w}_i$ is an i.i.d. random Gaussian vector.

Intermezzo: Randomized trace estimation

Given a matrix A, we can estimate the trace

$$\mathrm{tr}\left(A^T A\right) \approx \sum_{i=1}^{k} \mathbf{w}_i A^T A \mathbf{w}_i,$$

where $\mathbf{w}_i$ is an i.i.d. random Gaussian vector.

Such techniques have been very succesfull in PDE-constrained optimization/inverse problems.

[Avron & Toledo 11; Haber et. al. '12]

Randomized linear algebra

First approach:

1. Compress the matrix by random projection:

$$\widetilde{R} = RW_k,$$

   where $\mathbb{E}\{W_k W_k^T\} = I_n$.

2. Compute k-SVD: $\widetilde{R} = \widetilde{U}_k \widetilde{\Sigma}_k \widetilde{V}_k^T$, and find

$$RR^T \approx \widetilde{U}_k \widetilde{\Sigma}_k^2 \widetilde{U}_k^T.$$

Cost: k PDE solves + k-SVD of $m \times k$ matrix.

Randomized linear algebra

Second approach:

1. Capture range of the matrix $Y = RW_k$ and find orthonormal basis L for Y.
2. Compute k-SVD of $L^T R = \widetilde{U}_k \widetilde{\Sigma}_k V_k^T$ and find

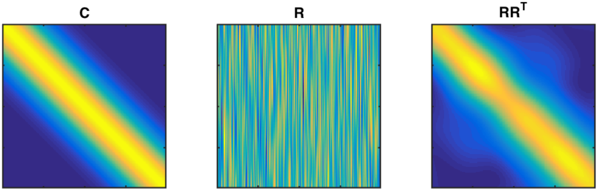
$$RR^T \approx L\widetilde{U}_k \widetilde{\Sigma}_k^2 V_k^T.$$

Cost: $2k$ PDE-solves + QR of $m \times k$ matrix + k-SVD of $k \times n$ matrix.

[Halko et. al. 11]

