SPE-173236-MS

Quantification of the Impact of Ensemble Size on the Quality of an Ensemble Gradient Using Principles of Hypothesis Testing

R.M. Fonseca, S.S. Kahrobaei, and L.J.T. van Gastel, Delft University of Technology (TU Delft);
O. Leeuwenburgh, TNO; J.D. Jansen, TU Delft

This paper was prepared for presentation at the SPE Reservoir Simulation Symposium held in Houston, Texas, USA, 23–25 February 2015.

Abstract

With an increase in the number of applications of ensemble optimization (EnOpt) for production optimization, the theoretical understanding of the gradient quality has received little attention. An important factor that influences the quality of the gradient estimate is the number of samples. In this study we use principles from statistical hypothesis testing to quantify the number of samples needed to estimate an ensemble gradient that is comparable in quality to an accurate adjoint gradient. We develop a methodology to estimate the necessary ensemble size to obtain an approximate gradient that is within a predefined angle compared to the adjoint gradient, with a predefined statistical confidence. The method is first applied to the Rosenbrock function (a standard optimization test problem), for a single realization, and subsequently for a case with uncertainty, represented by multiple realizations (robust optimization). The maximum allowed error applied in both experiments is a 10° angle between the directions of the EnOpt gradient and the exact gradient. For the single-realization case we need, depending on the perturbation size, 900, 5 and 3 samples to estimate a "good" gradient with 95% confidence at 50 points in the optimization space for 50 different random sequences. For the robust case, the conventional EnOpt approach is to couple one model realization with one control sample, which leads to a computationally efficient technique to estimate a mean gradient. However, our results show that in order to be 95% confident the original one-to-one model realization to control sample ratio formulation is not sufficient. To achieve the required confidence requires a ratio of 1:1100, i.e. each model realization is paired with 1100 control samples using the original formulation. However, using a modified formulation we need a ratio of 1:10 to stay within the maximum allowed error for 95% of the points in space, though a 1:1 ratio is sufficient for 85% of the points. We also tested our methodology on a reservoir case for deterministic and robust cases, where we observe similar trends in the results. Our results provide insight into the necessary number of samples required for EnOpt, in particular for robust optimization, to achieve a gradient comparable to an adjoint gradient.

Introduction

Multiple studies have shown the successful application of various optimization algorithms to maximize hydrocarbon recovery or net present value (NPV) over the producing life of a hydrocarbon reservoir. For

such problems, gradient-based techniques, in terms of accuracy and computational efficiency, are the most successful and widely applied. The adjoint method provides the most accurate gradient and is computationally the most efficient method. However, the adjoint method requires access to a reservoir simulator's source code which for commercial simulators is practically impossible. Additionally it requires a considerable amount of time to implement and is not very flexible in adaptation to different control types. These limitations of the adjoint method have led to the development of alternative gradient-based techniques. One such alternative technique, Ensemble Optimization (EnOpt), inspired by the Ensemble Kalman Filter (EnKF) method was first introduced by Lorentzen et al. (2006) and Nwaozo (2006).

Chen (2008) proposed the now standard formulation of the EnOpt method which uses an ensemble of randomly perturbed control vectors to approximate a gradient of the objective function with respect to some specific controls. The major advantages of EnOpt are its ease of implementation, flexibility to adapt to different control types and ability to be used with any reservoir simulator. The major drawback of this method, relative to the adjoint method, is its computational inefficiency and inaccuracy of the gradient approximation. Nonetheless, recently many studies such as Chen et al. (2009), Chen and Oliver (2010), and Leeuwenburgh et al. (2010) have demostrated the applicability of EnOpt for large-scale production optimization problems. Most of these papers have focused on deterministic optimization problems starting from a single reservoir model. However, in reality the geological and reservoir modeling process is fraught with uncertainties since a reservoir is modeled using uncertain interpretations based on uncertain data sources such as seismics, well logs etc. Incorporating these uncertainties into the optimization framework is vital to achieve results of any practical significance.

Van Essen et al. (2009) introduced a 'robust optimization' methodology in conjunction with the adjoint method to include the effect of uncertainties into the optimization framework. They used an ensemble of equi-probable reservoir models with differing geology and maximized the expectation of the objective function over this ensemble of models. Chen (2008) introduced this robust optimization concept within the ensemble optimization framework. They proposed the use of an ensemble of controls of equal size as the ensemble of geological models. Coupling of one member from the control ensemble with one member of the geological ensemble, a mean gradient can be approximated with the EnOpt formulation. This formulation, while computationally very attractive for robust optimization, has received scant attention with respect to its theoretical understanding. Recently Fonseca et al. (2014) demonstrated a case wherein the original formulation for ensemble-based robust optimization leads to inferior results and suggested a modified gradient formulation.

For EnOpt the two main inputs which influence the quality of the approximate gradient are the covariance matrix used to create the ensemble of perturbed controls and the number of control samples created, i.e. the ensemble size. The effect of the covariance matrix has been investigated recently in Fonseca et al. (2013) and a theoretical foundation for the use of a varying covariance matrix has been provided in Stordal et al. (2014). Sarma and Chen (2014) have investigated the applicability of different sampling techniques to improve the quality of a gradient estimate. However none of those studies have performed a detailed investigation into the effect of ensemble size on the estimated ensemble gradient quality.

In this study we aim to quantify the ensemble size required to approximate a gradient comparable to the adjoint gradient especially for robust optimization problems using principles from hypothesis testing and statistical analysis. In this paper we first provide an introduction of the hypothesis testing methodology and the different test statistics used. This will be followed by a detailed set of experiments on the widely used Rosenbrock function for cases with and without model uncertainty. Finally we test the proposed methodology on a medium-sized reservoir model, again with and without geological uncertainty

Theory

The two most commonly used objective functions for production optimization are ultimate recovery or an economic objective such as Net Present Value (NPV). In this work we chose the objective function J to be the NPV, defined in the usual fashion as

$$J=\sum_{k=1}^{K}\left(\frac{\left\{\left[(q_{o,k})\cdot r_o-(q_{wp,k})\cdot r_{wp}\right]-\left[(q_{wi,k})\cdot r_{wi}\right]\right\}\cdot\Delta t_k}{(1+b)^{t_k/\tau_t}}\right), \tag{1}$$

where $q_{o,k}$ is the oil production rate in bbl/day, $q_{wp,k}$ is the water production rate in bbl/day, $q_{wi,k}$ is the water injection rate in bbl/day, r_o is the price of oil produced in \$/bbl, r_{wp} is the cost of water produced in \$/bbl, r_{wi} is the cost of water injected in \$/bbl, Δt_k is the difference between consecutive time steps in days, b is the discount factor expressed as a fraction per year, t_k is the cumulative time in days corresponding to time step k, and τ_t is the reference time period for discounting, typically one year. Gradient-based optimization requires the gradient $\mathbf{g} = (dj/d\mathbf{u})^T$ which is used within an optimization algorithm to iteratively optimize the objective function. For a detailed description of various available optimization algorithms see, e.g., Nocedal and Wright (2006). Usually the elements of the control vector are required to stay within upper and lower bounds, and different approaches for such bound control problems are available. Moreover, in addition to these constraints on the inputs, there may be constraints on the outputs of the simulator, which are much more difficult to handle. However, in this paper we are only considering the quality of the gradients under the presence of simple bound constraints.

Ensemble Optimization (EnOpt)

Ensemble-Based Deterministic Formulation

In this section we outline the standard formulation of the EnOpt algorithm as proposed by Chen et al. (2009). We take $\mathbf{u}$ to be a single control vector containing all the control variables to be optimized. This vector has length N equal to the product of the controllable well parameters (number of well settings like bottom hole pressures, rates or valve settings) and the number of control time steps. Chen et al. (2009) sample the initial mean control vector from a Gaussian distribution while, at later iteration steps the final control vector of the previous iteration is taken as the mean control. However the initial controls can also be chosen by the user, as will be done in our experiments.

$$\mathbf{u}=\begin{bmatrix}u_1 & u_2 & \cdots & u_N\end{bmatrix}^T. \tag{2}$$

To estimate the EnOpt gradient, a multivariate, Gaussian distributed ensemble $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M\}$ is generated with a distribution mean $\mathbf{u}$ and a predefined distribution covariance matrix $\tilde{\mathbf{C}}$ where M is the ensemble size. During the iterative optimization process, $\mathbf{u}$ is updated until convergence, whereas $\tilde{\mathbf{C}}$ is, traditionally, kept constant. [An alternative procedure, in which $\tilde{\mathbf{C}}$ is updated during the optimization process, is decribed in Fonseca et al. (2013)]. In our implementation of EnOpt the ensemble members $\mathbf{u}_i$, $i = 1, 2, \ldots, M$, are created using

$$\mathbf{u}_i=\bar{\mathbf{u}}+\tilde{\mathbf{C}}^{1/2}\mathbf{z}_i, \tag{3}$$

with

$$\bar{\mathbf{u}}=\frac{1}{M}\sum_{i=1}^{M}\mathbf{u}_i. \tag{4}$$

We use a Cholesky decomposition to calculate $\tilde{\mathbf{C}}^{1/2}$, and draw $\mathbf{z}_i$ from a univariate Gaussian distribution. To estimate the gradient, a mean-shifted ensemble matrix is defined as

$$\mathbf{U}=\begin{bmatrix}\mathbf{u}_1-\bar{\mathbf{u}} & \mathbf{u}_2-\bar{\mathbf{u}} & \cdots & \mathbf{u}_M-\bar{\mathbf{u}}\end{bmatrix}. \tag{5}$$

[Note that in earlier publications we used the transposed version of **U**. We modified our notation to bring it in line with that of textbooks such as Conn et al. (2009).] A mean-shifted objective function vector is defined as

$$\mathbf{j} = \begin{bmatrix} J_1 - \bar{J} & J_2 - \bar{J} & \cdots & J_M - \bar{J} \end{bmatrix}^T, \tag{6}$$

where the expectation of the objective function is given by

$$\bar{J} = \frac{1}{M}\sum_{i=1}^{M} J_i. \tag{7}$$

The approximate gradient as proposed by Chen (2008) and Chen et al. (2009) is given by

$$\mathbf{g} = \mathbf{C}_{uu}^{-1}\mathbf{c}_{uJ}\ , \tag{8}$$

where

$$\mathbf{C}_{uu} = \frac{1}{M-1}(\mathbf{U}\mathbf{U}^T) \tag{9}$$

and

$$\mathbf{c}_{uJ} = \frac{1}{M-1}(\mathbf{U}\mathbf{j}) \tag{10}$$

are ensemble (sample) covariance and cross-covariance matrices respectively. (Note that $\mathbf{c}_{uJ}$ is a one-dimensional matrix, i.e. a vector.) For the usual case where $M < N$, matrix $\mathbf{C}_{uu}$ is rank-deficient, and Chen (2008) and Chen et al. (2009) therefore propose not to use expression (8) but, instead, to use

$$\mathbf{g}' = \mathbf{C}_{uu}\mathbf{C}_{uu}^{-1}\mathbf{c}_{uJ} = \mathbf{c}_{uJ}, \tag{11}$$

or

$$\mathbf{g}'' = \mathbf{C}_{uu}\mathbf{c}_{uJ}\ . \tag{12}$$

Alternatively, the pre-multiplication in equation (12) can be performed with $\tilde{\mathbf{C}}$, leading to

$$\mathbf{g}''' = \tilde{\mathbf{C}}\mathbf{c}_{uJ}. \tag{13}$$

All three expressions (11), (12) and (13) can be interpreted as modified (regularized or smoothed) approximate gradients. In the present paper we use a straight gradient, i.e. expression (8), computed as the underdetermined least squares solution

$$\mathbf{g} = (\mathbf{U}\mathbf{U}^T)^{\dagger}\mathbf{U}\mathbf{j}\ , \tag{14}$$

where the superscript † indicates the Moore-Penrose pseudo inverse, which is conveniently computed using a singular value decomposition (SVD); see, e.g., Strang (2006). Moreover, we use smoothed and double-smoothed versions of equation (14):

$$\mathbf{g}'''' = \tilde{\mathbf{C}}(\mathbf{U}\mathbf{U}^T)^{\dagger}\mathbf{U}\mathbf{j}\ , \tag{15}$$

$$\mathbf{g}''''' = \tilde{\mathbf{C}}\tilde{\mathbf{C}}(\mathbf{U}\mathbf{U}^T)^{\dagger}\mathbf{U}\mathbf{j}\ , \tag{16}$$

