Minimization-based sampling from the posterior distribution for inverse problems

Dean Oliver

21 August 2017

University of Potsdam

uni Research

Numerical models for forecasting and decisions

Given prior distribution $\mathcal{N}(\mu, C_X)$ on model parameters $x \in \mathbb{R}^{N_x}$ and observations $d^{\mathrm{o}} \in \mathbb{R}^{N_d}$ with forward map $g : \mathbb{R}^{N_x} \to \mathbb{R}^{N_d}$ for unknown x^* and unknown measurement errors $\eta \sim \mathrm{N}(0, C_D)$, i.e.

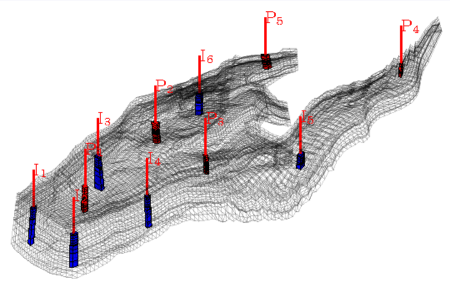
$$d^{\mathrm{o}} = g(x^*) + \eta$$

we wish to generate samples from the posterior distribution

$$\pi_X(x|d^{\mathrm{o}}) = \frac{\pi_{XD}(x, d^{\mathrm{o}})}{\pi_D(d^{\mathrm{o}})} = \frac{\exp(-L(x))}{\pi_D(d^{\mathrm{o}})}$$

with

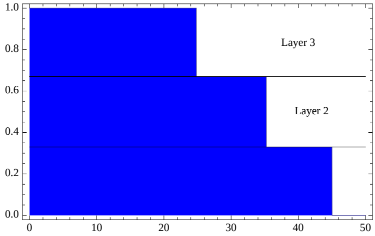
$$L(x) = \frac{1}{2}(x-\mu)^T C_X^{-1}(x-\mu) + \frac{1}{2}(g(x)-d^{\mathrm{o}})^T C_D^{-1}(g(x)-d^{\mathrm{o}}).$$



- Parameters ($N_x \approx 10^5$–10^7) are coefficients of PDEs describing flow and transport.
- Parameters generally modeled as correlated Gaussian
- Observations of the state (e.g. pressure or saturation) spatially sparse or low resolution ($N_d \approx 10^4 - 10^6$)
- Likelihood function evaluation is expensive (0.1–10 hour)

Figure from SINTEF

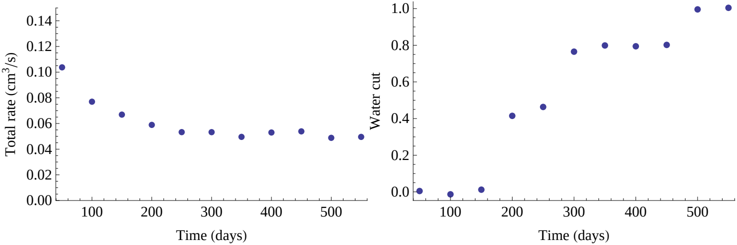
Simple three-layer flow problem[1]



- Three parameters (k_1, k_2, and k_3) to be estimated.
- Water injected at constant pressure into all three layers.
- Fluids are produced at constant pressure from all three layers.
- No vertical communication between layers.

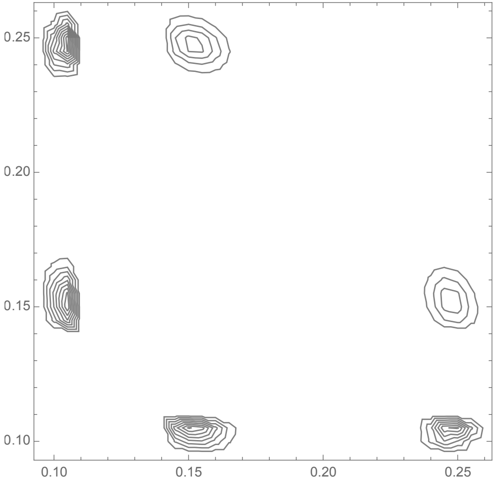
[1] Oliver et al. (2011)

Observations at the outlet face



The total flow rate exhibits a steady decline, but the water cut (fraction of the produced fluid that is water) increases in discrete steps followed by periods of slow continuous increase.

