

Metropolized Randomized Maximum Likelihood for sampling from multimodal distributions

Dean Oliver

Uni Research CIPR

22 February 2016

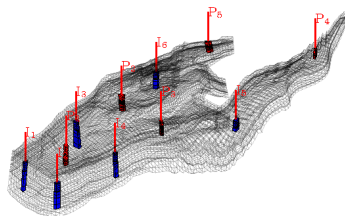


Problem statement

Sample from a target distribution π that is a product of a prior distribution (Gaussian) and a likelihood function with additive Gaussian errors

- π may be multimodal
- Dimension of model space may be relatively large (targeted at geoscience problems).

Motivation



Numerical models are used for forecasting and decisions.

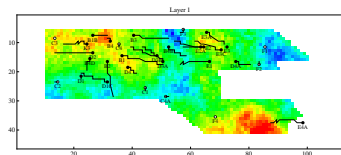
- Many parameters (10^5 – 10^6).
- Observations are generally sparse (1 km apart, but daily observations for several years).
- Likelihood function evaluations expensive (0.1–10 hour)

Figure source: <http://www.sintef.no/>

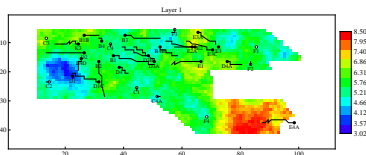
Current methodology

Currently use iterative ensemble smoothers for reservoir “history matching”. Multiple approximations:

1. Based on Randomized Maximum Likelihood without weighting
2. Updates are computed from correlations between data and model variables
3. Ignore model error
4. ...



realization 1



realization 2

Background

Several methods use optimization to place candidate states in regions of high probability.

Quasi-Linear Estimation did not apply a MH test to correct the sampling for nonlinear (Kitanidis, 1995).

Randomized Maximum Likelihood applied MH test to obtain correct sampling but required marginalization (Oliver et al., 1996).

Randomize-Then-Optimize MH acceptance test limited the application to distributions with a single mode (Bardsley et al., 2014).

In this paper, the RML method is modified so that marginalization is not required. Applicable to multimodal distributions.

Simulation methods based on minimization

Many methods can be used to generate realizations from multi-normal distributions. Two are particularly interesting because of the contrast in approaches.

- One approach is to **generate a “rough” field** (an unconditional simulation) with the same covariance as the true field, then to **subtract a smooth correction** that forces the simulated field to pass through the data.
- A second approach is to **compute a “smooth” estimate** that passes through the data, then use the LU decomposition of the estimation error covariance to **add a stochastic component** to the estimate.

It is known that these methods are equivalent for data without errors (Krzanowski, 1988) and for data with errors (Oliver, 1996).

Smooth plus rough — Gauss-linear problem

The maximum a posteriori estimate minimizes

$$\begin{aligned} m_{\text{map}} &= \operatorname{argmin}(m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}) \\ &\quad + (Gm - d_{\text{obs}})^T C_D^{-1} (Gm - d_{\text{obs}}), \\ &= m_{\text{prior}} + C_M G^T (G C_M G^T + C_D)^{-1} (d_{\text{obs}} - G m_{\text{prior}}). \end{aligned}$$

Samples from posterior can be generated

$$m_i = m_{\text{map}} + LZ_i \quad \text{for } i = 1, \dots, N$$

where $Z_i \sim N[0, I]$, $LL^T = C_{M'}$, and

$$C_{M'} = (C_M^{-1} + G^T C_D^{-1} G)^{-1}$$

Rough plus smooth — Gauss-linear problem

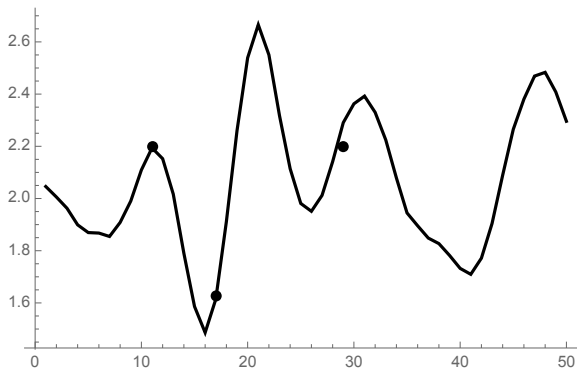
In this algorithm, we first generate unconditional realizations of the model (and data), then make a correction to the unconditional model.