Equation (14) was also described in Dehdari and Oliver (2012), while Do and Reynolds (2013) recently demonstrated that it is akin to what is known as a 'Simplex gradient' in, e.g., Conn et al. (2009). Do and Reynolds (2013) also provided theoretical connections between various ensemble methods such as simulataneous perturbation stochastic approximation (SPSA), Simplex gradient, EnOpt etc. Moreover, they proposed a modification to the gradient formulation which uses the current control vector $\mathbf{u}^l$ and the

corresponding objective function value j^l to calculate the control and objective function anomalies **U** and **j**:

$$\mathbf{U} = \left[\mathbf{u}_1 - \mathbf{u}^\ell \quad \mathbf{u}_2 - \mathbf{u}^\ell \quad \cdots \quad \mathbf{u}_M - \mathbf{u}^\ell\right], \tag{17}$$

$$\mathbf{j} = \left[J_1 - J^\ell \quad J_2 - J^\ell \quad \cdots \quad J_M - J^\ell\right]^T, \tag{18}$$

where the superscript l is the optimization iteration counter.

Equations (11-16) can all be used to estimate a gradient-based on either the original [equations (5) and (6)] or the modified [equations (17) and (18)] formulations. Thus we can estimate as many as twelve different gradient formulations for deterministic cases. Further varieties will emege when considering robust optimization.

Ensemble-Based Robust Formulation

Chen (2008) proposed a computationally efficient robust ensemble algorithm in which she used an ensemble of controls of equal size as the ensemble of geological models. Coupling one member from the control ensemble with one member of the geological ensemble a mean gradient is approximated; see Chen (2008) for details. Therefore the ensemble size M of the controls is equal to the number of geological models, i.e. a 1:1 ratio. Hence only M simulation runs are needed to approximate the 'robust' gradient of the objective function. Recently Stordal et al. (2014) reached a similar conclusion starting from a different mathematical viewpoint. However the theoretical understanding of using this 1:1 ratio is still incomplete. As an alternative to this formulation, Fonseca et al. (2014) propose a modified formulation for the robust gradient which no longer uses the mean-shifted control samples and objective values, equations (5) and (6). Instead, in equation (5) the control sample mean $\bar{\mathbf{u}}$ is replaced by the control vector of the current iteration step, $\mathbf{u}^l$:

$$\mathbf{U} = \left[\mathbf{u}_1 - \mathbf{u}^\ell \quad \mathbf{u}_2 - \mathbf{u}^\ell \quad \cdots \quad \mathbf{u}_M - \mathbf{u}^\ell\right], \tag{19}$$

The new formulation replacing equation (6) is

$$\mathbf{j} = \left[J_1 - J_1^\ell \quad J_2 - J_2^\ell \quad \cdots \quad J_M - J_M^\ell\right]^T, \tag{20}$$

Note that equation (19) is identical to equation (17) as used in the deterministic modified expression of Do and Reynolds (2013), but that equation (20) is different from equation (18). This modified gradient formulation [based on equations (19) and (20)] will also be tested in our set of experiments. It behaves distinctly different compared to the original robust formulation [based on equations (5) and (6)]. First, because the subtractions in the objective function values in equation (20) are with respect to the individual objective function values J_i^ℓ and not with respect to the mean. Second, because for bound-constrained control problems, $\mathbf{u}^l$ and $\bar{\mathbf{u}}$ may be shifted with respect to each other. Thirdly, because the effect of outliers, which may strongly influence the mean value our least-squares approach to estimate the gradient, is reduced in the modified formulation.

Note: all the different gradient formulations (11-16) for deterministic optimization are also applicable to the robust case. Together with the robust modified formulation [equations (19) and (20)] this leads to a total of 18 potential robust gradient formulations for the 1:1 ratio (i.e. one control perturbation for each geological realization) approach. However, another distinction can be made if we use other ratios. E.g., Raniolo et al. (2013) suggest the use of 20 control perturbations for every model realization. For every model realization, using the 1:20 ratio, they estimate an individual gradient, whereafter they take the mean of the individual gradients to obtain the robust gradient. This formulation will hereafter be referred to as the 'Mean of Individual Gradients' (MIG). Alternatively, one can combine all the controls and objective function anomalies to estimate a single robust gradient, i.e. not estimate individual gradients for every model realization. This approach will hereafter be referred to as the 'Hotch-Potch Gradient' (HPG). This

additional disctinction leads to a total of 30 potential formulations [2 times 18 minus 6 because for the MIG approach there is no difference between using equations (18) and (20)].

Adjoint method

The adjoint method has been investigated extensively for use in data assimilation and production optimization. Detailed derivations for the production optimization case can be found in, e.g., Brouwer and Jansen (2004), Sarma et al. (2005), Kraaijevanger et al. (2007) and Jansen (2011). The adjoint method is the most accurate and computationally efficient method for computing a gradient. Computation of the gradient only requires one forward simulation and one fast backward computation. Therefore the number of simulation runs is independent of the number of controls. However for robust optimization using the adjoint requires running the forward and backward simulation for every geological realization, thus to compute the robust gradient, the same number of simulation runs will be performed as required for the robust EnOpt gradient using the 1:1 ratio. For our experiments we assume that the adjoint gradient is the exact gradient, which the EnOpt method tries to approximate. In this study the adjoint module available in the Shell in-house simulator was used (Kraaijevanger et al. 2007).

Hypothesis testing

We use principles from hypothesis testing to validate the research goal of this paper, namely to test if the approximate EnOpt gradient is comparable in quality to the adjoint gradient. To be able to determine the difference in gradients we compute the angle between them by using the dot product:

$$\cos(\theta) = \frac{\mathbf{g}_{adj} \cdot \mathbf{g}_{ens}}{\|\mathbf{g}_{adj}\| \|\mathbf{g}_{ens}\|}. \tag{21}$$

Another measure that describes the difference in direction of two vectors is the length of the difference between the normalized adjoint and EnOpt gradients, defined as

$$\Delta = \left\| \frac{\mathbf{g}_{adj}}{\|\mathbf{g}_{adj}\|} - \frac{\mathbf{g}_{ens}}{\|\mathbf{g}_{ens}\|} \right\|. \tag{22}$$

To eliminate the effect of a difference in gradient magnitude between both methods the gradient vectors are made into unit vectors by dividing them by their norm. When Δ goes to zero, the two gradients will point in the same direction, just like if the angle goes to zero. These two test parameters can be used to cross-validate each other, because a difference of 10° corresponds to a dimensionless length difference of 0.175. Thus the two equivalent null hypotheses used are

$$\begin{aligned} H_0 : \quad & \theta \le 10^\circ , \\ : \quad & \Delta \le 0.175 . \end{aligned} \tag{23}$$

The statistical inference method used is based on pre-defined confidence intervals for the testing parameters defined above.

Confidence intervals

Creating a confidence interval is a method to define a range at and the certainty that the true value of an estimated parameter lies within it, based on the knowledge of the sampling distribution (Dekking et al., 2005). In our numerical experiments we create a dataset of our parameters, given in equations (21) and (22). The parameter of interest θ is the maximum allowable deviation of the EnOpt gradient with regard to the exact or adjoint gradient. As it is virtually impossible to achieve a 100% confidence, we apply a confidence level of $\gamma = 0.95$. The general definition of a confidence interval assumes a two-sided interval, i.e. an upper and a lower limit. However it is also possible to have a one-sided interval. As the test parameters used for the numerical experiments are absolute values of deviations we only want to find the confidence interval of the maximum deviation, thus the upper limit. Using a one-sided interval, the

confidence interval is just the integral of the probability distribution, i.e. the cumulative density function (CDF).

Beta distribution

The EnOpt method samples random points from a normal distribution with a user-defined standard deviation. The distribution of the test parameters (θ and Δ) are, however, not normally distributed due to the non-linearity of the objective function and the function for the test statistics. In order to determine a confidence interval the distribution of the underlying parameters needs to be known. The two parameters θ and Δ are, by definition, ratios of two different gradients, and the beta distribution is suitable to fit data sets that are ratios. The beta distribution forms a class of continuous probability distributions, parameterized by two shape parameters a and b that define the shape of the distribution. AbouRizk et al. (1994) outline several methods to determine these shape parameters, of which we have chosen to use the maximum likelihood estimator. AbouRizk et al. (1994) also demonstrate that there is virtually no difference in the results when using different fitting methods. Note that the beta distribution is always bounded in the interval [0,1]. However, in this study, because we use the cosine function, our data is bounded between [-1 1]. Thus in order to use the beta distribution for angles greater than 90 degrees (i.e. $\cos(\theta) < 0$) we simply reset the angle to 90 degrees (i.e. $\cos(\theta) = 0$).

Methodology using Beta Distributions

The methodology proposed in this work to either accept or falsify our null hypothesis and estimate the necessary ensemble size to accept our null hypothesis is as follows:

- Sample points in the control space and compute the adjoint and EnOpt gradients at each point.
- Estimate the fitting parameters a and b of the beta distribution using the data.
- Plot the CDF for the beta distribution and compute the confidence interval for the predefined error.
- Repeat for varying ensemble sizes.
- Read-off the ensemble size to achieve the desired 95% confidence interval

We have chosen the 95% confidence interval to test our methodology, however any different confidence interval can be chosen within the same workflow. Varying the desired confidence interval will automatically vary the ensemble size needed to accept our hypothesis. In essence an ensemble size is found that gives an accurate gradient approximation in a number of times equal to the confidence interval, i.e. if the numerical experiment would be repated many times, 95% of the experiments would result in an approximate gradient within the error margin compared to the true (adjoint) gradient. For a more detailed explanation of confidence intervals see, e.g. Dekking et al. (2005).

Methodology with Traditional Hypothesis Testing Principles

In addition to using the beta distribution to test our null hypothesis we can also use a traditional hypothesis testing approach to either accept or falsify our hypothesis and estimate the necessary ensemble size as follows:

- Sample points in the control space and compute the adjoint and EnOpt gradients at each point.
- Count the number of points that satisfy the null hypothesis
- Compute the confidence interval for the predefined maximum allowable error.
- Repeat for varying ensemble sizes until the confidence interval is achieved.

In this paper we compare the results of the two methodologies for the different test statistics in the following section.

