



The University of Tulsa  
Petroleum Reservoir Exploitation Projects (TUPREP)

## Assisted History Matching

Al Reynolds

Tutorial Lecture on Assisted History Matching  
IPAM Workshop III, Data Assimilation, Uncertainty Quantification and  
Optimization for Subsurface Flow  
March 22, 2017

- Linear Inverse Problems
- Regularized Least Squares Problems
- Bayesian Formulation Framework for Uncertainty Quantification
- Randomized Maximum Likelihood
- Gradient-Based and Ensemble-Based Optimization
- AHM Examples

# Key References

- Nocedal, J. and S. J. Wright, Numerical Optimization, Springer, New York, 2006.
- Strang, G., Linear Algebra and Its Applications, Academic Press, New York, 1976.
- Oliver, D. S., A. C. Reynolds, and N. Liu, Inverse Theory for Petroleum Reservoir Characterization and History Matching, Cambridge University Press, Cambridge, UK, 2008.
- Vogel, C. R., Computational Methods for Inverse Problems, SIAM, 2002.
- Jansen, J. D., Adjoint-based optimization of multi-phase flow through porous media, Computers & Fluids, 46, 40-51, 2011.
- Oliver, D. S. and Y. Chen, Recent progress on reservoir history matching: a review, Computational Geosciences, 15, 185-221, 2011.
- Aanonsen, S. I., G. Naevdal, D. S. Oliver, A. C. Reynolds, and B. Valles, Review of ensemble Kalman filter in petroleum engineering, SPE Journal, 14(3), 393-412, 2009.

## Other References

- Tarantola, A., Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, Philadelphia, USA, 2005.
- Menke, W., Geophysical Data Analysis: Discrete Inverse Theory, revised edn., Academic Press, San Diego, 1989.
- Chavent, G. M., M. Dupuy, and P. Lemonnier, History matching by use of optimal control theory, SPE Journal, 15(1), 74-86, 1975.
- Chen, W. H., G. R. Gavalas, J. H. Seinfeld, and M. L. Wasserman, A new algorithm for automatic history matching, SPE Journal, 14(6), 593-608, 1974.
- Carter, R. D., L. F. Kemp, A. C. Pierce, and D. L. Williams, Performance matching with constraints, SPE Journal, 14(2), 187-196, 1974. [Jacquard(1964)]
- Evensen, G., Data Assimilation: The Ensemble Kalman Filter, Springer, Berlin, 2007.
- Burgers, G., P. van Leeuwen, and G. Evensen, Analysis scheme in the ensemble Kalman filter, Monthly Weather Review, 126(6), 1719-1724, 1998.

- Inverse problem: given specific observations (observed data) of the state variables (or data that are a direct function of state variables), infer information about model parameters (includes uncertainty). Our focus is on discrete inverse problems (system characterized by a finite number of parameters).
- History matching is a discrete inverse problem.

- The number of data is always finite, and the data always contain measurement errors.
- It is impossible to correctly estimate all the parameters of a model from inaccurate, inconsistent, and insufficient data using a forward model that may contain modeling errors, but reducing the number of parameters can yield an unrealistic reduction in the level of uncertainty which can be highly misleading.
- We always have some prior information about the plausibility of models. In the case of history matching, it includes a geological prior model (hopefully stochastic) constructed from log, core and seismic data and information on depositional environment.

- In gradient based assisted history matching, calculation of sensitivities must be done efficiently, i.e., using an adjoint solution.
- Probabilistic estimates or bounds are often the most meaningful. For nonlinear problems, this is usually best accomplished using Markov chain Monte Carlo but its cost is often prohibitive. (Recent work has the potential to reduce the costs.)
- The ultimate goal of inverse theory (and history matching) is to make informed decisions on investments, surveillance operations for data acquisition, and optimal reservoir management. Good decisions and risk management require that the uncertainty in reservoir description and future performance, and the consequences of actions can be accurately characterized.

# Introduction to Gradients

$m$  a column vector,  $f(m)$  a real valued function.

$g = g(m) = [g_1(m) \ g_2(m) \ \cdots \ g_{N_d}]$  is a  $1 \times N_d$  matrix, i.e., a row vector.

$$m = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_{N_m} \end{bmatrix} \quad \nabla_m = \begin{bmatrix} \frac{\partial.}{\partial m_1} \\ \vdots \\ \frac{\partial.}{\partial m_{N_m}} \end{bmatrix}.$$

$$\nabla_m m^T = \begin{bmatrix} \frac{\partial.}{\partial m_1} \\ \vdots \\ \frac{\partial.}{\partial m_{N_m}} \end{bmatrix} [m_1 \ m_2 \ \cdots \ m_{N_m}] = I_{N_m}.$$



# Introduction to Gradients

$g = g(m) = [g_1(m) \ g_2(m) \ \cdots \ g_{N_d}]^T$  is a  $N_d \times 1$  matrix. .

$$G^T \equiv \nabla_m g(m)^T = \begin{bmatrix} \frac{\partial \cdot}{\partial m_1} \\ \vdots \\ \frac{\partial \cdot}{\partial m_{N_m}} \end{bmatrix} [g_1(m) \ g_2(m) \ \cdots \ g_{N_d}] =$$
$$\begin{bmatrix} \frac{\partial g_1}{\partial m_1} & \frac{\partial g_2}{\partial m_1} & \cdots & \frac{\partial g_{N_d}}{\partial m_1} \\ \frac{\partial g_1}{\partial m_2} & \frac{\partial g_2}{\partial m_2} & \cdots & \frac{\partial g_{N_d}}{\partial m_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_1}{\partial m_{N_m}} & \frac{\partial g_2}{\partial m_{N_m}} & \cdots & \frac{\partial g_{N_d}}{\partial m_{N_m}} \end{bmatrix} = [\nabla g_1 \ \nabla g_2 \ \cdots \ \nabla g_{N_d}]$$

If  $A = A(m)$  is  $1 \times p$  and  $B = B(m)$  is  $p \times 1$ , then

$$\nabla_m(AB) = (\nabla_m A)B + (\nabla_m(B^T))A^T.$$

Thus if entries of  $B$  are independent of  $m$ ,

$$\nabla_m(AB) = (\nabla_m A)B.$$

If  $m$  is  $N_m \times 1$  and  $B$  is  $N_m \times 1$ ,

$$\nabla_m(m^T B) = (\nabla_m m^T)B + (\nabla_m(B^T))m = B + (\nabla_m(B^T))m,$$

or if  $B$  independent of  $m$ ,  $\nabla_m(m^T B) = B$ . Similarly if  $A$  is  $N_m \times 1$  and independent of  $m$ ,  $\nabla_m(A^T m) = A$ .

# Introduction to Gradients

Let  $A$  be  $N_d \times N_d$  with entries independent of the  $N_m \times 1$  vector  $m$ , let  $G$  be  $N_d \times M$  with entries independent of  $m$ , let  $d$  be a constant  $N_d$  dimensional column vector and define

$$O(m) = (Gm - d)^T A (Gm - d) = (m^T G^T - d^T) A (Gm - d).$$

$$\begin{aligned} \nabla_m O(m) &= [\nabla_m (m^T G^T - d^T)] A (Gm - d) + \\ &\quad [\nabla_m ((m^T G^T - d^T) A^T)] (Gm - d) = \\ &\quad G^T A (Gm - d) + G^T A^T (Gm - d) \end{aligned}$$

If  $A$  is symmetric,

$$\nabla_m O(m) = 2G^T A (Gm - d)$$

## Definition

A real symmetric  $N \times N$  matrix  $A$  is said to be positive definite if for every real  $N$ -dimensional column vector  $x \neq 0$ ,  $x^T A x > 0$

## Theorem

Let  $A$  be a  $N \times N$  real matrix.

- (i) If  $A$  is real symmetric ( $A^T = A$ ), then the eigenvalues  $\{\lambda_i\}_{i=1}^N$  of  $A$  are all real and  $A$  has a set of orthonormal eigenvectors,  $\{x_i\}_{i=1}^N$ .
- (ii)  $A$  is positive definite if and only if all eigenvalues of  $A$  are positive.
- (iii) If  $A$  is positive definite, then  $A$  is nonsingular. Moreover, if  $\{\lambda_i, x_i\}_i^N$  are the eigenpairs of  $A$ , then  $\{1/\lambda_i, x_i\}_i^N$  are eigenpairs of  $A^{-1}$ .

Positive semi-definite defined similarly. Real SPD matrices are important because they arise when solving self-adjoint pde's by discretization methods and in regularized least squares problems.

# Null Space of a Linear Operator

All vectors are column vectors unless explicitly stated otherwise. Let  $G$  be a known, or computable,  $N_d \times N_m$  matrix,  $G : R^{N_m} \rightarrow R^{N_d}$  and consider the forward model

$$d = Gm \quad (1)$$

Inverse problem: given noisy data  $d_{\text{obs}}$ , solve  $Gm = d_{\text{obs}}$ .

Note  $\nabla_m d^T = G^T$  and the entries of  $G$  are  $g_{i,j} = \frac{\partial d_i}{\partial m_j}$ .  $G$  is the sensitivity matrix.

## Definition

The null space of  $G$  is the set of all vectors in  $S(m) \equiv R^{N_m}$  that satisfy  $Gm = 0$ .

The null space of  $G$  is denoted by  $N(G)$ . Let  $\dim N(G) = \ell > 0$ . If  $m_0$  is a solution of  $Gm_0 = d_{\text{obs}}$  and  $\{m_k\}_{k=1}^{\ell}$  is a basis for  $N(G)$ , then for any scalars,  $\alpha_j, j = 1, 2, \dots, \ell$ ,  $m = m_0 + \sum_{j=1}^{\ell} \alpha_j m_j$  is a solution of Eq. 1. In fact, all solutions of the inverse problem of Eq. 1 are of this form

## Definition

The range of the  $N_d \times N_m$  matrix  $G$  (denoted  $\mathbf{R}(G)$ ) is the set of all vectors  $d$  in  $S(d) = R^{N_d}$  such that there exists at least one  $m$  in  $S(m) = R^{N_m}$  which satisfies  $Gm = d$ .

- The dimension of the range of the  $N_d \times M$  matrix  $G$  is denoted by  $\dim \mathbf{R}(G)$  and is called the **rank** of  $G$ . The rank of  $G$  is equal to the number of linearly independent columns of  $G$  which is also equal to the number of linearly independent rows of  $G$ .
- $\mathbf{R}(G) \leq \min\{N_d, N_m\}$ . If  $\mathbf{R}(G) = \min\{N_d, N_m\}$ ,  $G$  is said to be of full rank.

