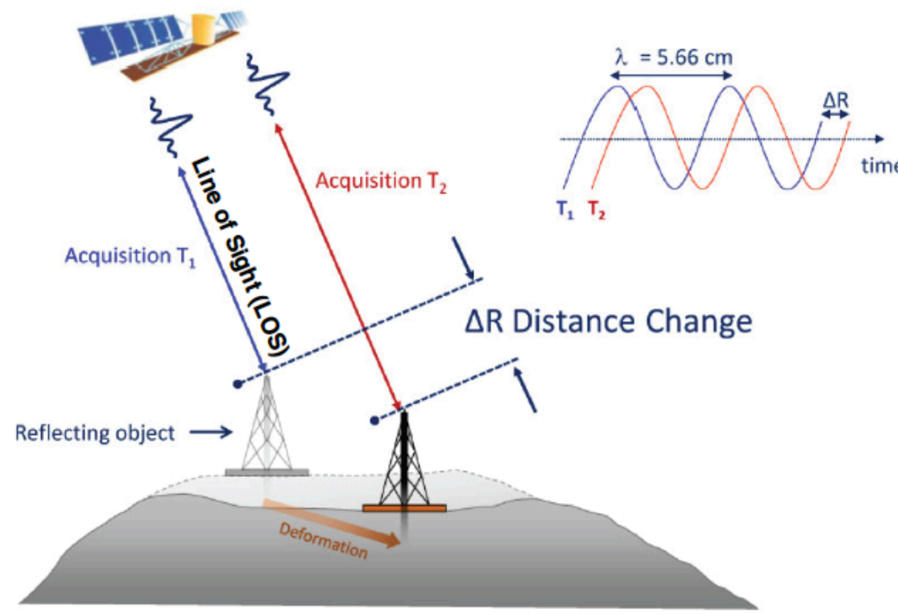
Subspace Driven Data Reduction Strategies for Linear Bayesian Inverse Problems

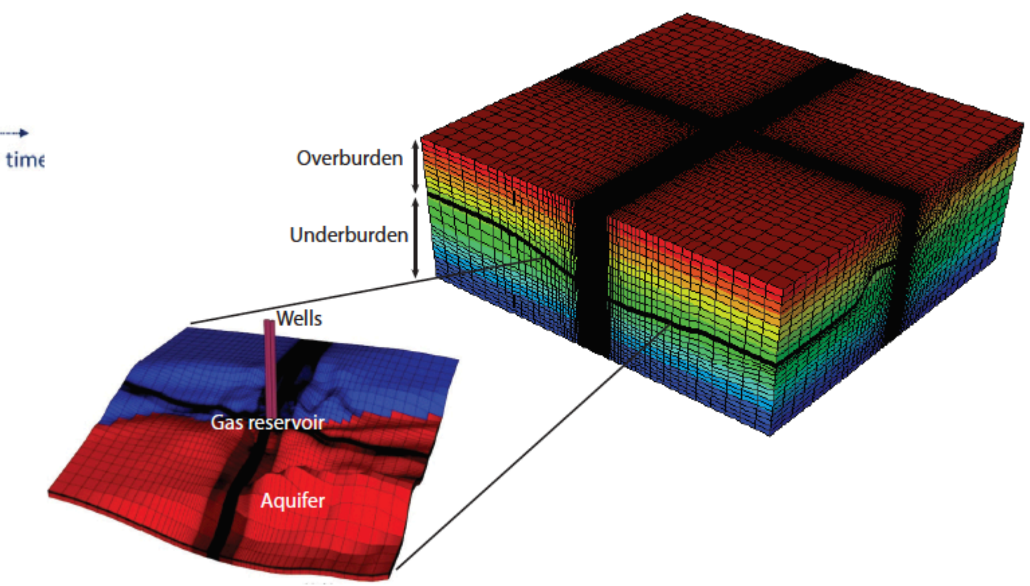
2017 SIAM Conference on Computational Science
Atlanta, Georgia, USA