Posteriori distribution for k_1 and k_2



Characterize uncertainty in reservoir predictions, conditional on observations

- Approaches based on 'best solution' and Hessian for uncertainty have not been useful
- Realistic problems are too big for MCMC
- Approximate sampling via 'randomized maximum likelihood' (Oliver et al., 1996; Oliver, 2014) or 'randomize-then-optimize' (Bardsley et al., 2014).

Generate proposals

The RML method draws samples (x_i', δ_i'), $i = 1, \dots, M$, from the Gaussian distribution

$$q_{X'\Delta'}(x', \delta') = c_q \exp\left(-\frac{1}{2}\left(x' - \mu\right)^T C_X^{-1}\left(x' - \mu\right) - \frac{1}{2}\left(\delta' - d^{\mathrm{o}}\right)^T C_D^{-1}\left(\delta' - d^{\mathrm{o}}\right)\right)$$

for given μ and d^{o} and then minimizes the cost functional

$$J_i(x) = \frac{1}{2}\left(x - x_i'\right)^T C_X^{-1}\left(x - x_i'\right) + \frac{1}{2}\left(g(x) - \delta_i'\right)^T C_D^{-1}\left(g(x) - \delta_i'\right)$$

to determine

$$x_i = \arg\min J_i(x).$$

Proposal density for RML

Minimisation leads to a map from (x', δ') to (x, δ) defined implicitly by

$$x' = x + C_X G^T C_D^{-1}(g(x) - \delta)$$

and

$$\delta' = \delta.$$

with transformed distribution $p_{X\Delta}$ given by

$$\begin{aligned} p_{X\Delta}(x, \delta) &:= q_{X'\Delta'}(x', \delta')\, J(x, \delta) \\ &= q_{X'}\left(x + C_X G^T C_D^{-1}(g(x) - \delta)\right)\, q_{D'}(\delta)\, J(x, \delta) \end{aligned}$$

Here $J(x, \delta)$ denotes the determinant of the Jacobian matrix for the inverse map $(x, \delta) \to (x', \delta')$ and $G := Dg(x)$.

Proposal density for RML

The Jacobian matrix is provided by

$$\begin{pmatrix} I + Db(x,\delta) & -C_X G^{\mathrm{T}} C_D^{-1} \\ 0 & I \end{pmatrix}$$

with $b(x,\delta) = C_X G^{\mathrm{T}} C_D^{-1}(g(x) - \delta)$.
To simplify we will use

$$V(x) = C_D^{-1} + C_D^{-1} G C_X G^T C_D^{-1}$$

and

$$\begin{aligned} \eta(x) &= -C_D^{-1}(g(x) - d^o) + C_D^{-1} G(x - \mu) \\ &= C_D^{-1}[G(x-\mu) - (g(x) - d^o)] \end{aligned}$$

Proposal density for RML

After some algebra, we obtain

$p_{X\Delta}(x, \delta) =$

$$\overbrace{A_0 \exp\left[-\frac{1}{2}(x-\mu)^T C_X^{-1}(x-\mu) - \frac{1}{2}(g(x)-d^o)^T C_D^{-1}(g(x)-d^o)\right]}^{\pi_X(x)}$$

$$\times A_1 |V|^{1/2} \exp\left[-\frac{1}{2}\left(\delta - g(x) - V^{-1}\eta(x)\right)^T V\left(\delta - g(x) - V^{-1}\eta(x)\right)\right]$$

$$\times A_2 |V|^{-1/2} \exp\left[\frac{1}{2}\eta(x)^T V^{-1}\eta(x)\right] J(x, \delta)$$