# Solutions of Linear Inverse Problems

- If  $N_d > N_m$ , then  $\text{rank}(G) \leq N_m < N_d$  so for some  $d_{\text{obs}} \in \mathbb{R}^{N_d}$ ,  $Gm = d_{\text{obs}}$  has no solution. Instead, try finding a least squares solution by minimizing  $\|Gm - d_{\text{obs}}\|_2^2$ . Setting the gradient of

$$O(m) \equiv \|Gm - d_{\text{obs}}\|_2^2 = (Gm - d_{\text{obs}})^T (Gm - d_{\text{obs}}) \quad (2)$$

equal to zero and rearranging gives

$$G^T Gm = G^T d_{\text{obs}} \quad (3)$$

So any least squares solution satisfies Eq. 3. If  $G$  is of full rank, then the  $N_m \times N_m$  matrix  $G^T G$  is also full rank, so Eq. 3 has a unique solution. In this case,  $G^T G$  which is always positive definite.

- Singular value decomposition can be applied to find the least squares solution of minimum norm.
- On the other hand, if  $N_m > N_d$ , then the rank of  $G^T G$  is  $\leq N_d < N_m$  so  $G^T G$  is singular. Some form of regularization is needed to pick a solution.

# Zero Order Regularization

Minimize the following objective function by setting its gradient w.r.t.  $m$  equal to zero.

$$O(m) = \frac{1}{2} \|Gm - d_{\text{obs}}\|_2^2 + \frac{1}{2} \lambda \|m - m_0\|_2^2. \quad (4)$$

$$\begin{aligned} \nabla O(m) &= G^T(Gm - d_{\text{obs}}) + \lambda(m - m_0) \\ &= (G^T G(m - m_0 + m_0 - d_{\text{obs}}) + \lambda I(m - m_0)) = 0 \end{aligned} \quad (5)$$

$$(\lambda I + G^T G)(m - m_0) = (G^T(d_{\text{obs}} - Gm_0)) \quad (6)$$

$$m_{\text{est}} = m_0 + (\lambda I + G^T G)^{-1}(G^T(d_{\text{obs}} - Gm_0)). \quad (7)$$

Solution is unique for any  $\lambda > 0$ , but if  $\lambda$  is small, the matrix  $(\lambda I + G^T G)$  can become very ill-conditioned. Note if  $m_0$  is a solution of the original inverse problem, then  $m_{\text{est}} = m_0$ . It is convenient to think of the problem as trying to find the closest model to  $m_0$  so that this model agrees well with the observed data;  $\lambda$  controls both the conditioning of the real SPD matrix  $(\lambda I + G^T G)$  and how strongly we wish to impose the constraint that  $m$  is close to  $m_0$ .



# First order regularization

Choose the “flattest” model that satisfies the data exactly or in a least squares sense. For the continuous inverse problem: find the flattest  $m(x)$  such that

$$d_{\text{obs},i} = \int_a^b G_i(x)m(x)dx.$$

the flattest solution is the function  $m(x) \in L^2[a, b]$  whose derivative is square integrable and minimizes

$$F(m) = \int_a^b \left( \frac{dm(x)}{dx} \right)^2 dx,$$

subject to the constraint that  $m(x)$  must satisfy the continuous inverse problem in a least-squares sense.

Second order regularization would replace the first derivative by the second derivative. We can also consider combinations of regularization terms, e.g., zero order plus second order.

# First Order Regularization

In the discrete linear inverse problem, difference operators replace derivatives. Thus, define the  $(N_m - 1) \times N_m$  matrix  $D$  by

$$D = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ 0 & 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}.$$

Letting  $d_{i,j}$  denote the element in the  $i$ th row and  $j$ th column of  $D$ ,  $d_{i,i} = -1$  and  $d_{i,i+1} = 1$  for  $i = 1, 2, \dots, N_m - 1$ , with all other entries of  $D$  equal to zero. Note that if  $\|Dm\|^2 = 0$ , then  $Dm = 0$  which implies  $-m_i + m_{i+1} = 0$  for  $i = 1, 2, \dots, N_m - 1$ , i.e., all components of  $m$  must be identical.

Minimize

$$O(m) = \frac{1}{2} [(m - m_0)^T A (m - m_0) + (Gm - d_{\text{obs}})^T C_D^{-1} (Gm - d_{\text{obs}})].$$

First order regularization:  $A = aI + D^T D$  with  $a = 0$ .

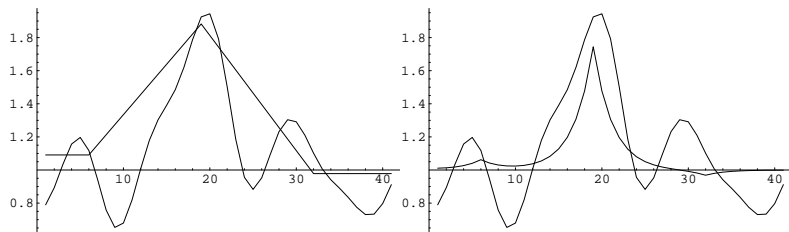
First order with a zeroth order component,  $a > 0$ ,  $A = aI + D^T D$ .

**Solution:**

$$m_{\text{est}} = m_0 + (A + G^T C_D^{-1} G)^{-1} G^T C_D^{-1} (d_{\text{obs}} - Gm_0).$$

# First Order Regularization, Example

Forty dimensional 1D inverse permeability field. Measurements at  $x = 6, 19, 32$  with fairly small measurement error,  $\sigma_d = 0.1$ .



**Figure:** Two estimates using the derivative as a measure of closeness to the prior. For the estimate on the left  $A = D^T D$ . On the right,  $A = D^T D + 0.2I$ .

Note for  $\alpha = 0$ , the solution linearly interpolates between measurement locations. In both cases, the solution is not differentiable at measurement locations.

# Singular Value Decomposition

(SVD)[Golub and van Loan(1989)], [Strang(1976)]

Let  $G$  be any  $N_d \times N_m$  real matrix. Then there exists an  $N_d \times N_d$  orthogonal matrix  $U$ , an  $N_m \times N_m$  orthogonal matrix  $V$  and an  $N_d \times N_m$  matrix  $\Lambda$  such that

$$G = U\Lambda V^T \quad (8)$$

$$\Lambda = \begin{bmatrix} \Lambda_q & O \\ O & O \end{bmatrix}$$

where  $\Lambda_q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_q)$  with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_q \geq 0$  and  $q = \min\{N_d, N_m\}$ . Note, in the following we choose  $r$  to be the largest integer such that  $\lambda_r > 0$  and then the preceding equation reduces to  $G = U_r \Lambda_r V_r^T$ . where the columns of  $U_r$  are equal to the first  $r$  columns of  $U$ , the columns of  $V_r$  are equal to the first  $r$  columns of  $V$  and  $\Lambda_r = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r)$

If  $G$  is a square real-symmetric positive definite matrix,  $V = U$  and the SVD decomposition is the standard Schur (eigenvalue-eigenvector) decomposition.

# Singular Value Decomposition (SVD)

- The  $N_d \times N_m$  matrix  $G$  maps  $R_{N_m}$  into  $R_{N_d}$ . Basic inverse problem is  $Gm = d_{\text{obs}}$  or in terms of the least squares problem  $G^T Gm = G^T d_{\text{obs}}$  where  $G^T G$  is  $N_m \times N_m$ . One way to show that  $G^T Gm = G^T d_{\text{obs}}$  always has at least one solution is to first show that the range of  $G^T G$  is identical to the range of  $G^T$ . To do so requires knowledge of the orthogonality of the four principle subspaces (see around slide 87).
- SVD solution of the least squares problem is

$$m_{\text{est}} = V \begin{bmatrix} \Lambda_r^{-1} & O \\ O & O \end{bmatrix} U^T d_{\text{obs}} = V_r \Lambda_r^{-1} U_r^T d_{\text{obs}}.$$

- We let  $u_i, i = 1, 2, \dots, N_d$  be the columns of  $U$ . These vectors form an orthonormal set and are referred to as the left singular vectors of  $G$ . The columns of  $V, v_j, j = 1, \dots, N_m$ , also form an orthonormal set and are referred to as the right singular vectors of  $G$ .

$$U^T U = I_{N_d} \quad U U^T = I_{N_d} \quad V^T V = V V^T = I_{N_m}.$$

# Singular Value Decomposition (SVD)

- The columns of  $U$  are a basis for  $R_{N_d}$  and the columns of  $V$  are a basis for  $R_{N_m}$ , so for any  $m \in R^{N_m}$  and any  $d_{\text{obs}} \in R^{N_d}$ , there exist two vectors of scalars,  $\alpha$  and  $\beta$  such that  $m = V\alpha = \sum_{j=1}^{N_m} \alpha_j v_j$  and  $d_{\text{obs}} = U\beta$ .
- With singular values of  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_r > 0$ , the frequency of  $v_j$  and  $u_j$  increase as  $j$  increases, so as we will see more clearly later, parameterizing  $m$  in terms of a linear combination of  $V_p = \{v_1, v_2, \dots, v_p\}$  where  $p < r$  where we preserve only the largest singular values may be appropriate. Then, the projection of  $m$  onto the span of  $V_p$  will preserve the smoothest (largest components) features of  $m$  provided we pick  $p$  appropriately, e.g., s.t.

$$\sum_{j=1}^p \lambda_j \geq 0.9 \sum_{j=1}^r \lambda_j.$$

- Among all solutions of the least squares problem,  $\min \| Gm - d_{\text{obs}} \|_2$ , the SVD solution is the solution of minimum norm.
- If  $G^T G$  is positive definite, then the SVD solution

$$m_{\text{est}} = (V_p \Lambda_p^{-1} U_p^T) d_{\text{obs}} \quad (9)$$

is the unique least squares solution.

- Eq. 9 is a solution of  $Gm = d_{\text{obs}}$  if and only if  $d_{\text{obs}}$  is in  $\text{span}\{u_1, u_2 \cdots u_p\}$ , which is always true if  $\text{rank } G = N_d$ . Note

$$Gm_{\text{est}} = (U_p \Lambda_p V_p^T)(V_p \Lambda_p^{-1} U_p^T) d_{\text{obs}} = U_p U_p^T d_{\text{obs}}, \quad (10)$$

where the final term in this equation is equal to  $d_{\text{obs}}$  if  $d_{\text{obs}}$  is in the span of the columns of  $U_p$ .