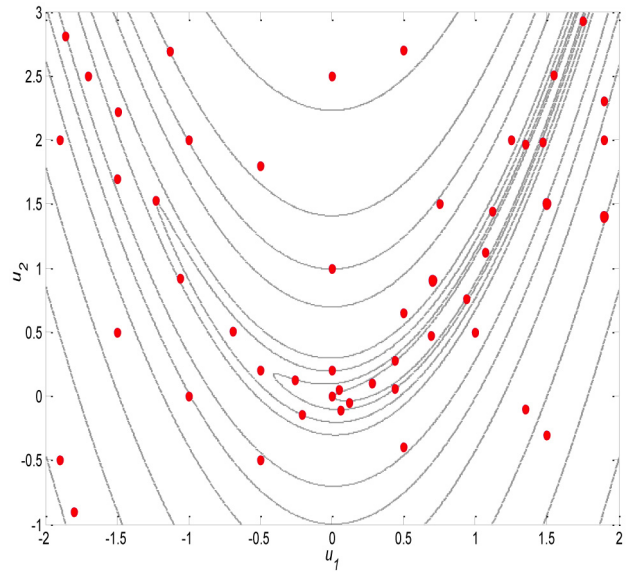

Figure 1—Contour plot of Rosenbrock function given by equation (24). Red dots are 50 points randomly distributed in space.

Numerical Example

Rosenbrock Function

The methodology is first applied to the non-linear Rosenbrock function named after the mathematician who first used it to demonstrate his optimization algorithm, see Rosenbrock (1960). This analytical function has since been used as a standard test case in mathematical optimization. The Rosenbrock function consists of a curved narrow valley which most algorithms have little difficulty finding. However once found, the difficulty lies in finding the global optimum which is situated inside the valley. Equation (24) is a slightly altered version, as it is multiplied by -1, making it a maximization problem opposed to a minimization problem:

$$J(u_1,u_2) = -100(u_2 - u_1^2)^2 - (1-u_1)^2 \,. \tag{24}$$

The Rosenbrock function has an optimal solution $J=0$ at $(u_1,u_2) = [1\ 1]$. This point lies on a long curved ridge; see Fig. 1 for a contour plot. Since it is an analytical function it is possible to compute the exact gradient. The red dots in Fig. 1 are 50 randomly distributed points in space which will be used to test our methodology. All the numerical experiments will be carried out over the same set of points for different scenarios. Since we are working with approximate gradient techniques the effect of different random sequences also needs to be accounted for. Therefore, all the experiments are repeated for 50 different random sequences. The results presented below are the mean values for the 50 points in space over the 50 different random sequences.

Deterministic Case

We first test the two hypothesis testing methodologies which consist of calculating the angle between the gradient estimated by EnOpt and the exact gradient for different ensemble sizes. The points are uniformly distributed so as to capture the effect of the spatial variability in the objective space on the gradient quality, with many points that are on the ridge or on the edge of the ridge. This will to some degree ensure that once the right ensemble size is determined the EnOpt gradient will be comparable to the true gradient almost everywhere in space. The other important factor in determining a high-quality ensemble gradient is the perturbation size used to generate the ensemble of controls for the gradient estimate. Fig. 2 is the mean angle over the 50 spatial points and 50 random seeds versus an increasing ensemble size which

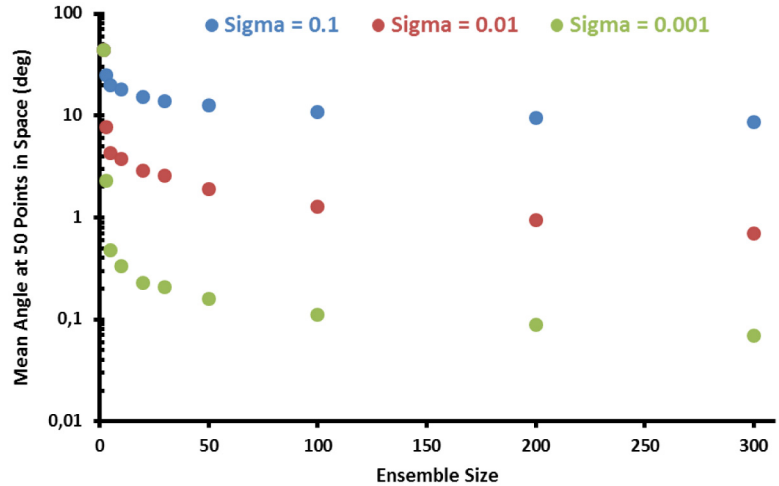

Figure 2—Illustration of the trend in the mean angle over the 50 points in space and 50 random seeds for an increasing ensemble size and varying perturbation sizes (σ). Lower values (green & red dots) of σ give significantly better results compared to higher values (blue dots).

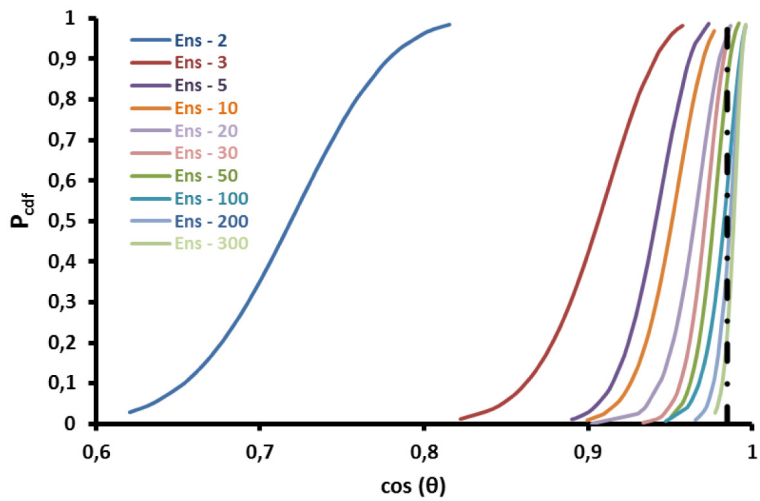

Figure 3—Illustration of the cumulative density functions of the beta distribution for varying ensemble sizes for σ = 0.1.

shows the impact of ensemble size for varying perturbation sizes σ. We observe that, as expected, as σ decreases the gradient quality increases for smaller ensemble sizes while for an increasing ensemble size the gradient quality always improves. The Rosenbrock function is a two-control problem, thus we would need a maximum of three (i.e. $n+1$, where n is the problem dimension) simulations to estimate a 'good' gradient with the finite difference method provided the perturbation size is sufficiently small. Fig. 2 is a reconfirmation that EnOpt, for a small perturbation, also would require an ensemble size equal to 3 to estimate a 'good' gradient. For the largest value of σ used we observe that even for an ensemble size of 300 we do not achieve a 'good' gradient at all the 50 points in space.

One of the methodologies we propose in this paper uses beta distributions to represent the calculated angles to estimate an ensemble size. Fig. 3 and Fig. 4 are the cumulative density function of the beta distributions for different ensemble sizes and varying perturbation sizes. While Fig. 2 represents a single mean value for a particular ensemble size, Fig. 3 and Fig. 4 show the spread in the angles at the different points for varying perturbation and ensemble sizes. We observe in Fig. 3, that for increasing ensemble sizes the curves shift towards an angle equal to zero degrees and the black line is the defined maximum allowable difference between the two gradients i.e. 10 degrees. We can read-off the achieved confidence interval as 1-P_{cdf} for the different ensemble size. We see that even for an ensemble size of 300 (rightmost

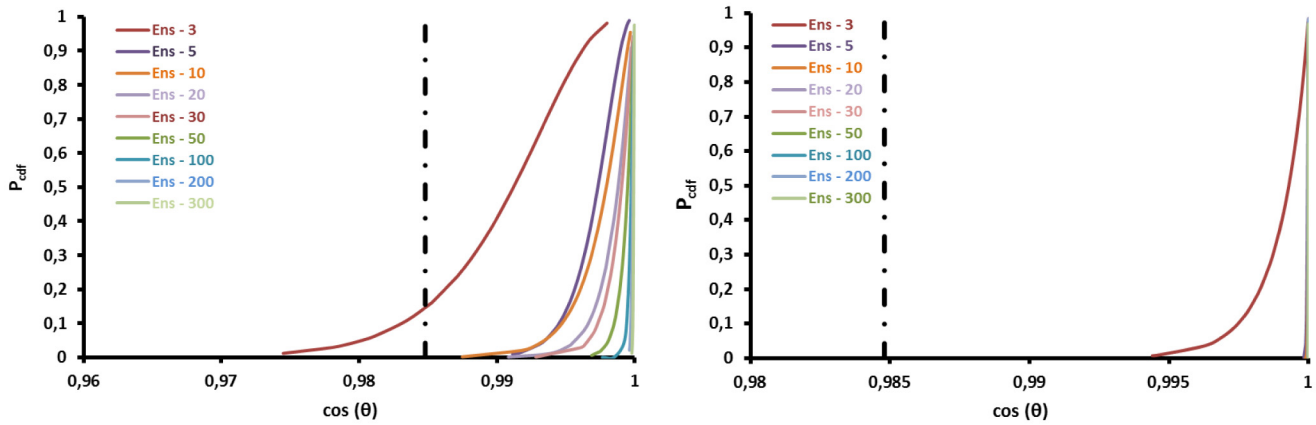

Figure 4—(a) Cummulative density functions for σ = 0.01 where for an ensemble size of 5 or higher the hypothesis is satisfied at all points. (b) Cumulative density functions for σ = 0.001, where for an ensemble size of 3 and higher the hypothesis is satisfied.

light green curve) we achieve a confidence interval of 78% which is much less than the 95% confidence interval which would satisfy the null hypothesis used in this paper. As a comparison, if we use the second methodology, i.e. a traditional hypothesis testing approach, we count that 38 of the 50 points satisfy our null hypothesis. Thus 38/50, i.e. 76% of the points, satisfy our hypothesis, and hence, with an ensemble size of 300 and a perturbation size (σ) equal to 0.1, we have a confidence interval of 76%. The two methodologies give almost the same results which is a cross validation of the beta distribution approach. Note: for an ensemble size equal to 2 (leftmost blue curve in Fig. 3) we cannot achieve good results and this is also the case for the different perturbation sizes.

The same exercise is repeated using the Δ norm instead of the angle as the data. We observe (results not shown) very similar results, i.e. similar confidence intervals, compared to the results shown above with exactly the same trend in the data. Note: this result is not surprising because essentially calculating an angle or the Δ norm with the same data should lead to the same results. To achieve the confidence interval defined in this work we would need approximately 900 samples when using $\sigma = 0.1$. Fig. 4 consists of two plots for different perturbation sizes. Fig. 4a (left-side plot) depicts, similar to the curves shown in Fig. 3, the distribution of the angles calculated for the 50 points using different ensemble sizes with $\sigma = 0.01$. Fig. 4b (right-side plot) shows the distribution of the calculated angles for the same 50 points with $\sigma = 0.001$. We observe that for $\sigma = 0.01$ with an ensemble size of 5 samples (Fig. 4a, purple curve) we achieve a 100% confidence interval while for $\sigma = 0.001$, i.e. one order of magnitude smaller, we achieve a 100% confidence interval with an ensemble size of 3 samples (Fig. 4b, red curve). These results are also cross-validated using the traditional hypothesis testing method as well as using the Δ norm.

The above figures show that the gradient quality is strongly affected by the perturbation size which in turn determines the ensemble size needed to achieve good quality gradients. We observe that, for large perturbation sizes, we need significantly larger ensemble sizes to achieve the 95% confidence interval. For significantly smaller perturbation sizes we would need an ensemble size of 3, i.e. for an n dimensional problem we need an ensemble size equal to $n+1$, just like if we were to estimate a finite difference gradient.

