Mathematical Geology, Vol. 34, No. 2, February 2002 (© 2002)

Optimization With the Gradual Deformation Method[1]

Mickaële Le Ravalec-Dupin[2] and Benoît Nœtinger[2]

Building reservoir models consistent with production data and prior geological knowledge is usually carried out through the minimization of an objective function. Such optimization problems are nonlinear and may be difficult to solve because they tend to be ill-posed and to involve many parameters. The gradual deformation technique was introduced recently to simplify these problems. Its main feature is the preservation of the spatial structure: perturbed realizations exhibit the same spatial variability as the starting ones. It is shown that optimizations based on gradual deformation converge exponentially to the global minimum, at least for linear problems. In addition, it appears that combining the gradual deformation parameterization with optimizations may remove step by step the structure preservation capability of the gradual deformation method. This bias is negligible when deformation is restricted to a few realization chains, but grows increasingly when the chain number tends to infinity. As in practice, optimization of reservoir models is limited to a small number of iterations with respect to the number of gridblocks, the spatial variability is preserved. Last, the optimization processes are implemented on the basis of the Levenberg–Marquardt method. Although the objective functions, written in terms of Gaussian white noises, are reduced to the data mismatch term, the conditional realization space can be properly sampled.

KEY WORDS: optimization, gradual deformation, convergence.

INTRODUCTION

History matching is an important problem in reservoir engineering. Given a flow model, it consists in approximating a reservoir by a grid of permeability or porosity values to explain observed production data as much as possible. The computation of the permeability or porosity values is performed through an optimization problem, that is by minimizing a nonlinear objective function.

The current literature identifies three main types of optimization approaches: gradient-based, enumerative, and randomized. Gradient-based methods have local scope. They seek for a local optimum in the neighborhood of the starting guess. Enumerative procedures compute the objective function at every point in the space, one at a time. Since many practical applications depend on a large

[1]Received 8 May 2000; accepted 11 January 2001.

[2]Institut Français du Pétrole, 1 et 4, avenue de Bois-Préau, 92852 Rueil-Malmaison Cedex, France; e-mail: mickaele.le-ravalec@ifp.fr or benoit.noetinger@ifp.fr

number of parameters, these methods are usually considered intractable. In the long run, pure random schemes cannot be expected to do better than enumerative approaches. They must be distinguished from the randomized methods such as simulated annealing (Farmer, 1992) or genetic algorithms (Holland, 1992) that use random choice as a tool in a directed search process. Unlike gradient-based methods, randomized approaches are theoretically capable to reach the global optimum, although they converge slowly.

In this paper, we focus on optimization problems specifically designed for the geostatistical treatment of permeability or porosity within reservoirs. A regain of interest in geostatistical inversion and practical needs have brought to fashion problems with a huge number of parameters (permeability or porosity values); discretization of reservoir models involves commonly more than 10^6 gridblocks. Additionally, prior geological knowledge about the reservoir structure is expressed as a geostatistical constraint that must be integrated into the conditioning process. Thus, conditional reservoir models have to be consistent with the observed production data, but also with the prior constraint. These issues motivated very specific developments in optimization. de Marsily and others (1984) elaborated the pilot point method involving a gradient-based search. Hu (2000a) introduced another geostatistical parameterization technique called the Gradual Deformation Method (GDM). As the pilot point method, the GDM can handle reservoir models with many parameters.

Our subject is concerned with the coupling of the gradual deformation technique with optimization processes. Such an approach is hybrid because it involves a random search crossed with gradient techniques. First, we give a brief recap of the GDM and its extension, the Enhanced GDM. The latter was developed to assess uncertainties in production forecasts. Then, we investigate whether and how fast optimizations based on the gradual deformation converge to the global optimum. In parallel, we analyze the changes in the spatial variability of realizations all along the optimization processes. Last, we consider a peculiar optimization approach whereby minimization is carried out through the Levenberg–Marquardt algorithm.

DEFINITION OF THE OBJECTIVE FUNCTION

Reservoir history matching consists in estimating a permeability or porosity grid, that is a reservoir model, from production data, and possibly prior geological knowledge. We focus on methods where the grid is considered as a realization of a random function.

Early approaches placed the emphasis on providing a good fit only between the simulated and the measured production data. Basically, a reservoir model is proposed as a starting guess. Then, a fluid flow is simulated to mimic the production

data. The suitability of the proposed model $\mathbf{y}$ is measured by an objective function expressed as

$$J(\mathbf{y}) = \sum_i w_i \left(d_{\mathrm{sim}_i} - d_{\mathrm{obs}_i}\right)^2 \tag{1}$$

where w_i are weighting coefficients, d_{obs_i} are the observable data, and d_{sim_i} are the simulated data. If g is the operator from the unconditional realization space to the data space, we have:

$$\mathbf{d}_{\mathrm{sim}} = g(\mathbf{y}) \tag{2}$$

After analyzing the computed data, the model is adjusted until it reproduces the observed data as much as possible. The actual reservoir model should be the minimizer of the objective function. However, such optimization problems tend to be ill-posed (Sun, 1994). First, an accurate solution may not exist when there are measurement errors or modeling errors. Second, several reservoir models may yield similar production data. Last, even if a solution exists and is unique, it is still inappropriate if it does not depend continuously on the input parameters. To circumvent these difficulties, extra information is added into the objective function. This process is called regularization (Neuman, 1973). The regularized objective function is written as

$$J(\mathbf{y}) = \sum_i w_i \left(d_{\mathrm{sim}_i} - d_{\mathrm{obs}_i}\right)^2 + \alpha \sum_i v_i \left(y_i - y_{0_i}\right)^2 \tag{3}$$

where $\mathbf{y}_0$ is the reservoir model described from prior geological knowledge, v_i are weighting coefficients, and α is a regularization coefficient. It penalizes departure from prior geological knowledge. When α tends to 0, the regularization term vanishes and the problem may be ill-posed. The solution is the same as the one derived from the basic objective function [Eq. (1)]. When α tends to infinity, the problem is well-posed, but the influence of the observed data on the solution becomes negligible. The regularized solution depends on α.

