Improved sampling strategies for ensemble-based optimization

Ramaswamy, K. R. ; Fonseca, R. M.; Leeuwenburgh, O.; Siraj, M.M.; Van den Hof, P.M.J.

DOI
10.1007/s10596-019-09914-8

Publication date
2020

Document Version
Final published version

Published in
Computational Geosciences

Computational Geosciences (2020) 24:1057–1069
https://doi.org/10.1007/s10596-019-09914-8

ORIGINAL PAPER

Improved sampling strategies for ensemble-based optimization

K. R. Ramaswamy[1] · R. M. Fonseca[2] · O. Leeuwenburgh[2,3] · M. M. Siraj[1] · P. M. J. Van den Hof[1]

Received: 10 December 2018 / Accepted: 23 October 2019 / Published online: 1 May 2020

Abstract

We are concerned with the efficiency of stochastic gradient estimation methods for large-scale nonlinear optimization in the presence of uncertainty. These methods aim to estimate an approximate gradient from a limited number of random input vector samples and corresponding objective function values. Ensemble methods usually employ Gaussian sampling to generate the input samples. It is known from the optimal design theory that the quality of sample-based approximations is affected by the distribution of the samples. We therefore evaluate six different sampling strategies to optimization of a high-dimensional analytical benchmark optimization problem, and, in a second example, to optimization of oil reservoir management strategies with and without geological uncertainty. The effectiveness of the sampling strategies is analyzed based on the quality of the estimated gradient, the final objective function value, the rate of the convergence, and the robustness of the gradient estimate. Based on the results, an improved version of the stochastic simplex approximate gradient method is proposed based on $UE(s^2)$ sampling designs for supersaturated cases that outperforms all alternative approaches. We additionally introduce two new strategies that outperform the $UE(s^2)$ designs previously suggested in the literature.

Keywords Robust optimization · Ensemble methods · Approximate gradient · Sampling strategies

1 Introduction

A continuous increase over recent decades in computing power, accompanied by improvements in numerical algorithms, has led to increasing use of simulation models to obtain optimal operating strategies for complex systems. Simulation of these models may be very computationally demanding and will therefore require highly efficient numerical optimization workflows. One domain in which computational demands are continuously challenging the efficiency of optimization workflows is the management of subsurface hydrocarbon reservoirs. This problem can be characterized by large-scale multiphase and compositional flow models; by high-dimensional control (input) spaces; and by large geological, economical, and operational uncertainty. The presence of significant uncertainty, even after years of data gathering, motivates the optimization of the expected value of the objective function, an approach that is sometimes referred to as robust optimization [35].

Controls may include the number of wells to be drilled (10–100s), their locations and trajectories, the drilling order, the well type (injector or producer), and operational controls such as well rates or pressures over a period of several years. The total number of variables to be optimized can easily be in the order of 1000s. Gradient-based methods have been shown to be the most efficient techniques to find optimal solutions for these complex problems [3, 16, 17, 32]. In many practical cases of interest, the types of controls (e.g., integer or categorical) and lack of access to the numeric model code prevent use of efficient gradient estimation by means of the adjoint method.

In such scenarios, approximate gradient methods that require a limited number of test simulations with perturbed controls as input have been proven to be quite useful. An advantage of this approach is that it treats the model as a black box, and therefore offers great flexibility in terms of the type of controls that can be considered. The main challenge associated with this approach is to ensure that

✉ O. Leeuwenburgh
olwijn.leeuwenburgh@tno.nl

1 Eindhoven University of Technology, Eindhoven, Netherlands

2 TNO, Utrecht, Netherlands

3 Faculty of Geosciences and Engineering, Delft University of Technology, Delft, Netherlands

the approximate gradients are accurate enough to enable sufficient increases in the objective function at reasonable computational cost. Since test simulations can be performed in parallel (assuming the availability of a parallel computing facility), this challenge translates into choosing the control perturbation set (ensemble) in such a way that the gradients can be estimated with minimal error.

Various methods exist for gradient approximation. Deterministic methods include finite differences and the simplex gradient [7], both of which are computationally unattractive for large numbers of controls since they require as many perturbation tests runs as there are controls. Stochastic approaches based on a limited number of random perturbations include simultaneous perturbation stochastic approximation (SPSA) [33] and stochastic noise reaction (SNR) [26], both of which are based on averaging, and ensemble optimization (EnOpt) [6] and a modified version coined stochastic simplex approximate gradient (StoSAG) [10, 11, 14], which are both based on least squares linear regression. Do and Reynolds [10] discussed the relationship between some of these methods in a deterministic context.

The sampling strategy (for example, the distribution) used to generate the ensemble of controls is extremely important. The SPSA method, for example, utilizes control perturbations sampled from the Bernoulli distribution instead of Gaussian sampling. The impact of sampling strategy on the performance of ensemble methods, on the other hand, has so far received rather little attention. The impact of ensemble size on the quality of the ensemble gradient was investigated for the Rosenbrock function and for an oil reservoir model [12]. It was also shown that the perturbation size (the standard deviation of a multivariate Gaussian distribution) has a significant impact on the gradient quality. A method called CMA-EnOpt to adaptively adjust the perturbation size through covariance matrix adaptation (CMA) was found to improve the performance of ensemble gradients [13]. The impact of alternative distributions was considered by Sarma and Chen [31] who investigated the impact of a quasi-random sampling method (Sobol sampling, [25]) that avoids clustering of samples on SNR gradient estimates. They found Sobol sampling to lead to a faster rate of convergence relative to Gaussian sampling when applied to a deterministic reservoir optimization problem. The performance of Sobol sampling strategies in a robust optimization context was not investigated.

Considering that the number of controls (N) for the problems of interest will normally be much larger than the feasible number of test simulations (M), we will be dealing here exclusively with the underdetermined (supersaturated) case. Specifically, we will address the question which sampling strategy for the supersaturated case leads to optimal performance of the approximate gradient estimation methods within large-scale nonlinear optimization problems under uncertainty. We investigate three categories of sampling: quasi-random (low-discrepancy) sequences, stratified sampling, and sampling designs motivated by optimality criteria. All sampling methods are applied in combination with the StoSAG gradient estimation method.

In the remainder of this paper, we first provide a brief review of ensemble optimization for both deterministic and robust cases in Section 2, followed by a discussion of the various sampling strategies used in this paper (Sobol sampling, Latin hypercube sampling (LHS), $UE(s^2)$-optimal supersaturated design) and the motivation for considering them in Section 3. Here we also introduce two new variants of $UE(s^2)$-optimal supersaturated design. Finally in Section 4, the sampling strategies are applied in conjunction with the StoSAG method first to the extended Rosenbrock optimization test function [9] and subsequently to a synthethic 3D reservoir model of realistic complexity (for both deterministic and robust cases) followed by a detailed analysis.

2 Ensemble-based gradient estimation

Chen [6] proposed a stochastic gradient estimation method for use within an ensemble-based optimization workflow referred to as EnOpt. Later modifications for deterministic and robust optimization problems [10, 11] addressed approximation errors associated with the use of ensemble means. The modified method has since been referred to as stochastic simplex approximate gradient (StoSAG) [14] to highlight the relationship with the simplex gradient [20]. While the simplex gradient estimation method is a full-rank deterministic method, the StoSAG method is a low-rank stochastic method based on random perturbations. With low-rank, we mean that the estimation typically involves fewer equations than unknowns. Consider the objective function $J(\boldsymbol{u}, \boldsymbol{m})$ of the control vector $\boldsymbol{u} = [u_1, \ldots, u_N]^\mathsf{T}$ and of model parameter vector $\boldsymbol{m}$. Given an ensemble of control perturbation vectors $\boldsymbol{U} = [\delta\boldsymbol{u}^1 \ldots \delta\boldsymbol{u}^M]^\mathsf{T}$ and corresponding objective function values anomalies $\boldsymbol{j} = [J(\boldsymbol{u}+\delta\boldsymbol{u}^1, \boldsymbol{m}) - J(\boldsymbol{u}, \boldsymbol{m}), \ldots, J(\boldsymbol{u}+\delta\boldsymbol{u}^M, \boldsymbol{m}) - J(\boldsymbol{u}, \boldsymbol{m})]^\mathsf{T}$, a first-order Taylor expansion of J around $\boldsymbol{u}$ leads to the linear system of equations

$$\boldsymbol{U}\boldsymbol{g} \approx \boldsymbol{j} \quad , \tag{1}$$

from which we wish to estimate the gradient $\boldsymbol{g} = \nabla_u J$. If the model is considered uncertain, one may choose to define an expected objective function $\overline{J}(\boldsymbol{u}) = \frac{1}{N_r}\sum_{i=1}^{N_r} J(\boldsymbol{u}, \boldsymbol{m}^i)$. An expected gradient can be obtained similarly as the sum of N_r individual gradients, which can be computed using Eq. 1 using M/N_r control perturbation vectors each. Theoretical and numerical studies [6, 14] in which a ratio equal to 1, that is $M = N_r$, was used, have shown that the expected gradient $\overline{\boldsymbol{g}} = \nabla_u \overline{J}$ can alternatively be estimated by solving a single system like (1) with $\boldsymbol{j}$ replaced by