Rosenbrock Function With Uncertainty

The quality of the 'robust' EnOpt gradient has not been investigated before, so uncertainty in the Rosenbrock function is introduced through equation (25). The uncertainty is introduced to mimic geological uncertainty.

$$J(u_1,u_2) = -100(c_1 \cdot u_2 - u_1^2)^2 - \sin(c_2)\cdot(1-u_1)^2. \tag{25}$$

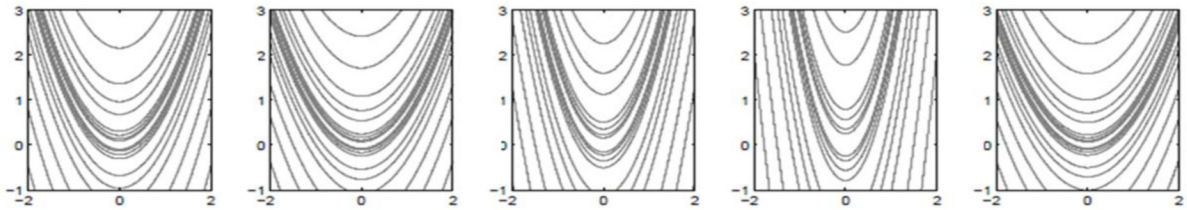

Figure 5—Contour plot of five realizations of the Rosenbrock function given by equation (25) for a case mimicking a low degree of uncertainty.

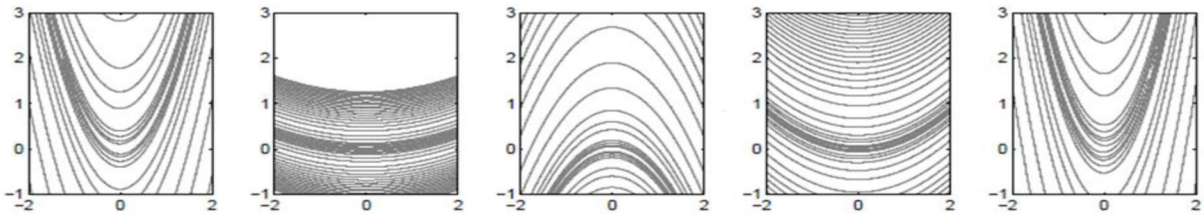

Figure 6—Contour plot of five realizations of the Rosenbrock function given by equation (25) for a case mimicking a higher degree of uncertainty.

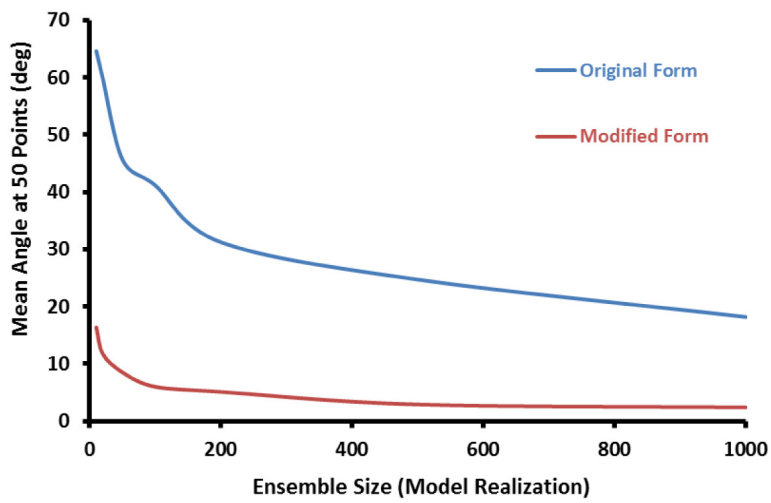

Figure 7—Comparison of the trend in the mean angle over the 50 spatial points with increasing ensemble sizes for the original form (blue) and modified form (red) using the computationally attractive 1:1 ratio.

The two constants c_1 and c_2 are drawn from a Normal distribution, where the standard deviation reflects the magnitude of the uncertainty introduced. This provides us a fast and accurate way to test the various 'robust' EnOpt gradient formulations to better understand these formulations. The value for c_1 mainly affects the steepness of the objective space, while c_2 rotates the space. Fig. 5 and Fig. 6 show five realizations each for cases which are representative of low and high uncertainty scenarios. This was done to investigate the impact of uncertainty on the gradient estimate.

In view of the results from the deterministic case, for the robust case we only use the angles as our testing parameter along with the traditional hypothesis approach to determine the ensemble size required to estimate a good quality 'robust' gradient. Chen (2008) first proposed the idea of the 1:1 ratio explained in the theory section to estimate a 'robust' gradient. The mathematical reasoning for the applicability of this 1:1 ratio provided by Chen (2008) is only applicable if we have very large ensembles for both the models and consequently the controls. In reality, however, we usually work with 100 models and in our experiments the original 1:1 formulation in many instances indeed produces results of insufficient quality. The poor results were reasoned to be the lack of a good quality gradient. Fig. 7 illustrates the effect of using the 1:1 ratio for an increasing ensemble size of model realizations and, consequently, also control

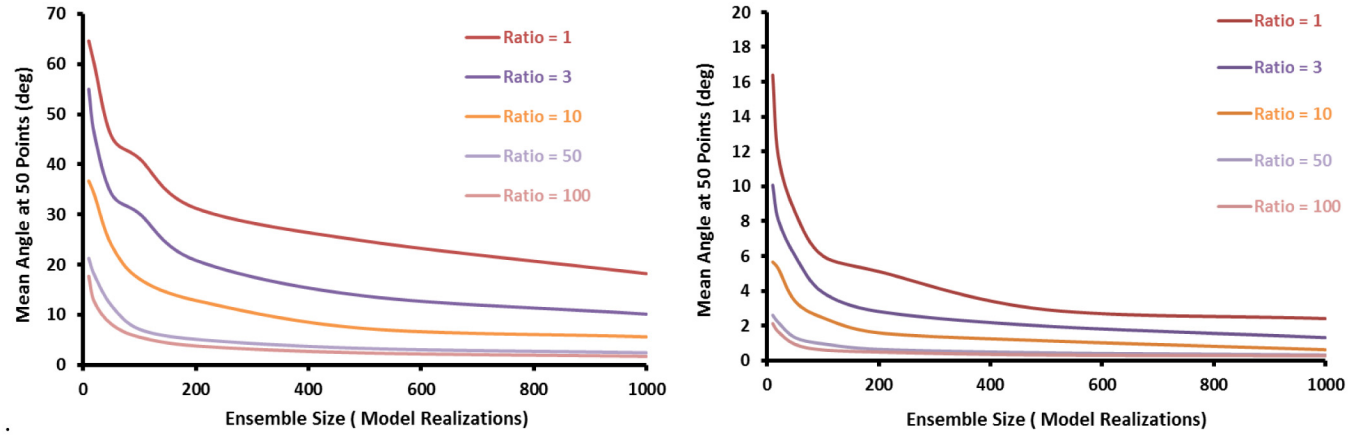

Figure 8—(a). Illustrates the effect of using higher ratios on the gradient quality when using the original formulation with the HPG formulation. (b) Illustrates the effect when using the modified formulation with the HPG formulation. Note the different vertical scales.

realizations. We observe, in line with the theory, that the mean angle significantly reduces for an increasing ensemble size. However, even for an ensemble size of 1000 we do not satisfy our null hypothesis when using the original formulation (blue curve). On the other hand the modified formulation based on equations (19) and (20) (red curve) which still retains the computational attractiveness of the original formulation performs significantly better. We observe that for an ensemble size of 1000 we have a mean angle of approximately 4 degrees and we have satisfied out null hypothesis and achieved the 95% confidence interval.

These results are based on the case representing the highest uncertainty in the models. In our reservoir optimization problem we usually use an ensemble size of 100 models to capture the uncertainty. For that size, the original formulation achieves a mean angle of 40 degrees while the modified formulation achieves a mean angle of 7 degrees. We recommend therefore that, when using the 1:1 ratio for robust optimization, the modified formulation based on equations (19) and (20) be applied.

Effect of Higher Ratios

The results in Fig. 7 were for the case representative of the highest model uncertainty using the 1:1 ratio. Some previous studies, e.g. Raniolo et al. (2013) and Li et al. (2013), did not achieve results of practical value with the 1:1 ratio and suggested the use of higher ratios (1:20 etc.) to find better gradient estimates. We investigate here the impact of varying the ratio between model and control perturbations on the quality of the gradient estimate. Note: the results are obtained with the HPG formulation using a perturbation size equal to 0.01. Fig. 8 consists of two plots, Fig. 8(a) which is the mean angle plotted against the ensemble size of model realization with the curves representing the ratio used for the gradient estimate with the original formulation, and Fig. 8(b) which displays the results using the modified formulation (note the difference in the vertical scale). In both plots we observe that an increased ratio gives better gradient estimates for the different ensemble sizes of model realizations. Once again, the modified formulation outperforms the original formulation also for larger ratios. Thus in general better gradient estimates can be achieved while using higher ratios. Note though that while it is better to use higher ratios, the computational costs of using these ratios must also be accounted for, especially for large-scale high-dimensional problems.

Effect of Uncertainty

The effect of uncertainty is investigated in conjunction with the ratios needed to satisfy the null hypothesis and achieve the desired confidence interval. Fig. 9 is an illustration of the impact that uncertainty has on the quality of the gradient estimate. Fig. 9(a) (left-side plot) depicts the results for the highest uncertainty case while Fig. 9(b) (right-side plot) depicts results for the lowest uncertainty scenario for both the original

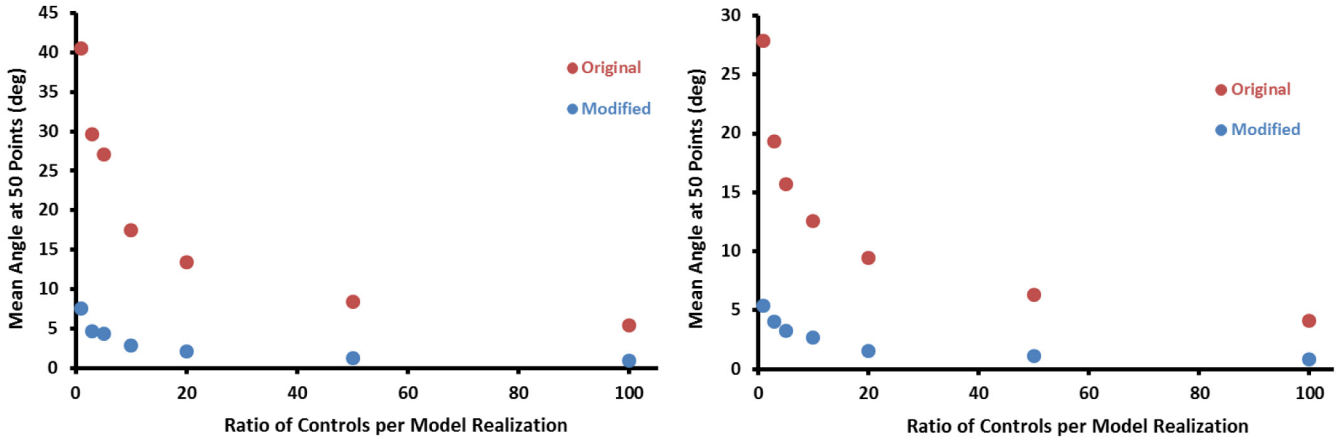

Figure 9—(a) shows the effect of a high degree of uncertainty on the mean angle for increasing ratios with the HPG formulation. (b) Illustrates the results for the lowest uncertainty case. Note: for both these plots an ensemble of 100 model realizations is used. The angles for the high uncertainty case are higher than the low uncertainty case. (Note the different vertical scales).