Extend the target distribution (Oliver, 2017)

Target distribution

$$\pi_X(x) \propto \exp\Big[-\frac{1}{2}(x-\mu)^T C_X^{-1}(x-\mu) - \frac{1}{2}(g(x)-d^o)^T C_D^{-1}(g(x)-d^o)\Big].$$

Introduce an extended target distribution

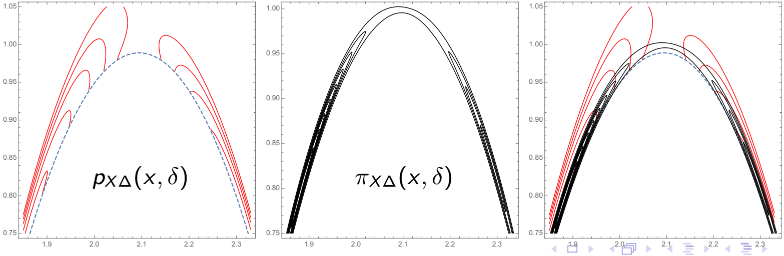
$$\begin{aligned}\pi_{X\Delta}(x,\delta) \propto \exp\Big[&-\frac{1}{2}(x-\mu)^T C_X^{-1}(x-\mu) \\ &- \frac{1}{2}(g(x)-d^o)^T C_D^{-1}(g(x)-d^o)\Big] \\ \exp\Big[&-\frac{1}{2\gamma(1-\gamma)}\left(\delta - g(x) + \gamma(g(x)-d^o)\right)^T \\ &\times C_D^{-1}\left(\delta - g(x) + \gamma(g(x)-d^o)\right)\Big].\end{aligned}$$

without changing target marginal density for model variable x.

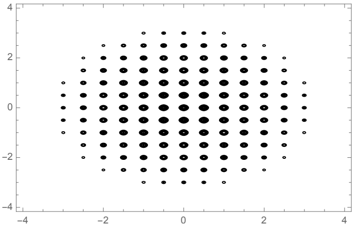
Importance weighting

$$w_i = \frac{\pi_{X\Delta}(x,\delta)}{p_{X\Delta}(x,\delta)}$$

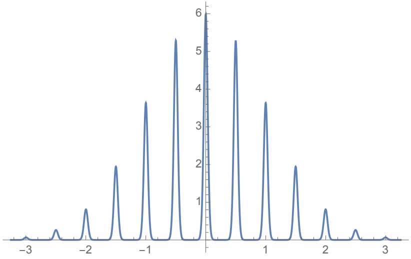
- Could have chosen an extended target to cancel in the case of linear g. Then $w_i = 1$ for linear.
- Potential problem when the map from (x', δ') to (x, δ) obtained from the condition $\Delta J_i = 0$ is not one-to-one.



Example with many modes



Target distribution for x_1, x_2.



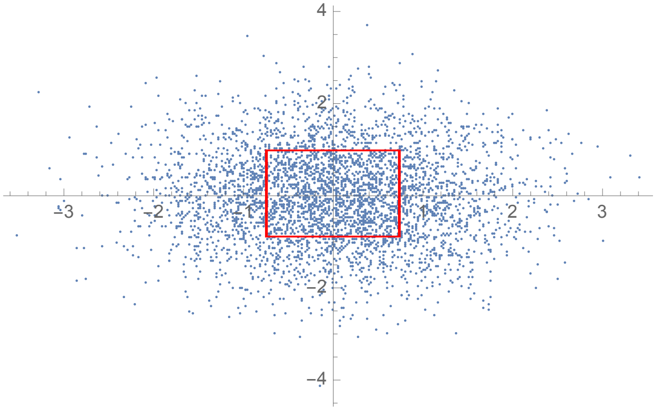
Target distribution for $x_1|x_2 = 0$.