$\tilde{\boldsymbol{j}} = [J(\boldsymbol{u}+\delta\boldsymbol{u}^1, \boldsymbol{m}^1) - J(\boldsymbol{u}, \boldsymbol{m}^1), \ldots, J(\boldsymbol{u}+\delta\boldsymbol{u}^M, \boldsymbol{m}^M) - J(\boldsymbol{u}, \boldsymbol{m}^M)]^\mathsf{T}$,

$$\boldsymbol{U}\overline{\boldsymbol{g}} \approx \tilde{\boldsymbol{j}} \quad . \tag{2}$$

In the following, we will refer to the optimization problem with $N_r = 1$ as deterministic, and to the case with $N_r > 1$ as robust optimization. For robust optimization with $M/N_r = 1$, we will use Eq. 2, and for higher ratios, we will use the mean of individual gradients computed each by solving Eq. 1. In all cases, the number of unknowns is N and the number of model simulations required to evaluate the required gradient is M.

Taking as example the deterministic case Eq. 1, the normal equations can be formulated by pre-multiplying with $\boldsymbol{U}^\mathsf{T}$, leading to

$$\boldsymbol{U}^\mathsf{T}\boldsymbol{U}\ \boldsymbol{g} \approx \boldsymbol{U}^\mathsf{T}\ \boldsymbol{j} \quad . \tag{3}$$

The matrix $\boldsymbol{U}^\mathsf{T}\boldsymbol{U}$ has dimension $N \times N$. Since the number of perturbations M that we can afford to evaluate (i.e., the number of equations) is typically less than N, the number of controls, the $N \times N$ matrix $\boldsymbol{U}^\mathsf{T}\boldsymbol{U}$ is rank deficient and its inverse does not exist. A unique solution is normally obtained by imposing a minimum norm constraint and can be computed from the generalized pseudoinverse as

$$\hat{\boldsymbol{g}} = \boldsymbol{U}^\dagger \boldsymbol{j} = (\boldsymbol{U}^\mathsf{T}\boldsymbol{U})^\dagger\, \boldsymbol{U}^\mathsf{T}\ \boldsymbol{j} \quad . \tag{4}$$

A regularized gradient estimate can be obtained by pre-multiplication of the above gradient by a positive definite matrix $\boldsymbol{B}$. A common choice is $\boldsymbol{B} = \boldsymbol{C}_u$ where $\boldsymbol{C}_u$ is a block-diagonal covariance matrix prescribing correlations between controls, for example, over time.

It can be shown [34] that if $\{\boldsymbol{z}^i\}_{i=1}^M$ with $\boldsymbol{z}^i = \boldsymbol{u} + \delta\boldsymbol{u}^i = \boldsymbol{u}^i$ is an i.i.d. sample from the multivariate Gaussian density $\mathcal{N}(\boldsymbol{u}, \boldsymbol{C}_u)$, the ensemble gradient (4) has the following convergence property (in the almost sure sense) for $M \to \infty$

$$\begin{aligned}
\hat{\boldsymbol{g}} = (\boldsymbol{U}^\mathsf{T}\boldsymbol{U})^\dagger \boldsymbol{U}^\mathsf{T}\boldsymbol{j} \ &= \ (\frac{1}{M}\boldsymbol{U}^\mathsf{T}\boldsymbol{U})^\dagger \frac{1}{M}\sum_{i=1}^{M}(\boldsymbol{z}^i \\
&\quad -\boldsymbol{u})(J(\boldsymbol{z}^i) - J(\boldsymbol{u})) \qquad (5)\\
&\quad \overset{a.s.}{\to} \boldsymbol{C}_u^{-1}\int (J(\boldsymbol{z}) - J(\boldsymbol{u}))\ (\boldsymbol{z} \\
&\quad -\boldsymbol{u})\ \mathcal{N}(\boldsymbol{z}|\boldsymbol{u}, \boldsymbol{C}_u)d\boldsymbol{z} \qquad (6)\\
&= \int J(\boldsymbol{z})\ \nabla_u \mathcal{N}(\boldsymbol{z}|\boldsymbol{u}, \boldsymbol{C}_u)d\boldsymbol{z} \qquad (7)
\end{aligned}$$

In other words, the ensemble gradient (4) is a Monte Carlo (i.e., random sampling-based) approximation of a probability-weighted integral of the function values $J(\boldsymbol{u}^i)$ over all possible values of $\boldsymbol{u}^i$. The convergence properties of such an approximation will depend strongly on the chosen sampling strategy [4].

3 Sampling strategies

In the context of estimation, the matrix $\boldsymbol{U}$ is known as the design matrix and the matrix $\boldsymbol{S} = (\boldsymbol{U}^\mathsf{T}\boldsymbol{U})$ as the information matrix. The choice for a set of samples is that for a particular design and can be motivated by the desired statistical properties of the solution of Eq. 3. These properties generally depend on properties of the matrix $\boldsymbol{S}$ or, equivalently, of its inverse, known as the dispersion matrix, and lead to a number of optimality criteria which will be discussed later in this section. If $M \geq N$ and the rank of $\boldsymbol{U}$ is equal to or greater than N, the solution (4) is the best linear unbiased estimator (BLUE) and has variance proportional to $\boldsymbol{S}^{-1}$. In the case that $M < N$, which is most relevant here, the solution (4) is the minimum bias estimator. If $M = N$ and the elements $S_{ij} = 0$ for all $i \neq j$, the design is called orthogonal.

3.1 Random sampling

Random sampling (or Monte Carlo sampling) is the conventional approach to generate control perturbations for ensemble-based gradient estimation. A generic approach to generating samples is to obtain random combinations of basis vectors that are obtained by factorization of a perturbation covariance matrix $\boldsymbol{C}_u$, for example, by Cholesky decomposition $\boldsymbol{C}_u = \boldsymbol{L}^\mathsf{T}\boldsymbol{L}$, such that $\delta\boldsymbol{u}^i = \boldsymbol{L}\,r^i$, where r^i is a number from a pseudo-random sequence as can be generated by random number generators available with any computer code. The standard distribution used for ensemble gradient estimation is the Gaussian distribution, i.e., $r \propto \mathcal{N}(0, \boldsymbol{C}_u)$. If perturbations are uncorrelated, $\boldsymbol{C}_u = \sigma^2\,\boldsymbol{I}_N$. In some cases, for example, when the controls represent long time series discretized in short intervals (as is typical for the oil reservoir well control problem), a regularized solution may be obtained by imposing time correlation between subsequent controls. In this case, $\boldsymbol{C}_u$ will be a block-diagonal matrix.