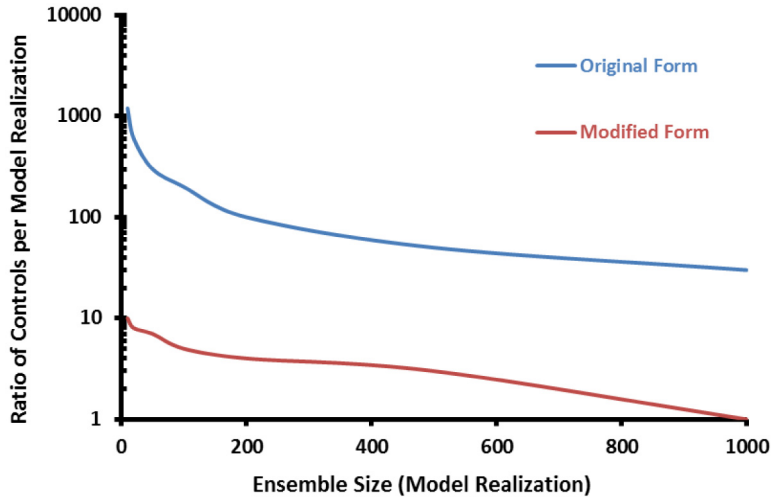

Figure 10—Hypothesis testing results using the traditional methodology for the original (blue) and modified (red) form using the highest uncertainty case with the HPG formulation. A lower ratio is needed for higher ensemble sizes of model realizations to satisfy the hypothesis and achieve the desired confidence interval.

and modified formulations. In both plots we observe, in accordance with all the results thus far, that the modified formulation results in significantly better gradients than the original formulation. Also the modified formulation gradient estimate is less sensitive to the effect of uncertainty. The ensemble size of model realizations was kept constant at 100 for this exercise and a perturbation size of 0.01 was used. We also noticed (results not shown) that for the original formulation using a larger perturbation size resulted in better mean angles, whereas for the modified formulation the gradient estimate improved for smaller perturbation sizes. We observe that, depending on the degree of uncertainty, the original formulation would need a ratio of 1:20 or 1:50 i.e. 2000 or 5000 function evaluations to find a mean angle less than 10 degrees. Note: if the mean angle is less than 10 degrees it does not guarantee that we have achieved the necessary confidence interval.

Hypothesis Test Results

Fig. 10 illustrates the ratio necessary to satisfy the null hypothesis used in this paper and achieve the desired 95% confidence interval for the different gradient formulations and the different ensemble sizes of model realizations. We observe that for a smaller ensemble size of model realizations, e.g. 10, we need significantly higher ratios (1:1100) when using the original formulation while for the same ensemble size,

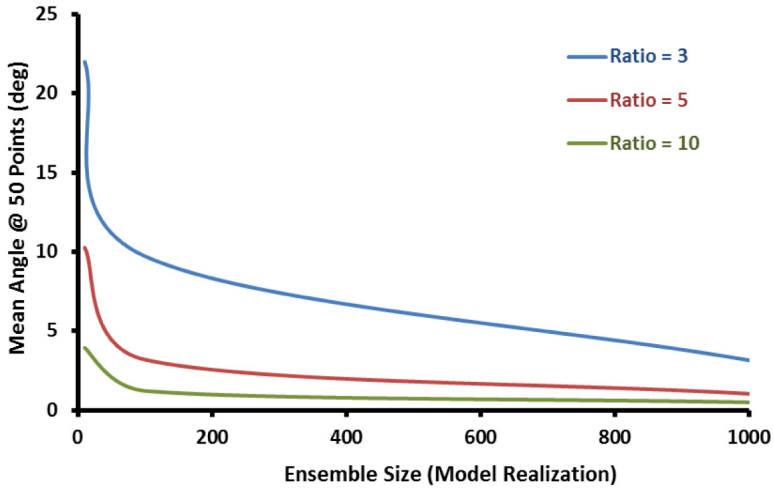

Figure 11—Illustration of the mean angle for increasing ensemble sizes and ratios. We observe a significant reduction in the ratio required even for small ensemble sizes compared to Fig (10) for the highest uncertainty case.

we would require a ratio of 1:10 with the modified formulation. We also observe that the required ratio decreases with an increase in the ensemble size of model realizations. All these results are for the highest uncertainty scenario. For the lowest uncertainty scenario we would require lower ratios: 1:800 with the original formulation for an ensemble size of 10 model realizations, and a 1:5 ratio for the modified formulation.

Mean of Individual Gradients (MIG)

The results presented in Figs. 8-10, which are obtained using the HPG formulation, illustrate that for small ensemble sizes we would require ratios as high as 1:1100 to satisfy the hypothesis and achieve the desired confidence interval. For a two-dimensional control problem these results are completely counter-intuitive. Thus we test the MIG formulation to estimate a 'robust' gradient when using a ratio other than 1:1. Note: This approach was followed by Van Essen et al. (2009) albeit using the adjoint formulation and also by Raniolo et al. (2013) using EnOpt gradients. Since we are dealing with a two dimensional problem we should need 3 function evaluations to estimate a gradient, for a sufficiently small perturbation size, as illustrated in Fig. 2. Using the MIG formulation we observe in Fig. 11 that for a perturbation size equal to 0.01 we would at best need a ratio of 1:5 to estimate a mean angle less than 10 degrees. Increasing the ratio will improve the gradient estimate, and the difference between the original and modified formulation is minor at best. These results are a marked improvement in terms of the ratio needed compared to the results with the HPG formulation. The results are based on the highest uncertainty case while the same trend is observed for the low uncertainty case.

The effect of perturbation size estimates from the MIG formulation is shown in Fig. 12. We observe, 'akin' to the deterministic case, that a decreasing perturbation size results in an increase in the quality of the gradient estimated. For a sufficiently small perturbation size ($\sigma = 0.001$) a ratio of 1:3 is sufficient to satisfy our null hypothesis and achieve the 95% confidence interval even for small ensemble sizes of model realizations. We observe that for an increasing perturbation size the ratio increases, although, it is significantly lower than the ratio required when using the HPG formulation.

In summary, for this case using the MIG formulation gives much better angles for smaller ratios especially when working with small ensembles of model realizations.

3D Reservoir Model: "Egg Model"

The 'Egg Model', first introduced by van Essen (2009), is a channelized reservoir with seven vertical layers. Fig. 13 is an illustration of the permeability field of a single model realization with the locations

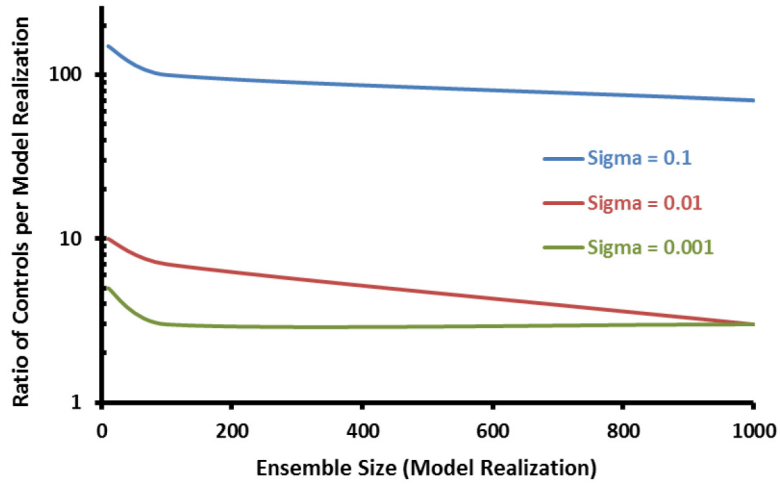

Figure 12—Hypothesis testing results using the traditional methodology for different perturbation sizes (σ) and ratios. Lower values would require significantly smaller ratio compared to the HPG formulation.

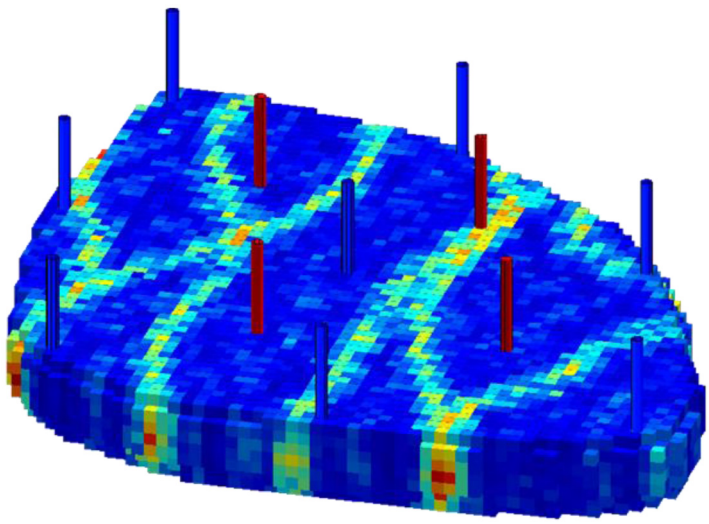

Figure 13—Reservoir model displaying the position of the injectors (blue) and producers (red)

of the eight (mainly peripheral) injection wells (blue) and four production wells (red) completed in all the layers. The model represents a channelized depositional system in the form of discrete permeability fields modeled with $60 \times 60 \times 7 = 25.200$ grid cells of which 18.553 cells are active. A detailed description of a standardized version of this Egg Model is given in Jansen et al. (2013). The reservoir and fluid properties used for all ensemble members are given in Table 1. No capillary pressures are included and the reservoir rock is assumed to be incompressible. The bottom hole pressures of the producers are constrained between 385 and 400 bar, while the injectors are rate-controlled between 1 and 79.5 m^3/day. The initial reservoir pressure is at 400 bars. Production of the field is simulated for a period of 3600 days or slightly less than 10 years. There are 40 control time steps of 90 days, thus using injection rates as controls we have of 40x8 = 320 controls, i.e. a 320 dimensional problem.

We use a commercial fully implicit black oil simulator (Eclipse, 2011) for the reservoir simulations to estimate the (robust) EnOpt gradient of J with respect to the controls $\mathbf{u}$. An in-house simulator is used to calculate the adjoint gradient, see Kraaijevanger et al. (2007) for details. Note that the model has been benchmarked for the different simulators, details of which can be found in Jansen et al. (2013).