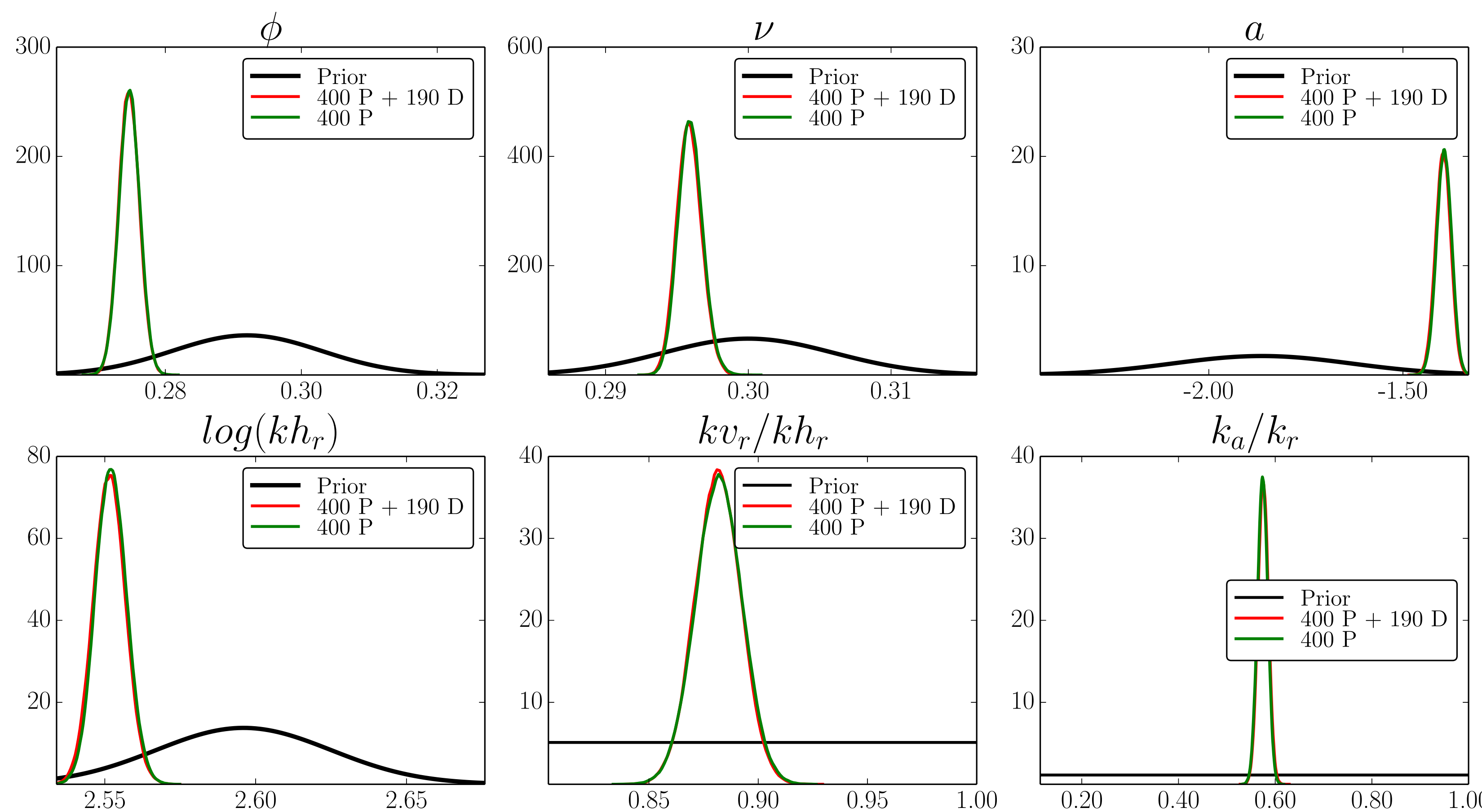
Jayanth Jagalur-Mohan
Youssef Marzouk

Uncertainty Quantification Group
Aerospace Computational Design Laboratory
Department of Aeronautics and Astronautics, MIT

Reservoir characterization using massive amount of flow and geodetic data - A motivating example







ϕ = Reservoir Porosity
ν = Poisson Ratio
a = Compressibility intercept
$log(kh_r)$ = Reservoir Horizontal Permeability
kv_r/kh_r = Ratio of Vertical to horizontal permeability
k_a/k_r = Ratio of Aquifer to reservoir permeability

Schematic diagrams courtesy of Jha *etal*, Int. J. Numer. Anal. Meth. Geomech. 2014

Outline

- Motivation ✓ Done!
- Notation and problem conceptualization
- Structure of optimal approximations of BLIP
- Criteria for quasi-optimal observation selection
- Illustrative examples

Linear Bayesian inverse problem

$$y = \mathcal{G}x + \epsilon, \quad x \sim \mathcal{N}(0, \Gamma_{Pr}), \quad \epsilon \sim \mathcal{N}(0, \Gamma_{Obs})$$

$x \in \mathbb{R}^N$ — Parameters to be inferred

$y \in \mathbb{R}^M$ — Observations

$G : \mathbb{R}^N \to \mathbb{R}^M$ — Linear forward operator

$\Gamma_{Pr}, \Gamma_{Obs} \succ 0$ — Non-singular covariances matrices

$$\boxed{\Gamma_{Pos} = \left(\Gamma_{Pr}^{-1} + \mathcal{G}^T \Gamma_{Obs}^{-1} \mathcal{G}\right)^{-1}}$$

$$\boxed{\mu_{Pos}(y) = \Gamma_{Pos} \mathcal{G}^T \Gamma_{Obs}^{-1} y}$$

Linear Bayesian inverse problem using reduced data

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1N} \\ G_{21} & G_{22} & \cdots & G_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ G_{i1} & G_{i2} & \cdots & G_{iN} \\ \vdots & \vdots & \vdots & \vdots \\ G_{M1} & G_{M2} & \cdots & G_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_M \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1N} \\ G_{21} & G_{22} & \cdots & G_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ G_{i1} & G_{i2} & \cdots & G_{iN} \\ \vdots & \vdots & \vdots & \vdots \\ G_{M1} & G_{M2} & \cdots & G_{MN} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_j \\ \vdots \\ x_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_i \\ \vdots \\ \epsilon_M \end{bmatrix}$$

$$y = \mathcal{G}x + \epsilon$$

$$\mathcal{P}y = \mathcal{P}\mathcal{G}x + \mathcal{P}\epsilon$$

$$\mathcal{P} : \mathbb{R}^M \to \mathbb{R}^{M'}, \qquad \mathcal{P} = \mathcal{I}(i_{M'}, :), \qquad \mathcal{P} = \begin{bmatrix} 0 & 1 & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & 1 & \cdots & 0 \end{bmatrix}, \qquad \mathcal{I} \in \mathbb{R}^{M \times M}, \qquad i_{M'} = (2, i, \cdots)$$

$$\boxed{\mu_{Pos}(y) = \Gamma_{Pos}\mathcal{G}^T\Gamma_{Obs}^{-1}y}$$

$$\boxed{\Gamma_{Pos} = \left(\Gamma_{Pr}^{-1} + \mathcal{G}^T\Gamma_{Obs}^{-1}\mathcal{G}\right)^{-1}}$$

$$\boxed{\widehat{\mu}_{Pos}(\mathcal{P}y) = \widehat{\Gamma}_{Pos}\left(\mathcal{P}\mathcal{G}\right)^T\left(\mathcal{P}\Gamma_{Obs}\mathcal{P}^T\right)^{-1}\mathcal{P}y}$$

$$\boxed{\widehat{\Gamma}_{Pos} = \left(\Gamma_{Pr}^{-1} + (\mathcal{P}\mathcal{G})^T\left(\mathcal{P}\Gamma_{Obs}\mathcal{P}^T\right)^{-1}(\mathcal{P}\mathcal{G})\right)^{-1}}$$

Remarks

- From a practical perspective we are interested in cases when $M, M' >> N$
- In the present context for the purpose of analysis we will assume that the rank of the forward operator, and its restricted counterpart is *N*.

A theoretical question: Given a constraint on the *number of observations we can afford to make in practice or utilize computationally*, can we specify which among the larger set result in the best posterior approximation ?

$$i_{M'} = \operatorname{argmin}_{\tilde{i}_{M'}} \mathfrak{D}\left(\begin{array}{c} \mu_{Pos} \\ \Gamma_{Pos} \end{array}, \begin{array}{c} \widehat{\mu}_{Pos} \\ \widehat{\Gamma}_{Pos} \end{array} \right)$$

Possible and reasonable choices for $\mathfrak{D}$ which render the problem combinatorial.