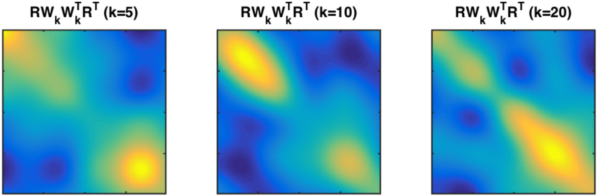
Randomized linear algebra

True covariance matrix



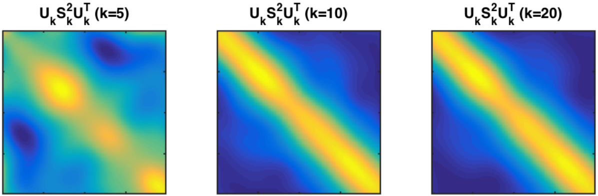
Randomized linear algebra

Stochastic approximation

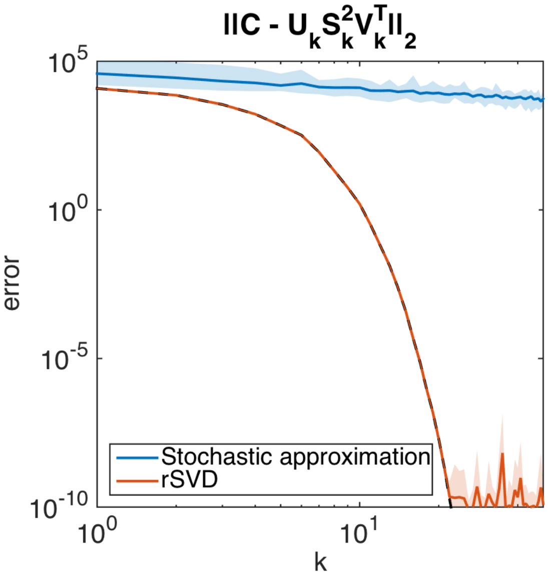


Randomized linear algebra

Randomized SVD



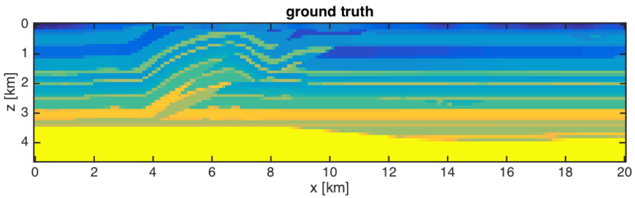
Randomized linear algebra

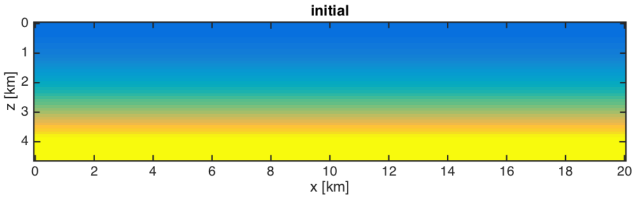


Numerical results

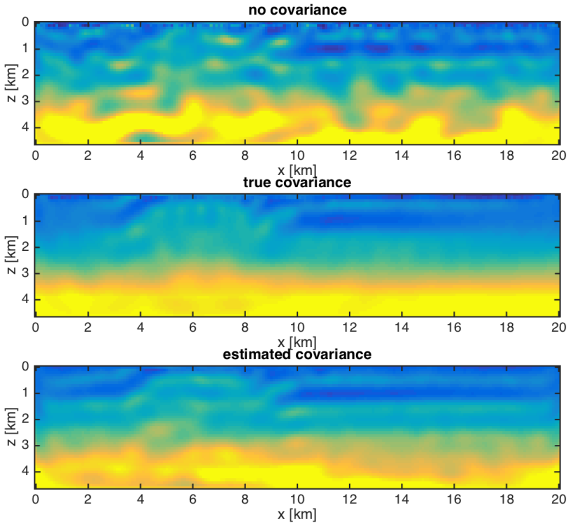
- PDE: 2D Helmholtz equation
- gradient-descent with Borzilai-Borwein steplength
- estimate covariance on-the-fly