Inverse problem theory provides a general framework for incorporating prior information into the objective function (Tarantola, 1987):

$$J(\mathbf{y}) = \frac{1}{2}(g(\mathbf{y}) - \mathbf{d}_{\mathrm{obs}})^{\mathrm{t}}\mathbf{C}_{\mathrm{D}}^{-1}(g(\mathbf{y}) - \mathbf{d}_{\mathrm{obs}}) + \frac{1}{2}(\mathbf{y} - \mathbf{y}_0)^{\mathrm{t}}\mathbf{C}_{\mathrm{Y}}^{-1}(\mathbf{y} - \mathbf{y}_0) \tag{4}$$

The first term on the right-hand side, called the likelihood constraint, measures the mismatch between the simulated and the observed data. The second one, referred

as the prior constraint, evaluates the discrepancy between reservoir model $\mathbf{y}$ and prior reservoir model $\mathbf{y}_0$. Covariance matrix $\mathbf{C}_D$ quantifies the experimental and theoretical uncertainties while covariance matrix $\mathbf{C}_Y$ characterizes the uncertainties in $\mathbf{y}_0$. Minimizing Eq. (4) leads to a solution $\mathbf{y}$ so that $g(\mathbf{y})$ is close to $\mathbf{d}_{obs}$ and $\mathbf{y}$ is close to $\mathbf{y}_0$. The minimizer is the maximum likelihood point for the posterior probability density function. This function measures the likelihood of built reservoir models. It depends on the likelihood and the prior probability density functions.

Computing J may be difficult, because two covariance matrices must be inverted. $\mathbf{C}_D$ is assumed to be diagonal: obtaining its inverse is straightforward. However, determining $\mathbf{C}_Y^{-1}$ is all the more difficult since $\mathbf{y}$ is discretized over many gridblocks. Two classes of methods have been developed to overcome this difficulty, both of them aiming at reducing the number of parameters. The first ones boil down to algebraic decompositions of covariance matrix $\mathbf{C}_Y$ (Reynolds and others, 1995). The second ones mix more intimately geostatistics with inverse problem resulting in geostatistical parameterization techniques. This led first to the pilot point method (de Marsily and others, 1984). A gradient-based optimization scheme is applied to a few selected points, termed the pilot points. Then, the optimal perturbation determined for these points is propagated to the whole reservoir model by kriging. The pilot point method allows both for considering many parameters and for honoring the prior constraint. However, it is subject to numerical artifacts and it cannot depart easily from the starting guess. More recently, Hu (2000a) designed the GDM on which we focus in the following sections.

GRADUAL DEFORMATION METHOD

The GDM was initially developed for gradually changing Gaussian-related stochastic reservoirs models while preserving their spatial variability (Hu, 2000a). Then, it was extended to non-Gaussian reservoir models simulated from sequential indicator (Hu, Blanc, and Nœtinger, 1999) and boolean (Hu, 2000b) algorithms.

The simplest gradual deformation scheme consists in combining two independent Gaussian random functions Y_1 and Y_2 with mean y_0 and identical covariances:

$$[Y(\theta) - y_0] = [Y_1 - y_0]\cos(\theta) + [Y_2 - y_0]\sin(\theta) \tag{5}$$

This relation ensures that $Y(\theta)$ is also a random Gaussian function with the same mean and the same covariance as Y_1 and Y_2. Given two independent realizations $\mathbf{y}_1$ and $\mathbf{y}_2$ of Y_1 and Y_2, a continuous realization chain $\mathbf{y}$ can be described varying the deformation parameter θ:

$$[\mathbf{y}(\theta) - \mathbf{y}_0] = [\mathbf{y}_1 - \mathbf{y}_0]\cos(\theta) + [\mathbf{y}_2 - \mathbf{y}_0]\sin(\theta) \tag{6}$$

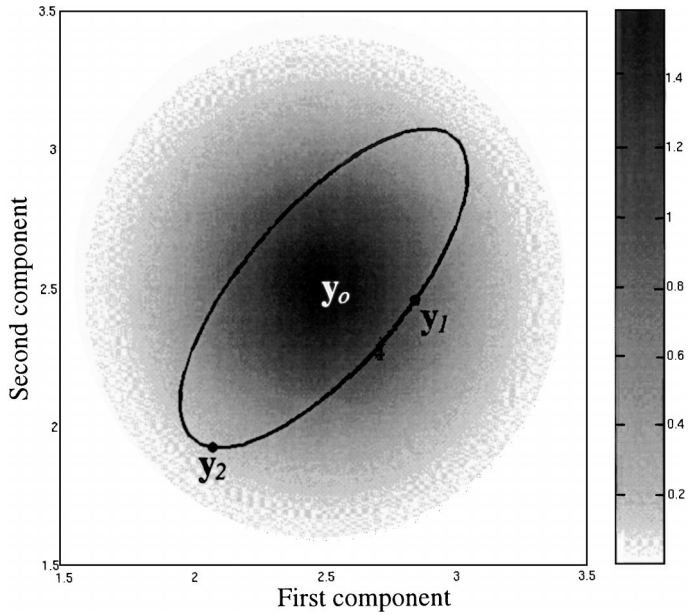

Figure 1. Realization chain built from two constituent realizations $\mathbf{y}_1$ and $\mathbf{y}_2$. These two primary realizations are constituted of two independent components drawn from N(2.5, 0.1). The center of the chain is mean realization.

As an example, the realization chain depicted in Figure 1 is derived from two starting realizations $\mathbf{y}_1$ and $\mathbf{y}_2$, both of them having two independent components randomly generated from the normal probability density function with mean 2.5 and variance 0.1.