Two model variables and two nonlinear observations.

$$g[x_1, x_2] = \begin{bmatrix} \sin[2\pi x_1] \\ \sin[2\pi x_2] \end{bmatrix}$$

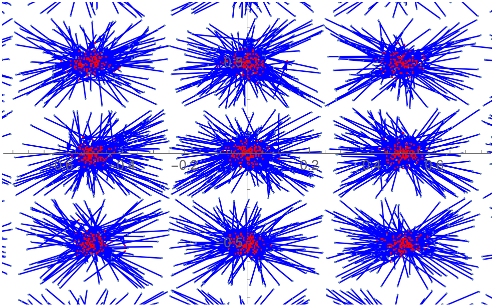
$\sigma_D = 0.2$, $\mu = (0.0, 0.0)$ and $\sigma_X = 1.$, $d^o = (0., 0.)$

Proposed transitions



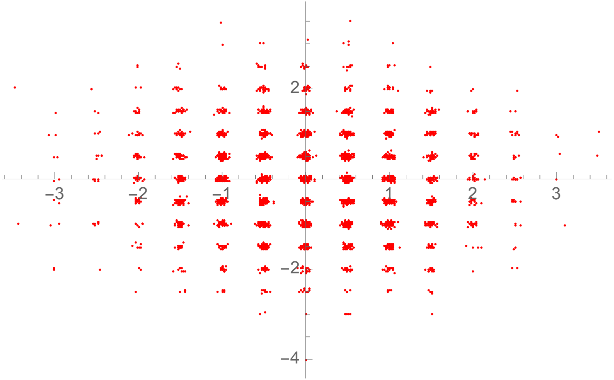
Sample independently from the prior distribution.

Proposed transitions



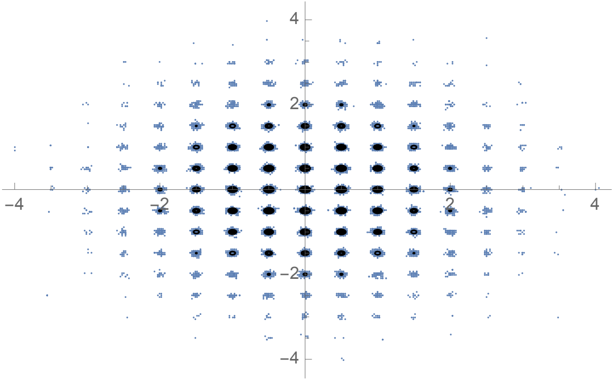
Solve a minimization problem which maps samples from the prior to samples from a proposal distribution.

Distribution of proposed transitions



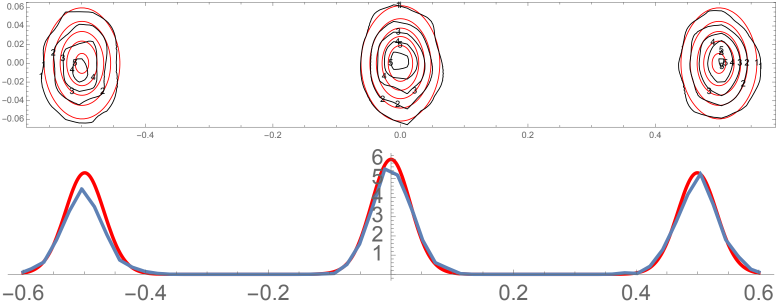
Apply Metropolis-Hastings test for samples x_i, δ_i.

MCMC samples



Samples from MH independence sampler with 40,000 elements.
Acceptance rate = 0.875.