3.2 Quasi-Monte Carlo sampling

Consider the N-dimensional half-open unit cube $\mathbb{I}^N = [0,1)^N$, $N \geq 1$. Let f be a real integrable function over $\mathbb{I}^N$. The Monte Carlo approximation of an integral of f over $\mathbb{I}^N$ is

$$\int_{I^N} f(x)dx \approx \frac{1}{M}\sum_{i=1}^{M} f(x_i) \tag{8}$$

where x_1, x_2, .., x_M are random points in $\mathbb{I}^N$ [28]. The strong law of large numbers ensures the convergence of

the approximation. Furthermore, the expected integration error is bounded by $O(\frac{1}{\sqrt{M}})$, with the interesting fact that it does not depend on dimension N. We have already seen that the ensemble gradient estimation is equivalent with a Monte Carlo integration (also known as quadrature). The quasi-Monte Carlo (QMC) method [23] is an alternative to the Monte Carlo (MC) method for calculating this approximation using quasi-random (deterministic) sequences with higher convergence rate than obtained with (pseudo) random sequences. The improved convergence originates from the uniformity of the quasi-random sampling distribution. The uniformity is quantified by the so-called discrepancy which measures the relative density of samples in each sub-volume of the sampled domain. Low-discrepancy sequences have good uniformity properties [4]. Examples of low-discrepancy quasi-random sequences are the Sobol and Halton sequences. More detailed discussion of quasi-random sequences and their properties is provided by, e.g., Niederreiter [24]. Given their low-discrepancy properties, which avoid clustering of samples in subvolumes, they are good candidates for generating space-filling designs [4]. In this work, we will present results obtained with the Sobol sequence which tends to produce lower correlations in high dimensions than Halton sampling [5, 23]. Successful application of Sobol sequences in problems of dimension ~ 300 has been reported in the literature [29].

3.3 Stratified sampling

A number of approaches that directly address the error variance of the Monte Carlo estimate are discussed by Caflisch [4]. Stratification is a variance reduction technique that, like low-discrepancy sampling, attempts to avoid the clustering of samples. Latin hypercube sampling (LHS) [22] is perhaps the best known stratified sampling method that is suitable for higher dimensions and settings where $M < N$ [28]. LHS divides the input (design) space equally into M strata (subdomains), an arrangement known as Latin squares, and places a sample randomly in each stratum. Theoretical considerations [22, 27] suggest that LHS can be much better than MC sampling and it cannot be much worse. However, it has been reported that LHS methods may produce clustering of samples in high dimensions [8].

Figure 1 shows examples of Gaussian and uniform (pseudo) random, quasi-random (Sobol), and stratified (LHS) sample distributions for a simple 2-control example and 100 samples, that is, $N = 2$ and $M = 100$. While this is different from the $M < N$ case of interest, the figure serves as a simple illustration of the motivation for considering sample distributions other than Gaussian. In order to enable comparison of the sample spread, the standard deviation was normalized to 1 in both directions for all four distributions. Gaussian sampling produces relatively dense sampling around the center, as expected. LHS appears

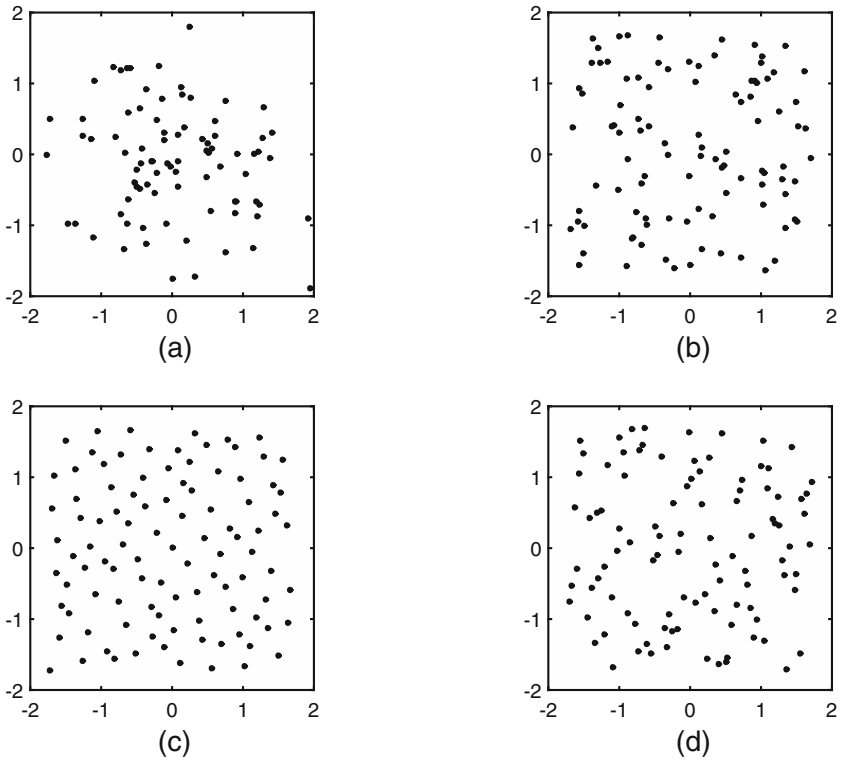

Fig. 1 Example of 2D sample distributions with standard deviation 1. **a** Gaussian. **b** Uniform. **c** Sobol. **d** LHS

to produce sampling distributions that are very similar to uniform sampling, at least for the case $M > N$, with some clustering and under-sampled intervals. Sobol sampling can be seen to produce a uniform space-filling sample distribution.

3.4 Optimal supersaturated designs

The theory of Design of Experiments (DOE) distinguishes saturated ($M = N$) and nonsaturated ($M > N$) designs. It furthermore defines a number of design criteria and techniques to obtain designs that satisfy these criteria. Here we are interested primarily in designs for the supersaturated case ($M < N$) based on the $E(s^2)$ criterion which defines an approximate orthogonality measure [2]. An extension, the so-called $UE(s^2)$-optimal supersaturated designs [19], adds an effective design optimality criterion (D-optimality).

For a supersaturated design, the information matrix $\boldsymbol{S}$ becomes rank deficient and hence its inverse does not exist. A natural approach in this case is to find a design that is nearly orthogonal, that is, the design in which the absolute values of the off-diagonal elements of the matrix $\boldsymbol{S}$ are small in some sense. Two alternative approaches have been suggested [2] to obtain near-orthogonal designs. The first is to choose a design with minimum $\max_{i \neq j} |s_{ij}|$ and among all such designs to choose one with the fewest s_{ij} that achieve this maximum. The second approach is to choose a design in which the sum of the squares of the off-diagonal elements is minimum, that is, a design that minimizes

$$E(s^2) = \frac{2}{(N-1)(N-2)} \sum_{i<j} s_{ij}^2 \quad , \tag{9}$$

which is called the $E(s^2)$ criterion. A design is $E(s^2)$-optimal if it satisfies the following conditions:

1. $s_{1j} = 0 \; \forall j = 2, ..., N$
2. among all those designs that satisfy 1, the design should minimize $E(s^2)$ given in Eq. 9.

There are various methods for construction of $E(s^2)$-optimal supersaturated designs [15], but we will consider only the methods using Hadamard matrices [21, 36].

Optimal designs are experimental designs for which the solution of the estimator satisfies particular statistical optimality criteria. Generally, these statistical criteria are formulated in terms of the (generalized) variance of the solution, for example, minimum trace of the covariance of $\hat{\boldsymbol{g}}$ (A-optimality), minimum maximum eigenvalue of the covariance of $\hat{\boldsymbol{g}}$ (E-optimality), or minimum product of non-zero eigenvalues of the covariance of $\hat{\boldsymbol{g}}$ (D-optimality) [1]. The $E(s^2)$ design can be made more theoretically strong and efficient by adding such traditional design optimality criteria. $UE(s^2)$-optimal designs are designs for the supersaturated case that are near-orthogonal but exchange the first constraint above for D-optimality. $UE(s^2)$-optimal supersaturated designs could therefore be described as producing minimum bias-minimum variance estimates (for details about algorithms for their construction from Hadamard matrices, we refer to Jones and Majumdar [19]). A brief summary of the general procedure and variants is provided here.

A Hadamard matrix $\boldsymbol{H} \in \mathbb{R}^{N \times N}$ is a square matrix whose columns are orthogonal to each other and for which holds that $\boldsymbol{H}\boldsymbol{H}^{\mathsf{T}} = \boldsymbol{H}^{\mathsf{T}}\boldsymbol{H} = N\,\boldsymbol{I}_N$ where $\boldsymbol{I}_N$ is the identity matrix of size $N \times N$. It consists of elements ± 1 and it is generally available for order N equal to 1, 2 and multiples of 4. Procedures for constructing a $UE(s^2)$-optimal design matrix $\boldsymbol{U} \in \mathbb{R}^{M \times N}$ with $M < N$ from Hadamard matrices are discussed by [19], who also review modern methods to construct Hadamard matrices of the required orders. Four situations can be distinguished based on the remainder of N when divided by 4 that are referred to as T_0, T_1, T_2, and T_3. In the following, $N = a(\text{mod } 4)$ means a is the remainder when N is divided by 4.