Numerical results

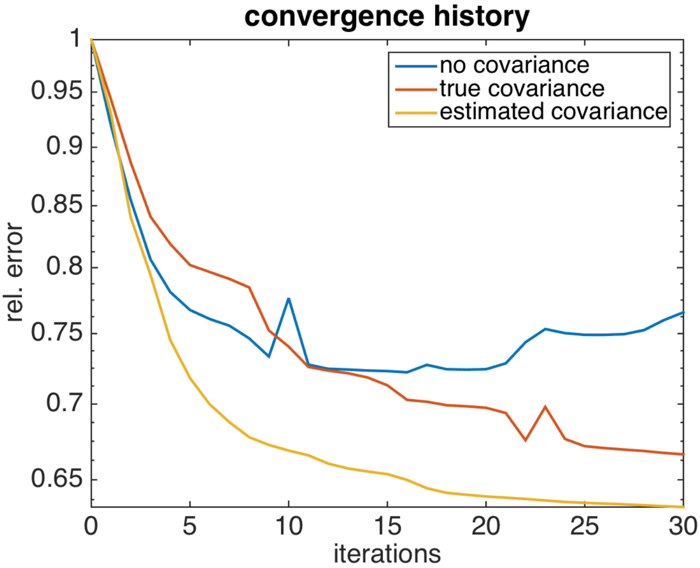




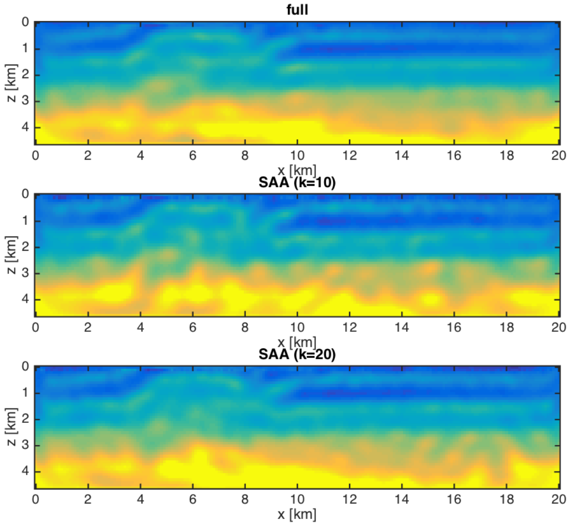
Numerical results



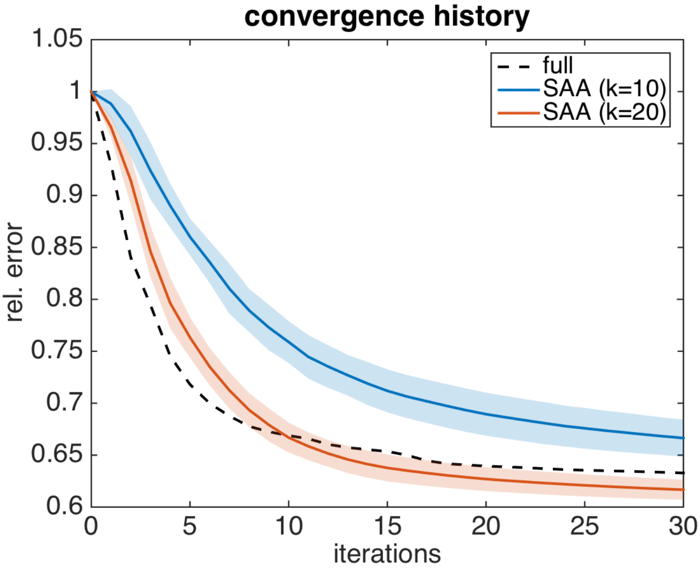
Numerical results



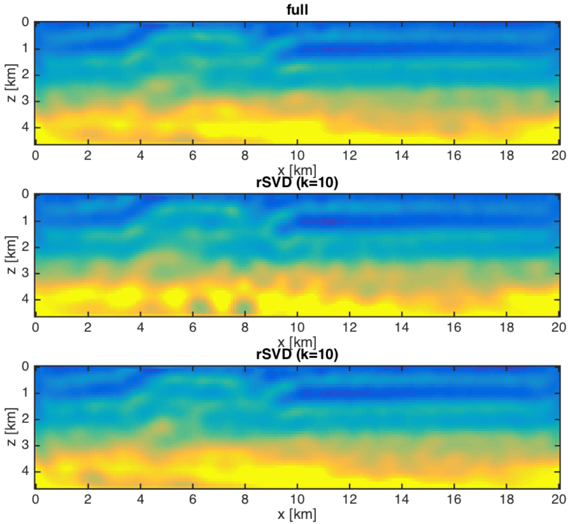
Numerical results



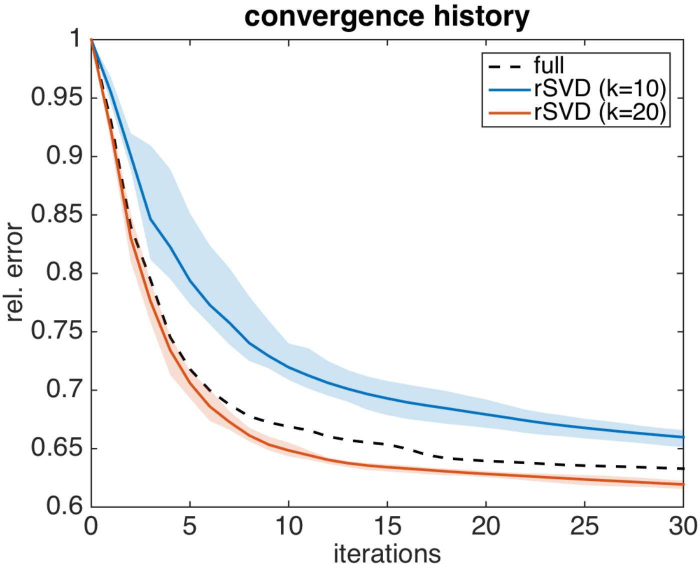
Numerical results



Numerical results



Numerical results



Conclusions

- We can exploit low-rank structure of the covariance matrix to reduce the # of PDE solves
- Low rank estimate of the covariance matrix can be computed on-the-fly using stochastic approximation or randomized SVD
- First results are promising, SA does remarkably well

Future work

- Adaptive estimation of the rank
- More sophisticated model for covariance matrix (diagonal + low rank)
- Exploit low rank structure of C to represent the GN Hessian