- Sample $m_u \sim N[m_{\text{prior}}, C_M]$.
- Sample $d_u \sim N[d_{\text{obs}}, C_D]$.
- Compute the model m_c that minimizes

$$\begin{aligned} m_c &= \operatorname{argmin}(m - m_u)^T C_M^{-1} (m - m_u) \\ &\quad + (Gm - d_u)^T C_D^{-1} (Gm - d_u), \\ &= m_u - C_M G^T [GC_M G^T + C_D]^{-1} (Gm_u - d_u) \end{aligned}$$

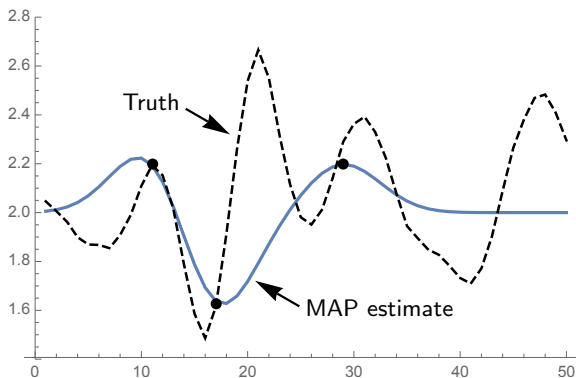
Does not require m_{map} or posteriori covariance.

Illustration: Truth



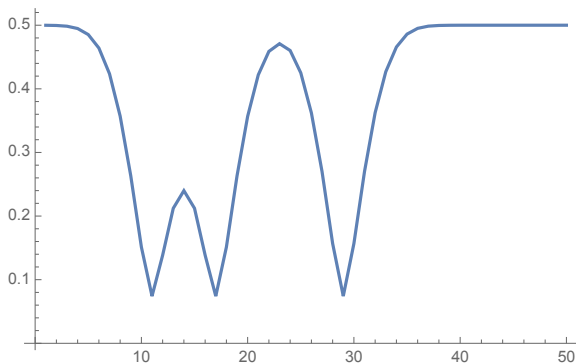
The truth with three (inaccurate) observations.

Smooth plus rough — Compute maximum a posteriori estimate



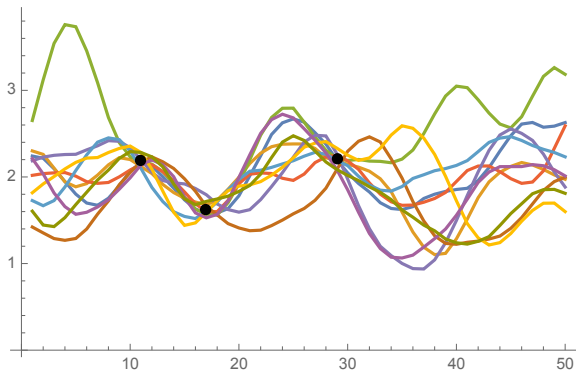
$$m_{\text{map}} = \operatorname{argmin} \left[(m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}) + (Gm - d_{\text{obs}})^T C_D^{-1} (Gm - d_{\text{obs}}) \right]$$

Smooth plus rough



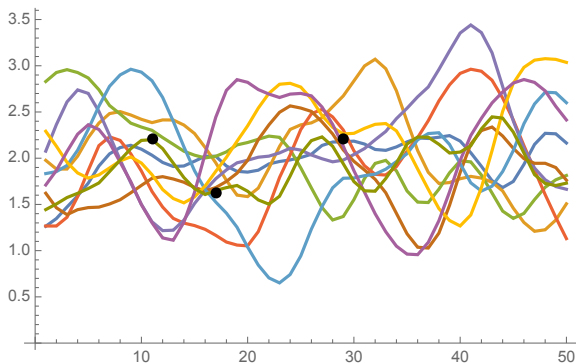
The MAP estimate of the standard deviation in the estimate (square root of posteriori variance).