1. T_0: If $N = 0(\text{mod } 4)$, $2 \leq M \leq N$-1. Start with a normalized Hadamard matrix of order N, $\boldsymbol{H}_N$. $\boldsymbol{U}$ can be formed by selection of any M rows of $\boldsymbol{H}_N$.
2. T_1: If $N = 1(\text{mod } 4)$, $2 \leq M \leq N$-1. Start with a normalized Hadamard matrix of order N-1, $\boldsymbol{H}_{N-1}$. Let $\boldsymbol{V}$ be a $M \times (N-1)$ matrix formed by any M rows of $\boldsymbol{H}_{N-1}$ and let $\boldsymbol{\phi}$ be an (arbitrary) $M \times 1$ vector with entries 1 or -1. $\boldsymbol{U} = (\boldsymbol{V}, \boldsymbol{\phi})$.
3. T_2: If $N = 2(\text{mod } 4)$, $2 \leq M \leq N$-2.
   (a) M is even, $M = 2p$. Start with a normalized Hadamard matrix of order $N-2$, $\boldsymbol{H}_{N-2}$. Let $\boldsymbol{U^*}$ be the $M \times (N-2)$ matrix formed by any M rows of $\boldsymbol{H}_{N-2}$. Let $\boldsymbol{X_1}$ be a $M \times 2$ matrix with each of the first p rows either (1,1) or (−1,−1) and each of the last p rows either (1,−1) or (−1,1). Then $\boldsymbol{U} = (\boldsymbol{U^*}, \boldsymbol{X_1})$.
   (b) M is odd, $M = 2p + 1$. Start with a normalized Hadamard matrix of order $N-2$, $\boldsymbol{H}_{N-2}$. Let $\boldsymbol{U^*}$ be the $M \times (N-2)$ matrix formed by any M rows of $\boldsymbol{H}_{N-2}$. Let $\boldsymbol{X_2}$ be a $M \times 2$ matrix with each of the first p rows either (1,1) or (−1, −1) and each of the last $p+1$ rows either (1, −1) or (−1, 1). Then $\boldsymbol{U} = (\boldsymbol{U^*}, \boldsymbol{X_2})$.
4. T_3: If $N = 3(\text{mod } 4)$, $2 \leq M \leq N$-1. Start with a normalized Hadamard matrix of order $N+1$, $\boldsymbol{H}_{N+1}$. Let $\boldsymbol{U^*}$ be the $M \times (N+1)$ matrix formed by any M rows of $\boldsymbol{H}_{N+1}$. Suppose the last column of $\boldsymbol{U}^*$ is denoted by ϕ and $\boldsymbol{U^*} = (\boldsymbol{U}, \boldsymbol{\phi})$. Thus $\boldsymbol{U}$ can be obtained.

Given that there is some freedom in constructing the Hadamard matrices, we will consider three variants for constructing $\boldsymbol{U}$ denoted as M1, M2, and M3 as explained below. M1 is the conventional approach, while M2 and M3 are new variants that we introduce here.

1. M1: This is the approach suggested by Jones and Majumdar [19]. In the construction of $UE(s^2)$-optimal designs of types T_0 to T_3, it is suggested to take M arbitrary rows of a Hadamard matrix. By choosing the rows randomly in each iteration of the optimization, variation in the samples can be achieved without loss of $UE(s^2)$ optimality. Thus, the method becomes stochastic.
2. M2: This approach is similar to type M1 except that the row of the Hadamard matrix containing only values of +1 is always picked. When type M1 is used, this row may not always be picked. Experiments presented below showed that in those instances, gradient quality was significantly reduced. The M2 variant avoids this but remains stochastic.
3. M3: In this approach, the first M rows of the Hadamard matrix (including the row with all values equal to +1) are always selected for each iteration of the optimization. This variant is therefore deterministic.

We finally note that the near-orthogonality and D-optimality of $UE(s^2)$-optimal designs do not hold for the case $N = 3(\text{mod } 4)$ where $M > \frac{(N+5)}{2}$. However, by choosing M properly, the design can be made D-optimal for this case as well [19].

4 Numerical experiments

The various sampling strategies and designs are first applied to gradient estimation in a simple toy problem for which exact gradients can be computed analytically. This enables us to determine the impact of the sampling strategy on the accuracy of the stochastic gradients.

We determine the gradient accuracy separately for two sets of control test points. In order to explain the composition of these two sets, we refer to the red line in Fig. 6 which represents a typical example of the evolution of the objective function as a function of iteration of an optimization process. The iterative objective function increase during optimization can be commonly characterized by fast improvements during early iterations when objective function values are far from the optimum (the objective function curve is steep), and very slow improvement towards convergence (the objective function curve is nearly flat). In our example (Fig. 6), the first stage would consist of roughly the first 10 iterations, while the second stage would consist of iterations 10 to 35. The separation between the two stages can simply be defined by, for example, a threshold rate of improvement in the objective function. We propose this procedure to identify points that are expected to lie relatively far away from the optimum (visited during the first stage) and points that are expected to lie in a part of the control space that is connected to the optimum through pathways along which the objective function is only very weak sloping (visited during the second stage). We are interested in determining the quality of gradient estimates during both stages of the optimization process separately since it may be expected that for a given perturbation size, the gradient quality is lower during the second stage than during the first stage. One hundred conventional robust optimization experiments were conducted with the extended Rosenbrock function for one hundred different starting controls.

4.1 Analytical toy problem

We use an extended version of the well-known Rosenbrock benchmark function which is characterized in 2D by a curved valley, with a minimum at coordinates $(1, 1)$ located in one of the two branches of the valley. In order to mimic the high dimensionality typically encountered in subsurface reservoir problems, we use the extended Rosenbrock function [9]. In addition to a large number of controls we want to investigate the impact of uncertainty in the model properties. Therefore, uncertainty is introduced to mimic the geological uncertainty following [12],

$$J(u_1, ..., u_N, c_1^j, c_2^j) = \sum_{i=1}^{N/2} -(\sin c_2^j)(1 - u_{2i-1})^2 - 100(c_1^j u_{2i} - u_{2i-1}^2)^2 , \quad (10)$$

where (c_1^j, c_2^j) with $j = 1, \ldots, 100$ are samples from $\mathcal{N}(0, \boldsymbol{I}_2)$ representing $N_r = 100$ model realizations. N is the number of controls which we set here to 320. The gradient of Eq. 10 can be derived analytically for any set of controls.

Using a simple objective function increase criterion solutions obtained at subsequent iterations of the optimization process were classified as belonging to either the first or second stage as discussed above. The gradient at each of these test points was estimated using different sampling strategies. For sampling strategies involving random numbers, the gradient computation was repeated 100 times, after which an average angle error α_{100} was computed by comparison with the analytical gradient direction and further averaging over all 50 test points. The standard deviation of the control perturbation magnitude was set to 0.01 in all cases. The angle error is estimated for different ratios M/N_r. If the ratio equals 1, we use the same number of perturbations as there are model realizations, while for a ratio of for example 3, three different perturbed controls are applied to each model realization. Figure 2 shows the angle error alpha 100 for 3 ratios and for different sampling strategies averaged over the first 50 test points and Fig. 3 shows the average angle error for the second set of 50 test points.

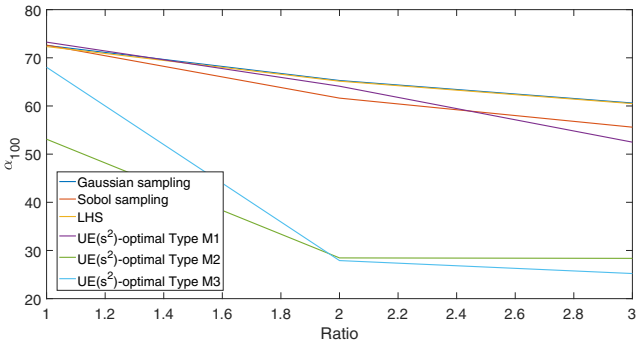

Fig. 2 Mean gradient direction error α_{100} at points with objective function values relatively far from the optimum for ratios M/N_r 1 to 3 and different sampling strategies using a standard deviation σ of 0.01. The true and estimated gradients represent the expected values over 100 different model realizations

For control points far away from the optimum, the Gaussian, Sobol, LHS, and $UE(s^2)$-M1 sampling strategies provide a similar gradient quality when the ratio equals 1 (Fig. 2). A slight difference is seen when the ratio is increased to 3 with $UE(s^2)$-M1 and Sobol performing slightly better on average than the Gaussian and LHS strategies. For a 1:1 ratio, $UE(s^2)$-M2 and $UE(s^2)$-M3 provide the best gradient quality, with angle errors that are 5° to 30° lower than for the other strategies.