# Effect of Measurement Errors and Small Singular Values

Let  $m_{\text{true}}$  be the true model and let  $d_{\text{true}} = Gm_{\text{true}}$ . Because a measurement of  $d_{\text{true}}$  is corrupted by measurement error, we end up generating a solution of  $G^T Gm = G^T d_{\text{obs}}$  where  $d_{\text{obs}} = d_{\text{true}} - \epsilon$  where  $\epsilon$  is the measurement error. The SVD solution satisfies

$$m_{\text{est}} = V_p (\Lambda_p^{-1} U_p^T) d_{\text{obs}} = V_p (\Lambda_p^{-1} U_p^T) (d_{\text{true}} - \epsilon). \quad (11)$$

Similarly

$$m_{\text{true}} = V_p \Lambda_p^{-1} U_p^T d_{\text{true}} \quad (12)$$

Subtracting Eq. 11 from Eq. 12 gives

$$(m_{\text{true}} - m_{\text{est}}) = V_p \Lambda_p^{-1} U_p^T \epsilon \equiv V_p \Lambda_p^{-1} \delta = V_p \left[ \frac{1}{\lambda_1} \delta_1, \dots, \frac{1}{\lambda_p} \delta_p \right]^T. \quad (13)$$

Small singular values amplify the measurement error and increase the error in the solution. In particular, the singular vectors of highest frequency are corrupted the most.

# Truncated Singular Value Decomposition, TSVD

To mitigate the highly negative effect of small singular values, one can truncate the singular values.

Order singular values such that  $\lambda_1 \geq \lambda_2, \dots$ , choose  $p$  such that

$$\sum_{j=1}^p \lambda_j \geq 0.9 \sum_{j=1}^{\min\{N_m, N_d\}} \lambda_j \quad \text{or such that} \quad \sum_{j=1}^p \lambda_j \geq 0.01 \lambda_1, \quad (14)$$

where the 0.9 and 0.01 values are just rule of thumb choices. Set

$$G = U_p \Lambda_p V_p^T \equiv G_p,$$

$$m_{\text{est}} = (V_p \Lambda_p^{-1} U_p^T) d_{\text{obs}}. \quad (15)$$

TSVD may be viewed as regularization.

References to TSVD in History Matching:

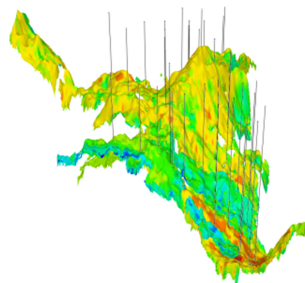
[Tavakoli and Reynolds(2010), Tavakoli and Reynolds(2011)],

[Shirangi and Emerick(2016)], and further extensions in progress.

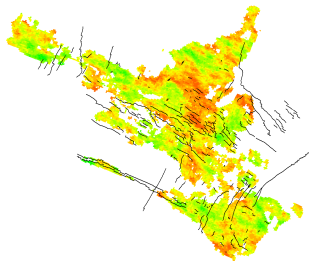
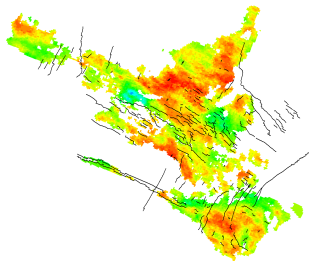
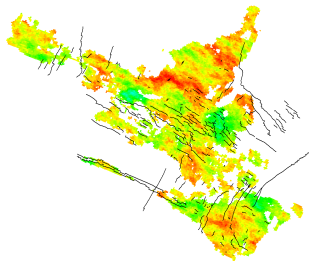
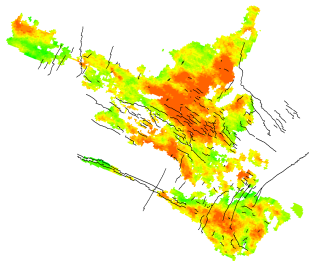
# ASSISTED HISTORY MATCHING - GOALS

- 1 Honor Observations.
- 2 Maintain geological realism.
- 3 Have predictive power not just in terms of matching future field production but in terms of fluid distributions.
- 4 Give at least some reasonable estimate of uncertainties for reservoir development and to manage risk - ideally characterize the posterior pdf.
- 5 Be compatible with company/institution simulation tools.
- 6 Be conceptually understandable to the user and management given a reasonable amount of training.
- 7 Be computationally feasible.

- Observed data:
  - 20 producers: oil rate, water rate, GOR, bottom-hole pressure.
  - 10 water injection: bottom-hole pressure.
- Initial ensemble:
  - 200 models.
  - Porosity and permeability ( $> 125,000$  active gridblocks).
  - Anisotropic ratio  $k_v/k_h$ .
  - Rock compressibility.
  - End point of water relative permeability curve.
- Data assimilation with ES-MDA (4 $\times$ ) with localization.

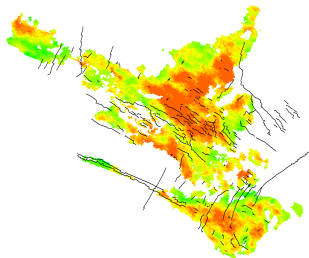


## Field Case 2: Model Plausibility – Permeability

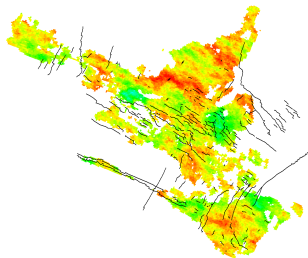


# Field Case 2: Model Plausibility – Permeability

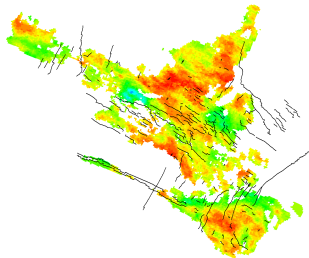
Prior # 200



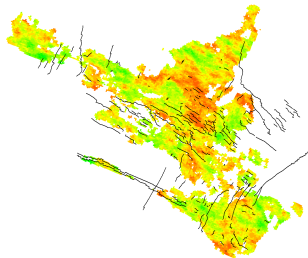
Post # 1



Post # 200

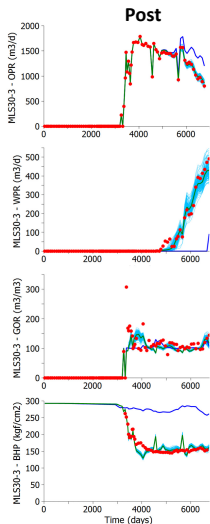
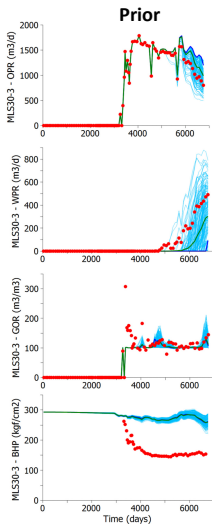


Prior # 1

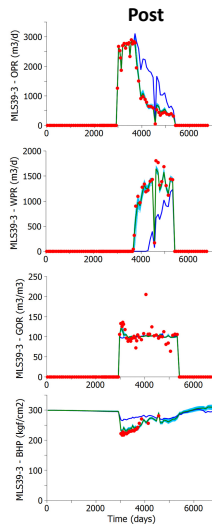
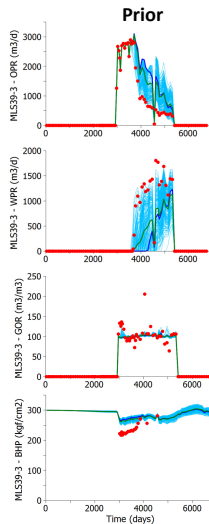


# Field Case 2: Well Data

## Well # 30



## Well # 39



- A random  $N_m$ -dimensional vector  $M$  is (multivariate) Gaussian or multinormal with mean  $m_{\text{prior}}$  and covariance  $C_M$  if the probability density for  $M$  is

$$f_M(m) = \frac{1}{(2\pi)^{N_m/2}} \frac{1}{\sqrt{\det(C_M)}} \exp\left(-\frac{1}{2}(m-m_{\text{prior}})^T C_M^{-1}(m-m_{\text{prior}})\right). \quad (16)$$

- The expectation or mean of  $M$  is  $m_{\text{prior}}$  and the autocovariance of  $M$  is  $C_M$ .
- Notation:  $M \sim N(m_{\text{prior}}, C_M)$ . We assume from this point on that  $M$  is an  $N_m$ -dimensional column vector and that  $D_{\text{obs}}$  is an  $N_d$ -dimensional column vector.



- $d = g(m)$  is the assumed theoretical relation between the model (a realization of  $M$ ) and the predicted data vector. In our work a realization of  $D_{\text{obs}}$  is given by

$$d_{\text{obs}} = g(m) + \epsilon$$

where  $\epsilon$  represents the random measurement error vector. If  $g(m)$  is incorrect, i.e., there is modeling error because of inexact physics, then  $\epsilon$  would be a sum of modeling and measurement errors. Any  $d_{\text{obs}}$  we actually measure then represents a realization of  $\epsilon$  added to  $g(m_{\text{true}})$ . Although  $d_{\text{obs}}$  depends on  $m$ , we have to assume  $\epsilon$  does not.

- Assume  $\epsilon \sim N(0, C_D)$  then  $f(d_{\text{obs}}|m)$  is Gaussian. Since  $E[d_{\text{obs}}|m] = g(m)$  and  $\text{cov}(d_{\text{obs}}|m) = C_D$ , it follows that

$$f_{D_{\text{obs}}, M}(d_{\text{obs}}|m) = \frac{1}{(2\pi)^{N_d/2} \sqrt{\det C_D}} \exp\left(-\frac{1}{2}(d_{\text{obs}} - g(m))^T C_D^{-1}(d_{\text{obs}} - g(m))\right)$$

# Gaussian Likelihood Function

$$f_{D_{\text{obs}},M}(d_{\text{obs}}|m) = \frac{1}{\sqrt{(2\pi)^{N_d} \det C_D}} \exp\left(-\frac{1}{2}(d_{\text{obs}} - g(m))^T C_D^{-1}(d_{\text{obs}} - g(m))\right)$$

Once an actual measurement  $d_{\text{obs}}$  is obtained, we no longer assume  $m$  is given. Instead we define the likelihood of any realization  $m$  of  $M$  by

$$L(m|d_{\text{obs}}) = \frac{1}{\sqrt{(2\pi)^{N_d} \det C_D}} \exp\left[-\frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}})\right]$$

A model which maximizes the likelihood is called a maximum likelihood estimate. Maximizing  $L$  is equivalent to minimizing

$$O(m) = \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}),$$

and at a minimum we must have

$$\nabla O(m) = \nabla(g(m)^T) C_D^{-1}(g(m) - d_{\text{obs}}) = 0,$$

which is a nonlinear system of equations and must be solved iteratively.