Smooth plus rough



$$m_i = m_{\text{map}} + C_{M'}^{1/2} Z_i \quad \text{for } i = 1, \dots, N$$

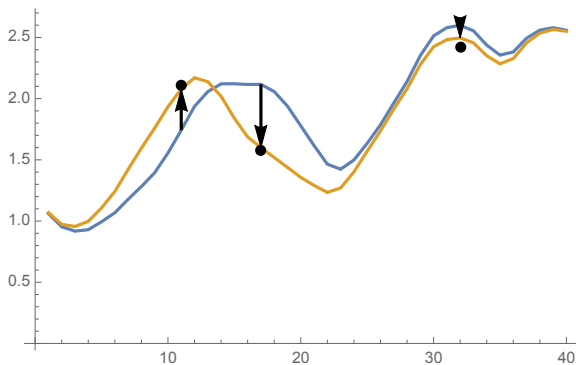
Alternative: Rough plus smooth



Ten unconditional realizations.

$$m_i = m_{\text{prior}} + C_M^{1/2} Z_i \quad \text{for } i = 1, \dots, N$$

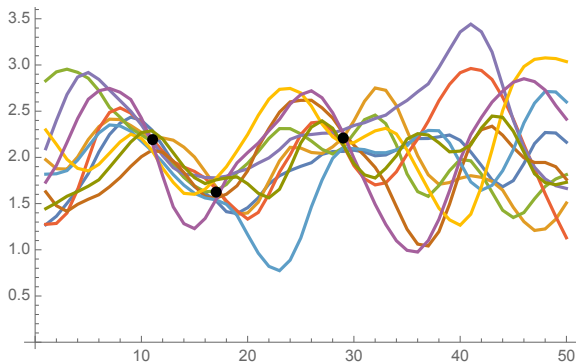
Illustration: Rough plus smooth



Add a smooth correction to an unconditional realization to make it conditional to data.

$$\delta m = -C_M G^T [G C_M G^T + C_D]^{-1} (G m_u - d_u).$$

Illustration: Rough plus smooth



The 10 conditional realizations resulting from smooth corrections to the unconditional realizations.

Smooth plus rough and rough plus smooth— Summary

Both algorithms are valid methods for sampling the posterior for linear problems with gaussian measurement errors and a gaussian prior PDF for the model variables.

Neither is correct for nonlinear problems, but the method of adding a smooth correction to unconditional realizations appeared to be more robust to nonlinearity.

Application to nonlinear problems

Let X be multivariate normal with mean μ and covariance C_x such that the prior probability density for X is

$$p(x) = c_p \exp \left(-\frac{1}{2} (x - \mu)^T C_x^{-1} (x - \mu) \right).$$

Observations $d^o = g(x) + \epsilon_d$ with $\epsilon_d \sim N(0, C_d)$ are assimilated resulting in a posterior density

$$\pi(x) \propto \exp \left(-\frac{1}{2} (x - \mu)^T C_x^{-1} (x - \mu) - \frac{1}{2} (g(x) - d^o)^T C_d^{-1} (g(x) - d^o) \right).$$

Standard RML for MCMC

1. Generate candidate state x_* :

1.1 Independently sample

$$x_{uc} \sim N[x_{pr}, C_x] \quad \text{and} \quad d_{uc} \sim N[d_{obs}, C_d].$$

1.2 Minimize a nonlinear least-squares functional:

$$x_* = \underset{x}{\operatorname{argmin}} \left[\underbrace{\frac{1}{2} (x - x_{uc})^T C_x^{-1} (x - x_{uc})}_{\text{Model parameter mismatch}} + \underbrace{\frac{1}{2} (g(x) - d_{uc})^T C_d^{-1} (g(x) - d_{uc})}_{\text{Sum of squared data mismatch}} \right]$$

(Need to compute the probability of proposing x_* .)