- $\mathbb{E}[||\mu_{Pos} - \widehat{\mu}_{Pos}||_S^2]$
  - *S* is a useful weighting matrix
- $d\left(\Gamma_{Pos}, \widehat{\Gamma}_{Pos}\right) = \sum_i ln^2(\sigma_i), \quad \sigma_i = \operatorname{eig}(\Gamma_{Pos}, \widehat{\Gamma}_{Pos})$
  - *d* is a metric on the manifold *S.P.D* matrices [#]

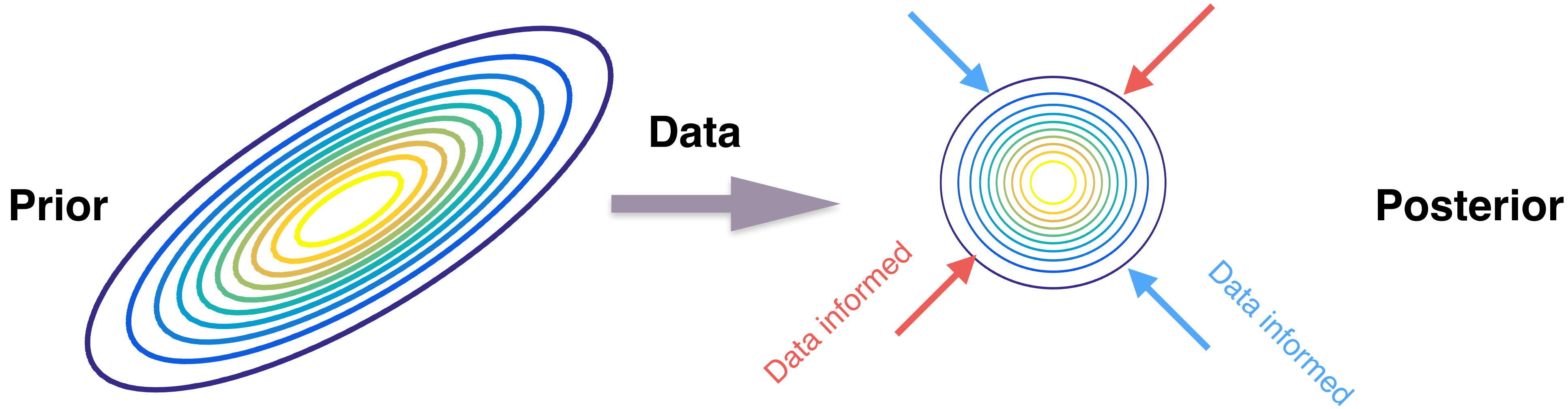
A practical question: Can we obtain a *quasi-optimal subset* of observations by solving some relevant *non-combinatorial problem*, but which approximates the posterior to a sufficient degree ? **Yes we can!**

[#] *Bhatia 2007 Princeton University Press; Förstner and Moonen 2003 Springer Berlin Heidelberg.*

Outline

- Motivation ✓ Done!
- Notation and problem conceptualization
- Structure of optimal approximations of BLIP
- Criteria for quasi-optimal observation selection
- Illustrative examples

Optimal low-rank approximations of Bayesian linear inverse problems



$$\Gamma'_{Pos} = \Gamma_{Pr} - K_r K_r^T, \qquad \overbrace{K_r K_r^T = \sum_{i=1}^{r} \frac{\sigma_i^2}{1+\sigma_i^2} v_i v_i^T}^{\text{Low rank update}}$$

$\left(\sigma_i^2, v_i\right)$ are the generalized eigenvalue-eigenvector pairs of the pencil $\left(H, \Gamma_{pr}^{-1}\right)$

$$\underbrace{\mathcal{G}^T \Gamma_{Obs}^{-1} \mathcal{G}}_{H} v_i = \sigma_i^2 \Gamma_{pr}^{-1} v_i, \qquad \text{with} \quad \sigma_i^2 \geq \sigma_{i+1}^2$$

H → Hessian of the negative log-likelihood

Spantini etal, 2015 SIAM J. Sci. Comput.

A generalized eigenvalue problem on the observation space

$$\mathcal{G}\Gamma_{Pr}\mathcal{G}^T u_i = \sigma_i^2 \Gamma_{Obs} u_i \quad \text{with} \quad \sigma_i^2 \geq \sigma_{i+1}^2$$

$\left(\sigma_i^2, u_i\right)$ are the generalized eigenvalue-eigenvector pairs of the pencil $\left(\mathcal{G}\Gamma_{Pr}\mathcal{G}^T, \Gamma_{Obs}\right)$

Remarks

- Both the *GEVPs* are related !

$$\Gamma'_{Pos} = \Gamma_{Pr} - K_r K_r^T, \qquad \overbrace{K_r K_r^T = \sum_{i=1}^{r} \frac{1}{1+\sigma_i^2} \breve{u}_i \breve{u}_i^T}^{\text{Low rank update}}, \qquad \breve{u}_i = \Gamma_{Pr}\mathcal{G}^T u_i$$

- The *GEVP* on the restricted observation space yields eigenvalues that are always lower in comparison with the full space.

$$(\mathcal{P}\mathcal{G})\,\Gamma_{Pr}\,(\mathcal{P}\mathcal{G})^T\,\widehat{u}_i = \widehat{\sigma}_i^2\left(\mathcal{P}\Gamma_{Obs}\mathcal{P}^T\right)\widehat{u}_i \quad \text{with} \quad \widehat{\sigma}_i^2 \geq \widehat{\sigma}_{i+1}^2$$

$$\sigma_i^2 \geq \widehat{\sigma}_i^2$$

Pf: An immediate consequence of Cauchy interlacing theorem for matrix pencils#

Notation and assumptions going forward: We will assume a block uncorrelated structure for the observations and perform our analysis on a two block system.