## Maximum Likelihood Estimate (MLE), Cont'd

- Referring to the preceding equation,  $G^T \equiv \nabla(g(m)^T)$  and  $G = (\nabla(g(m)^T))^T$  is the sensitivity matrix. The  $(i, j)$  entry of  $G$  is the sensitivity of the  $i$ th predicted data to the  $j$ th model parameter. Accurate efficient computation of  $G$  requires an adjoint solution.
- Gauss-Newton with  $\ell$  as the iteration index and a line search: For  $\ell = 0, 1, 2, \dots$  until convergence

$$(G_\ell^T C_D^{-1} G_\ell) \delta m^{\ell+1} = -G_\ell^T C_D^{-1} (g(m_\ell) - d_{\text{obs}}) \quad (17)$$

$$\alpha_\ell \approx \arg \min_{\alpha} \{O(m^\ell + \alpha \delta m^{\ell+1})\} \quad (18)$$

$$m^{\ell+1} = m^\ell + \alpha_\ell \delta m^{\ell+1} \quad (19)$$

End(For)

- The matrix  $(G_\ell^T C_D^{-1} G_\ell)$  may be badly ill-conditioned or even singular so we usually need regularization.

# Maximum Likelihood Estimate (MLE) with Levenberg-Marquardt

Levenberg-Marquardt [Levenberg(1944), Marquardt(1963)]. Precursor of trust region methods which are generally preferred to line-search methods.

$$O(m) = \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}), \quad (20)$$

Levenberg-Marquardt with  $\ell$  as the iteration index: For  $\ell = 0, 1, 2, \dots$  until convergence Solve Eq. 21

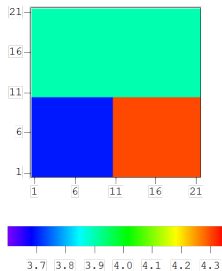
$$(\lambda_\ell I + G_\ell^T C_D^{-1} G_\ell) \delta m^{\ell+1} = -G_\ell^T C_D^{-1}(g(m_\ell) - d_{\text{obs}}) \quad (21)$$

for  $\delta m_{\ell+1}$ . Set  $m_c^{\ell+1} = m^\ell + \delta m_{\ell+1}$ .

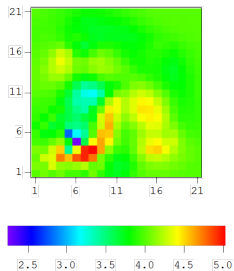
If  $O(m_c^{\ell+1}) < O(m^\ell)$ , then set  $m^{\ell+1} = m_c^{\ell+1}$  and set  $\lambda_{\ell+1} = \beta \lambda_\ell$  for  $\beta > 1$  and go to the next iteration. Otherwise, set  $\lambda_\ell = \gamma \lambda_\ell$  for  $\gamma < 1$  and repeat the iteration (Eq. 21). Often, one chooses  $\beta = 10$  and  $\gamma = 1/10$ .

# Levenberg-Marquardt vs Gauss-Newton

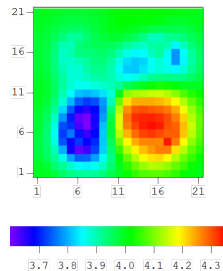
This example is from [Li et al.(2003)Li, Reynolds, and Oliver].



**True model**

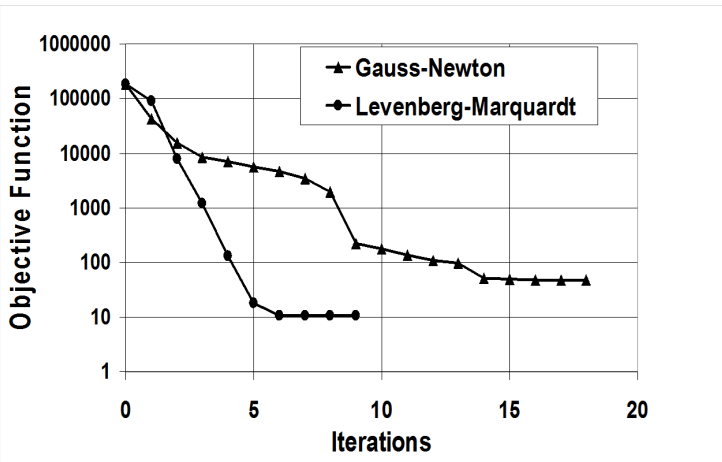


**Gauss-Newton**



**Levenberg-Marquardt**

# Levenberg-Marquardt vs Gauss-Newton



$$f_{M,D_{\text{obs}}}(m, d_{\text{obs}}) = f_{D_{\text{obs}}|M}(d_{\text{obs}}|m)f_M(m) = f_{M|D_{\text{obs}}}(m|d_{\text{obs}})f_{D_{\text{obs}}}(d_{\text{obs}}).$$

- Bayes theorem immediately follows and implies that

$$f(m|d_{\text{obs}}) = \frac{f(d_{\text{obs}}|m)f(m)}{f(d_{\text{obs}})}.$$

The theorem says nothing; it is the way we interpret it to define probabilities that is useful. If we have an specific observation ( $d_{\text{obs}}$ ), then we define the posterior pdf by

$$f(m|d_{\text{obs}}) = aL(m|d_{\text{obs}})f(m)$$

were  $a$  is the normalizing constant and  $f(m)$  is referred to as the prior pdf.

# Posterior PDF for Gaussian Prior and Gaussian Measurement/Modeling Errors

- Assume the prior model is multivariate Gaussian,  $N(m_{\text{prior}}, C_M)$ ,

$$f(m) = \frac{1}{(2\pi)^{N_d/2} \sqrt{\det C_M}} \exp\left(-\frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}})\right)$$

so that the product of the the prior and the likelihood inserted in Bayes' theorem gives

$$f(m|d_{\text{obs}}) = a \exp\left(-\frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}})\right) \times \\ \exp\left(-\frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}})\right) = a \exp(-O(m))$$

$$O(m) = \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}) + \\ \frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}})$$



$$f(m|d_{\text{obs}}) = a \exp(-O(m))$$

$$O(m) = \frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}}) + \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}).$$

- “The” maximum a posteriori (MAP) estimate is the model  $m$  that minimizes  $O(m)$ , or equivalently, maximizes  $f(m|d_{\text{obs}})$ . In general, a MAP estimate is not unique. If  $f(m|d_{\text{obs}})$  is multimodal, any gradient based optimization algorithm we use will estimate one of the modes. Which mode is obtained depends on the initial guess. For future reference, we note we can obtain a conditional mode by minimizing  $O(m)$  but the ensemble Kalman filter is designed to estimate a conditional mean. For a multimodal distribution, a conditional mean may not be very meaningful.

Minimize

$$O(m) = \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}) + \frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}})$$

For  $\ell = 0, 1, 2, \dots$  until convergence

$$(C_M^{-1} + G_\ell^T C_D^{-1} G_\ell) \delta m^{\ell+1} = -(C_M^{-1}(m^\ell - m_{\text{prior}}) + G_\ell^T C_D^{-1}(g(m_\ell) - d_{\text{obs}})) \quad (22)$$

$$\alpha_\ell \approx \arg \min_{\alpha} \{O(m^\ell + \alpha \delta m^{\ell+1})\} \quad (23)$$

$$m^{\ell+1} = m^\ell + \alpha_\ell \delta m^{\ell+1} \quad (24)$$

End(For)

Difficult to work with inverses of large covariance matrices, so we generally rewrite Eq. 26 using matrix inversion lemmas to obtain:

For  $\ell = 0, 1, 2, \dots$  until convergence

$$\delta m^{\ell+1} = m_{\text{prior}} - m^{\ell} - C_M G_{\ell}^T \left( C_D + G_{\ell} C_M G_{\ell}^T \right)^{-1} \\ \times \left( g(m^{\ell}) - d_{\text{obs}} - G_{\ell} (m^{\ell} - m_{\text{prior}}) \right).$$

$$\alpha_{\ell} \approx \arg \min_{\alpha} \{O(m^{\ell} + \alpha_{\ell} \delta m^{\ell+1})\} \quad (25)$$

$$m^{\ell+1} = m^{\ell} + \alpha_{\ell} \delta m^{\ell+1} \quad (26)$$

End(For)

- The entry in the  $i$ th row and the  $j$ th column of the sensitivity matrix  $G_\ell$  is

$$g_{i,j} = \frac{\partial g_i(m^\ell)}{\partial m_j} \quad (27)$$

- Generating all these sensitivity coefficients with the adjoint method requires  $N_d$  adjoint solutions, [Li et al.(2003)Li, Reynolds, and Oliver]. Generation of each adjoint solution is similar to one backward reservoir simulation run but generally requires on the order of 1/4 the time that a forward simulation run requires.
- Formulation of the adjoint problem for a fully-implicit reservoir simulator is easy because the main matrix involved in the discrete adjoint formulation is just the transpose of the Jacobian used in the final Newton iteration in the forward solution, [Li et al.(2003)Li, Reynolds, and Oliver].

- Adjoint can easily be used to compute sensitivity (derivative) of any predicted data w.r.t. any primary parameter that appear as a coefficient in the discretized systems of PDE's; sensitivities of data to any other parameter requires a known relation between that parameter and primary parameters so a chain rule can be applied.
- If  $N_d$  is large, the generation of  $N_d$  adjoint solutions is not feasible. In this case, we cannot afford Gauss-Newton or its LM form.
- On the other hand computing  $\nabla O(m^\ell)$  requires only one adjoint solution ([Zhang and Reynolds(2002)] which makes quasi-Newton methods feasible [Nocedal and Wright(1999)].

In quasi-Newton, e.g., LBFGS, one solves

$$\delta m^{\ell+1} = -H_{\ell}^{-1}O(m^{\ell}) \quad (28)$$

where at each iteration the new inverse Hessian is computed from the old one using a rank one update. Initial guess for  $H^{-1}$  is  $C_M$ . We generally use a trust-region quasi-Newton.

The quasi-Newton methods we have used emanate from the work of [Zhang and Reynolds(2002)]. Also see [Gao and Reynolds(2006)].

Sensitivity of the  $i$ th data to the  $j$  model parameter is  $\partial d_i / \partial m_j$ . This reflects how much a predicted data changes if we change  $m$ . But, in a Bayesian setting, it is not the best indicator to understand how a parameter will be changes to match the data. It is better to look at dimensionless sensitivities,

$$g_{D,i,j} = \frac{\partial d_i}{\partial m_j} \frac{\sigma_{m,j}}{\sigma_{d,i}}.$$

This generalizes to

$$G_D = C_D^{-1/2} G C_M^{1/2}.$$

One way to do regularization effectively comes from using TSVD of the dimensionless sensitivity coefficient matrices for an “ideal” parameterization in a Gauss-Newton or Levenberg-Marquardt algorithm; see

[Tavakoli and Reynolds(2010), Tavakoli and Reynolds(2010)],  
[Shirangi and Emerick(2016)], and further extensions in progress.