Standard RML for MCMC (2)

2. Compute probability of proposing (x_*, d_{uc}) .

2.1 Probability of proposing (x_{uc}, d_{uc})

$$f(x_{uc}, d_{uc}) \propto \exp\left(-\frac{1}{2}(x_{uc} - \mu)^T C_x^{-1}(x_{uc} - \mu) - \frac{1}{2}(d_{uc} - d_{obs})^T C_d^{-1}(d_{uc} - d_{obs})\right)$$

2.2 Probability of proposing (x_*, d_{uc})

$$q(x_*, d_{uc}) = f(x_{uc}(x_*, d_{uc}), d_{uc}) |\det J|$$

where

$$x_{uc} = x_* + C_x G(x_*)^T C_d^{-1}(g(x_*) - d_{uc})$$

from the necessary condition for x_* to be a minimum.

$$G(x_*) = \partial g / \partial x|_{x_*}.$$

Standard RML for MCMC (3)

3. Compute marginal probability of proposing x_* .

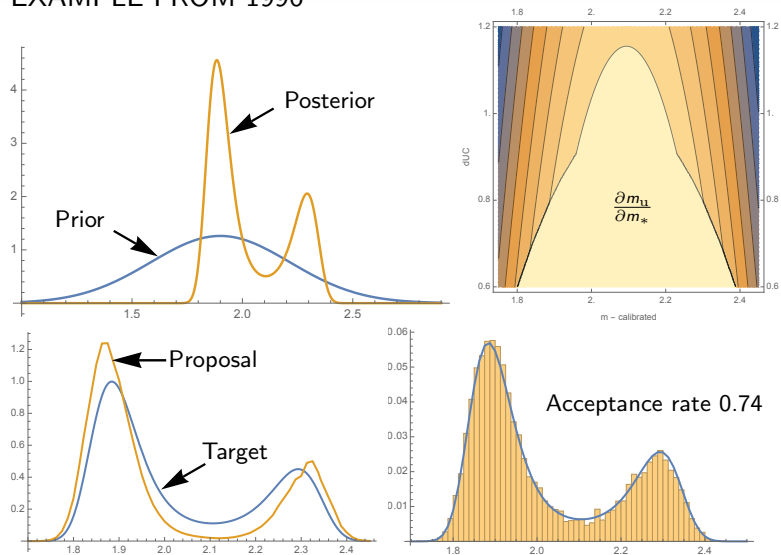
$$q_x(x_*) = \int_D q(x_*, d_{uc}) dd_{uc}.$$

4. Accept proposed state x_* with probability

$$\alpha(x, x_*) = \min \left(1, \frac{\pi(x_*)q_x(x)}{\pi(x)q_x(x_*)} \right).$$

else retain state x (Metropolis-Hastings).

EXAMPLE FROM 1996



Comments on Randomized Maximum Likelihood

- The proposal density was typically close to the target density (identical for linear observations).
- Acceptance rate was high in multimodal experiments
- MH acceptance ratio was difficult to compute
- In practice — ignored MH test (accepted all transitions)
- Formed the basis for several iterative variants of ensemble Kalman filter-like methods

Augmented state RML¹

1. Augment the state with data variables
2. Modify the target joint pdf such that the marginal pdf for model variables is identical to the true posterior pdf for model variables
3. Modify the proposal density to improve MH acceptance

¹Oliver (2015)

Modify the target joint pdf

Define the target joint probability for the augmented state is

$$\pi(x, d) \propto \exp \left[-\frac{1}{2} (x - \mu)^T C_x^{-1} (x - \mu) - \frac{1}{2\gamma} (g(x) - d)^T C_d^{-1} (g(x) - d) - \frac{1}{2(1 - \gamma)} (d - d_{\text{obs}})^T C_d^{-1} (d - d_{\text{obs}}) \right].$$

Modify the target joint pdf

Define the target joint probability for the augmented state is

$$\pi(x, d) \propto \exp \left[-\frac{1}{2}(x-\mu)^T C_x^{-1}(x-\mu) - \frac{1}{2\gamma}(g(x)-d)^T C_d^{-1}(g(x)-d) - \frac{1}{2(1-\gamma)}(d-d_{\text{obs}})^T C_d^{-1}(d-d_{\text{obs}}) \right].$$

The marginal target density for model variable x can be shown to be

$$\pi(x) \propto \exp \left[-\frac{1}{2}(x-\mu)^T C_x^{-1}(x-\mu) - \frac{1}{2}(g(x)-d_{\text{obs}})^T C_d^{-1}(g(x)-d_{\text{obs}}) \right]$$

independent of γ .