The gradual deformation principle can be extended to more than two independent realizations. Let us consider $S + 1$ independent realizations with mean $\mathbf{y}_0$ and identical covariances, a realization chain is built from

$$[\mathbf{y}(\theta_i, i \in [1, S]) - \mathbf{y}_0] = \prod_{i=1}^{S} [\mathbf{y}_1 - \mathbf{y}_0] \cos(\theta_i) + \sum_{i=1}^{S} \sin(\theta_i) \prod_{j=i+1}^{S} \cos(\theta_j)[\mathbf{y}_{i+1} - \mathbf{y}_0] \quad (7)$$

where θ_i are the deformation parameters (Roggero and Hu, 1998). Again, $\mathbf{y}$ is characterized by mean $\mathbf{y}_0$ and the same covariance as the combined realizations.

The main advantages of the GDM are the decrease in the number of parameters and the preservation of the spatial variability while changing continuously the

reservoir model. Whatever the number of gridblocks, the whole reservoir model is modified by varying a few deformation parameters. Because of these properties, it is attractive to integrate the gradual parameterization technique into optimization processes. As reservoir models can be perturbed while being consistent with the prior constraint, Hu (2000a) suggested reducing the objective function to the data mismatch term [Eq. (1)]. It results in an optimization problem simpler to the one pointed out by Eq. (4). The basic idea is to calibrate a reservoir model **y** to production data by tuning the deformation parameters. A first realization chain is built from the initial guess and randomly drawn complementary realizations [Eq. (7)]. Then, a search process is implemented to determine the deformation parameters minimizing Eq. (1). At this point, the fit may not be good enough. A new starting realization chain is constructed from the optimal realization identified at the end of the previous search process. Sometimes, it may be impossible to determine deformation parameters decreasing the objective function. In these conditions, the starting guess is unchanged. Then, new complementary realizations are drawn randomly. A search process is run again to investigate the resulting realization chain and minimize further the objective function. This approach is iterated until a satisfactory matching is achieved.

SAMPLING THE CONDITIONAL REALIZATION SPACE WITH THE GDM

Because of the simplicity of the previously described gradual deformation scheme, several optimizations can be run to get as many constrained reservoir models. All of them can be used to infer the posterior probability density function and thus to assess uncertainties. However, to get reliable estimates, we must ensure that the set of constrained reservoir models reflects properly the posterior probability density function. Le Ravalec-Dupin, Hu, and Nœtinger (2000) focused on a few numerical experiments following the randomized sampling method (Oliver, He, and Reynolds, 1996). They showed that GDM-based optimizations, when performed as explained above, do not correctly sample the constrained realization space. To alleviate this drawback, they added the prior constraint into the objective function [Eq. (4)] although the gradual deformation parameterization was applied during the whole optimization process. But then, the inverse of covariance matrix $\mathbf{C}_Y$ had to be computed. They applied the following variable change referring to the moving average method (Le Ravalec, Nœtinger, and Hu, 2000; Oliver, 1995):

$$y = y_0 + f * z \tag{8}$$

y is a Gaussian random field with mean y_0 and covariance C_Y. Function f results from the decomposition of C_Y as $C_Y = f * f \cdot z$ is a field of independent

normal deviates with mean zero and variance one, that is a Gaussian white noise. Substituting Eq. (8) into Eq. (4) results in the following expression for the objective function:

$$J(\mathbf{z}) = \frac{1}{2}(\mathsf{g}(\mathbf{z}) - \mathbf{d}_{\text{obs}})^{\text{t}}\mathbf{C}_{\text{D}}^{-1}(\mathsf{g}(\mathbf{z}) - \mathbf{d}_{\text{obs}}) + \frac{1}{2}\mathbf{z}^{\text{t}}\mathbf{z} \tag{9}$$

g is the operator mapping the Gaussian white noise space into the data space. Clearly, accounting for the prior constraint in the objective function gets much easier.

This approach, called the Enhanced GDM, still allows for decreasing the number of parameters and ensures consistency with the prior constraint. In addition, full formulation of the objective function renders the problem well-posed and leads to proper sampling of the constrained realization space (Le Ravalec-Dupin, Hu, and Nœtinger, 2000).

Basically, the sampling process is as follows. First, an unconditional observation data vector $\mathbf{d}_{\text{us}}$ and an unconditional Gaussian white noise $\mathbf{z}_{\text{us}}$ are randomly drawn. Then, they are substituted into Eq. (9) in place of $\mathbf{d}_{\text{obs}}$ and the maximum of the prior probability density function, that is, $\mathbf{0}$.

$$J(\mathbf{z}) = \frac{1}{2}(\mathsf{g}(\mathbf{z}) - \mathbf{d}_{\text{us}})^{\text{t}}\mathbf{C}_{\text{D}}^{-1}(\mathsf{g}(\mathbf{z}) - \mathbf{d}_{\text{us}}) + \frac{1}{2}(\mathbf{z} - \mathbf{z}_{\text{us}})^{\text{t}}(\mathbf{z} - \mathbf{z}_{\text{us}}) \tag{10}$$

Last, a conditional Gaussian white noise is obtained by minimizing this intermediate objective function. To generate n conditional realizations, the process is repeated n times.

CONVERGENCE RATES

In this section, we investigate the convergence of the objective function when using the gradual deformation parameterization. Such optimization processes mix a random search scheme with gradient-based computations. They exploit the global perspective of the Monte Carlo sampling as well as the local scope of gradient techniques. In a sense, the Monte Carlo sampling finds the holes and the gradient techniques go to their bottoms.

In the random approach, the global minimum is supposedly reached provided the search is long enough. However, for most practical problems, this is too long. Our purpose is to point out the benefits due to the combination with gradient-based computations.

Prior Constraint

We consider a basic problem characterized by the following objective function:

$$J = \frac{1}{2}\mathbf{z}^{\mathrm{t}}\mathbf{z} \tag{11}$$

where $\mathbf{z}$ is a Gaussian white noise. This objective function is no more than the prior constraint as expressed in Eq. (9).