Gradient errors are significantly larger for points closer to the optimum as seen in Fig. 3. Otherwise, the results are more or less consistent with those for the early stage except that the differences between $UE(s^2)$-optimal designs of types M2 and M3 and Gaussian, Sobol, LHS, and $UE(s^2)$-optimal design of type M1 are relatively smaller. In conclusion, for the high-dimensional Rosenbrock function with uncertainty, $UE(s^2)$-optimal designs of types M2 and M3 provide significantly better expected gradients than the other sampling strategies, especially in the early stages of the optimization process when solutions are far from the optimum. In the following, we will only present robust optimization results for a ratio of 1.

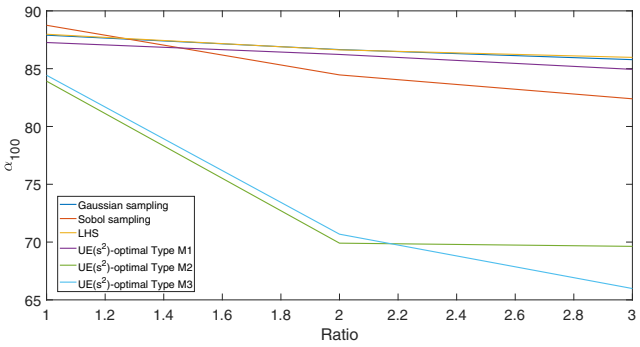

Fig. 3 Mean gradient direction error α_{100} at points with objective function values relatively close to the optimum for ratios M/N_r 1 to 3 and different sampling strategies using a standard deviation σ of 0.01. The true and estimated gradients represent the expected values over 100 different model realizations

4.2 Oil reservoir case

In this section, we will investigate the impact of the different sampling strategies on an optimization process for a small, but realistically complex, reservoir test case. The 3D reservoir model used here is the "Egg" benchmark model [18, 35]. Figure 4 shows the permeability field of one model realization and the position of eight injection wells (blue) and four production wells (red). The egg model is a channelized reservoir model with seven vertical layers and a total of 18,553 active cells.

The permeability values are not conditioned to values at the wells, and the porosity is assumed to be constant. The producers are operated at constant bottom hole pressure, while the injectors are rate-controlled between 10 and 79.5 m^3/day. Production of the field is simulated for a period of 3600 days that is discretized into 40 control time intervals of 90 days. This results in a total of $40 \times 8 = 320$ injection rate controls. The objective function used in this work is the undiscounted net present value (NPV), i.e., the sum of revenues and costs induced over the production period. We use an oil price of 126 \$/m^3 and costs of 19 \$/m^3 and 6 \$/m^3 for water production and injection respectively. The fully implicit black oil simulator OPM flow is used for the model simulations and the objective function is computed based on the simulator output. We investigate both the deterministic and robust optimization cases, where the model realizations are taken from a set of 100 permeability realizations. Six of these realizations are shown in Fig. 5. More details on the model can be found in Jansen et al. [18].

4.2.1 Deterministic optimization

In deterministic optimization, there is no uncertainty in the model, and therefore only a single model realization

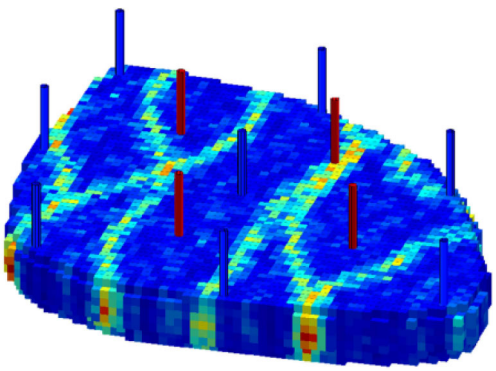

Fig. 4 Example permeability field of the Egg reservoir model with 8 injector wells (blue) and 4 producer wells (red)

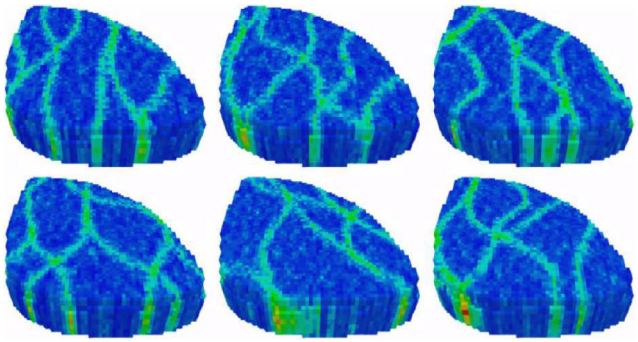

Fig. 5 Six randomly chosen model realizations, taken from [18], characterizing the uncertainty in the permeability

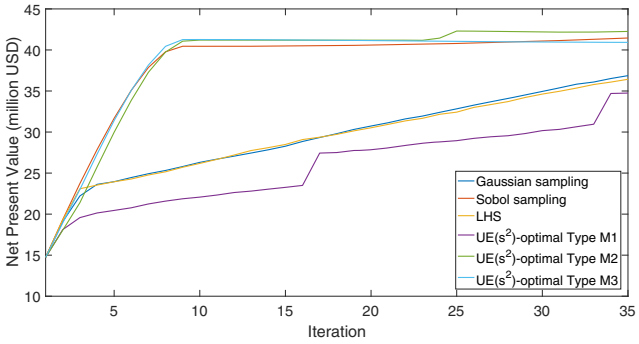

Fig. 7 NPV as a function of iteration for different sampling strategies. Gradients are estimated from Eq. 1 with $N_r = 1$ and $M = 100$

is used in this section (i.e., $N_r = 1$). All optimization experiments are run for a fixed number of iterations (35) and use a steepest ascent update with a normalized gradient (that is, the norm of the gradient vector is 1) and a fixed step size of 0.1. The convergence will thus be affected primarily by the quality of the gradient. The initial control vector consists of equal values of 79.5 in units of m^3/day which corresponds to the maximum injection rate. M =100 perturbation vectors are generated to estimate the gradients by solving (4). Figures 6 and 7 show the objective function curves over all iterations for different sampling strategies with the number of perturbation vectors M=100 and M = 30 respectively.

The curves for $UE(s^2)$ designs of types M2 and M3 flatten after 14 iterations. The curve for Sobol sampling approaches the same final objective function value at a slightly slower rate. These methods also produce high convergence rates and final objective function values when the ensemble size is very small (30). From the results of the Rosenbrock function, it was observed that $UE(s^2)$ designs of type M1 provides inferior gradient quality compared with M2 and M3 at poor control points (steep section of the objective function curve). This is also observed in Figs. 6 and 7. The curve for M1 shows iteration intervals for which convergence is extremely slow, alternated by intervals with steep increases in the objective function. Upon inspection, it was discovered that these intervals correspond to iterations in which the row of the Hadamard matrix containing only +1 values was either not included (slow improvement) or was included (fast improvement). This behavior actually motivated the creation of the new schemes M2 and M3 in this paper. In general, both M2 and M3 perform slightly better than Sobol sampling which tends to produce objective function curves that flatten a bit earlier. Since the curves for Gaussian and LHS sampling have not yet flattened after 35 iterations, it is not possible from these results to draw conclusions about the final objective function value that can be reached. Given the high computational cost associated with simulating large and complex reservoir models, it is reasonable to consider the performance of different methods for a limited number of iterations (or function evaluations). It appears that LHS does not perform better than Gaussian sampling for the number of controls considered in these experiments.