Target joint pdf

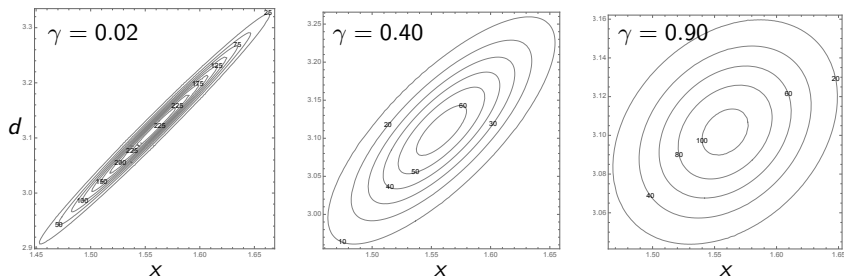


Figure 1: Dependence of the joint density for (x, d) on magnitude of the modelization error γ .

Generate independent candidates for MCMC

1. Draw unconditional samples from the prior distribution of model and data variables,

$$x_{\text{uc}} \sim N[\mu, C_x] \quad \text{and} \quad d_{\text{uc}} \sim N[d_{\text{obs}}, C_d].$$

2. Candidate joint states are obtained by minimizing a nonlinear least squares function

$$\begin{aligned} (x_*, d_*) = \operatorname{argmin}_{x, d} & \left[(x - x_{\text{uc}})^T C_x^{-1} (x - x_{\text{uc}}) \right. \\ & + \frac{1}{\rho} (g(x) - d)^T C_d^{-1} (g(x) - d) \\ & \left. + \frac{1}{(1 - \rho)} (d - d_{\text{uc}})^T C_d^{-1} (d - d_{\text{uc}}) \right] \end{aligned}$$

Proposal probability density

The inverse transformation is straightforward,

$$x_{\text{uc}} = x_* + \frac{1}{\rho} C_x G^T C_d^{-1} (g(x_*) - d_*)$$

and

$$d_{\text{uc}} = \frac{1}{\rho} d_* - \left(\frac{1 - \rho}{\rho} \right) g(x_*).$$

So the joint proposal density is

$$\begin{aligned} q(x_*, d_*) &= p(x_{\text{uc}}(x_*, d_*), d_{\text{uc}}(x_*, d_*)) |\det J| \\ &= c_p \exp \left[-\frac{1}{2} (x_{\text{uc}}(x_*, d_*) - \mu)^T C_x^{-1} (x_{\text{uc}}(x_*, d_*) - \mu) \right. \\ &\quad \left. - \frac{1}{2} (d_{\text{uc}}(x_*, d_*) - d_{\text{obs}})^T C_d^{-1} (d_{\text{uc}}(x_*, d_*) - d_{\text{obs}}) \right] |\det J| \end{aligned}$$

Dependence of proposal density on ρ

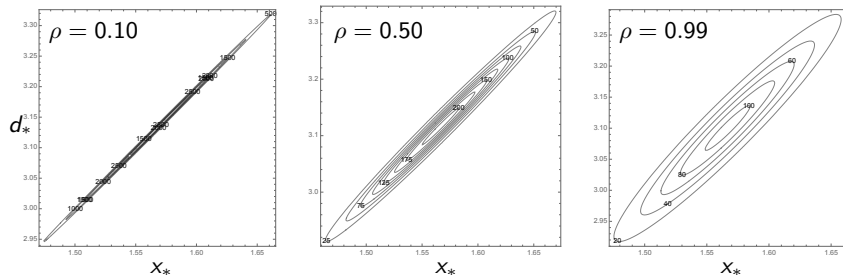
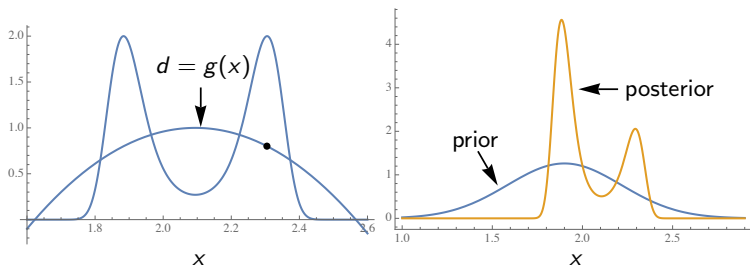