# MAP Estimate in Linear-Gaussian Case

Assuming Gaussian prior and  $g(m) = Gm$  where the sensitivity matrix  $G$  is independent of  $m$ .

- In the linear-Gaussian case, the Gauss-Newton method converges in one iteration for any initial guess.
- The posterior pdf,  $f(m|d_{\text{obs}})$ , is the Gaussian  $N(m_{\infty}, C_{\text{MAP}})$  where

$$m_{\infty} = m_{\text{MAP}} = m_{\text{prior}} - (C_M^{-1} + G^T C_D^{-1} G)^{-1} G^T C_D^{-1} (Gm_{\text{prior}} - d_{\text{obs}}), \quad (29)$$

$$C_{\text{MAP}} = C_{\infty} = H^{-1} = \left( C_M^{-1} + G^T C_D^{-1} G \right)^{-1} = C_M - C_M G^T (C_D + G C_M G^T)^{-1} G C_M \quad (30)$$

where the last equality follows from an inverse matrix lemma.

$$f(m|d_{\text{obs}}) = a \exp \left( -\frac{1}{2} (m - m_{\infty})^T [C_{\text{MAP}}]^{-1} (m - m_{\infty}) \right), \quad (31)$$

$$m \sim N(m_{\infty}, C_{\text{MAP}}) \quad (32)$$



- The MAP estimate may be too smooth to represent the reservoir property fields. In the linear case, the MAP estimate is the mean of the posterior pdf and means tend to be very smooth.
- In the linear case, we could sample the posterior  $N(m_\infty, C_{\text{MAP}})$  by

$$m_{c,j} = m_\infty + C_{\text{MAP}}^{1/2} Z_j; \quad Z_j \text{ a sample from } N(0, I)$$

- How do we sample the posterior pdf in the nonlinear case? Start with randomized maximum likelihood (RML), [Oliver et al.(1996)Oliver, He, and Reynolds]. For a simple proof that RAML samples correctly in the linear Gaussian case, see [Reynolds et al.(1999)Reynolds, He, and Oliver].

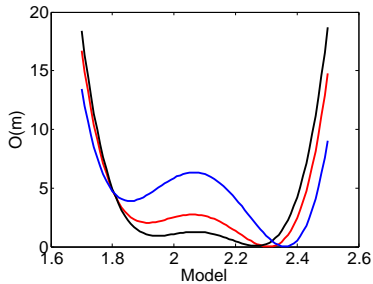
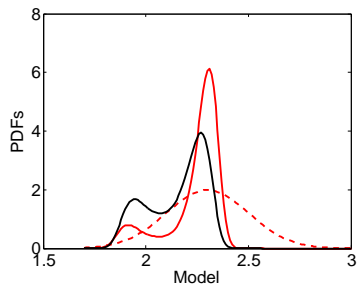
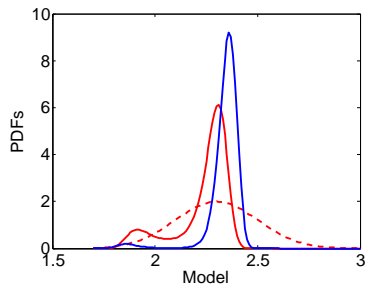
- We can use RML for both the linear and nonlinear cases but can only show it samples correctly for the linear-Gaussian case.

$$f(m|d_{\text{obs}}) = a \exp(-O(m)) \quad O(m) = \frac{1}{2}(m - m_{\text{prior}})^T C_M^{-1}(m - m_{\text{prior}}) + \frac{1}{2}(g(m) - d_{\text{obs}})^T C_D^{-1}(g(m) - d_{\text{obs}}). \quad (33)$$

- To generate a single sample of the posterior with RML, generate a sample  $d_{uc,j}$  from  $N(d_{\text{obs}}, C_D)$  and a sample  $m_{uc,j}$  from  $N(m_{\text{prior}}, C_M)$ . The corresponding RML sample of the posterior is defined as “the” model  $m_{c,j}$  obtained by minimizing

$$O_{r,j} = \frac{1}{2}(m - m_{uc,j})^T C_M^{-1}(m - m_{uc,j}) + \frac{1}{2}(g(m) - d_{uc,j})^T C_D^{-1}(g(m) - d_{uc,j}). \quad (34)$$

# Explanation of RML



- Oliver, Dean S., Metropolized Randomized Maximum Likelihood for sampling from multimodal distributions, SIAM JUQ, 2016.
- Li, X. and A. C. Reynolds, Generation of a proposal distribution for efficient MCMC characterization of uncertainty in reservoir description and forecasting, in Proceedings of the SPE Reservoir Simulation Conference, Montgomery, Texas, 20-22 February, SPE 182684, 2017.

# Equivalence between ES and Gauss-Newton

Following [Reynolds et al.(2006)Reynolds, Zafari, and Li], we derive the ensemble smoother as an approximation to RML-Gauss-Newton equation with full step where all ensemble members are updated with the same average sensitivity matrix.

$$m_j^{\ell+1} = m_{uc,j} - C_M G_{\ell,j}^T \left( C_D + G_{\ell,j} C_M G_{\ell,j}^T \right)^{-1} \times \left( g(m_j^\ell) - d_{uc,j} - G_{0,j}(m_j^\ell - m_{uc,j}) \right). \quad (35)$$

If we do a single iteration with  $\ell = 0$  (initial guess), set  $m_j^a \equiv m_j^1$ ,  $m_j^f = m_{uc,j}$  and  $d_j^f = g(m_{uc,j})$ , then

$$m_j^a = m_j - C_M G_{0,j}^T \left( C_D + G_{0,j} C_M G_{0,j}^T \right)^{-1} \times \left( d_j^f - d_{uc,j} \right). \quad (36)$$

Assuming that  $\overline{d^f} = d^f(\overline{m^f}) = g(\overline{m^f})$ , a first-order Taylor series expansion gives

$$d_j^f - \overline{d^f} = g(m_{uc,j}) - g(\overline{m^f}) = G(\overline{m^f})(m_j^f - \overline{m^f}) \equiv \overline{G}(m_j^f - \overline{m^f}) \quad (37)$$

# Equivalence between ES and Gauss-Newton

$$\begin{aligned}\tilde{C}_{MD}^f &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_{uc,j} - \bar{m}^f)(d_j^f - \bar{d}^f)^T \\ &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j^f - \bar{m}^f)(\bar{G}(m_j^f - \bar{m}^f))^T = C_M \bar{G}^T\end{aligned}\tag{38}$$

$$\begin{aligned}\tilde{C}_{DD}^f &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (d_j^f - \bar{d}^f)(d_j^f - \bar{d}^f)^T \\ &= \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (\bar{G}(m_j^f - \bar{m}^f))(\bar{G}(m_j^f - \bar{m}^f))^T = \bar{G} C_M \bar{G}^T\end{aligned}\tag{39}$$

Replacing  $G_{0,j}$  by  $\bar{G}$  in 36

$$m_j^a = m_j - C_M \bar{G}^T \left( C_D + \bar{G} C_M \bar{G}^T \right)^{-1} \times \left( d_j^f - d_{uc,j} \right). \quad (40)$$

Using Eqs. 38 and 39 in Eq. 40 gives the following equation which is the ensemble smoother:

$$m_j^a = m_j - \tilde{C}_{MD}^f \left( C_D + \tilde{C}_{DD}^f \right)^{-1} \times \left( d_j^f - d_{uc,j} \right). \quad (41)$$

$$\tilde{C}_{MD}^f = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j^f - \bar{m}^f)(d_j^f - \bar{d}^f)^T$$

$$\tilde{C}_{DD}^f = \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (d_j^f - \bar{d}^f)(d_j^f - \bar{d}^f)^T$$

The ensemble smoother

$$\begin{aligned} m_j^a &= m_j^f - \tilde{C}_{MD}^f \left( C_D + \tilde{C}_{DD}^f \right)^{-1} \left( d_j^f - d_{uc,j} \right) \\ &= m_j^f - \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j^f - \bar{m}^f)(d_j^f - \bar{d}^f)^T (x_j) = \\ m_j^f - \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j^f - \bar{m}^f)(d_j^f - \bar{d}^f)^T x_j &= m_j^f - \frac{1}{N_e - 1} \sum_{j=1}^{N_e} (m_j^f - \bar{m}^f) a_j. \end{aligned} \tag{42}$$



- Since  $\overline{m}^f$  is a linear combination of the ensemble of the initial ensemble of models, Eq. 42 indicates that every  $m_j^a$  is a linear combination of the models in the initial ensemble, i.e., we are trying to find a linear combination of the initial ensemble members that matches data. Thus, if the models contain large structural features, it will be difficult to capture the geology by assimilating data with the ES.
- Avoids explicit computation of large covariance matrices.
- Easy to implement with any reservoir simulator; the simulator truly becomes a black box.
- Generates multiple realizations and a rough characterization of uncertainty.

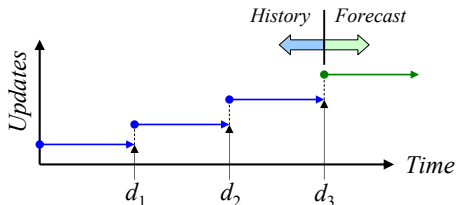
- Generate an ensemble of realizations from the prior geological model  $m_j^0 = m_j^f$  for  $j = 1, 2, \dots, N_e$ .
- For  $j = 1, 2, \dots, N_e$ , sample  $d_{uc,j}$  from  $N(d_{obs}, C_D)$ .  
( $d_{uc,j} = d_{obs} + C_D^{-1} Z_{d,j}$ .)
- Update equation:

$$m_j^a = m_j^f + \tilde{C}_{MD}^f \left( \tilde{C}_{DD}^f + C_D \right)^{-1} \left( d_{uc,j} - d_j^f \right), \quad (43)$$

where  $d_j^f = d^f(m_j^f) = g(m_j^f)$ ,  $\tilde{C}_{MD}^f$  is the estimated covariance between the forecast models and the forecast data and  $\tilde{C}_{DD}^f$  is the auto-covariance of predicted data.

# Ensemble Kalman Filter (EnKF)

- Sequential data assimilation method.