Compare sampling to exact distribution

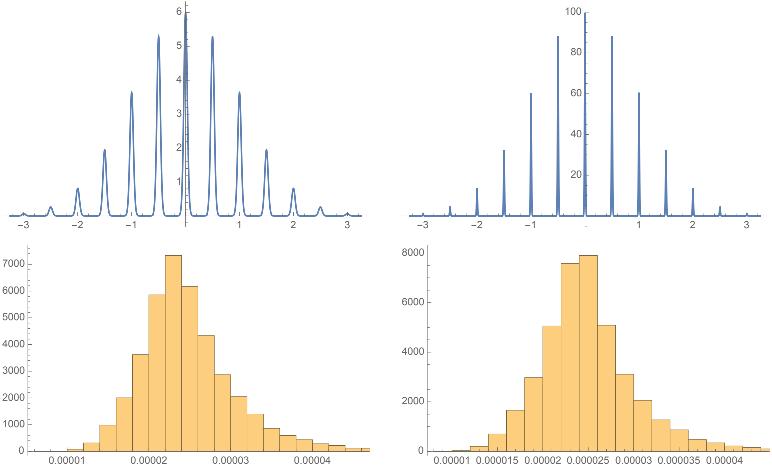


Red is true model density. Black is density estimated by kernel smoothing (bandwidth 0.01) of 4200 samples in the regions of three central peaks.

Computational effort

- Each proposal required 15 function evaluations using a modified Levenberg-Marquardt and x' to initialize the minimization.
- Computation of the Jacobian of the mapping for MH required an additional 5 function evaluations.
- The acceptance rate for MH is 0.873 so the cost is approximately 23 functions evaluations per independent sample from the target distribution.

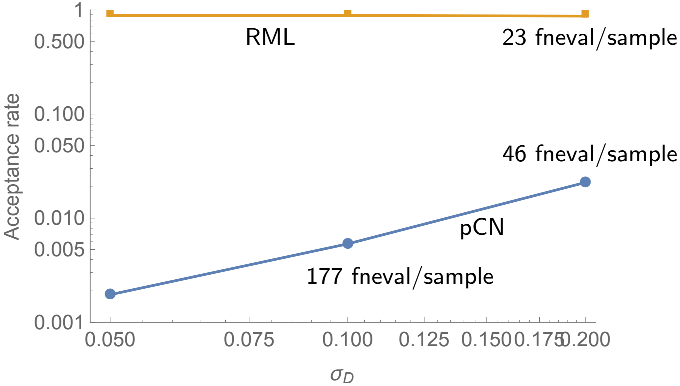
'Curse of dimensionality'



$\sigma_D = 0.20$ $\sigma_D = 0.05$

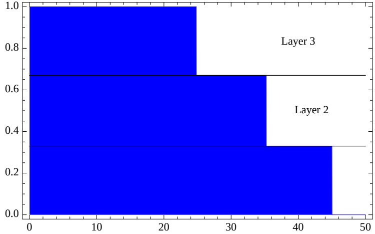
Distribution of particle weights after updating ($N_e = 40,000$)

'Curse of dimensionality'



As the distance between the prior and the posterior increases (as σ_D gets smaller), the acceptance rate (or effective sample size) for RML is nearly constant.

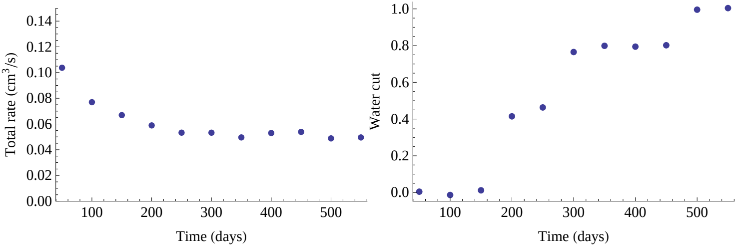
Simple flow problem with multi-modal pdf[2]



- Three parameters (k_1, k_2, and k_3) to be estimated.
- Water injected at constant pressure into all three layers.
- Fluids are produced at constant pressure from all three layers.
- No vertical communication between layers.