We aimed at minimizing Eq. (11) using the simplest gradual deformation scheme, that is by combining sequentially two sets of independent normal deviates. At step $(n+1)$, $\mathbf{z}^{(n+1)}$ can be expressed as a function of $\mathbf{z}^{(n)}$ and $\mathbf{z}_{\mathrm{a}}$:

$$\mathbf{z}_2^{(n+1)}(\theta) = \mathbf{z}_2^{(n)}\cos(\theta) + \mathbf{z}_{\mathrm{a}}\sin(\theta) \tag{12}$$

$\mathbf{z}^{(n)}$ is the optimal Gaussian white noise identified when exploring the nth realization chain; it is also the starting realization for the $(n+1)$th chain. $\mathbf{z}_{\mathrm{a}}$ is a randomly drawn Gaussian white noise. Subscript 2 indicates that the gradual deformation process applies to two constituent Gaussian white noises.

The gradual deformation principle ensures that any $\mathbf{z}^{(n+1)}$ resulting from Eq. (12) is a Gaussian white noise provided $\mathbf{z}^{(n)}$ and $\mathbf{z}_{\mathrm{a}}$ are independent Gaussian white noises. In other words, the mean of the objective function should equal $M/2$, where M is the number of components of $\mathbf{z}$. However, the minimal value for the objective function is clearly zero. The question addressed here is whether this minimal value is reached when performing optimization on the basis of the gradual deformation parameterization.

We calculate the objective function against the number of explored realization chains. At step $(n+1)$, the objective function is written as

$$J_2^{(n+1)}(\theta) = \frac{1}{2}\left[\mathbf{z}_2^{(n+1)}(\theta)\right]^{\mathrm{t}}\mathbf{z}_2^{(n+1)}(\theta). \tag{13}$$

The optimization process relies on the estimation of the optimal θ values minimizing the objective function for the successive realization chains. Integrating Eq. (11) into Eq. (12) leads to

$$J_2^{(n+1)}(\theta) = r\cos(2\theta + \varphi) + C \tag{14}$$

with

$$r^2 = A^2 + B^2$$
$$\cos(\varphi) = A/r$$

$$\begin{aligned}
\sin(\varphi) &= -B/r \\
A &= (1/4)\left[\mathbf{z}_2^{(n)^{\mathrm{t}}}\mathbf{z}_2^{(n)} - \mathbf{z}_{\mathrm{a}}^{\mathrm{t}}\mathbf{z}_{\mathrm{a}}\right] \\
B &= (1/2)\left[\mathbf{z}_{\mathrm{a}}^{\mathrm{t}}\mathbf{z}_2^{(n)}\right] \\
C &= (1/4)\left[\mathbf{z}_2^{(n)^{\mathrm{t}}}\mathbf{z}_2^{(n)} + \mathbf{z}_{\mathrm{a}}^{\mathrm{t}}\mathbf{z}_{\mathrm{a}}\right]
\end{aligned} \tag{15}$$

At step $(n+1)$, the smallest value of the objective function is $C - r$. As $\mathbf{z}_2^{(n)^{\mathrm{t}}}\mathbf{z}_2^{(n)} = 2J_2^{(n)}$ and $\langle \mathbf{z}_{\mathrm{a}}^{\mathrm{t}}\mathbf{z}_{\mathrm{a}} \rangle = M$, we obtain $\langle C \rangle = (1/4)(2J_2^{(n)} + M)$. Last, it can be shown that $\langle (\mathbf{z}_{\mathrm{a}}^{\mathrm{t}}\mathbf{z}_2^{(n)})^2 \rangle = 2J_2^{(n)}$. Thus, the mean value for the minimum of the objective function at step $(n+1)$ is

$$\left\langle J_2^{(n+1)} \right\rangle = \frac{J_2^{(n)}}{2} + \frac{M}{4} - \sqrt{\frac{1}{16}\left(2J_2^{(n)} - M\right)^2 + \frac{J_2^{(n)}}{2}} \tag{16}$$

As far as M is finite, the resulting sequence is decreasing, monotonous, and positive. It can be easily shown that it converges to 0. Figure 2 compares this sequence with experimental computations performed for given starting $\mathbf{z}$ realizations. The experimental results are derived from the minimization of the objective function for successively explored realization chains. The experimental results (Fig. 2, left) are in a good agreement with the theoretical ones (Fig. 2, right). The objective function computed for the successively generated Gaussian white noises $\mathbf{z}_{\mathrm{a}}$ is also plotted in Figure 2 (left). The mean value for this objective function is $M/2$. None

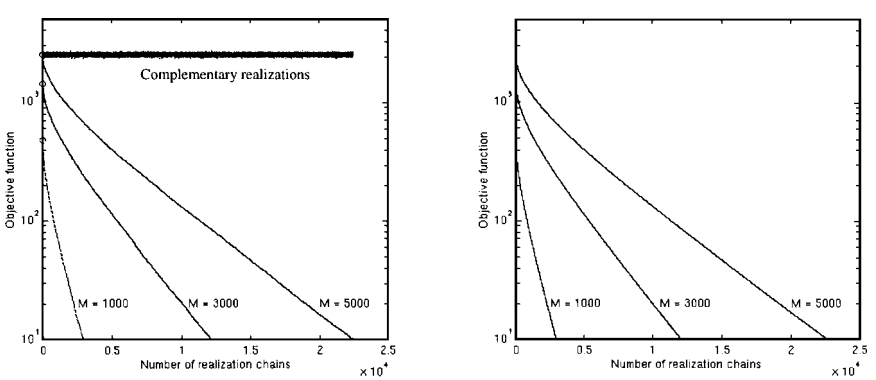

Figure 2. Objective function against the number of investigated realization chains for Gaussian white noises with M components. Comparison of the experimental calculations (left) with the theoretical ones (right). Objective function for the complementary realization is added in the left plot (points on the top).

of the drawn $\mathbf{z}_a$ produces a strong decrease in the objective function. The iterative objective function decrease is fully driven by the gradual combination process. Last, because the objective function converges to 0, Eq. (14) and (15) show that θ tends toward 0 as the number of investigated realization chains increases.