- Sequential data assimilation with EnKF requires to updating both model parameters and primary variables (parameter-state-estimation problem).

$$y^n = \left[ \begin{array}{c} \text{model param.} \\ \phi^T, (\ln k)^T, \dots, \underbrace{(p^n)^T, (S_w^n)^T, (S_g^n)^T, (R_s^n)^T}_{\text{primary variables}} \end{array} \right]^T$$

- In practice, we may observe **inconsistency** between the updated parameters and states.

- Explicit computation of large covariances are avoided. Reasonably robust and can be automated in workflows.
- ES faster and easier to implement than EnKF; no inconsistency between states and parameters, no overwriting reservoir simulation input files, complete parallelization.
- EnKF and ES are effectively the same as updating each ensemble member by doing one iteration of the Gauss-Newton method using the same average sensitivity matrix for all ensemble members.
- To match data, we can only adjust  $N_e$  coefficients. Thus the initial ensemble is important. Moreover, there is a possibility of ensemble collapse particularly when measurement error is low. Every time one perfect datum is assimilated, you lose one degree of freedom in the ensemble, Lorenc (2003).
- Approximation of covariances by a small number of ensemble members leads to spurious correlations. Ensemble collapse and spurious correlations can be ameliorated by covariance localization.

- **Comment:** For history-matching with gradient-based methods some form of damping at early iterations is usually necessary to avoid getting trapped in an excessively rough model which does not give a very good history match (Li et al., SPEJ, 2003, Gao and Reynolds, SPEJ 2006.)

# Multiple Data Assimilation with ES (or EnKF)

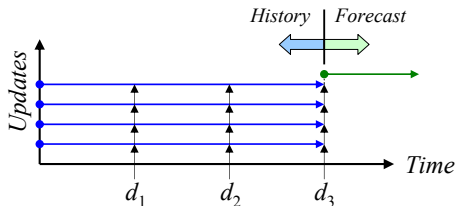
- Assimilate the same data multiple times with inflated data covariance ( $C_D$ ).
- Single and multiple data assimilations are equivalent for the linear-Gaussian case as long as the inflation coefficients of  $C_D$  satisfy the following condition:

$$\sum_{i=1}^{N_a} \frac{1}{\alpha_i} = 1^* \quad (\text{ex. } \alpha_i = N_a \text{ for } i = 1, \dots, N_a).$$

- Samples correctly for the linear case.
- For the nonlinear case, we replace a single (and potentially large) correction by  $N_a$  smaller corrections.

\*Proof in *Computers & Geosciences* 2013

- Parameter estimation method.



- Update equation:

$$m_j^a = m_j^f + \tilde{C}_{MD}^f \left( \tilde{C}_{DD}^f + \alpha_i C_D \right)^{-1} \left( d_{uc,j} - d_j^f \right).$$

- $\tilde{C}_{MD}^f$  and  $\tilde{C}_{DD}^f$  calculated from the updated ensemble at each iteration.
- Same data assimilated  $N_a$  times with  $\alpha_i$  satisfying:

$$\sum_{i=1}^{N_a} \frac{1}{\alpha_i} = 1 \quad (\text{ex. } \alpha_i = N_a \text{ for } i = 1, \dots, N_a).$$

- 1 Choose the number of data assimilations,  $N_a$ , and the coefficients  $\alpha_i$  for  $i = 1, \dots, N_a$ . Generate the initial ensemble denoted by  $\{m_j^a\}_{j=1}^{N_e}$
- 2 For  $i = 1$  to  $N_a$ :
  - 1 Set  $m_j^f = m_j^a$  for  $j = 1, 2, \dots, N_e$ .
  - 2 Run the ensemble from time zero.
  - 3 For each ensemble member, perturb the observation vector using

$$d_{uc,j} = d_{obs} + \sqrt{\alpha_i} C_D z_d,$$

where  $z_d \sim N(0, I_{N_d})$ .

- 4 Compute covariances and update the ensemble using

$$m_j^a = m_j^f + \tilde{C}_{MD}^f \left( \tilde{C}_{DD}^f + \alpha_i C_D \right)^{-1} \left( d_{uc,j} - d_j^f \right).$$

for  $j = 1, 2, \dots, N_e$ .



- ES-MDA was proposed by Emerick and Reynolds in 2011. Motivated by analogy of ES/EnKF with a single iteration of Gauss-Newton iteration (Reynolds et al., 2005, 2006) and the need to provide regularization and damping at early iteration.
- Avoids statistical inconsistencies between updated models parameters and states that can occur with EnKF.
- Truly black box and completely parallelizable.

- When gridblock permeabilities, porosities, can be described as Gaussian random fields, assisted history matching is relatively straightforward. But in many (most) situations, we really to estimate the distribution of facies together with the distribution of the rock property fields within the facies.
- For random looking facies distributions, pluri-Gaussian methods have proven potential: [Liu and Oliver(2003)] and more recent work by Dean Oliver and his colleagues.
- Sebacher, B., R. Hanea, and A. Stordal, The adaptive plurigaussian simulation (APS) model versus the truncated plurigaussian simulation (TPS) model used in the presence of hard data, in Geostatistics Valencia 2016. For more complicated models such as channelized systems:
- Kernel PCA: Sarma, P., L. J. Durlofsky, and K. Aziz, Kernel principal component analysis for efficient differentiable parameterization of multipoint geostatistics, *Mathematical Geosciences*, 2008.

- Discrete Cosine Transforms: B. Jafarpour and D. B. McLaughlin, Efficient Permeability Parameterization With the Discrete Cosine Transform, 2007 RSC, paper SPE 106453.
- Hai X. Vo and Louis J. Durlofsky, “Data Assimilation and Uncertainty Assessment for Complex Geological Models using a New PCA-Based Parameterization,” *Computational Geosciences*, 2015. (Combines optimized principal component analysis (O-PCA) with RML history matching that can be applied to non-Gaussian models.
- History matching of multi-facies channelized reservoirs using ES-MDA with common basis DCT, Zhao, Yu and Forouzanfar, Fahim and Reynolds, Albert C, *Computational Geosciences*, 2016.

## Step 1 - Sampling

- Generate the prior ensemble using MPS algorithm or object-based modeling.

## Step 2 - Common basis DCT

- Implement DCT for the facies field,  $\mathbf{m}_{\text{facies},j}$ , of the  $j$ th ensemble member to obtain  $\Phi_j$  and  $\mathbf{v}_j$ .
- Construct common set of basis function,  $\bar{\Phi}$ , and recalculate the corresponding coefficients,  $\mathbf{v}'_j$ .  
Note: we use  $\mathbf{v}_j$  to replace  $\mathbf{v}'_j$  in the following text.
- Assembly the model parameters vector  $\mathbf{y}$  by

$$\mathbf{y} = \left[ \mathbf{v}^T, \mathbf{m}_{\text{shale}}^T, \mathbf{m}_{\text{levee}}^T, \mathbf{m}_{\text{sand}}^T \right]^T$$

## Step 3 - ES-MDA & Post-processing

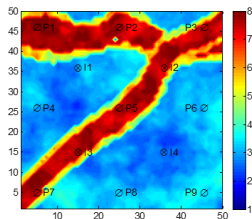
- Update the model parameters vector of the  $j$ th ensemble member,  $\mathbf{y}_j$ , to obtain  $\mathbf{v}_j^{\text{updated}}$ .
- Reconstruct the continuous facies field,  $\tilde{\mathbf{m}}_{\text{facies},j}^{\text{updated}}$ , with common basis set,  $\bar{\Phi}$ , and updated coefficients,  $\mathbf{v}_j^{\text{updated}}$ .
- Obtain the discrete facies field,  $\mathbf{m}_{\text{facies},j}^{\text{updated}}$ , by implementing the post-processing technique, [Vo and Durlofsky(2015)].
- Renew the petrophysical property fields of the entire reservoir, then start next iteration of ES-MDA.

# HM of channelized reservoir: 3D case

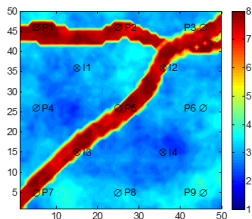
- Reservoir dimension:  $50 \times 50 \times 5$  with  $\Delta x = \Delta y = 100$  ft and  $\Delta z = 20$  ft
- 3 facies: shale (0,  $k=20$ mD,  $\ln(\bar{k})=2.996$ ,  $\sigma(\ln(k))=0.3$ ),  
levee (1,  $k=200$ mD,  $\ln(\bar{k})=5.298$ ,  $\sigma(\ln(k))=0.3$ ),  
sand (2,  $k=2000$ mD,  $\ln(\bar{k})=7.601$ ,  $\sigma(\ln(k))=0.3$ )
- Two-phase (oil and water) flow
- Models are generated using object-based modeling
- Initial pressure: 5000 psi, initial water saturation: 0.2
- 4 injectors and 9 producers
- Known facies hard data at all well locations
- History matching period: 10 years, prediction period: 5 years
- $N_e = 200$ ,  $N_c = 100$  (retained number for each layer),  $N_a = 8$

# HM of channelized reservoir: 3D case

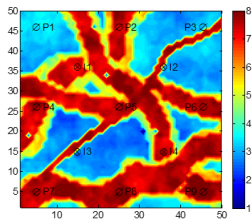
- Two geological zones: zone 1 (layers 1 and 2), zone 2 (layers 3 through 5)



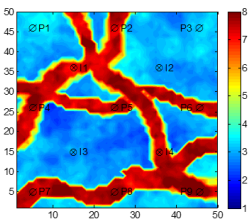
(a) True layer 1



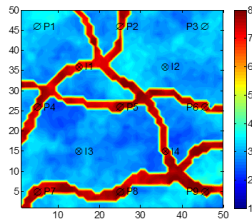
(b) True layer 2



(c) True layer 3



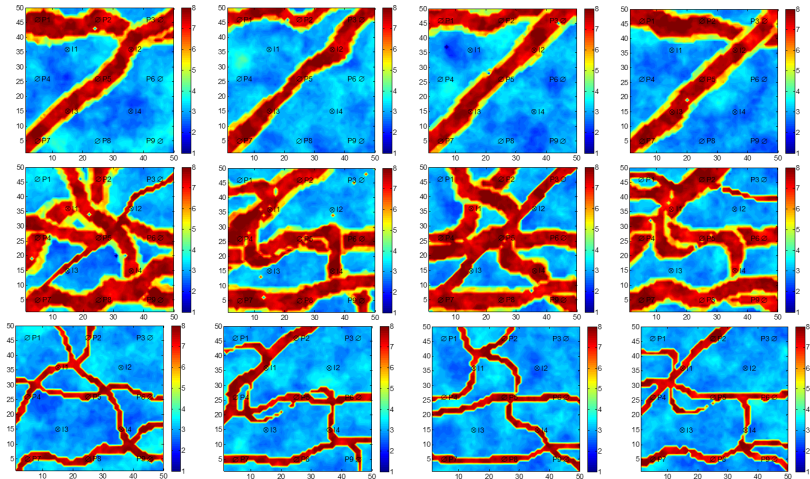
(d) True layer 4



(e) True layer 5

# HM of channelized reservoir: 3D case

- True model and three **prior** realizations: layer 1 (first row), layer 3 (second row) and layer 5 (third row)