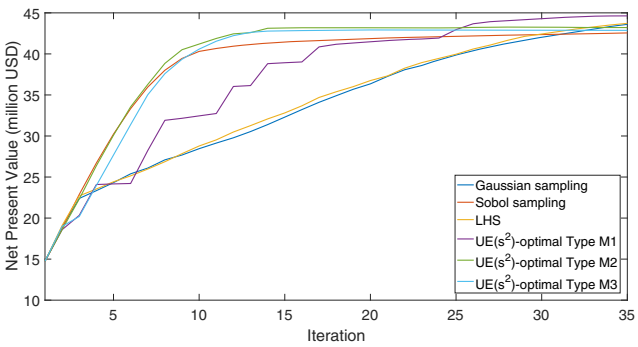

Fig. 6 NPV as a function of optimization iteration using Eq. 1 with $N_r = 100$ and M = 100 for different sampling strategies

The optimal control strategies for one of the injectors obtained after 35 iterations are shown in Fig. 8.

The choice of sampling method clearly has a significant impact on the character of the resulting control strategy. While Sobol sampling produces a strategy with frequent and large changes in the injection rate, the $UE(s^2)$ designs tend to produce fairly smooth low-rate profiles. Highly dynamic control strategies are generally undesirable from an operational point of view. Regularization of the gradients is often proposed as a means to produce smooth control profiles. One way to achieve this is by imposing correlations over time between the control perturbations, i.e., between the samples. The impact of this approach on the optimization process for different sampling methods is illustrated in Fig. 9, where a correlation length of 15 control intervals was applied (the total number of intervals is 40).

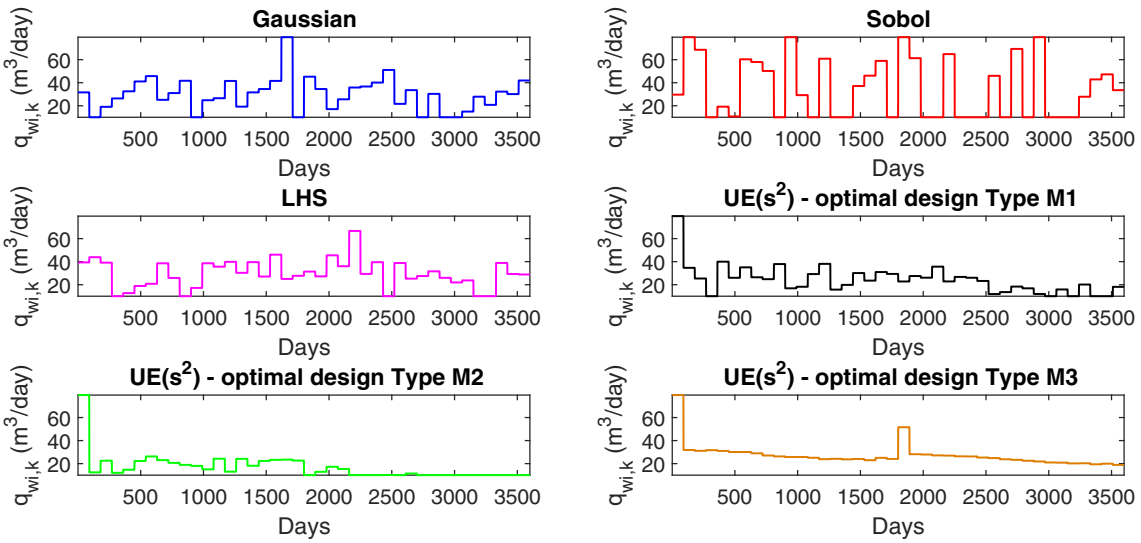

Fig. 8 Control strategy representing the injection rate of one of the injectors as a function of time at the final iteration (35) obtained with 6 different sampling strategies, $N_r = 1$ and $M = 100$

Correlation clearly benefits the convergence properties for all sampling methods except Sobol sampling. Gaussian sampling, LHS, and $UE(s^2)$ designs all produce very similar objective function profiles. Gaussian sampling and LHS produce the highest final objective function values, while the values obtained for $UE(s^2)$ are nearly identical to those obtained without induced correlation. The convergence rate for Sobol sampling on the other hand has decreased notably. We conclude that this latter result must be related to the loss of uniformity of Sobol distributions after smoothing.

4.2.2 Robust optimization

In this section, the optimization is aimed at maximizing the expected NPV as evaluated over $N_r = 100$ equiprobable realizations of the model with different permeability fields as illustrated in Fig. 5. The gradient of the expected NPV is computed directly using the formulation of Eq. 2 based on 100 control perturbation vectors that are paired on a 1:1 ratio basis to the model realizations. The perturbation standard deviation, random seed, and initial controls are the same for all experiments and identical to those used in the deterministic case. The optimization process is performed for a fixed number of 25 iterations (gradient evaluations). A lower value than used for the deterministic case was chosen to limit the computational cost; 100 simulations are now required to determine the expected objective function value for a proposed control update, whereas only 1 simulation is required in a deterministic setting. The results from experiments without and with time correlations between controls are shown in Figs. 10 and 11 respectively.

The results indicate that the performance of the different sampling methods in the robust optimization case is similar

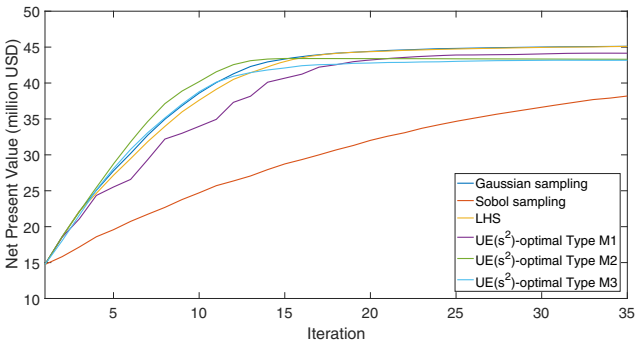

Fig. 9 NPV as a function of iteration for different sampling strategies. Gradients are estimated from Eq. 1 with $M = 100$ and perturbation smoothing

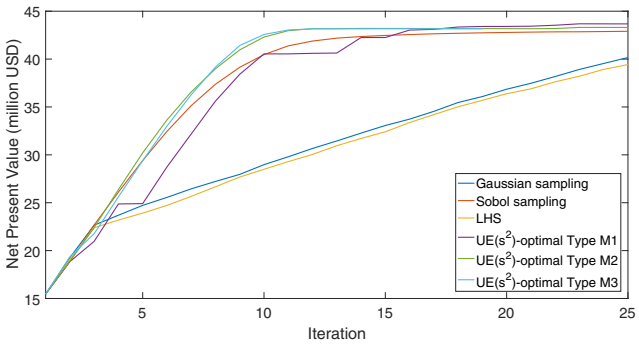

Fig. 10 Expected NPV evaluated over 100 model realizations as a function of iteration for different sampling strategies. Gradients are estimated from Eq. 2 with $M = N_r = 100$

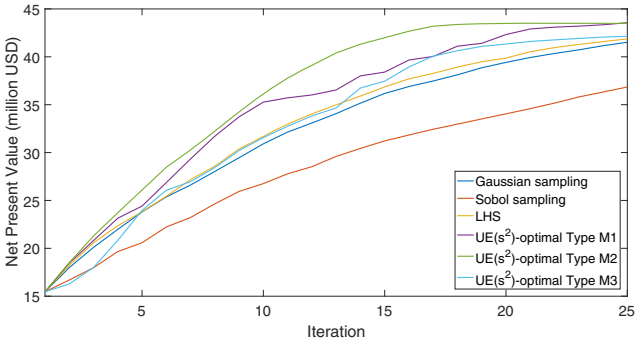

Fig. 11 Expected NPV evaluated over 100 model realizations as a function of iteration for different sampling strategies. Gradients are estimated from Eq. 2 with $M = N_r = 100$ and perturbation smoothing

to that in the deterministic case. The main differences are observed if time correlation is imposed on the samples. While in the deterministic setting all methods except Sobol performed similarly, in the setting with model uncertainty, $UE(s^2)$ designs of type M2 clearly perform better than all other methods. Sobol sampling still performs worse than all other methods. The use of time correlation leads to improved objective function values when Gaussian, LHS and $UE(s^2)$ sampling of type M1 is used, and to reduced objective function values when $UE(s^2)$ sampling of type M3 or Sobol sampling is used. The results for $UE(s^2)$-M2 are hardly affected.