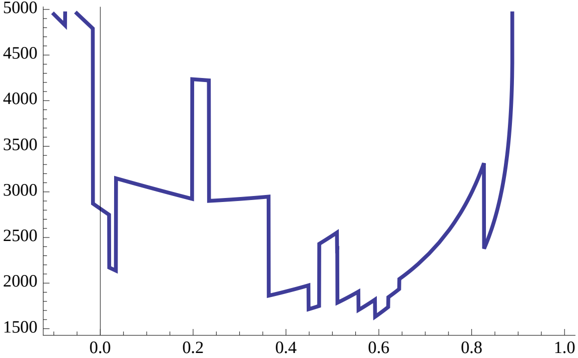
[2]Oliver et al. (2011)

Observations at the outlet face



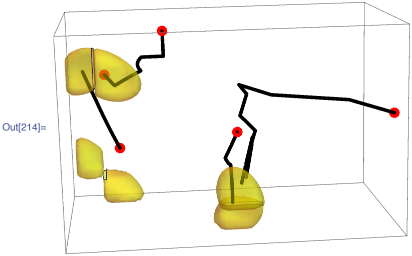
The total flow rate exhibits a steady decline, but the water cut increases in discrete steps followed by periods of slow continuous increase.

Objective function



The objective function along the steepest descent direction for a random starting point.

Paths from prior samples to posterior samples



Yellow regions have significant posterior probability. Red dots are samples from prior distribution. Black curve shows minimization path. Approximately 65% got stuck at local minima.

Rejecting poorly calibrated samples

Objective function:

$$J_i(x) = \frac{1}{2}(x - x_i')^T C_X^{-1}(x - x_i') + \frac{1}{2}(g(x) - \delta_i')^T C_D^{-1}(g(x) - \delta_i')$$

and $x_i = \operatorname{argmin} J_i(x)$.

Model diagnostics[3]:

$$\hat{J}_i = J_i(x_i)$$

$$\hat{J}_d = \frac{1}{2}(g(x_i) - \delta')^T C_D^{-1}(g(x_i) - \delta')$$

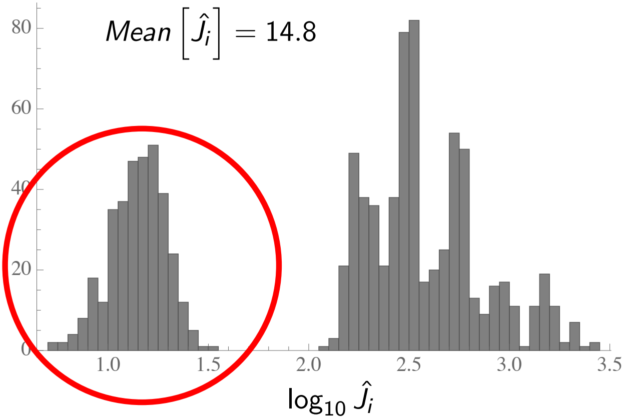
$$\hat{J}_d^o = \frac{1}{2}(g(x_i) - d^o)^T C_D^{-1}(g(x_i) - d^o)$$

[3] Tarantola (1987); Bennett (1992); Talagrand (1999); Desroziers and Ivanov (2001)