The asymptotic behavior of the objective function is characterized assuming that the objective function gets much smaller than M at step n:

$$J_2^{(n+1)} \approx J_2^{(n)} \left(1 - \frac{1}{M}\right) \tag{17}$$

This optimization technique exhibits an exponential convergence rate:

$$J_2 \propto \exp\left(-\frac{n}{M}\right) \tag{18}$$

It was found experimentally that the convergence rate is speeded up when the reservoir is divided into subregions (Le Ravalec-Dupin and others, 2000) and when the gradual deformation process is based upon the combination of more than two realizations (Roggero and Hu, 1998). In the first case, a deformation parameter θ is attributed to every subregion. The improvement results from the decrease of components associated to a single deformation parameter. If the reservoir model is divided into S subregions with the same number of gridblocks, the convergence rate becomes

$$J_2 \propto \exp\left(-\frac{nS}{M}\right) \tag{19}$$

Let us investigate more precisely the second deformation scheme. We consider the same problem as above [Eq. (11)], but the gradual deformation process is developed for three Gaussian white noises instead of two. Equation (12) is rewritten as

$$\begin{aligned} \mathbf{z}_3^{(n+1)}(\theta_a, \theta_b) &= \mathbf{z}_3^{(n)} \cos(\theta_a)\cos(\theta_b) + \mathbf{z}_a \sin(\theta_a)\cos(\theta_b) + \mathbf{z}_b \sin(\theta_b) \\ &= \left[\mathbf{z}_3^{(n)} \cos(\theta_a) + \mathbf{z}_a \sin(\theta_a)\right]\cos(\theta_b) + \mathbf{z}_b \sin(\theta_b) \end{aligned} \tag{20}$$

Subscript 3 stands for "combining three realizations." The realization chain depends on three realizations, that is two deformation parameters denoted θ_a and θ_b. $\mathbf{z}^{(n)}$ is the starting Gaussian white noise at step $(n+1)$ while $\mathbf{z}_a$ and $\mathbf{z}_b$ are two randomly drawn Gaussian white noises. The first bracketed term on the right-hand side is equivalent to Eq. (12). It is convenient to define

$$\mathbf{z}_3^{(n+1)}(\theta_a) = \mathbf{z}_3^{(n)} \cos(\theta_a) + \mathbf{z}_a \sin(\theta_a) \tag{21}$$

Introducing Eq. (21) into Eq. (20) provides

$$\mathbf{z}_3^{(n+1)}(\theta_a, \theta_b) = \mathbf{z}_2^{(n+1)}(\theta_a)\cos(\theta_b) + \mathbf{z}_b \sin(\theta_b) \tag{22}$$

Again, this expression is similar to Eq. (12), but the first $\mathbf{z}$ vector depends now on deformation parameter θ_a. An approach similar to the one discussed above leads to

$$J_3^{(n+1)}(\theta_a, \theta_b) = r(\theta_a)\cos(2\theta_b + \varphi(\theta_a)) + C(\theta_a) \tag{23}$$

Again, the mean objective function expresses as

$$\left\langle J_3^{(n+1)} \right\rangle = \frac{J_2^{(n+1)}}{2} + \frac{M}{4} - \sqrt{\frac{1}{16}\left(2J_2^{(n+1)} - M\right)^2 + \frac{J_2^{(n+1)}}{2}} \tag{24}$$

This expression is close to Eq. (16), but the sequence is defined relatively to the number of combined realizations. We verify that the greater the number of combined realizations, the smaller the objective function (Fig. 3).

The convergence rate is strongly influenced by the number of combined realizations. If the gradual deformation process involves $(S + 1)$ realizations, that is

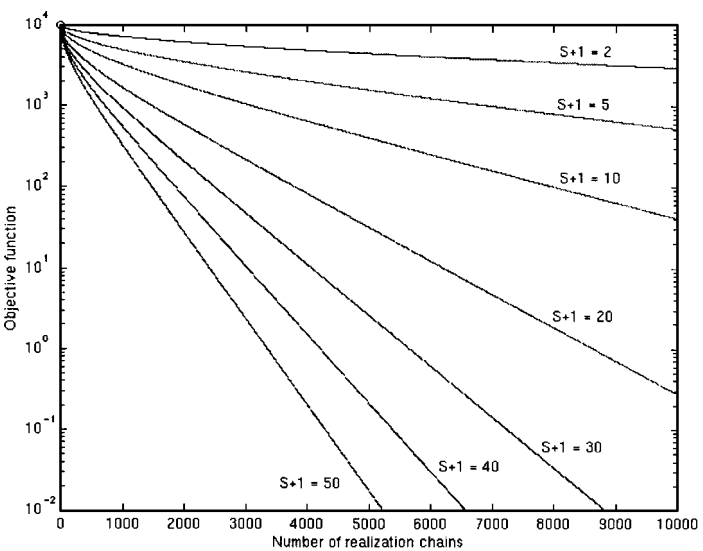

Figure 3. Objective function against the number of investigated realization chains for different gradual deformation schemes. $S + 1$ is the number of realizations that are combined together to form a chain.

S deformation parameters, the convergence rate is given by

$$J \propto \exp\left(-\frac{nS}{M}\right) \tag{25}$$

Combining $(S+1)$ realizations or dividing the reservoir into S subregions is equivalent in terms of convergence.

General Formulation

The general problem consists in minimizing Eq. (9). For simplicity, the operator g is assumed to be linear. In such condition, it is shown heuristically that

$$\left\langle J_2^{(n+1)}\right\rangle - \left\langle J_2^{(n)}\right\rangle = -\frac{1}{2}\frac{\left\langle \left(\mathbf{z}_2^{(n)} - \tilde{\mathbf{z}}\right)^{\mathrm{t}}\mathbf{C}^{-1}\mathbf{C}^{-1}\left(\mathbf{z}_2^{(n)} - \tilde{\mathbf{z}}\right)\right\rangle}{\left\langle \mathrm{Tr}(\mathbf{C}^{-1})\right\rangle} \tag{26}$$

$\tilde{\mathbf{z}}$ is the minimizer of Eq. (9) and $\mathbf{C}$ is the covariance matrix describing uncertainties in conditional reservoir models. It is given by (Tarantola, 1987)

$$\mathbf{C}^{-1} = \mathbf{G}^{\mathrm{t}}\mathbf{C}_{\mathrm{D}}^{-1}\mathbf{G} + \mathbf{I} \tag{27}$$

where sensitivity matrix $\mathbf{G}$ consists of the derivatives of g and $\mathbf{I}$ is the identity matrix. As $(\mathbf{C}^{-1}\mathbf{C}^{-1}) \leq \mathbf{C}^{-1}$, the following convergence rate can be pointed out:

$$J_2 \propto \exp\left(-\frac{n}{\left\langle \mathrm{Tr}(\mathbf{C}^{-1})\right\rangle}\right) \tag{28}$$

$\mathrm{Tr}(\mathbf{C}^{-1})$ reduces to M when the objective function equals the prior constraint term [Eq. (11)].