Figure 2: Dependence of the joint density for proposed transitions (x_*, d_*) on magnitude of ρ .

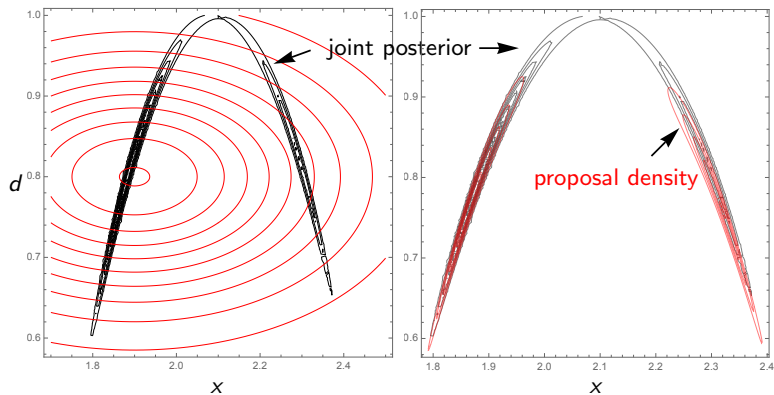
Increasing ρ provides a wider proposal density.

Bimodal example: problem definition



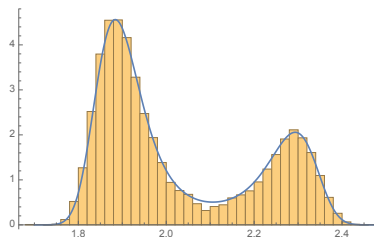
Observation operator is quadratic: two identical peaks in the likelihood.

Bimodal example: augmented state space

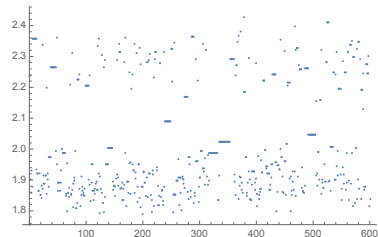


$$\gamma = 0.02 \quad \text{and} \quad \rho = 0.65$$

Marginal distribution of model state from Metropolized RML



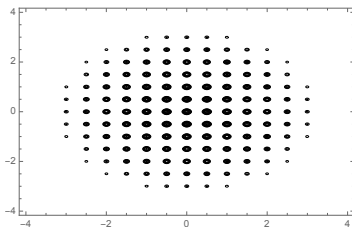
Distribution of samples of x



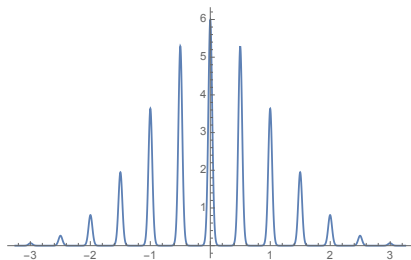
Markov chain for x

Acceptance rate for independent proposals is 64% (almost independent of ρ for $0.5 \leq \rho \leq 0.8$).

Example with many modes



True posterior pdf for x .



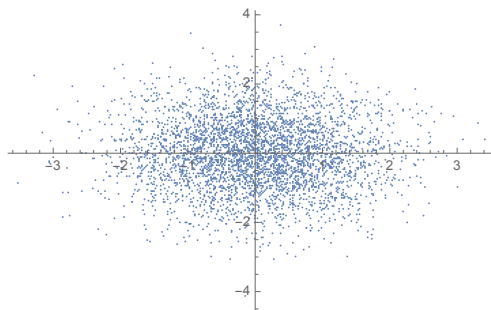
Mapping x_{UC} to x_*

Two model variables and two nonlinear observations.