Rejecting poorly calibrated samples

$$E\left[\hat{J}_i\right] \approx \tfrac{1}{2} N_D = 11$$

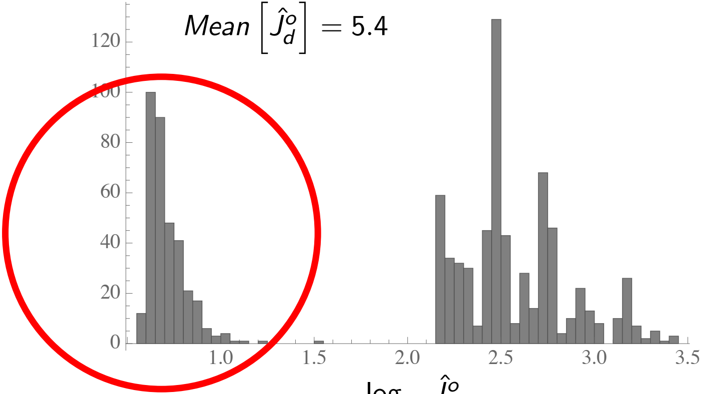
$$Mean\left[\hat{J}_i\right] = 14.8$$



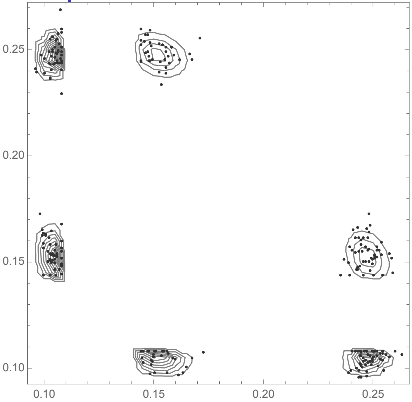
Rejecting poorly calibrated samples

$$E\left[\hat{J}_d^o\right] \approx \frac{1}{2} N_D = 11$$

$$Mean\left[\hat{J}_d^o\right] = 5.4$$



Estimated marginal pdf from minimization-based sampling



Samples from RML are shown by black dots. True marginal distribution for permeabilities of layers 1 and 2 are shown by contours. Used Levenberg-Marquardt with accurate derivatives for minimization. The joint distribution has six peaks, which were all identified. True model had permeabilities (0.10, 0.15, 0.25.)

Iterative method for sampling via minimization

Solving $\nabla J_i(x) = 0$,

$$\delta x_i^\ell = -(x_i^\ell - x_i') - C_x G_\ell^T \left(C_D + G_\ell C_X G_\ell^T \right)^{-1} \left(g(x_i^\ell) - \delta_i - G_\ell (x_i^\ell - x_i') \right).$$

where $G_\ell = \nabla g(x_i^\ell)$.

At the first iteration ($\ell = 1$), when $x_i^\ell = x_i'$

$$\delta x_i^1 = -C_X G_1^T \left(C_D + G_1 C_X G_1^T \right)^{-1} \left(g(x_i') - \delta_i \right).$$

Regularized ensemble-based iterative updating[4]

RML required computation of the Jacobian **G** or the gradient of the objective function. Not easy to get derivatives for reservoir simulators.

Ensemble-based Levenberg-Marquardt iterative updates (iterative ES):

$$\delta \mathbf{x}_i^\ell = -\left[(1+\lambda_\ell) P_\ell^{-1} + G_\ell^T C_D^{-1} G_\ell\right]^{-1} C_X^{-1}(x_i^\ell - x_i') \\ - \Delta x_\ell \Delta d_\ell^T \left[(1+\lambda_\ell)(N_e - 1) C_D + \Delta d_\ell \Delta d_\ell^T\right]^{-1} (g(x_i^\ell) - \delta_i)$$

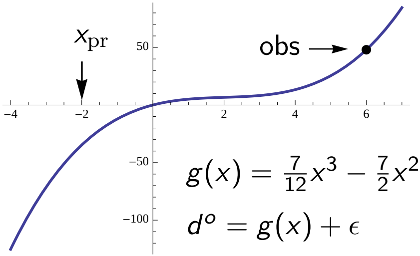
where x_i' is the $i$the sample from the prior distribution and Δx_ℓ is the matrix of mean removed model variables at the ℓth iteration.

[4] Chen and Oliver (2013)

Regularized ensemble-based iterative updating

- First iteration is exactly the same as would be obtained with the ensemble smoother (except that $C_D \to (1+\lambda)C_D$).
- The initial value for λ is typically quite large in reservoir flow problems ($\lambda_1 \sim 10^4$).
- The gradient of the objective function is not modified — only the approximation to the Hessian.
- For sampling the posterior, a different objective function is used for each realization.