DISCUSSION

Integrating the gradual deformation into optimization procedures leads to hybrid optimization methods. They imply random searches because of the randomly drawn complementary realizations. However, those ones are not generated anyhow: they depend on prior geological information. Thus, the resulting realization chains explore a space with a priori highly suitable realizations. As investigation is performed randomly, the global minimum can be achieved provided the search is carried for a sufficiently longer time. However, we showed that such optimization processes converge to the global minimum at least for linear problems. This improvement results from the combination of random search directions with gradient based computations. For every realization chain, optimal deformation parameters

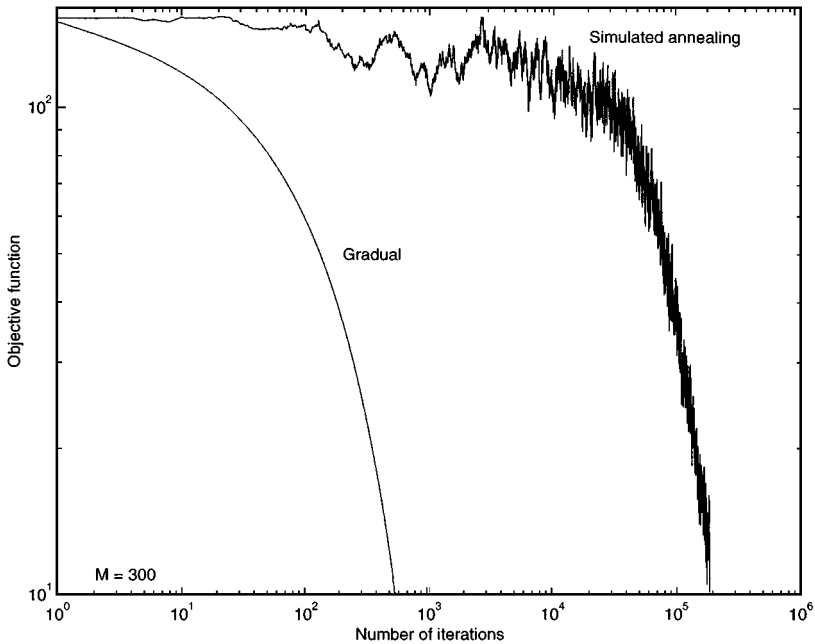

Figure 4. Evolution of the objective function against the number of investigated realizations when optimization is performed on the basis of the GDM and simulated annealing. Gaussian white noises are constituted of 300 components.

are determined using gradients, making the search process more efficient. This result is shown in Figure 2 (left). Over all the complementary realizations, none could significantly reduce the objective function. However, a smooth decrease was observed for the objective function: it was fully driven by the optimal combination computed for the successive chains. For comparison, Eq. (11) is also minimized on the basis of simulated annealing that is a randomized technique without any gradient. Figure 4 shows that the objective function reaches the value 10 in about 500 iterations with a gradual deformation based optimization and in more than 100,000 iterations with simulated annealing.

Besides, the studied example [Eq. (10)] point out that gradual deformation based optimizations can achieve global minima for linear problems. It also raises questions about the preservation of the spatial variability. A primary idea was that realizations built from gradual deformation exhibit the same spatial variability as the independent constituent realizations. Thus, combining Gaussian white noises should yield a Gaussian white noise. However, the described optimization process converged to a zero realization whose variability differs a lot from the expected one. Similarly, we matched a onedimensional Gaussian covariance realization by combining successively exponential covariance realizations (Fig. 5). Again, their

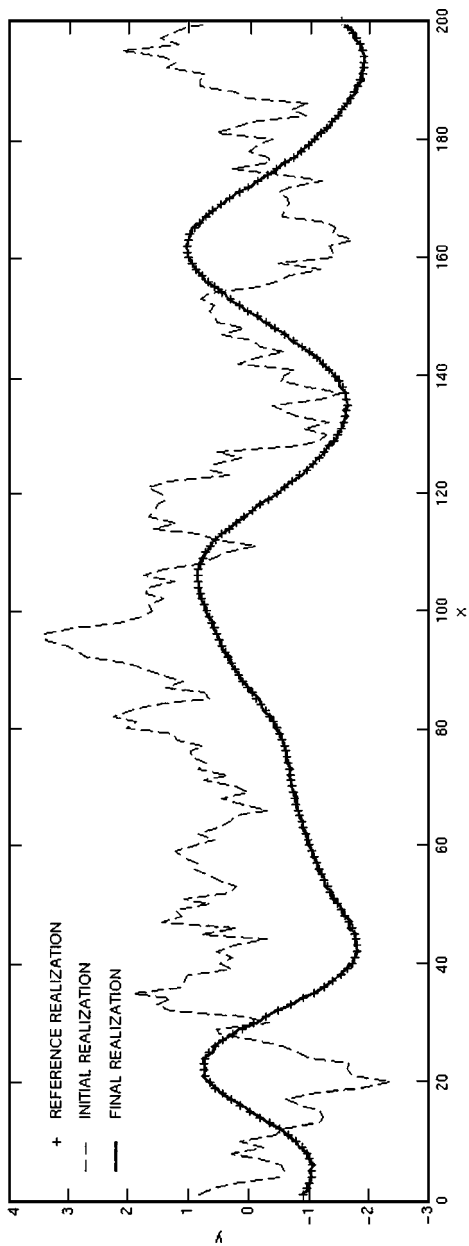

Figure 5. Matching of a Gaussian covariance realization combining sequentially exponential covariance realizations. The combination is implemented with respect to the gradual deformation method.