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

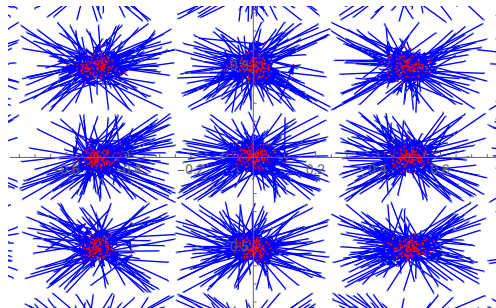
$\sigma_D = 0.2$, $x_{pr} = (0.0, 0.0)$ and $\sigma_X = 1.$, $d_{obs} = (0., 0.)$

Proposed transitions



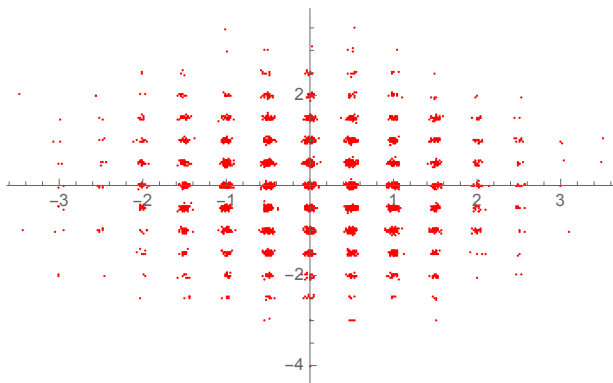
Sample independently from the prior distribution.

Proposed transitions



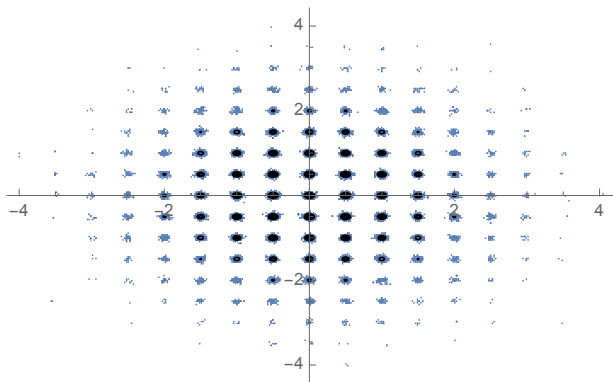
Solve a minimization problem which maps samples from the prior to samples from a proposal distribution.

Distribution of proposed transitions



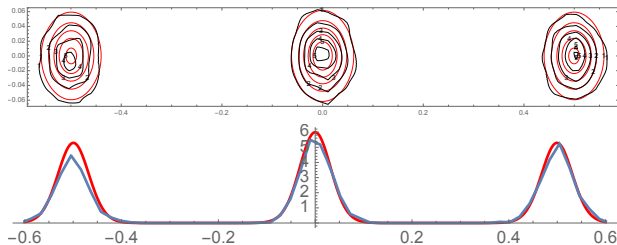
Need to apply Metropolis-Hastings test for samples of x_* , d_* .

MCMC samples



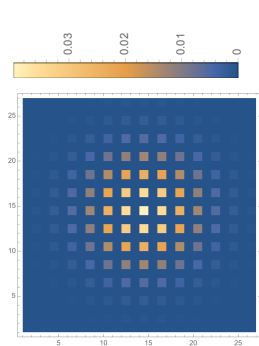
Samples from MH independence sampler with 40,000 elements.
Acceptance rate = 0.875 ± 0.002 .

Compare sampling to exact pdf

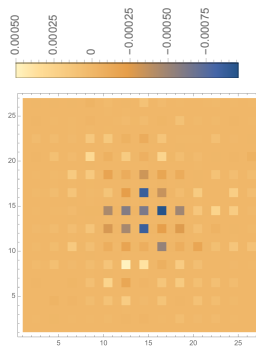


Red is true model density. Black is density estimated by kernel smoothing (bandwidth 0.01) of 4200 samples in the regions of three central peaks.