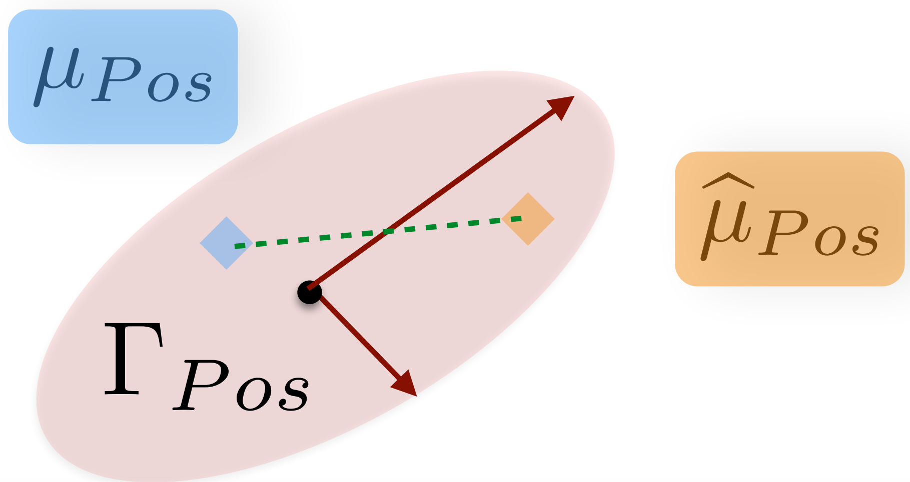
$$\mathcal{G} = \begin{bmatrix} \mathcal{G}_{(1)} \\ \mathcal{G}_{(2)} \end{bmatrix}, \text{ with } \quad \mathcal{P}\mathcal{G} = \mathcal{G}_{(1)}, \quad \text{and } \Gamma_{Obs} = \begin{bmatrix} \Gamma_{Obs,(1)} & \\ & \Gamma_{Obs,(2)} \end{bmatrix}.$$

Kressner *etal*. 2014 Numerical algorithms

Outline

- Motivation ✓ Done!
- Notation and problem conceptualization
- Structure of optimal approximations of BLIP
- Criteria for quasi-optimal observation selection
- Illustrative examples

Analyzing the error in the posterior mean approximation



We will examine the squared error loss in the norm weighted by the precision of the posterior covariance.

$$\mathbb{E}[||\mu_{Pos} - \widehat{\mu}_{Pos}||^2_{\Gamma^{-1}_{Pos}}]$$

$$\boxed{\mu_{Pos}(y) = \left(\Gamma_{Pr} - KK^T\right)\mathcal{G}^T\Gamma^{-1}_{Obs}y}$$

$$\boxed{\widehat{\mu}_{Pos}(y_{(1)}) = \left(\Gamma_{Pr} - \widehat{K}\widehat{K}^T\right)\mathcal{G}^T_{(1)}\Gamma^{-1}_{Obs,(1)}y_{(1)}}$$

$$\mu_{Pos}(y) - \widehat{\mu}_{Pos}(y_{(1)}) = \left(\Gamma_{Pr} - KK^T\right)\mathcal{G}^T_{(2)}\Gamma^{-1}_{Obs,(2)}y_{(2)} + \left(\widehat{K}\widehat{K}^T - KK^T\right)\mathcal{G}^T_{(1)}\Gamma^{-1}_{Obs,(1)}y_{(1)}$$

Error due to neglecting $y_{(2)}$

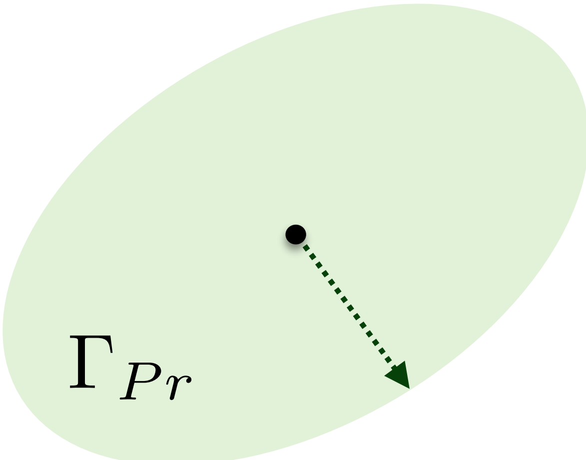
Error due to Eigenspace mismatch

$$\mathbb{E}[||(\Gamma_{Pr} - KK^T)\mathcal{G}^T_{(2)}\Gamma^{-1}_{Obs,(2)}y_{(2)}||^2_{\Gamma^{-1}_{Pos}}] = ||\Gamma^{\frac{1}{2}}_{Obs,(2)}U_{(2)}\Sigma||^2_F$$

Remarks

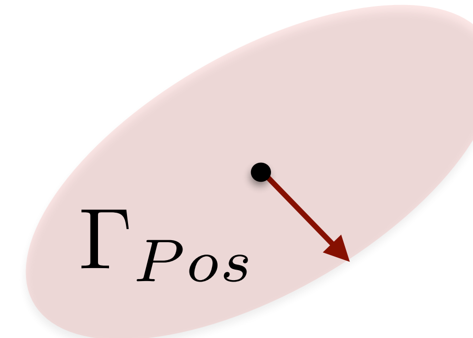
- U are the previously defined generalized eigenvectors on the observation space, and Σ is square root of the diagonal matrix of eigenvalues.
- The term $\Gamma^{\frac{1}{2}}_{Obs}U$ is *L2* orthogonal and is actually related to the whitened forward operator. It induces an appropriate weight for each observation much like leverage scores.

Bounding the cumulative change in generalized Rayleigh quotients



$$\mathcal{R}(z) = \frac{z^T \Gamma_{Pos} z}{z^T \Gamma_{Pr} z}$$

$$\text{argmin}_{z \in \mathbb{R}^N} \mathcal{R}(z) = \frac{1}{1+\sigma_1^2}$$



$$\boxed{\mathcal{G}^T \Gamma_{Obs}^{-1} \mathcal{G} V = \Gamma_{Pr}^{-1} V \Sigma^2}$$

$$\boxed{\mathcal{G}_{(1)}^T \Gamma_{Obs,(1)}^{-1} \mathcal{G}_{(1)} \widehat{V} = \Gamma_{Pr}^{-1} \widehat{V} \widehat{\Sigma}^2}$$