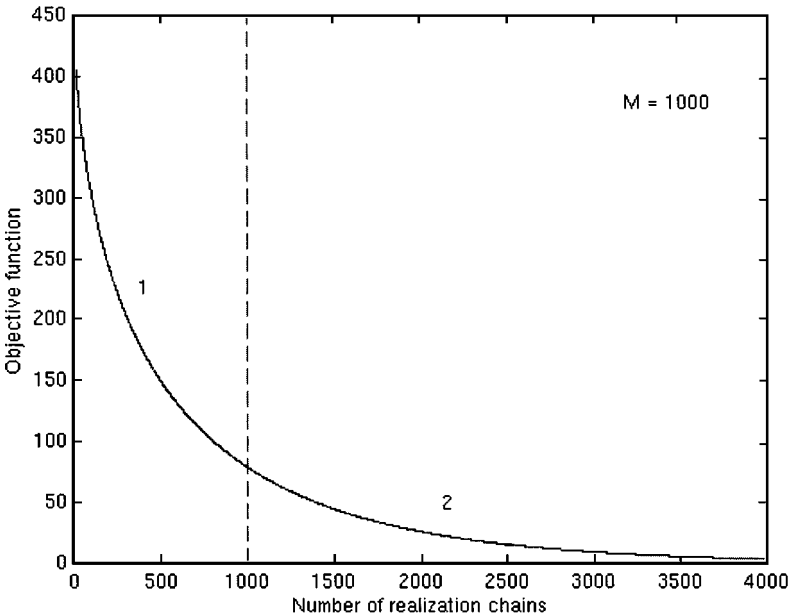

Figure 6. Behaviors displayed by the objective function against the number of investigated realization chains.

spatial structures are very different. Gaussian covariance realizations are smooth while exponential covariance realizations look more erratic. As can be seen in Figure 5, the final exponential covariance realization could reproduce the spatial structure of the Gaussian covariance realization.

Strictly speaking, optimizations based on the gradual deformation method do not allow for preserving the spatial structure, but this remark deserves some details. The discussed method can converge to any kind of minimizer provided the number n of investigated realization chains tends to infinity and the number of components M of the realization is finite [Eq. (16)]. As shown in Figure 6, the objective function displays two distinct behaviors. For $n > M$, the objective function becomes negligible with respect to M and decreases exponentially. The spatial structure is dictated by the minimizer. If its spatial variability is not the same as the prior one, the iterative gradual deformation process will not preserve the prior structure. For $n < M$, the objective function decreases faster, but the spatial structure stays close to the prior one.

These results may look dubious since the basic principle for gradual deformation is the spatial structure preservation. As stated above, Eq. (6) ensures that $\mathbf{y}(\theta)$ has the same mean and covariance as $\mathbf{y}_1$ and $\mathbf{y}_2$ whatever θ, provided $\mathbf{y}_1$ and $\mathbf{y}_2$ are independent. In practice, realizations are finite and never perfect.

Thus, they do not have the same mean, the same covariance and above all, they are never independent. These numerical artifacts are negligible for small n, but grow increasingly with n. They finally yield to departure from the prior spatial structure when the minimizer is not consistent with the prior spatial variability. Equation (16) shows that the spatial structure can be modified except when M is infinite.

These results do not mean that the gradual deformation parameterization should be avoided. Indeed, it still allows for reducing strongly the number of parameters to be optimized and for investigating successively chains constituted of a priori highly suitable realizations. In addition, we observed experimentally that for reservoir optimizations, the number of investigated realization chains is much smaller than the number of gridblocks. In such conditions, it may be considered that optimizations based on the gradual deformation parameterization preserve the spatial variability of the starting reservoir model.

Last, we emphasized that expressing the objective function against $\mathbf{z}$ instead of $\mathbf{y}$ is suitable [Eq. (9)]. The prior constraint becomes simpler because the covariance matrix for Gaussian white noises $\mathbf{z}$ is the identity matrix. This formulation also conciliates two common ideas in reservoir optimization. The prior term may be regarded as an objective the same way as the data mismatch term, but also as a regularization term. The Taylor expansion of the objective function J around realization $\mathbf{z}_i$ is:

$$J(\mathbf{z}_i + \Delta\mathbf{z}) = J(\mathbf{z}_i) + \nabla J^{\mathrm{t}} \Delta\mathbf{z} + \frac{1}{2} \Delta\mathbf{z}^{\mathrm{t}} \mathbf{H} \Delta\mathbf{z} \tag{29}$$

Higher order terms are ignored. Gradient vector ∇J and Hessian matrix $\mathbf{H}$ are evaluated at point $\mathbf{z}_i$. Differentiating this equation leads to an iterative determination of the minimizer (Sun, 1994):

$$\mathbf{z}_{n+1} = \mathbf{z}_n - \mathbf{H}_n^{-1} \nabla J_n \tag{30}$$

Subscript n indicates that vectors or matrices are calculated at point $\mathbf{z}_n$. If Eq. (9) was restricted to the data mismatch term, ∇J and $\mathbf{H}$ would equal:

$$\begin{aligned} \nabla J_n &= \mathbf{G}_n^{\mathrm{t}} \mathbf{C}_{\mathrm{D}}^{-1} (\mathbf{G}_n \mathbf{z}_n - \mathbf{d}_{\mathrm{obs}}) \\ \mathbf{H}_n &= \mathbf{G}_n^{\mathrm{t}} \mathbf{C}_{\mathrm{D}}^{-1} \mathbf{G}_n \end{aligned} \tag{31}$$

where $\mathbf{G}_n$ are the derivatives of g estimated at point $\mathbf{z}_n$. The last equality holds because second-order terms are assumed to be very small when $\mathbf{z}_n$ is not too far

from the minimizer. In such conditions, Eq. (30) is rewritten as:

$$\mathbf{z}_{n+1} = \mathbf{z}_n - \left(\mathbf{G}_n^{\mathrm{t}} \mathbf{C}_{\mathrm{D}}^{-1} \mathbf{G}_n\right)^{-1} \nabla J_n \tag{32}$$