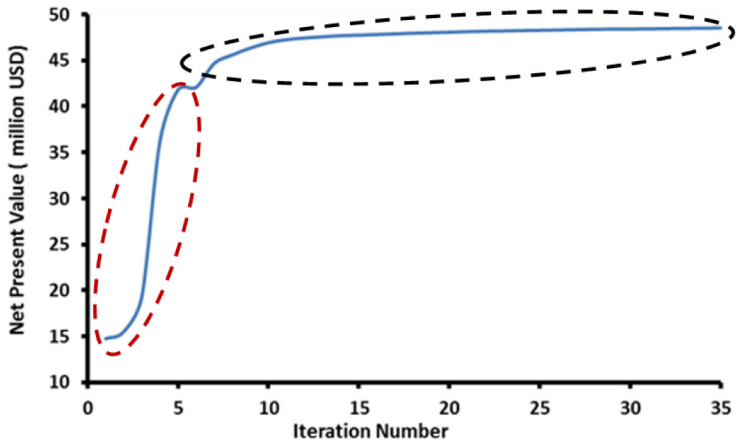

Figure 14—Illustration of the optimization process (blue curve) using adjoint gradients. The curve is divided into two parts, a red ellipse (indicative of the steeper region of the curve) and a black ellipse (indicative of the flatter region of the curve)

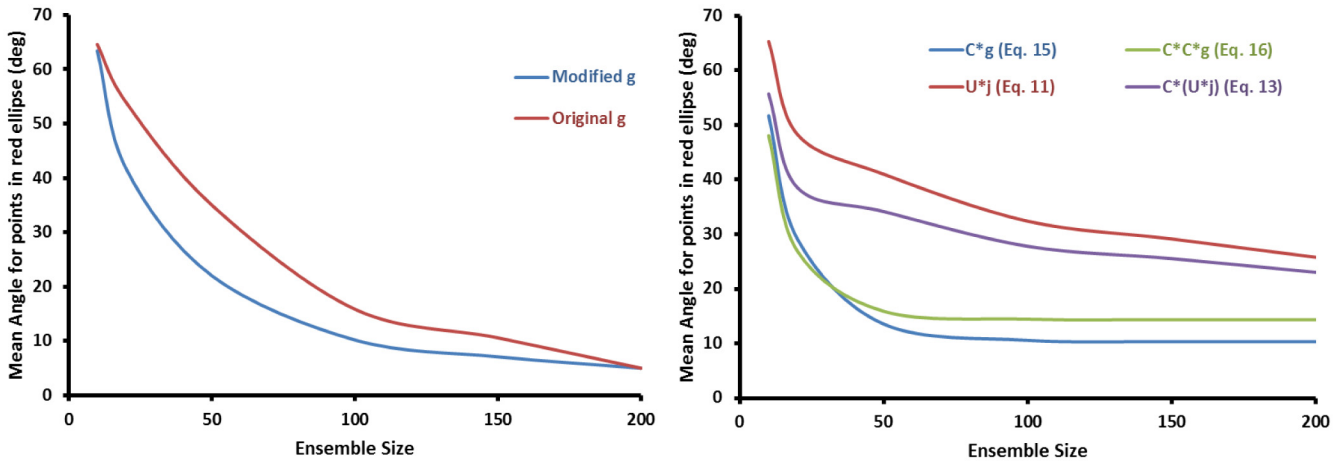

Figure 15—(a) illustrates the performance of the two different formulations used to estimate the 'straight gradient' g given by equation (14), while (b) displays the effect of using different versions of the 'smoothed' gradient and their relative behavior.

Deterministic case

The methodology presented above is first tested on a single realization of the Egg Model and thus we are dealing with a deterministic case. When working with a 320-control problem it is, of course, impossible to visualize the objective function space. Randomly creating points to evaluate our methodology is also risky since we will not be sure that our points have covered all features of the objective function space, due to the 'emptiness' of a 320-dimensional space. Thus, in order to obtain 'relevant' points which capture the nature of the objective function space we perform an adjoint- based optimization from two different initial strategies. Fig. 14 shows the optimization process for 35 iterations for an initial strategy of constant rates equal to 79.5 m^3/day. Thus we have 35 points to test the methodology, i.e., instead of using a large number of randomly distributed control points, as was done for the Rosenbrock function, we now test the gradient quality for 35 poins along a pre-defined control trajectory. Fig. 14 also shows two dashed regions, a steep region, indicated with a red ellipse, in which there exists significant scope for optimization, and a flatter region, indicated with the black ellipse, which is indicative of a peak, ridge or plateau in the objective function space. While using the same hypothesis, we investigate the two regions separately, because different parameters are required to achieve the necessary confidence intervals. Note: the apparent decrease in the objective function value at iteration number 7 is an artifact of the plotting algorithm.

The Steep Region Fig. 15 displays the results for the points encapsulated by the red ellipse in Fig. 14.

It depicts the decrease in the mean angle for an increasing ensemble size with a constant perturbation size, $\sigma = 0.01$. The results are the mean of 5 experiments with different random seeds. The left subplot of Fig. 15 illustrates the mean angle of the different points in the red ellipse. We observe that we need an ensemble size of approximately 150 to satisfy the null hypothesis. We also observe that increasing the ensemble size leads to higher-quality gradients. Thus for a 320-dimensional problem with approximately 150 samples we can estimate a high-quality gradient. We also tested (results not shown) that for increasing perturbation sizes the quality of the gradient estimate decreases, while for smaller perturbation sizes we also obtain inferior gradient estimates, most likely due to numerical round-off errors. E.g., with $\sigma = 0.1$ (i.e. 10 times larger), and σ=0.0001, (i.e. 100 times smaller), we would need 400 and 300 samples respectively to satisfy the null hypothesis and achieve the desired confidence interval. To obtain these results we have used a time-correlated covariance matrix with a correlation length equal to 20 which was arbitrarily chosen. The results presented in Fig. 15(a) are obtained with equation (14) and the modified formulation suggested by Do and Reynolds (2013) [equations (17) and (18)]. This performs better than using equation (14) with the original formulation [equations (5) and (6)]. Irrespective of the gradient formulation used, an increasing ensemble size leads to higher-quality gradients and the modified and original formulation gradients converge for larger ensemble sizes. This is trend is also observed for different perturbation sizes and correlation lengths (results not shown), although the ensemble size needed to achieve the desired confidence intervals will then be different. Fig. 15(b) shows the effect of smoothing on the different gradient estimates. We observe that smoothing of the gradient, i.e. pre-multiplication of the gradient by the covariance matrix, has a marginally negative impact on the gradient quality for this set of points and a $\sigma = 0.01$. We also observe that single smoothing of the straight gradient obtained from equation (14), i.e. equation (15), performs better than using the smoothed gradient given by equation (11). We also note that a double smoothing of the straight gradient, i.e. equation (16), gives better results for smaller ensemble sizes and inferior results for larger ensemble sizes. Note: pre-multiplication of the gradient given by the cross-covariance vector, i.e. equation (13), achieves better results than those obtained with equation (11).

The Flatter Region The general trend in the results observed for the steeper region is also observed for the flatter region. However, estimating the straight gradient which gives good estimates for the steeper region does not achieve good gradients in the flatter part. Smoothing of the gradients, on the other hand, achieves much-higher-quality gradients in this flatter region. Note, in both regions the modified formulation of Do and Reynolds (2013) performs best. A major difference in gradient estimation for the flatter region is the sensitivity of the gradient estimate quality to the choice of the correlation length used. Although we do not achieve the 95% confidence interval for this region we observe that we need smaller perturbation sizes, $\sigma = 0.001$, and correlation lengths varying from 8 to 12 control time steps in conjunction with an ensemble size equal to 300 to satisfy the hypothesis at most of the points. It is virtually impossible to know a-priori the correct correlation length, and the lengths used here were obtained through an exhaustive trial and error approach. Similar to the results for the steeper region, we observe that a smoothed gradient given by equation (15) achieves a better angle compared to directly estimating a smoothed gradient using equation (11). Fig. 16 shows the quality of the different smoothed gradient formulations. When using equation (11), then a further single smoothing, i.e. equation (13), which is akin to a double smoothing of equation (14), i.e. equation (16), achieves better quality gradients for all ensemble sizes. Note: irrespective of the different gradient estimates used, the gradient quality improves with an increase in ensemble size. We also observe that for this region with ensemble sizes higher than 300 we do not observe a significant improvement in gradient quality. Also note that, irrespective of ensemble size, the angles are never greater than 90 degrees, i.e. within the first quadrant with respect to the adjoint gradient. This is particularly important, because it implies that even for smaller ensemble sizes we are able to approximate, in general, a roughly ‘correct’ up-hill direction.

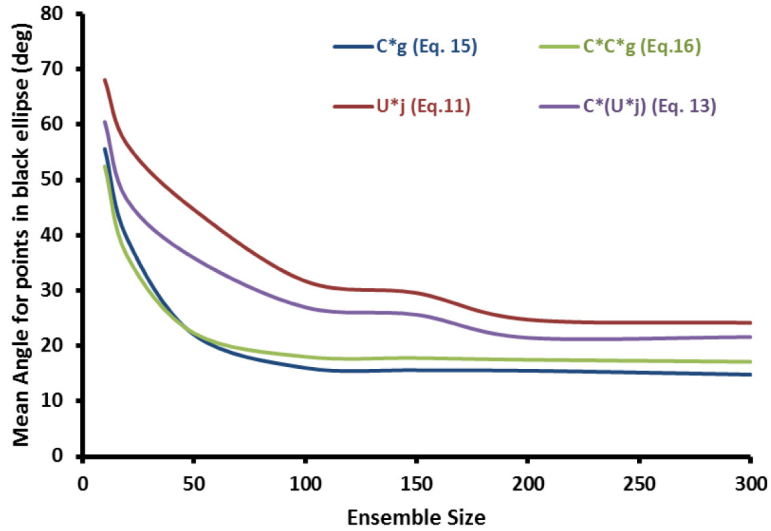

Figure 16—Comparison of the effect of smoothing and different gradient estimates for the flatter region (black ellipse) of Fig. (14).

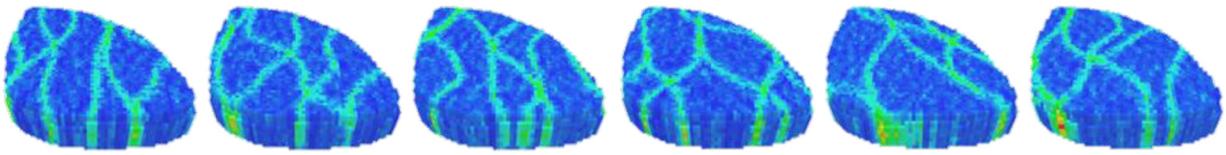

Figure 17—Six randomly chosen realizations displaying the uncertainty in the geological models; see Jansen et al. (2013) for details.

Robust case

The main purpose of this study is to quantify the quality of the 'robust' ensemble gradient and determine the optimal ratio required to estimate a high-quality gradient for a realistic reservoir test case. Thus, to test our methodology we use an ensemble of 100 equi-probable geological models, six of which are displayed in Fig. 17. The uncertainty is captured through the different permeability fields and the different directions and orientations of the channels, see Jansen et al. (2013) for details. The hypothesis tests are performed to estimate a 'robust' ensemble gradient using all hundred realizations of this model. The controls and the fluid model as well as the other properties are exactly the same as for the deterministic case. To generate the points to test our methodology, the same approach is used as described above. That is, we first perform robust optimization using the adjoint method as described in Van Essen et al. (2009), and then assess the quality of the gradient along this control trajectory. The result of the robust optimization is displayed in Fig. 18. Note that the optimization was limited to 25 iterations due to the computational complexity involved. Again, for the analysis, we divide the optimization process into two regions using the same reasoning as explained above.

The Steep Region Following the analysis provided for the deterministic case we first consider the steeper part of the optimization curve encapsulated by the red ellipse as shown in Fig. 18. We have five points (control strategies) which lie in this region for which the original and modified formulation using the computationally attractive 1:1 ratio have been compared. We observe that for all the five points we are able to satisfy our null hypothesis, i.e. the angle is less than 10 degrees, when using the modified robust formulation [based on equations (19) and (20)]. However if we use the 1:1 ratio with the original formulation [based equations (5) and (6)] we never satisfy our hypothesis at any of the points. With the original formulation we observe angles between 80 and 90 degrees for all the points when using equation (14). (As for the deterministic case, all results are the mean of 5 experiments with different random seeds.) If we use the'smoothed' versions of the gradient as given by equations (11, 13, 15 and 16), we observe

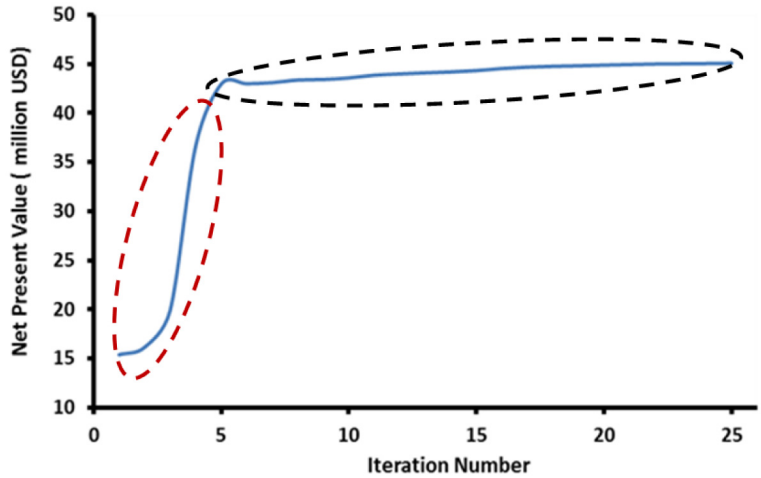

Figure 18—Illustration of the expected objective function value over 100 realizations with the robust optimization process (blue curve) using adjoint gradients. The curve is divided into two parts, a red ellipse (indicative of the steeper region of the curve) and a black ellipse (indicative of the flatter region of the curve).