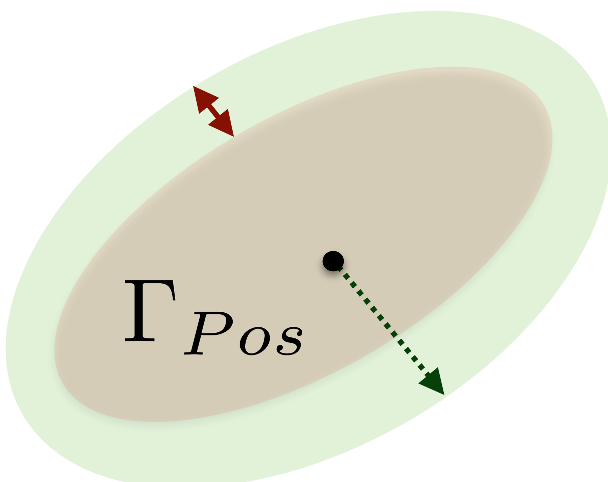
$$\sum_{i=1}^{N} \left| \frac{1}{1+\sigma_i^2} - \frac{1}{1+\widehat{\sigma}_i^2} \right|^2 \leq \sum_{i=1}^{N} |\sigma_i^2 - \widehat{\sigma}_i^2|^2 = ||\text{diag}(\sigma_1^2 - \widehat{\sigma}_1^2, \cdots, \sigma_N^2 - \widehat{\sigma}_N^2)||_F^2$$

The *Hoffman-Wielandt*[#] inequality allows us to bound the euclidean distance between the Eigenspectrums

$$\begin{aligned} ||\text{diag}(\sigma_1^2 - \widehat{\sigma}_1^2, \cdots, \sigma_N^2 - \widehat{\sigma}_N^2)||_F^2 &\leq ||\Gamma_{Pr}^{\frac{1}{2}} \left(\mathcal{G}^T \Gamma_{Obs}^{-1} \mathcal{G} - \mathcal{G}_{(1)}^T \Gamma_{Obs,(1)}^{-1} \mathcal{G}_{(1)} \right) \Gamma_{Pr}^{\frac{1}{2}} ||_F^2 \\ &\leq ||\Gamma_{Pr}^{\frac{1}{2}} \mathcal{G}_{(2)}^T \Gamma_{Obs,(2)}^{-1} \mathcal{G}_{(2)} \Gamma_{Pr}^{\frac{1}{2}} ||_F^2 \\ &\leq ||\Gamma_{Obs,(2)}^{\frac{1}{2}} \widehat{U}_{(2)} \Sigma ||_F^2 \end{aligned}$$

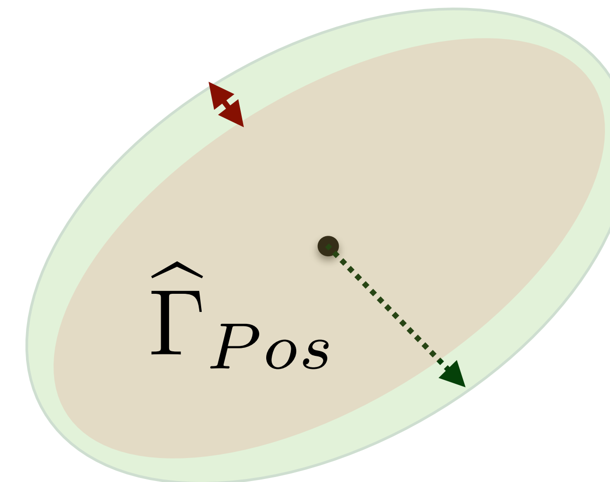
[#]*Mirsky 1960 Q. J. Math.; Li 1993 Linear Algebra Appl.*
Shah 1960, Ann. Math. Stat. ; Eccleston and Hedayat 1974, Ann. Statist

Bounding the volume defined by product of change in the generalized Rayleigh quotient



$$\Delta(z) = \frac{z^T \Gamma_{Pr} z - z^T \Gamma_{Pos} z}{z^T \Gamma_{Pr} z}$$

$$\text{argmax}_{z \in \mathbb{R}^N} \Delta(z) = \frac{\sigma_1^2}{1+\sigma_1^2}$$



$$\boxed{\mathcal{G}^T \Gamma_{Obs}^{-1} \mathcal{G} V = \Gamma_{Pr}^{-1} V \Sigma^2}$$

$$\boxed{\mathcal{G}_{(1)}^T \Gamma_{Obs,(1)}^{-1} \mathcal{G}_{(1)} \widehat{V} = \Gamma_{Pr}^{-1} \widehat{V} \widehat{\Sigma}^2}$$

$$\prod_{i=1}^{N} \frac{\sigma_i^2}{1+\sigma_i^2} - \prod_{i=1}^{N} \frac{\widehat{\sigma}_i^2}{1+\widehat{\sigma}_i^2} \leq \frac{\prod_{i=1}^{N} \sigma_i^2 - \prod_{i=1}^{N} \widehat{\sigma}_i^2}{\prod_{i=1}^{N} (\sigma_i^2)} = \frac{\det\left(\Gamma_{Pr} \mathcal{G}^T \Gamma_{Obs}^{-1} \mathcal{G}\right) - \det\left(\Gamma_{Pr} \mathcal{G}_{(1)}^T \Gamma_{Obs,(1)}^{-1} \mathcal{G}_{(1)}\right)}{\det\left(\Gamma_{Pr} \mathcal{G}^T \Gamma_{Obs}^{-1} \mathcal{G}\right)}.$$

We can bound the RHS by using some standard results for relative perturbation of the determinant[#]

$$\Rightarrow \quad \prod_{i=1}^{N} \frac{\sigma_i^2}{1+\sigma_i^2} - \prod_{i=1}^{N} \frac{\widehat{\sigma}_i^2}{1+\widehat{\sigma}_i^2} \leq \left(1 + \|\frac{1}{\sigma_{min}} \Gamma^{\frac{1}{2}}_{Obs,(2)} U_{(2)} \Sigma\|_2^2\right)^N - 1.$$