The resulting sequence is known as the Gauss–Newton algorithm. When $\mathbf{H}_n$ is near singular, the Gauss–Newton direction cannot be calculated. The Levenberg–Marquardt method avoids this difficulty. The previous relation is turned into

$$\mathbf{z}_{n+1} = \mathbf{z}_n - \left(\mathbf{G}_n^{\mathrm{t}} C_{\mathrm{D}}^{-1} \mathbf{G}_n + \alpha \mathbf{I}\right)^{-1} \nabla J_n \tag{33}$$

where α is a penalty coefficient and $\mathbf{I}$ is the identity matrix. If α is set to 1, Eq. (33) is just the same as the one we would have obtained at once if we had considered the whole objective function given by Eq. (9). In these conditions, there is no need to include explicitly the prior constraint term into the objective function as far as the optimization process is implemented with the Levenberg–Marquardt method. It is appropriate because J is expressed as a function of the Gaussian white noise $\mathbf{z}$ instead of the realization $\mathbf{y}$.

CONCLUSION

We showed that at least for linear problems, optimizations based on gradual deformation converge exponentially to the global minimum. This result cannot be extended to nonlinear problems. However, if the search is iterated long enough, the method is theoretically capable to achieve the global minimum. Actually, gradual deformation based optimizations take advantage of the crossing of random search with gradient computations. Random search allows for investigating new domains with minima and for going out of them when the minima are local while gradients lead to the minimum in a given domain.

Additionally, we pointed out that optimizations based on the GDM do not preserve the spatial structure. When the number of investigated realization chains is small compared to the number of components of the reservoir model as it is usually for reservoir optimization, the prior spatial structure is approximately kept, although fluctuations can occur. For greater chain numbers, changes may be more significant: they depend on the spatial structure of the minimizer. This unexpected behavior depends on the realizations used to describe reservoir models. As they are finite and not perfect, they introduce a bias yielding to departure from the prior spatial structure.

Last, to ensure the proper sampling of the posterior probability density function, Le Ravalec-Dupin, Hu, and Nœtinger (2000) suggested adding the prior constraint into the objective function. When running the optimization process on the basis of the Levenberg–Marquardt algorithm, this additional term can be canceled

as far as the objective function depends not on the reservoir model itself, but on its underlying Gaussian white noise.

ACKNOWLEDGMENTS

The authors thank the Institut Français du Pétrole for permission to publish their results and L. Y. Hu for motivating discussions.

REFERENCES

de Marsily, G., Lavedan, C., Boucher, M., and Fasanino, G., 1984, Interpretation of interference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, *in* Verly, G., David, M., Journel, A. G., and Maréchal, A., eds., Geostatistics for natural resources characterization: Reidel, Dordrecht, The Netherlands, v. 2, p. 831–849.

Farmer, C. L., 1992, Numerical rocks, *in* King, P. R., ed., Mathematics of oil recovery: Oxford University Press, Oxford, UK, p. 437–447.

Holland, J. H., 1992, Adaptation in natural and artificial systems: Ann Harbor, The University of Michigan Press, Cambridge, MA, 211 p.

Hu, L. Y., 2000a, Gradual deformation and iterative calibration of Gaussian-related stochastic models: Math. Geol., v. 32, no. 1, p. 87–108.

Hu, L. Y., 2000b, Gradual deformation of non-Gaussian stochastic simulations: 6th Geostatistics Congress, Cape Town, South Africa, 10–14 April.

Hu, L. Y., Blanc, G., and Nœtinger, B., 1999, Gradual deformation and iterative calibration of sequential stochastic simulations: S. J. Lippard, A. Naess, and R. Sinding–Larsen, Proceedings of IAMG 1995, Trondheim, Norway, 6–11 Aug., 1999, vol. 2, p. 493–498.

Le Ravalec, M., Nœtinger, B., and Hu, L.-Y., 2000, The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning Gaussian simulations: Math. Geol., v. 32, no. 6, p. 701–723 .

Le Ravalec-Dupin, M., Hu, L.-Y., and Nœtinger, B., 2000, Sampling the conditional realization space using the gradual deformation method: 6th Geostatistics Congress, Cape Town, South Africa, 10–14 April.

Le Ravalec-Dupin, M., Nœtinger, B., Hu, L.-Y., and Blanc, G., 2000, Conditioning to dynamic data: An improved zonation approach: Pet. Geosci., vol. 7, p. S9–S16, 2001.

Neuman, S. P., 1973, Calibration of distributed parameter groundwater flow models viewed as a multiobjective decision process under uncertainty: Water Resour. Res., v. 9, no. 4, p. 1006–1021.

Oliver, D. S., 1995, Moving averages for Gaussian simulation in two and three dimensions: Math. Geol., v. 27, no. 8, p. 939–960.

Oliver, D. S., He, N., and Reynolds, A. C., 1996, Conditioning permeability fields to pressure data: 5th ECMOR, Leoben, Austria, 3–6 Sept., p. 259–268.

Reynolds, A. C., He, N., Chu, L., and Oliver, D. S., 1995, Reparameterization techniques for generating reservoir descriptions conditioned to variograms and well-test pressure data: SPE annual technical conference and exhibition, Dallas, TX, 22–25 Oct., SPE 30588, p. 609–624.

Roggero, F., and Hu, L.-Y., 1998, Gradual deformation of continuous geostatistical models for history matching: SPE annual technical conference and exhibition, New Orleans, LA, SPE 49004.

Sun, N.-Z., 1994, Inverse problems in groundwater modeling: Kluwer, Dordrecht, The Netherlands, 337 p.

Tarantola, A., 1987, Inverse problem theory—Methods for data fitting and model parameter estimation: Elsevier, Amsterdam, The Netherlands, 613 p.