(a) True

(b) Realization 1

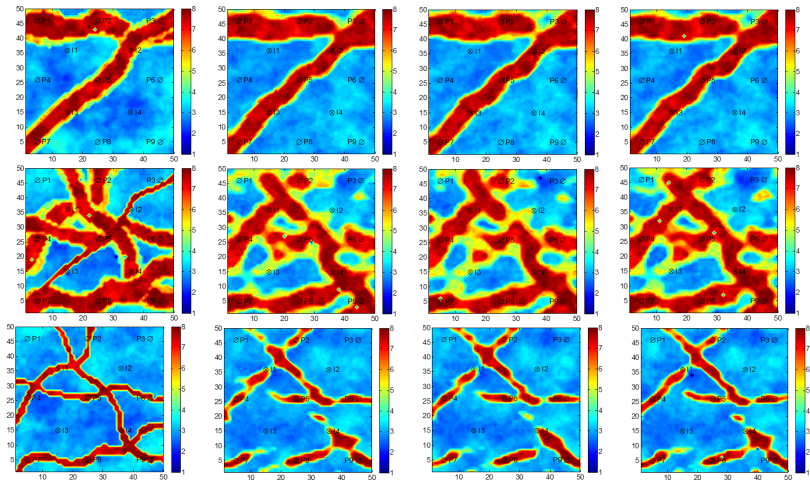
(c) Realization 2

(d) Realization 3



# HM of channelized reservoir: 3D case

- True model and three **posterior** realizations: layer 1 (first row), layer 3 (second row) and layer 5 (third row)



(a) True

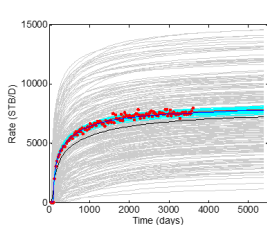
(b) Realization 1

(c) Realization 2

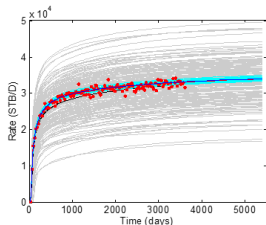
(d) Realization 3

# HM of channelized reservoir: 3D case

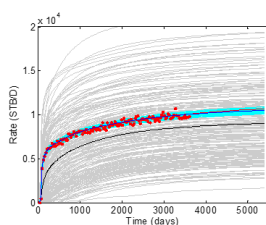
## ● Data match results



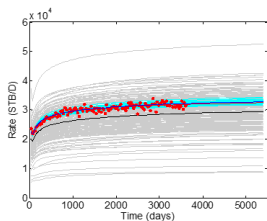
(a) P1 water flow rate



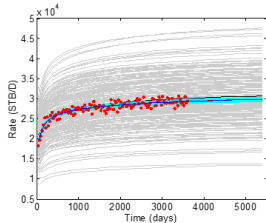
(b) P5 water flow rate



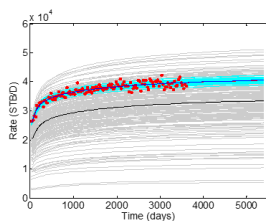
(c) P8 water flow rate



(d) I2 water injection rate



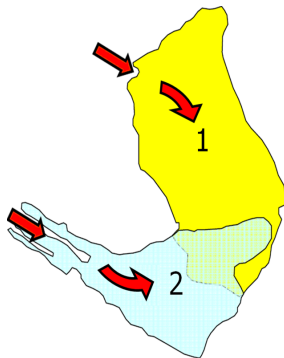
(e) I3 water injection rate



(f) I4 water injection rate

# Field Case 1

- Upper zone of a turbidite reservoir in Campus Basis.
- **Current model:** manual history matching.
- Very clean sandstones with high permeability (1–20 Darcy).
- **Observed data:**
  - **3D seismic** (P-impedance) before production.
  - **Time-lapse (4D) seismic** after 5 years of production.
  - 10 years of **production data** (7.6 years for history matching and 2.4 for forecast).

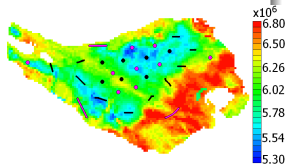


- **Initial ensemble:**
  - 200 models.
  - Porosity using sequential Gaussian simulation (SGS).
  - NTG using SG co-simulation with porosity.
  - Permeability with SGS using well testing interpretation data as “hard data.”
- Data assimilation with **EnKF** and **ES-MDA**.
- Both methods with **localization**.

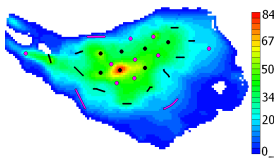
# Observed 3D Seismic

- Inverted P-impedance data.
- Average P-impedance over all layers of the model.

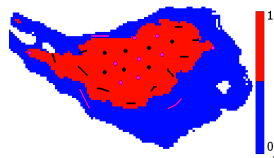
### P-impedance ( $\text{kg/m}^2\text{s}$ )



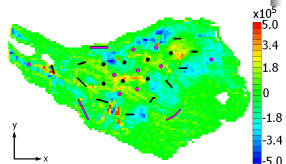
### Reservoir thick (m)



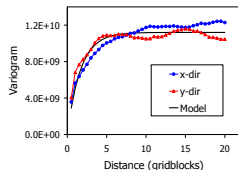
### Active data



### Residual ( $\text{kg/m}^2\text{s}$ )



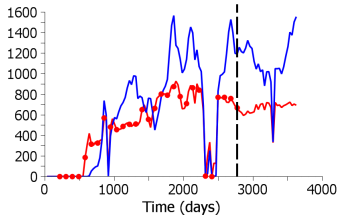
### Variograms (residual)



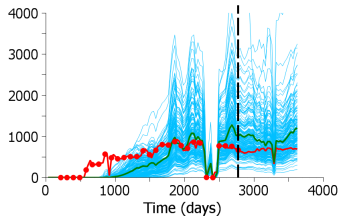
- Spatially correlated noise with range of 500 meters.

# Water Rate (P-86)

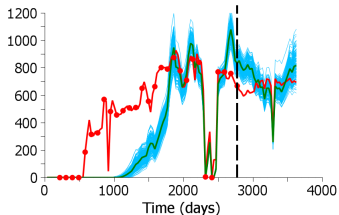
## Manual history matching



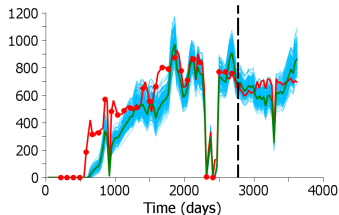
## Initial



## EnKF

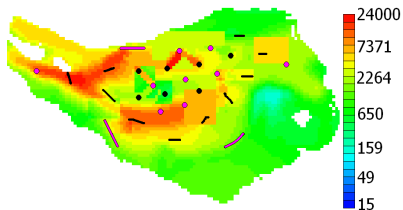


## ES-MDA (4x)

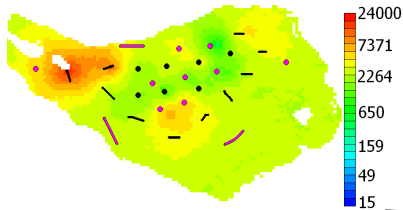


# Permeability – Ensemble Mean

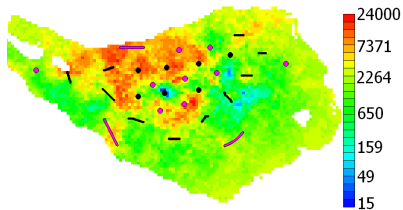
Manual history matching



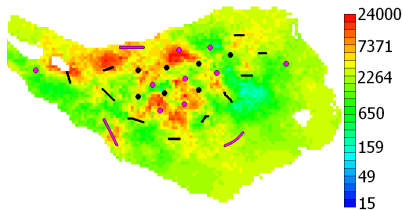
Prior mean



EnKF



ES-MDA (4x)



- The CPU time of ES-MDA (4×) was **only 4% higher** than the CPU time of EnKF.



- Randomized maximum likelihood with adjoint gradients and quasi-Newton trust region method samples modes of the distribution but does not give a strictly correct sampling. Proposal: Find multiple modes (local minima) with gradient-based RML, cluster if necessary and then build a Gaussian mixture model to use as a proposal distribution. See my talk at Workshop III, Data Assimilation, Uncertainty Reduction and Optimization for Subsurface Flow, May 23, 2017.
- Common industry practice: Reduce model to a small number of parameters; use ED to build a proxy models commonly a set of response surfaces. Response surface used as the forward model when proposing new models in a MCMC procedure (Chevron).

THIS ENDS MY TALK.

However, there are a few additional slides on matrix inversion lemmas, probability and one very basic slide on linear algebra.

First matrix inversion formula is

$$(C_M^{-1} + G^T C_D^{-1} G)^{-1} G^T C_D^{-1} = C_M G^T (C_D + G C_M G^T)^{-1},$$

Second matrix inversion lemma is

$$(C_M^{-1} + G^T C_D^{-1} G)^{-1} = C_M - C_M G^T (C_D + G C_M G^T)^{-1} G C_M,$$

# Derivation of First Matrix Inversion Lemma

Start with the following identity:

$$G^T C_D^{-1} (C_D + G C_M G^T) = (C_M^{-1} + G^T C_D^{-1} G) (C_M G^T)$$

Because  $C_D$  and  $C_M$  are real-symmetric positive definite,  $(C_D + G C_M G^T)$  and  $(C_M^{-1} + G^T C_D^{-1} G)$  are also real-symmetric positive definite and hence nonsingular.

Premultiplying the preceding equation by  $(C_M^{-1} + G^T C_D^{-1} G)^{-1}$  and postmultiplying by  $(C_D + G C_M G^T)^{-1}$  gives

$$(C_M^{-1} + G^T C_D^{-1} G)^{-1} G^T C_D^{-1} = C_M G^T (C_D + G C_M G^T)^{-1},$$

which we refer to as the first matrix inversion lemma (identity). We will use it to derive the second matrix inversion lemma on the next slide.