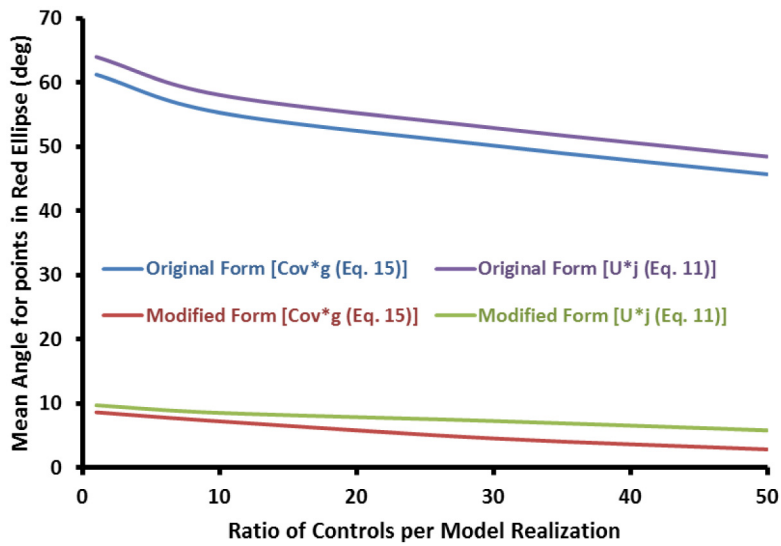

Figure 19—Illustrates the difference in gradient quality between the original and modified formulations for varying ratios of controls per model realization using the HPG ratio formulation.

much better angles as shown in Fig. 19, though the hypothesis is still not satisfied. Thus smoothing of the gradient leads to better estimates of the ensemble gradient, and we observe a similar trend as for the deterministic case in that the smoothed version of equation (14), i.e. equation (15), achieves better results than equation (11). We also observe that using a relatively larger perturbation size ($\sigma = 1$) for this region leads to the best results. These results highlight the advantage of using the modified formulation with the 1:1 ratio because the maximum contribution (largest increase in objective function) is from within this region so achieving good angles at acceptable computational costs is more important for this region. Note : when using non-smoothed gradients, i.e. equation (14), the gradient quality is inferior (results not shown) compared to using smoothed gradients. Note: while the 1:1 ratio satisfies our hypothesis, using higher ratios leads to higher-quality gradients for both gradient formulations with the HPG formulation as depicted in Fig. (19).

For the Rosenbrock function we observed significant differences in gradient quality when using the HPG and MIG ratio formulations. For this case we observe a similar trend in the results. Fig. 20 consists of two plots which show the difference in gradient quality for the two ratio formulations. Fig. 20(a) depicts

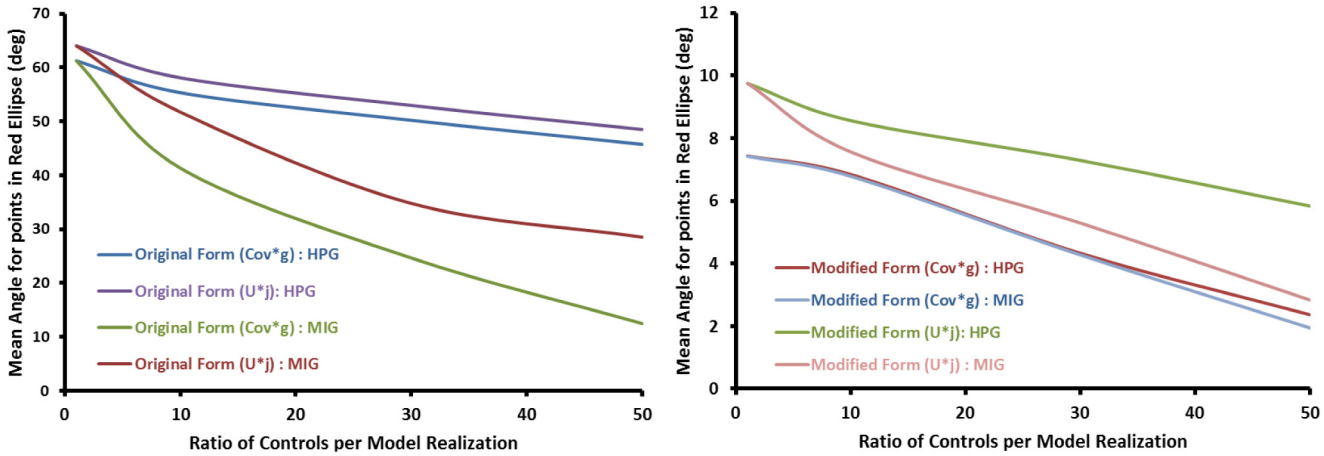

Figure 20—(a) Difference in gradient quality for the two ratio formulations (HPG and MIG) for the original gradient formulation while (b) highlights the gradient quality when using the modified gradient formulation. Note the different vertical scales

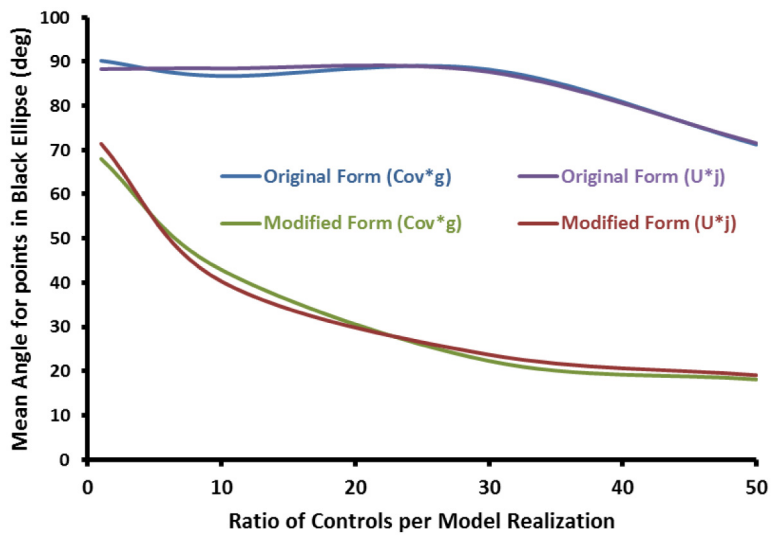

Figure 21—Illustrates the difference in gradient quality between the original and modified formulations for varying ratios of controls per model realization using the HPG formulation.

the results obtained when using the original formulation, where, similar to the Rosenbrock results, we always achieve significantly better gradient estimates with the MIG formulation. We also observe, as expected, that as the ratio increases the gradient quality improves. Fig. 20(b) depicts the results when using the modified formulation. The results indicate that the modified formulation is less sensitive to the different ratio formulations, although the MIG formulation still performs better. Irrespective of the ratio formulation the modified form, satisfies the hypothesis at all the points, even for the 1:1 ratio.

The Flatter Region The 1:1 ratio with the modified formulation [based on equations (19) and (20)] leads to very high quality gradient estimates for the steeper part of the optimization curve. However, for the flatter part, though still better than the original formulation [based equations (5) and (6)], the gradient quality is not as high. An increase in the ratios significantly improves the gradient quality as shown in Fig. 21. We also observe, slightly different to our previous observations that the different versions of the 'smoothed' gradient, given by equations (11, 13, 15 and 16), give very similar results for both the modified and original formulations. Note: similar to the deterministic case we use a smaller perturbation size ($\sigma = 0.01$) and a correlation length that varies between 8 and 17. It requires a substantial number of

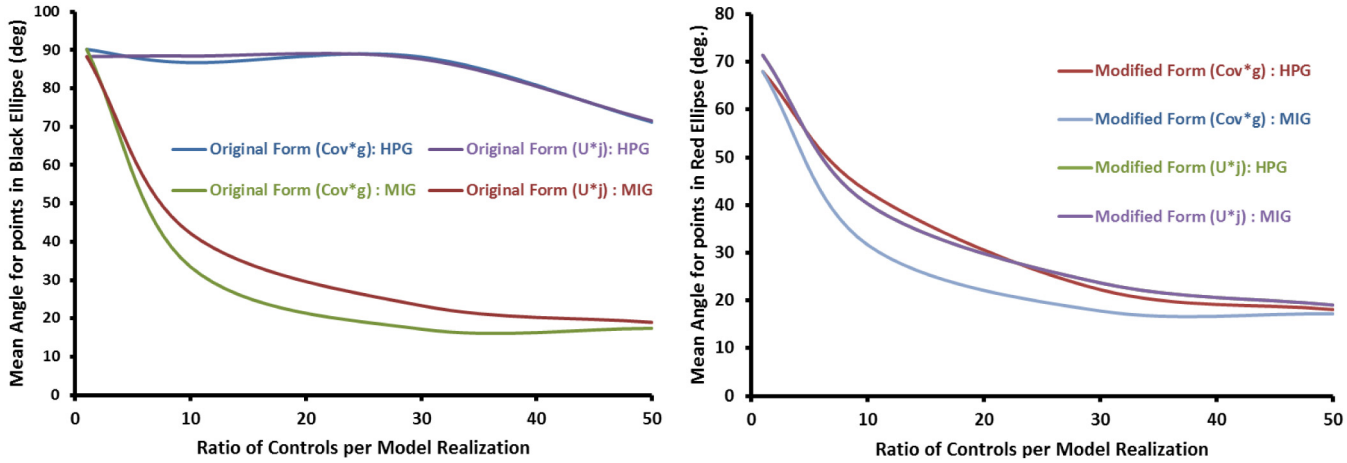

Figure 22—(a) shows the difference in gradient quality for the two ratio formulations (HPG and MIG) for the original gradient formulation while (b) highlights the results when using the modified gradient formulation.

simulations to perform an extensive search for the 'correct' correlation length which has a strong impact on the gradient quality.

Similar to the results shown above for the steeper region, the different ratio formulations show a similar trend for the points in flatter region. Fig. 22 consists of two plots which show the difference in gradient quality for the different ratio formulations. Fig. 22 (a) depicts the results obtained when using the original formulation and 22 (b) depicts the results when using the modified formulation. We observe that the same trend in the results exists, although on the flatter part of the curve we have satisfied our hypothesis for only a few points for this robust case. We could argue that for the robust case and, considering our problem dimension, the difference between 10 and 20 degrees is marginal. Thus if we were to have a hypothesis which had an allowable error margin of 20 degrees then, with the MIG ratio formulation and both the original or modified gradient formulation, we would satisfy our hypothesis at most of the points for a ratio of 1:50, i.e. we would need to apply 50 control samples per geological realization to estimate a 'good' gradient. This would imply that, to estimate a gradient at one point within the optimization process we would need to perform 5000 reservoir simulations since the ensemble size is 100. This is computationally challenging and therefore, since the increase in objective function value for the flatter region is less than 1% we suggest the use of the modified formulation with a 1:1 ratio.