[#] *Ipsen and Rehman 2008 SIAM J. Matrix Anal. Appl.*

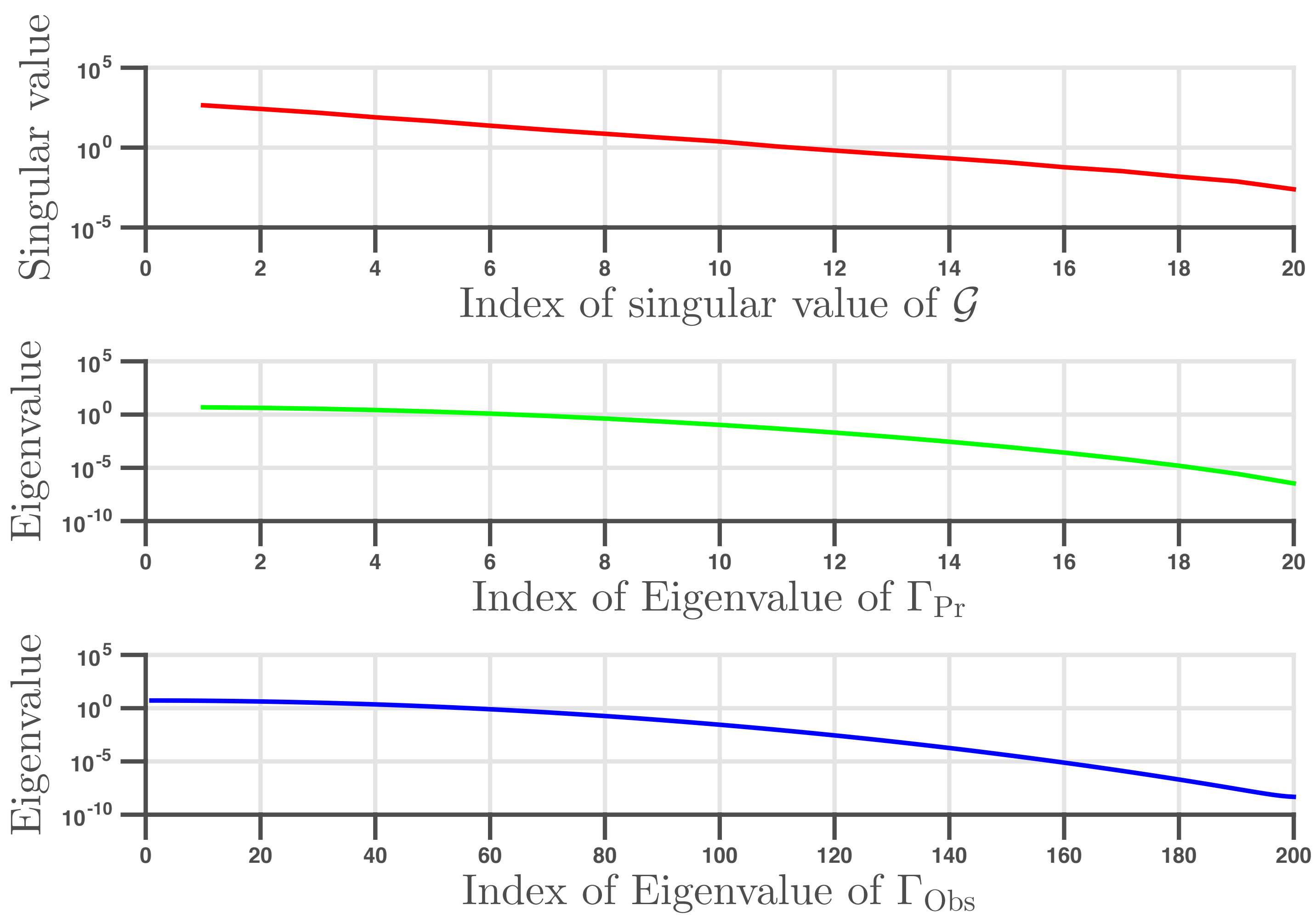
Outline

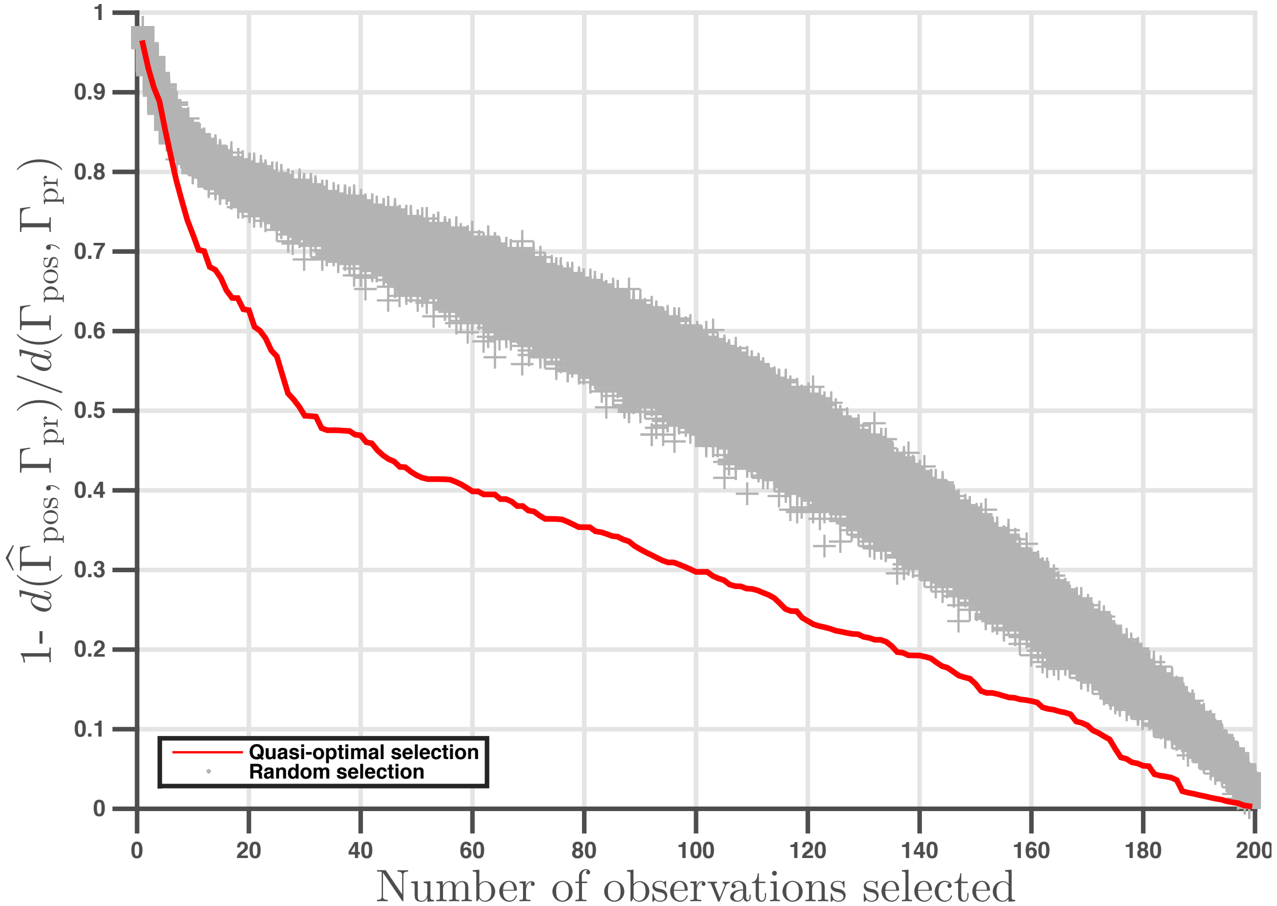
- Motivation ✓ Done!
- Notation and problem conceptualization
- Structure of optimal approximations of BLIP
- Criteria for quasi-optimal observation selection
- Illustrative examples