The optimal control strategies obtained after 25 iterations are shown in Fig. 12 for all sampling strategies. A similar behavior can be observed as in the deterministic case. When using Sobol sampling, the water injection rate jumps between near-minimum and near-maximum values. This is close to what is known as a bang-bang strategy, which is an optimal strategy for certain linear problems and is characterized by solutions that attain only the minimum and maximum allowable control values. The solutions obtained with Gaussian sampling and LHS tend to vary around an intermediate average control value, while the $UE(s^2)$ solutions consistently suggest near-minimum injection rates. The solutions for this well are characteristic also for those of the other wells, with $UE(s^2)$-based injection rates mostly in the range of 10–30 m^3/day.

4.3 Sensitivity of results

The optimization experiments presented here were performed with the same constant perturbation size which is the standard approach in approximate gradient applications. It has been observed in experiments with different perturbation sizes [12] that smaller perturbations may be preferred during the later stages of the optimization process. One way to improve the quality of gradient estimates at different stages of the optimization process is to adjust the perturbation size [30]. This was found to be an effective solution when the traditional Gaussian sampling strategy is used. Here we investigate to what extend the sampling strategy affects the sensitivity to perturbation size. Experiments were performed with the Egg reservoir model with different sampling strategies for both deterministic and robust settings. Figure 13 shows robust optimization results for three different fixed perturbation sizes and all sampling strategies considered in this paper. The results suggest that the performance for $UE(s^2)$ designs of type M2 is much less sensitive

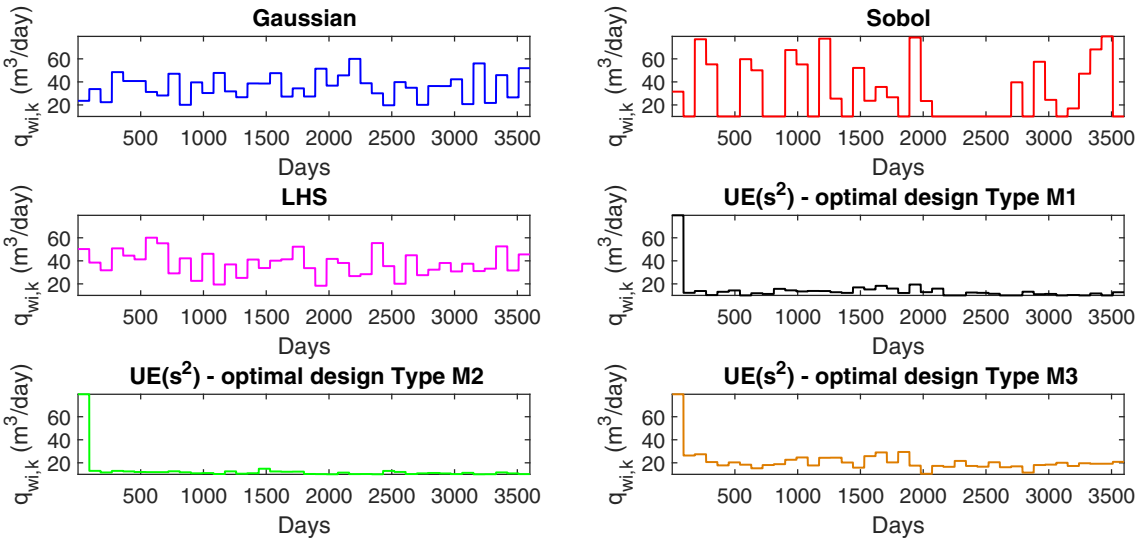

Fig. 12 Robust control strategy representing the injection rate of one of the injectors as a function of time at the final iteration (25) obtained with 6 different sampling strategies and $M = N_r = 100$

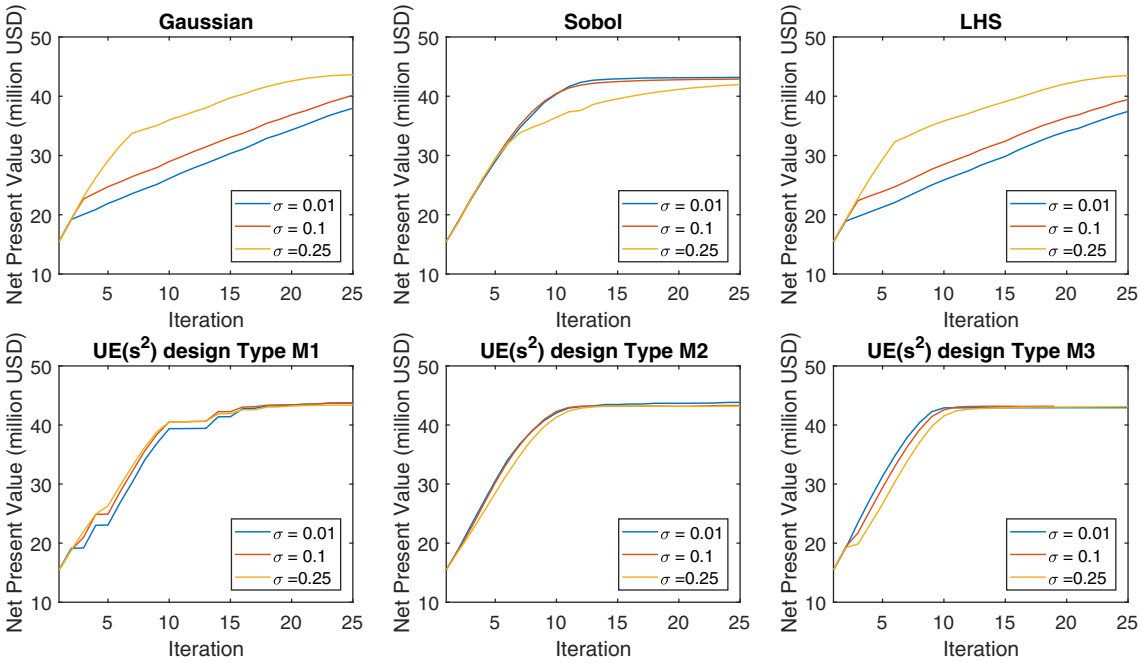

Fig. 13 Expected NPV evaluated over 100 model realizations as a function of iteration for three robust optimization experiments with different perturbation sizes (σ)

to the perturbation size than that for Gaussian and Sobol sampling or LHS designs.

Some of the considered sampling strategies, including Sobol sampling and $UE(s^2)$ designs of type M3, are deterministic and therefore will produce the same result each time. Strategies based on pseudo-random numbers (Gaussian, LHS), or on random selection of perturbations from a fixed set ($UE(s^2)$ designs of types M1 and M2) may produce different results for different random number generator seeds. In Fig. 14, we present robust optimization results obtained with $UE(s^2)$ designs of type M1 and type M2 for five different initial random seeds. The sensitivity of the gradient quality and convergence with respect to the initial seed is very large for $UE(s^2)$ designs of type M1, but almost negligible for designs of type M2. This is another benefit of the $UE(s^2)$-type M2 designs proposed here.

5 Conclusions

The standard practice of using Gaussian sampling to generate random perturbations for use in approximate gradient estimation procedures is compared against various alternative sampling strategies. The alternative strategies include two space-filling designs, namely Sobol sampling and LHS, based on low-discrepancy concepts as achieved by quasi-Monte Carlo approaches and stratification respectively. A second class of methods is based on the $E(s^2)$ near-orthogonality concept for supersaturated designs and D-optimal reduction of the generalized variance of the gradient estimate ($E(s^2)$-optimal designs). Two new variants of $E(s^2)$-optimal designs were proposed. The sampling strategies were applied to high-dimensional analytical test problem to evaluate their impact on the gradient quality.

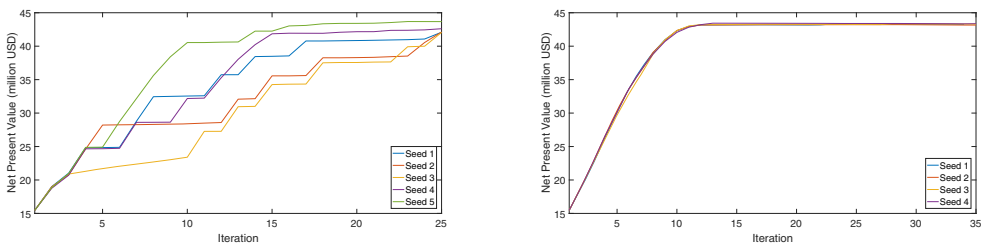

Fig. 14 Expected NPV evaluated over 100 model realizations as a function of iteration for 5 robust optimization experiments with different random seeds using M1 (left) and M2 (right) sampling designs

In a second example, they were applied to an oil reservoir case with realistic complexity in terms of number of controls and uncertainty in parameter values to test their impact on optimization performance. The main conclusions can be summarized as follows.