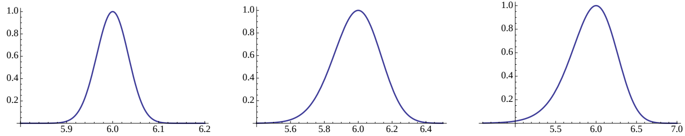
Simple 1-variable nonlinear problem



$$g(x) = \frac{7}{12}x^3 - \frac{7}{2}x^2 + 8x$$

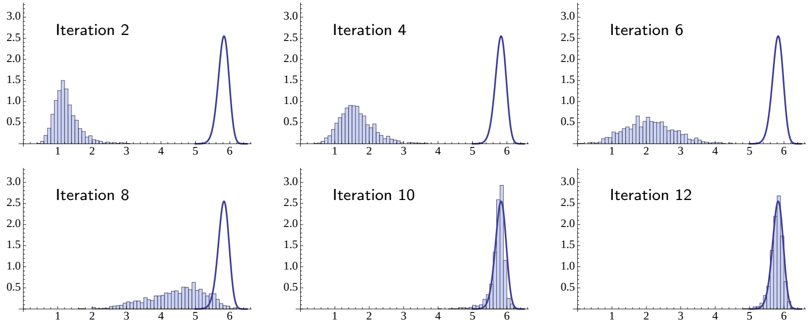
$$d^o = g(x) + \epsilon$$

The variable x is distributed as $N[-2, 1]$.



var $d = 1$ var $d = 16$ var $d = 64$

Validation: 1 variable problem (var $d = 1$)



Iterative ensemble smoother

Minimization for sampling

- Quite robust to nonlinearity (e.g. multimodal posterior distributions)
- Not as robust with respect to prior distribution
- When prior is nongaussian, can sometimes introduce latent Gaussian variables
- The use of ensemble-based methods can increase limitations on uncertainty quantification
- Assumed that the cost function to be minimized was “correct” — will almost certainly be invalidated with sufficient data.

Minimization for sampling

- Quite robust to nonlinearity (e.g. multimodal posterior distributions)
- Not as robust with respect to prior distribution
- When prior is nongaussian, can sometimes introduce latent Gaussian variables
- The use of ensemble-based methods can increase limitations on uncertainty quantification
- Assumed that the cost function to be minimized was “correct” — will almost certainly be invalidated with sufficient data.

Acknowledgements

Primary support has been provided by the cooperative research project “4D Seismic History Matching” which is funded by industry partners Eni, Petrobras, and Total, as well as the Research Council of Norway (PETROMAKS).

References I

Bardsley, J., Solonen, A., Haario, H., and Laine, M. (2014). Randomize-Then-Optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1895–A1910.

Bennett, A. F. (1992). *Inverse Methods in Physical Oceanography*. Cambridge University Press.

Chen, Y. and Oliver, D. S. (2013). Levenberg-Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification. *Comput. Geosci.*, 17(4):689–703.

Desroziers, G. and Ivanov, S. (2001). Diagnosis and adaptive tuning of observation-error parameters in a variational assimilation. *Q. J. R. Meteorol. Soc.*, 127(574, B):1433–1452.

References II

Oliver, D. S. (2014). Minimization for conditional simulation: Relationship to optimal transport. *Journal of Computational Physics*, 265(0):1–15.

Oliver, D. S. (2017). Metropolized randomized maximum likelihood for improved sampling from multimodal distributions. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):259–277.

Oliver, D. S., He, N., and Reynolds, A. C. (1996). Conditioning permeability fields to pressure data. In *Proceedings of the European Conference on the Mathematics of Oil Recovery, V*, pages 1–11.

Oliver, D. S., Zhang, Y., Phale, H. A., and Chen, Y. (2011). Distributed parameter and state estimation in petroleum reservoirs. *Computers & Fluids*, 46(1):70–77.

References III

Talagrand, O. (1999). A posteriori verification of analysis and assimilation algorithms. In *Proceedings of Workshop on Diagnosis of Data Assimilations Systems, 2–4 Nov 1998, ECMFW, Reading, UK*, pages 17–28.

Tarantola, A. (1987). *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, Amsterdam.