# Derivation of the Second Matrix Inversion Lemma

We derived the first matrix inversion lemma:

$$(C_M^{-1} + G^T C_D^{-1} G)^{-1} G^T C_D^{-1} = C_M G^T (C_D + G C_M G^T)^{-1},$$

Now note that

$$\begin{aligned} I_{N_m} &= (C_M^{-1} + G^T C_D^{-1} G)^{-1} (C_M^{-1} + G^T C_D^{-1} G) \\ &= (C_M^{-1} + G^T C_D^{-1} G)^{-1} C_M^{-1} + (C_M^{-1} + G^T C_D^{-1} G)^{-1} (G^T C_D^{-1}) G. \end{aligned}$$

Using the first matrix inversion lemma, we can rewrite the preceding equation as

$$I_{N_m} = (C_M^{-1} + G^T C_D^{-1} G)^{-1} C_M^{-1} + C_M G^T (C_D + G C_M G^T)^{-1} G.$$

Postmultiplying by  $C_M$  and rearranging

$$(C_M^{-1} + G^T C_D^{-1} G)^{-1} = C_M - C_M G^T (C_D + G C_M G^T)^{-1} G C_M,$$

which we refer to as the second matrix inversion lemma. Note that if  $G$  is the sensitivity matrix, then the LHS of the preceding equation is  $C_{\text{MAP}}$ .

# Orthogonality of Four Principal Subspace associated with $N_d \times N_m$ matrix $G$

- (i) Every row of  $G$  is orthogonal to  $N(G)$ . This means that  $\mathbf{R}(G^T)$  and  $N(G)$  are orthogonal subspaces of  $R^{N_m}$  and  $\mathbf{R}(G^T) \cap N(G) = \{0\}$ . Moreover since  $\dim \mathbf{R}(G^T) + \dim N(G) = N_m$ ,  $\mathbf{R}(G^T) \cup N(G) = R^{N_m}$ .
- (ii) Every column of  $G$  is perpendicular to  $N(G^T)$  which means that  $\mathbf{R}(G)$  and  $N(G^T)$  are orthogonal subspaces of  $R^{N_d}$  and  $\mathbf{R}(G) \cap N(G^T) = \{0\}$ .  
Similar to item (i), we also have that  $\mathbf{R}(G) \cup N(G^T) = R^{N_d}$ .

**Definition:**  $X$  is a continuous random variable if the probability of  $X$  taking the value  $x$  for any  $x$  is zero. We write this as

$$P(X = x) = 0, \quad -\infty < x < \infty. \quad (44)$$

When this is the case, it is often more useful to speak of the probability density, or the probability density function (pdf). A pdf is any nonnegative function  $f$  such that

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (45)$$

A continuous random vector  $X = [X_1, X_2, \dots, X_n]^T$  is defined similarly. A nonnegative function  $f_X(x)$  is a (joint) probability density function for  $X$  if

$$\int_{R^n} f_X(x) dx = 1. \quad (46)$$

The probability that a realization  $x$  of  $X$  lies within the region  $a_1 \leq X_1 < b_1, \dots$ , and  $a_n \leq X_n < b_n$  is

$$\begin{aligned} &P(a_1 \leq X_1 < b_1 \text{ and } \dots \text{ and } a_n \leq X_n < b_n) \\ &= \int_{a_1}^{b_1} \dots \int_{a_n}^{b_n} f_X(x_1, \dots, x_M) dx_1 \dots dx_M \quad (47) \end{aligned}$$



The *marginal probability* density for the random variable  $X_i$  is found by integrating the joint probability density over the remaining variables. For example, the marginal density for the random variable  $X_1$  is

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_X(x_1, x_2, \dots, x_n) dx_2 dx_3 \dots dx_n. \quad (48)$$

The *conditional probability* density of  $X$  given  $Y$  is defined by

$$f_{X|Y}(x|y) = \frac{f_{XY}(x, y)}{f_Y(y)}. \quad (49)$$

In this case, we are computing the probability density that  $X$  is equal to  $x$  when  $Y$  is equal to  $y$ . For reservoir characterization problems, we might be interested in the conditional probability density for the permeability at a particular location given a set of well test data.

From Eq. 49 we immediately see that

$$f_{XY}(x, y) = f_{X|Y}(x|y)f_Y(y) = f_{Y|X}(y|x)f_X(x). \quad (50)$$

When we are interested in the conditional probability of a variable given measurements of another variable, we rearrange the equations slightly to obtain:

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{f_Y(y)}. \quad (51)$$

This result is called Bayes' theorem. It provides the basis for computing the conditional probability of reservoir parameters from a combination of inaccurate measurements and a marginal probability density for reservoir parameters.

- The expected value of a function  $g$  of a random vector  $X$  whose probability density is  $f_X(x)$  is

$$E[g(X)] = \int_{R^n} g(x)f_X(x) dx. \quad (52)$$

- The expected value (or *expectation*) of  $X$  is also called the mean value and is often denoted  $\mu_X$  or simply  $\mu = \int_{R^n} x f_X(x) dx$ .

Let  $X$  be a vector of random variables, i.e.,

$$X = [X_1 \quad X_2 \quad \cdots X_n]^T \quad (53)$$

and let  $\mu_X$  be the expected value of  $X$ . The auto-covariance of the elements of the random vector  $X$  is defined as follows,

$$\begin{aligned} C_{XX} &= E[(X - \mu_X)(X - \mu_X)^T] \\ &= E[XX^T - \mu_X X^T - X \mu_X^T + \mu_X \mu_X^T] \\ &= E[XX^T] - E[\mu_X X^T] - E[X \mu_X^T] + E[\mu_X \mu_X^T] \quad (54) \\ &= E[XX^T] - \mu_X \mu_X^T - \mu_X \mu_X^T + \mu_X \mu_X^T \\ &= E[XX^T] - \mu_X \mu_X^T \end{aligned}$$

- This is the covariance matrix given by

$$C_X = \begin{bmatrix} E[X_1X_1] - \mu_1\mu_1 & \cdots & E[X_1X_n] - \mu_1\mu_n \\ E[X_2X_1] - \mu_2\mu_1 & \cdots & E[X_2X_n] - \mu_2\mu_n \\ \vdots & \vdots & \vdots \\ E[X_nX_1] - \mu_n\mu_1 & \cdots & E[X_nX_n] - \mu_n\mu_n \end{bmatrix} \quad (55)$$

- The components  $X_i$  and  $X_j$  are uncorrelated if

$$E[X_iX_j] = \mu_i\mu_j \quad (56)$$

in which case the covariance matrix is a diagonal matrix with the variances of the  $X_i$  on its diagonal.







If  $X$  and  $Y$  are random column vectors of dimension  $n$ , then the covariance between  $X$  and  $Y$  is

$$\begin{aligned}\text{cov}(X, Y) &= C_{X,Y} = E[(X - \mu_X)(Y - \mu_Y)^T] \\ &= [E[(X_i - \mu_{X_i})(Y_j - \mu_{Y_j})]] = [\text{cov}(X_i, Y_j)]\end{aligned}\quad (57)$$






Letting  $\rho_{i,j}$  denote the correlation coefficient between  $X_i$  and  $X_j$ , i.e.,  $\rho_{i,j} = (\text{cov}[X_i, Y_j]) / (\sigma_{X_i} \sigma_{Y_j})$  and letting  $D_x$  and  $D_y$ , respectively, be diagonal matrices with  $\sigma_{X_i}$  and  $\sigma_{Y_i}$ , respectively as the  $i$ th diagonal entry, we have

$$C_{X,Y} = D_x [\rho_{i,j}] D_y \quad (58)$$

where  $[\rho_{i,j}]$  is the correlation matrix.

-  Gao, G. and A. C. Reynolds, An improved implementation of the LBFGS algorithm for automatic history matching, *SPE Journal*, **11**(1), 5–17, 2006.
-  Golub, G. H. and C. F. van Loan, *Matrix Computations*, 2nd edn., The Johns Hopkins University Press, Baltimore, 1989.
-  Jacquard, P., Théorie de l'interprétation des mesures de pression, *Revue de L'Institut Français du Pétrole*, **19**(3), 297–334, 1964.
-  Levenberg, K., A method for the solution of certain non-linear problems in least squares, *Quarterly of Applied Mathematics*, **2**, 164–168, 1944.
-  Li, R., A. C. Reynolds, and D. S. Oliver, History matching of three-phase flow production data, *SPE Journal*, **8**(4), 328–340, 2003.
-  Liu, N. and D. S. Oliver, Conditional simulation of truncated random fields using gradient methods, in *Proceedings of the IAMG 2003 Portsmouth Meeting*, 2003.



-  Marquardt, D. W., An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Indust. Appl. Math.*, **11**(2), 431–441, 1963.
-  Nocedal, J. and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.
-  Oliver, D. S., N. He, and A. C. Reynolds, Conditioning permeability fields to pressure data, in *Proceedings of the European Conference for the Mathematics of Oil Recovery*, 1996.
-  Reynolds, A. C., N. He, and D. S. Oliver, Reducing uncertainty in geostatistical description with well testing pressure data, in *Reservoir Characterization—Recent Advances*, (edited by R. A. Schatzinger and J. F. Jordan), pp. 149–162, American Association of Petroleum Geologists, 1999.
-  Reynolds, A. C., M. Zafari, and G. Li, Iterative forms of the ensemble Kalman filter, in *Proceedings of 10th European Conference on the Mathematics of Oil Recovery, Amsterdam, 4–7 September, 2006*.



Shirangi, M. G. and A. A. Emerick, An improved TSVD-based Levenberg-Marquardt algorithm for history matching and comparison with Gauss-Newton, *Journal of Petroleum Science and Engineering*, **143**, 258–271, 2016.



Strang, G., *Linear Algebra and Its Applications*, Academic Press, New York, 1976.



Tavakoli, R. and A. C. Reynolds, History matching with parameterization based on the SVD of a dimensionless sensitivity matrix, *SPE Journal*, **15**(12), 495–508, 2010.



Tavakoli, R. and A. C. Reynolds, Monte Carlo simulation of permeability fields and reservoir performance predictions with SVD parameterization in RML compared with EnKF, *Computational Geosciences*, **15**(1), 99–116, 2011.



Vo, H. X. and L. J. Durlofsky, Data assimilation and uncertainty assessment for complex geological models using a new pca-based parameterization, *Computational Geosciences*, **19**(4), 747–767, 2015.



Zhang, F. and A. C. Reynolds, Optimization algorithms for automatic history matching of production data, in *Proceedings of 8th European Conference on the Mathematics of Oil Recovery, Freiberg, Germany, 3–6 September, 2002.*