An example using randomly generated forward operator, with the Prior and Observation covariances defined using squared exponential kernels.

$$y = \mathcal{G}x + \epsilon, \quad x \sim \mathcal{N}(0, \Gamma_{Pr}), \quad \epsilon \sim \mathcal{N}(0, \Gamma_{Obs})$$

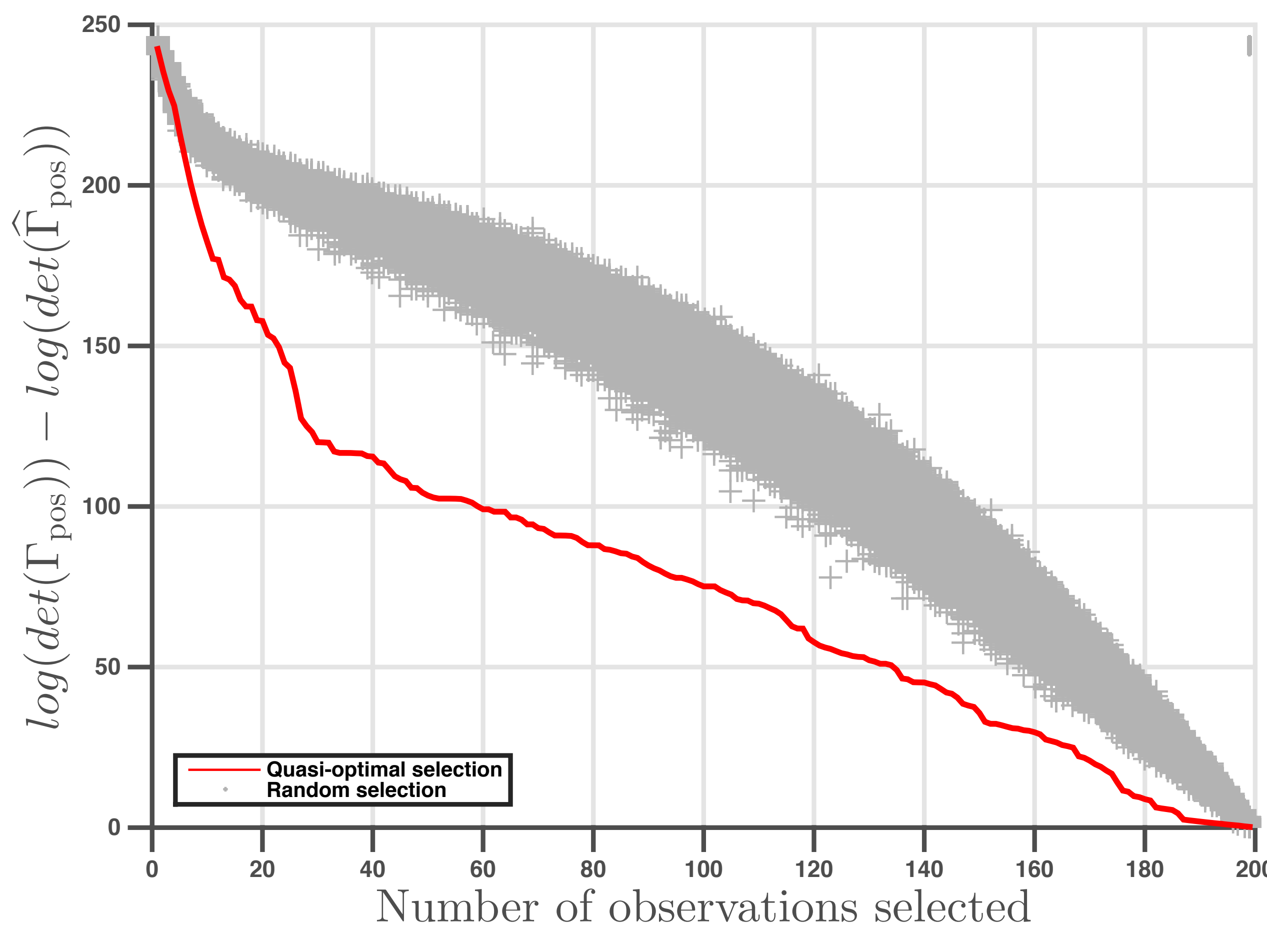
$$x \in \mathbb{R}^{20}, \quad y \in \mathbb{R}^{200}, \quad \mathcal{G} \in \mathbb{R}^{200\times 20}$$