Conclusions

- The two test parameters i.e. the angle θ and the Δ norm, give very similar results. For the relatively simple Rosenbrock function we need an ensemble size equal to 900, 5 or 3 to satisfy the null hypothesis and achieve the desired confidence interval depending on the perturbation size used ($\sigma = 0.1$, $\sigma = 0.01$ and $\sigma = 0.001$ respectively) to generate the ensemble of controls.
- Including uncertainty within the Rosenbrock function we find that the modified formulation [based on equations (19) and (20)] with the computationally attractive 1:1 ratio outperforms the original formulation [based equations (5) and (6)], irrespective of the degree of uncertainty. However the degree of uncertainty does play a role in the quality of the gradient estimate.
- The modified formulation with the 1:1 ratio is less sensitive to the degree of uncertainty and the ensemble size of model realizations compared to the original

formulation. Higher-quality gradients can be obtained with an increasing ensemble size of model realizations.

- In general, irrespective of the gradient formulation, using higher ratios (i.e. 1:10, 20, 50 etc.,) always leads to higher-quality gradient estimates.
- When working with higher ratios significantly better results are obtained through the use of the 'Mean of Individual Gradients' (MIG) formulation. Additionally the ratios required are smaller; for the Rosenbrock case, depending on the perturbation size, a ratio of 1:3 is sufficient to satisfy the hypothesis and meet the desired confidence interval.
- The deterministic case using the "Egg Model" showed a similar trend in the results as observed with the Rosenbrock function. The covariance matrix, i.e. the perturbation size, has a significant impact on the gradient quality.
- The analysis for the "Egg Model" case was divided into two regions: 1) a 'steeper' region, where with higher perturbation size ($\sigma = 0.01$) and lower ensemble sizes (150 samples) the hypothesis was satisfied at all points, and 2) a 'flatter' region where a smaller perturbation size ($\sigma = 0.001$) and a higher ensemble size (300 samples) were needed.
- In the flatter region, in addition to the perturbation size, a correlated covariance is needed where the choice of correlation length has a significant impact on the gradient quality. Through a trial and error procedure we observe that correlation lengths between 8 and 13 give the best results and lead to satisfying the hypothesis at most points.
- Using the modified formulation suggested by Do & Reynoldss (2013) [based on equations (17) and (18)] leads to better results at smaller ensemble sizes and leads to the same gradient as expected at higher ensemble sizes.
- Explicitly estimating a straight gradient given in equation (14) followed by a single pre-multiplication for smoothing, i.e. equation (15), achieves better results than using the cross-covariance vector, i.e. equation (11) as the smoothed gradient.
- For the robust "Egg Model" case, we observe a similar trend in the results as for the uncertain Rosenbrock function. The modified formulation using the 1:1 ratio achieves significantly better results compared to the original formulation at any point along the optimization curve
- In the flatter part of the optimization curve the modified formulation with the 1:1 ratio, although performing better than the original formulation, never satisfies the hypothesis. However, in the steeper part, the hypothesis is satisfied for all the points. Using relatively larger perturbation sizes in the steeper part of the curve gives higher-quality gradient estimates.
- The use of a ratio other than 1:1 is thus necessary to achieve good gradients along the flatter region of the curve. Like for the deterministic case the perturbation sizes also needs to be reduced and using the MIG formulation always performs better than the HPG formulation

- As the results have shown, the developed methodology can, in theory, be used to quantify the ensemble size required to achieve a high-quality gradient. However, all the angles obtained in this work are always less than 90 degrees which suggests that with ensemble methods, irrespective of the ensemble size (in our case), we estimate the 'correct' uphill direction.
- We recommend to use, out of the 30 potential robust gradient formulations identified in our paper, the single 'smoothed' modified formulation, i.e. equation (15) based on equations (19) and (20), using the 1:1 ratio for recovery optimization under uncertainty.

Nomenclature

b	= discount rate
$\mathbf{c}_{uJ}$	= ensemble cross-covariance vector
$\tilde{\mathbf{C}}$	= distribution covariance matrix
$\mathbf{C}_{\mathrm{uu}}$	= ensemble covariance matrix
g	= gradient vector
g'	= single-smoothed gradient vector
g''	= double smoothed gradient vector with ensemble covariance matrix $\mathbf{C}_{\mathrm{uu}}$
g'''	= double smoothed gradient vector with distribution covariance matrix $\tilde{\mathbf{C}}$
$\mathrm{g}''\,''$	= single smoothed gradient vector, using $\tilde{\mathbf{C}}$ and a straight gradient vector
$\mathrm{g}''\,'''$	= double smoothed gradient vector, using $\tilde{\mathbf{C}}$ and a straight gradient vector
l	= iteration counter
$\mathbf{j}$	= vector of mean-shifted objective function values
J	= objective function value
$\bar{J}$	= mean objective function value
k	= time step counter
K	= total number of time steps
M	= number of ensemble members
N	= number of control variables
q	= flow rates
r	= price per unit volume
t	= time
u	= control variable
$\mathbf{u}$	= vector of control variables
$\bar{\mathbf{u}}$	= ensemble mean
$\mathbf{U}$	= matrix of ensemble mean-shifted control vectors
σ	= perturbation size
τ	= reference time for discounting

Subscripts

o	= oil
w	= water
wp	= produced water
wi	= injected water

Acknowledgements

This research was carried out within the context of the Integrated Systems Approach to Petroleum Production (ISAPP) knowledge centre and the Recovery Factory (RF) program. ISAPP is a joint project of TNO, Delft University of Technology (TU Delft), ENI, Statoil and Petrobras. RF is a joint project of Shell Global Solutions International and TU Delft. The authors would like to acknowlegde Menno Bloemsma for the useful discussions about hypothesis testing. The authors acknowledge Schlumberger for providing multiple academic Eclipse licenses for this work and Shell for providing a license for their in-house reservoir simulator.

References

AbouRizk, S., Halpin, D., and Wilson, J. 1994. Fitting Beta Distributions Based on Sample Data. *Journal of Construction Engineering and Management* **120** (2) 288–305.

Brouwer, R. and Jansen, J.D. 2004. Dynamic Optimization of Water Flooding With Smart Wells Using Optimal Control Theory. *SPEJ* **9**(4) 391–402.

Chen, C., Li, G. and Reynolds, A. 2012. Robust Constrained Optimization of Short- and Long-Term Net Present Value for Closed-Loop Reservoir Management. *SPEJ* **17**(3) 849–864.

Chen, Y. 2008. *Efficient Ensemble based Reservoir Management*. PhD Thesis, University of Oklahoma, USA.

Chen, Y. and Oliver, D.S. 2010. Ensemble-Based Closed-Loop Optimization Applied to Brugge Field. *SPEREE* **13** (1) 56–71. DOI: 10.2118/118926-PA

Chen, Y., Oliver, D.S. and Zhang, D. 2009. Efficient Ensemble-Based Closed-Loop Production Optimization. *SPEJ* **14** (4) 634–645. DOI: 10.2118/112873-PA

Conn, A.R., Scheinberg, S., Vicente, L.N. 2009: *Introduction to Derivative-Free Optimization*. SIAM, Philadelphia.

Dekking, F., Kraaikamp, C., Lopuhaä, H., and Meester, L. 2005. *A Modern Introduction tp Probability and Statictics: Understanding Why and How*. Springer London, DOI: 10.1007/1-84628-168-7

Do, S.T. and Reynolds, A.C. 2013. Theoretical connections between optimization algorithms based on an approximate gradient. *Computational Geosciences* **17** (6) 959–973. DOI: 10.1007/s10596-013-9386-9.

Eclipse. 2011. http://www.slb.com/services/software/reseng/eclipse.

Fonseca, R.M., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2013. Improving the Ensemble Optimization Method through Covariance Matrix Adaptation (CMA-EnOpt). Paper SPE 163657 presented at SPE Reservoir Simulation Symposium, The Woodlands, Texas, U.S.A, 18-20 February.

Fonseca, R.M., Stordal, A.S., Leeuwenburgh, O. Van den Hof, P.M.J. and Jansen, J.D. 2014. Robust ensemble-based multi-objective optimization. Proc. 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV), Catania, Italy, 8-11 September.

Jansen, J.D. 2011. Adjoint-based Optimization of Multi-Phase Flow Through Porous Media–A review. *Computers & Fluids* **46** (1) 40–51. DOI: 10.1016/j.compfluid.2010.09.039.

Jansen, J.D., Fonseca, R.M., Kahrobaei, S.S., Siraj, M.M., Van Essen, G.M. and Van den Hof, P.M.J. 2013. *The "Egg Model". Research note and data set*. Delft University of Technology, The Netherlands. http://repository.tudelft.nl/view/ir/uuid:1b85ee17-3e58-4fa4-be79-8328945a4491/(report);

Kraaijevanger, J. F. B. M., Egberts, P. J. P., Valstar, J. R. and Buurman, H. W. 2007. Optimal Waterflood Design Using the Adjoint Method. Paper SPE 105764 presented at the SPE Reservoir Simulation Symposium, Houston, USA, 26-28 February. DOI: 10.2118/105764-ms.

Li, L., Jafarpour, B. and Mohammad-Khaninezhad, M.R. 2012. A simultaneous perturbation stochastic approximation algorithm for coupled well placement and control optimization under geologic uncertainty. *Computational Geosciences* **17**(1) 167–188.

Lorentzen, R.J., Berg, A., Naevdal, G. and Vefring, E.H. 2006. A New Approach for Dynamic Optimization of Waterflooding Problems. Paper SPE-99690-MS presented at the Intelligent Energy Conference and Exhibition, Amsterdam, The Netherlands, 11-13 April. DOI: 10.2118/99690-MS.

Nocedal, J. and Wright, S.J., 2006: *Numerical optimization,* 2nd ed., Springer, New York

Nwaozo, J. 2006. *Dynamic optimization of a wateflooding reservoir*. MSc Thesis, University of Oklahoma, USA.

Oliver, D.S., Reynolds, A.C. and Liu, N. 2008. *Inverse theory for petroleum reservoir characterization and history matching,* Cambridge University Press, Cambridge.

Raniolo, S., Dovera, L., Cominelli, A., Callegaro, C. and Masserano, F. 2013. History matching and polymer injection optimization in a mature field using the ensemble Kalman filter. In Proc. 17th European Symposium Improved Oil Recovery, St. Petersburg, Russia, 16-18 April.

Rosenbrock, H. 1960. An Automatic Method for Finding the Greatest or least Value of a Function. *The Computer Journal*. **3** 175–184

Sarma, P., Aziz, K. and Durlofsky, L.J. 2005. Implementation of adjoint solution for optimal control of smart wells. Paper SPE 92864 presented at SPE Reservoir Simulation Symposium, The Woodlands, Texas, U.S.A, 31 January-2 February.

Sarma, P., and Chen, W. 2014. Improved Estimation of the Stochastic Gradient with Quasi-Monte Carlo Methods. Proc. 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV), Catania, Italy, 8-11 September.

Strang, G. 2006: *Linear algebra and its applications,* 4th ed., Thomson Brooks/Cole, Pacific Grove

Stordal, A.S., Szklarz, S.P. and Leeuwenburgh, O. 2014. A closer look at ensemble based optimization in reservoir management. Proc. 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV), Catania, Italy, 8-11 September.

Van Essen, G.M., Zandvliet, M.J., Van den Hof, P.M.J., Bosgra, O.H. and Jansen, J.D. 2009. Robust waterflooding optimization of multiple geological scenarios. *SPEJ* **14**(1), 202–210.