- Sobol sampling and $UE(s^2)$ designs outperform random sampling and stratified experimental designs in terms of gradient quality and convergence properties in all cases when no smoothing is performed on the samples prior to gradient estimation.
- When samples are smoothed over time, the performance of Sobol sampling strongly deteriorates.
- The sampling strategy is found to have a significant impact on the character of the resulting control strategy. Sobol sampling tends to produce highly dynamic strategies, while $UE(s^2)$ designs produce fairly smooth strategies, also when no smoothing is explicitly applied.
- The new $UE(s^2)$ design referred to here as M2 was observed to outperform the optimal supersaturated design method previously suggested (M1), as well as a third variant (M3), in terms of performance of the optimization and in terms of sensitivity to the perturbation size and initial random seed.
- $UE(s^2)$-optimal supersaturated designs perform well in all situations that were investigated for both deterministic and robust cases and are therefore recommended for gradient approximation schemes where the number of samples is less than the number of unknowns.

References

1. de Aguiar, P.F., Bourguignon, B., Khots, M.S., Massart, D.L., Phan-Than-Luu, R.: D-optimal designs. Chemom. Intell. Lab. Syst. **30**(2), 199–210 (1995)
2. Booth, K.H.V., Cox, D.R.: Some systematic supersaturated designs. Technometrics **4**, 489–495 (1962)
3. Brouwer, D.R., Jansen, J.D.: Dynamic optimization of waterflooding with smart wells using optimal control theory. SPE J. **9**(4), 391–402 (2004)
4. Caflisch, R.E.: Monte Carlo and quasi-Monte Carlo methods. Acta Numerica **7**, 1–49 (1998). https://doi.org/10.1017/S0962492900002804
5. Cavazzuti, M.: Optimization methods: from theory to design scientific and technological aspects in mechanics. Springer, ISBN: 978-3-642-31186-4 (2013)
6. Chen, Y.: Efficient ensemble based reservoir management/ PhD Thesis. University of Oklahoma, USA (2008)
7. Custodio, A., Vicente, L.: Using sampling and simplex derivatives in pattern search methods. SIAM J. Optim. **18**(2), 537–555 (2007)
8. Diego, G., Loyola, R., Pedergnana, M., Garcia, S.G.: Smart sampling and incremental function learning for very large high dimensional data. Neural Netw. **78**, 75–87 (2016)
9. Dixon, L.C.W., Mills, D.J.: Effect of rounding errors on the variable metric method. J. Optim. Theory Appl. **80**(1), 175–179 (1994)
10. Do, S., Reynolds, A.: Theoretical connections between optimization algorithms based on an approximate gradient. Comput. Geosci. **17**(6), 959–973 (2013). https://doi.org/10.1007/s10596-013-9368-9
11. Fonseca, R.M., Stordal, A.S., Leeuwenburgh, O., den Hof, P.M.J.V., Jansen, J.D.: Robust ensemble-based multi-objective optimization. In: Proceedings of the 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV), 8-11 September, Catania (2014)
12. Fonseca, R.M., Kahrobaei, S., Van Gastel, L.J.T., Leeuwenburgh, O., Jansen, J.D.: Quantification of the impact of ensemble size on the quality of an ensemble gradient using principles of hypothesis testing. Paper SPE 173236-MS presented at the SPE Reservoir Simulation Symposium, Houston, USA, 23-25 February (2015)
13. Fonseca, R.M., Leeuwenburgh, O., Van den Hof, P.M.J., Jansen, J.D.: Improving the ensemble optimization method through covariance matrix adaptation (cma-enopt). SPE J. **20**(1), 155–168 (2015). https://doi.org/10.2118/163657-PA
14. Fonseca, R.M., Chen, B., Jansen, J.D., Reynolds, A.: A stochastic simplex approximate gradient (stosag) for optimization under uncertainty. Int. J. Numer. Methods Eng. **109**(13), 1756–1776 (2016)
15. Gilmour, S.G.: Factor screening via supersaturated designs. In: Dean, A., Lewis, S. (eds.) Screening: methods for experimentation in industry, drug discovery and genetics, vol. 8, pp. 169–190. Springer, New York (2006)
16. Van den Hof, P.M.J., Jansen, J.D., Van Essen, G., Bosgra, O.H.: Model based control and optimization of large scale physical systems-challenges in reservoir engineering. Chinese Control and Decision Conference, June 17-19 2009. Guilin, China (2009)
17. Jansen, J.D., Bosgra, O.H., Van den Hof, P.M.J.: Model-based control of multiphase flow in subsurface oil reservoirs. J. Process. Control. **18**(9), 846–855 (2008)
18. Jansen, J.D., Fonseca, R.M., Kahrobaei, S.S., Siraj, M.M., Van Essen, G.M., Van den Hof, P.M.J.: The egg model (research note). http://data.4tu.nl/repository/uuid:916c86cd-3558-4672-829a-105c62985ab2, online accessed July 2017 (2013)
19. Jones, B., Majumdar, D.: Optimal supersaturated designs. J. Am. Stat. Assoc. **109**(508), 1592–1600 (2014). https://doi.org/10.1080/01621459.2014.938810
20. Kelly, C.T.: Detection and remediation of stagnation in the nelder-mead algorithm using a sufficient decrease condition. SIAM J Optim **10**(1), 43–55 (1999)
21. Lin, D.K.J.: A new class of supersaturated designs. Technometrics **35**, 28–31 (1993)
22. McKay, M., Beckman, R., Conover, W.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics **42**(1), 55–61 (1979)
23. Morokoff, W.J., Caflisch, R.E.: Quasi-Monte Carlo integration. Comput. Phys. **122**(2), 218–230 (1995)
24. Niederreiter, H.: Quasi-Monte Carlo methods and pseudo-random numbers. Bull. Am. Math. Soc. **84**(6), 957–1041 (1978)
25. Niederreiter, H.: Low-discrepancy and low-dispersion sequences. J. Number Theory **30**, 51–70 (1988)
26. Okano, H., Koda, M.: An optimization algorithm based on stochastic sensitivity analysis for noisy objective landscapes. SPE J. **79**(2), 245–252 (2003)

27. Owen, A.B.: Monte carlo variance of scrambled net quadrature. SIAM J. Numer. Anal. **34**(5), 1884–1910 (1997)
28. Owen, A.B.: Monte Carlo theory, methods and examples. http://statweb.stanford.edu/owen/mc (2013)
29. Paskov, S.H., Traub, J.: Faster evaluation of financial derivatives. J. Portfolio Manag. **22**, 113–120 (1995)
30. Raniolo, S., Dovera, L., Cominelli, A., Callegaro, C., Masserano, F.: 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 (2013)
31. Sarma, P., Chen, W.: Improved estimation of the stochastic gradient with quasi-Monte Carlo methods. In: Proc 14th European Conference on the Mathematics of Oil Recovery (ECMOR XIV), Catania, Italy, 8-11 September (2014)
32. Sarma, P., Chen, W., Aziz, K., Durlofsky, L.: Production optimization with adjoint models under non-linear control-state path inequality constraints. SPE Reserv. Eval. Eng., 326–339 (2008)
33. Spall, J.: Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. IEEE Trans. Autom. Control **37**(3), 332–341 (1992)
34. Stordal, A.S., Szklarz, S.P., Leeuwenburgh, O.: A theoretical look at ensemble-based optimization in reservoir management. Math. Geosci. **48**(4), 399–417 (2016). https://doi.org/10.1007/s11004-015-9598-6
35. Van Essen, G., Zandvliet, M., Van den Hof, P., Bosgra, O., Jansen, J.D.: Robust waterflooding optimization of multiple geological scenarios. SPE J. **14**(1), 202–210 (2009)
36. Wu, C.F.J.: Construction of supersaturated designs through partially aliased interactions. Biometrika **80**, 661–669 (1993)