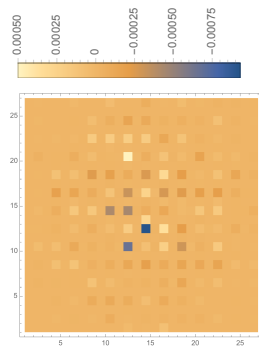
Benefit of Metropolization



True prob



Error RML w MH



Error RML wo MH

Small difference in total absolute error (0.0178 vs 0.0169).

Summary

- A new augmented variable independence Metropolis sampler
 - Minimization to place proposals in regions of high probability
 - Relatively high acceptance rate, even in multimodal distributions (rapid mixing).
- Open issues
 - Computation of the Jacobian determinant in high dimensions
 - Requirement for obtaining global minimum
 - Generalization to nongaussian priors
 - Applicability with EnKF-like methods

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).

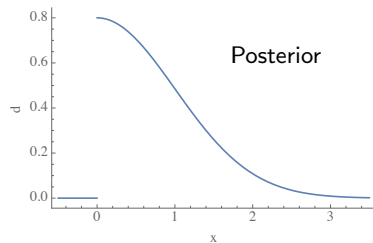
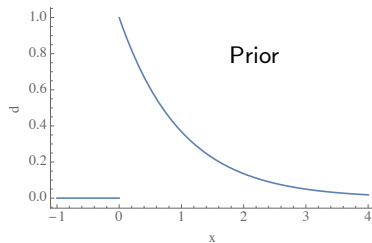
References I

- 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.
- Kitanidis, P. K. (1995). Quasi-linear geostatistical theory for inversing. *Water Resour. Res.*, 31(10):2411–2419.
- Krzanowski, W. J. (1988). *Principles of Multivariate Analysis: A User's Perspective*. Clarendon Press, Oxford. 563 p.
- Oliver, D. S. (1996). On conditional simulation to inaccurate data. *Math. Geology*, 28(6):811–817.
- Oliver, D. S. (2015). Metropolized Randomized Maximum Likelihood for sampling from multimodal distributions. *ArXiv e-prints*, (arXiv:1507.08563).

References II

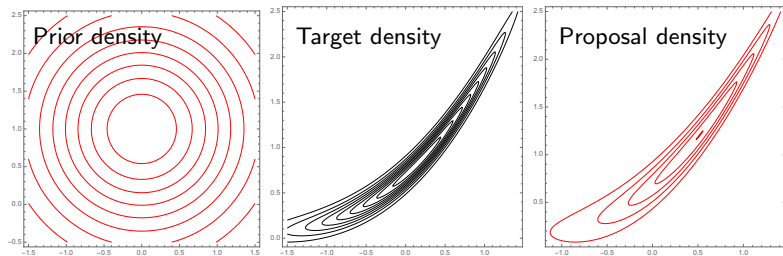
Oliver, D. S., He, N., and Reynolds, A. C. (1996). Conditioning permeability fields to pressure data. In *European Conference for the Mathematics of Oil Recovery, V*, pages 1–11.

Example: non-Gaussian prior



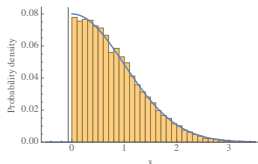
Prior and posterior model variable distributions for Example 3.

The Jacobian of transformation

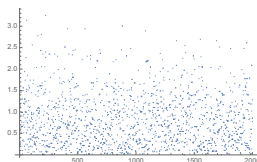


Joint prior, posterior, and proposal distributions for Gaussian model variable and data variable. Model variable on horizontal axis.

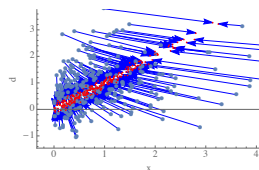
Example: non-Gaussian prior



Distribution samples



First 2000 elements



(x_{UC}, d_{UC}) to (x_*, d_*)

Results from Metropolized RML with variable transformation.

Acceptance rate is 74%.