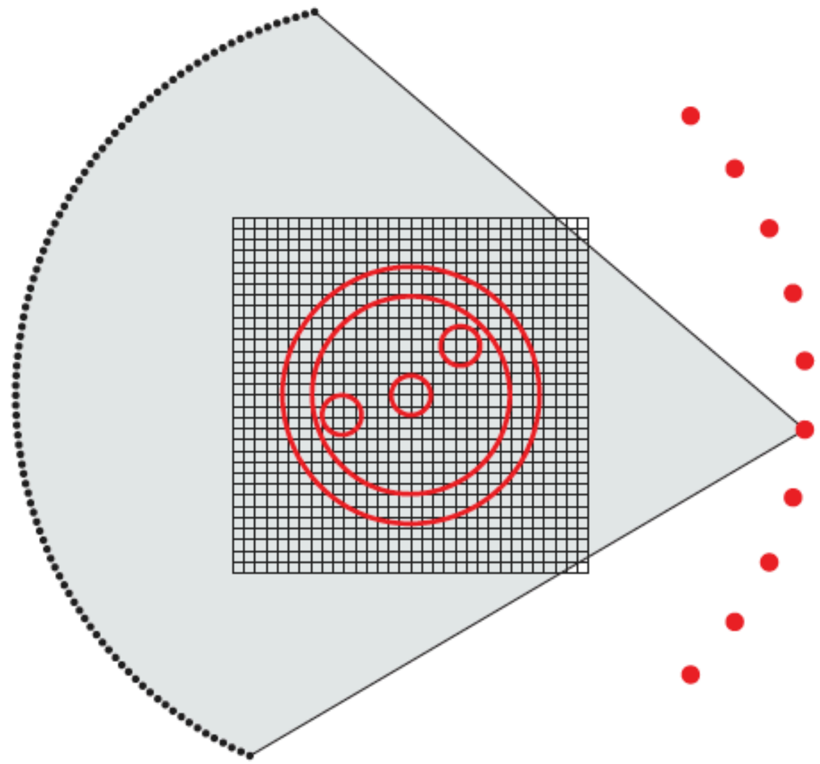


Top figure: Distance between the approximate posterior to the full posterior specified in relation to the distance between the full posterior and the prior. 2000 random samples for each value of reduced G.

Bottom figure: Difference between the log determinant of the actual posterior and the full posterior, *à la* D-optimalty. 2000 random samples for each value of reduced G.



Synthetic X-ray tomography example



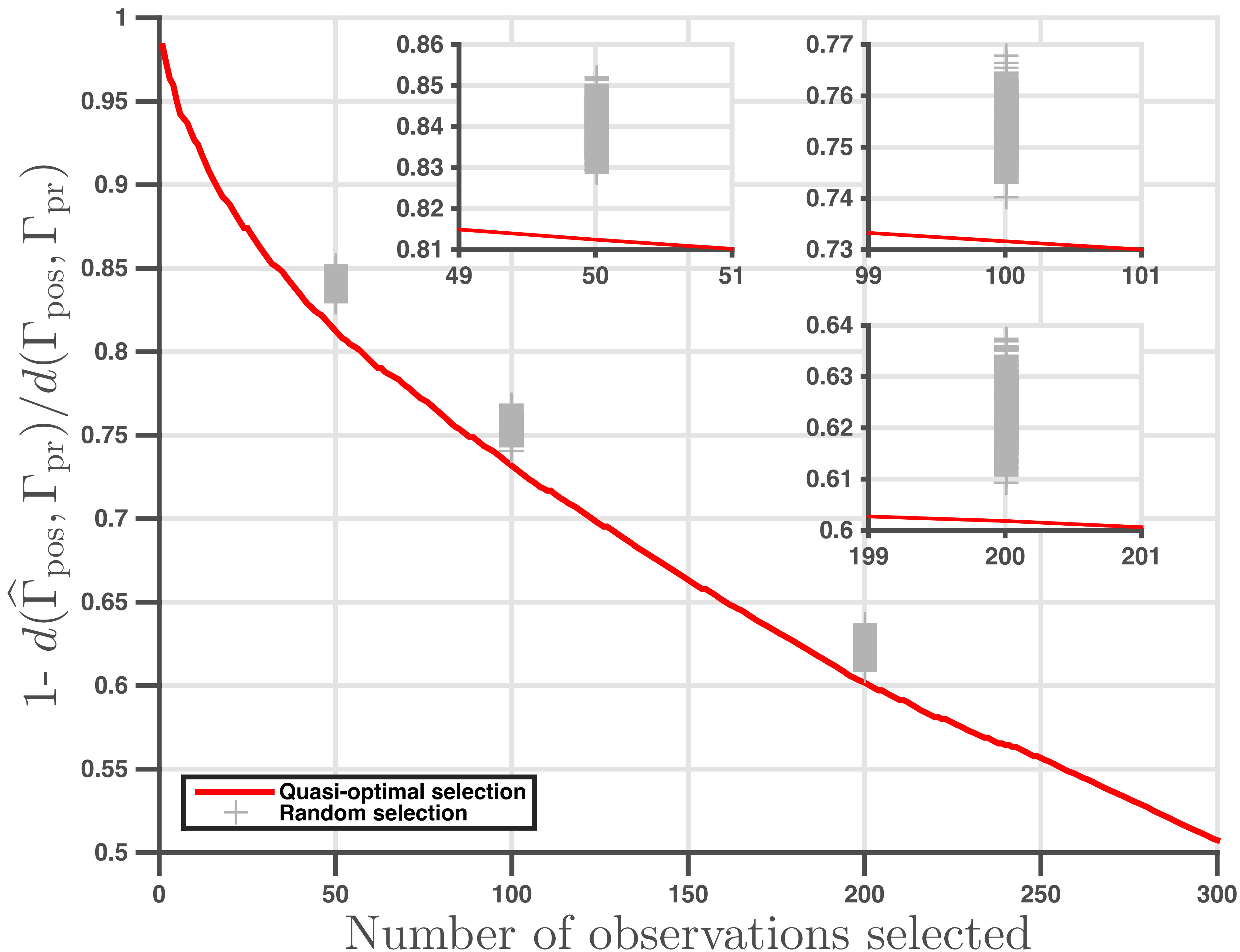
Problem setup#

- The true object consists of three circular inclusions, each of uniform density, inside an annulus.
- 10 X-ray sources are positioned on one side of a circle, and each source sends a fan of 100 X-rays that are measured by detectors on the opposite side of the object.
- The unknown density is estimated on a discretized domain with the prior specified as GRF

Figure courtesy of *Spantini etal*, 2015 SIAM J. Sci. Comput.

Courtesy of J. Heikkinen, 2008 Master's thesis, Lappeenranta Univ. of Tech.

Synthetic X-ray tomography example



Takeaway

- Quasi-optimal subset of observations can be sought leveraging information from certain subspaces on the observation side.
- The achieved ranking of observations provide a heuristic to solve the inverse problem using reduced data while approximating the posterior to a sufficient degree.

Looking ahead

- Investigating the change in Eigenspace **V** of the relevant GEVP using appropriate perturbation results.
- A more clear explanation for each entry in the Eigenvectors **U,** as relating to accounting for a reduction in variance in the posterior or an appropriate measure of information.
- Testing on more relevant challenging problems.

"Subspace driven Data reduction strategies for linear Bayesian inverse problems.", Jagalur-Mohan.J, Marzouk.Y, *in preparation*

Thank you!

SIAM Conference on
Computational Science and Engineering

February 27-March 3, 2017
Hilton Atlanta, Atlanta, Georgia, USA