

## Chapter 2

# From a Mathematical Point of View

Our objective is to simulate the facies at a set  $T$  of  $N$  target points given a data set  $D$ . The target is usually a large 2D or 3D grid, often with several hundred thousand nodes. In petroleum, the data points come from wells interpreted in lithofacies typically each foot whereas in mining studies data are core samples from drill-holes. In both cases there are usually a large number of data points too.

Let  $x$  and  $y$  denote the target points and the data locations respectively. Suppose that there are  $n_F$  facies. We model the facies via their indicators, that is, functions which take the value 1 at point  $x$  if  $x$  belongs to facies  $F_i$  and 0 otherwise. To simulate the facies, we have to sample from the conditional distribution:

$$P(1_{F_i}(x), x \in T | 1_{F_i}(y), i \in \{1, \dots, n_F\}, y \in D).$$

The two main difficulties are the availability of models for the conditional or even the unconditional distribution, and the dimension of this distribution. One classical way of solving these problems is to use indicator simulations based on kriging. But this approach has three drawbacks:

- Indicator kriging only gives a poor approximation of the conditional expectation.
- Indicator variograms and cross variograms are extremely difficult to model because of the relationships between them. This will be explained in Chap. 3.
- Often we need to address more complex problems where we have external information that is indirectly related to the facies and is defined at a different scale (for example, from seismics).

For all these reasons we chose to define the indicators by truncating a continuous multivariate random function  $Z(x)$  with  $n$  components. Let  $C_i(x)$  be a partition of  $\mathbb{R}^n$ . A point  $x$  belongs to the  $i^{\text{th}}$  facies  $F_i$  if and only if  $Z(x) \in C_i(x)$ , so

$$F_i(x) \triangleq Z(x) \in C_i(x).$$

The advantages of this model are twofold:

- Many more multivariate distributions are available for modelling continuous random function than random sets.

- The relationships between the  $n$  components  $Z_j(x)$  of  $Z(x)$  are simpler than those between the facies. In particular their variograms and cross variograms are simpler to model. Furthermore the facies covariances and cross-covariances can be obtained once we know the distributions of  $Z$ .

Despite these advantages, there are still difficulties modelling and simulating these continuous random functions. Adding extra conditioning information will usually be difficult. This is why we chose to restrict the random functions that we use to be surjective functions of a of multi-gaussian random functions. The fact that we have the freedom to select the functions, gives us access to many more types of continuous random functions than the classical ones for which we know the distribution.

The problem is tractable because, as we will see below:

- With a few hypotheses on the function, we can transform the problem in terms of multi-gaussian random functions.
- Once the gaussian values at the conditioning points have been simulated, we can use the classical conditional gaussian simulation framework first developed by Matheron and Journel (see Chilès and Delfiner 1999; Lantuéjoul 2002a, b).

Another advantage of using underlying gaussians is that linear functionals such as partial derivatives or convolutions of a gaussian are still gaussian. This makes it is easier to formulate and solve the simulation problem even with additional information. Finally it is easier to find models for variograms because any negative definite function can be used as the variogram of a gaussian. Care is required when modelling the cross-variograms, but some models such as the linear coregionalisation model are known to ensure consistency. As the formulas for the variograms and cross-variograms can be derived for linear transforms of gaussians, this model is by construction consistent.

Taking advantage of these properties we now define a general setting for these simulations. As we want to be able to handle additional categorical variables and seismic attributes in addition to lithofacies, we start by defining three standardized multi-gaussian vectors,  $Z(x)$ ,  $Y(x)$  and  $S(x)$ , which are in general correlated and which are defined below. The lithofacies are known at  $m_Z$  sample points, the categorical variables are known at  $m_Y$  points and the seismic attributes are available at a third set of  $m_S$  points. Our objective is to simulate  $Z(x)$  given these three types of data. Let

- $Z(x)$ ,  $x \in \mathbb{R}^d$  be an  $n$ -variate standardized multi-gaussian vector (i.e., the marginal and joint distributions are normal). This will be used to define the facies. The  $m_Z$  sample locations where the facies are known are denoted by  $x_Z$ . If the sample  $x_\alpha \in x_Z$  belongs to the  $i^{\text{th}}$  facies, then  $Z(x_\alpha) \in C_i^Z(x_\alpha)$ .
- $Y(x)$ ,  $x \in \mathbb{R}^d$  be an  $m$ -variate standardized multi-gaussian vector which is correlated to  $Z(x)$ . It will be used to define another categorical variable by truncation. We have information on it at  $m_Y$  points  $x_Y$ . If the sample  $x_\beta \in x_Y$  belongs to the  $j^{\text{th}}$  category, then  $Y(x_\beta) \in C_j^Y(x_\beta)$ . Note that although we use the

same symbol for the constraints on  $Z$  and  $Y$ , the constraints are generally not in the same space unless  $n = m$ .

- $S(x)$ ,  $x \in \mathbb{R}^d$  be a  $p$ -variate standardized multi-gaussian vector correlated to  $Z$  and  $Y$  that is known at some sample locations  $x_S$ .

We assume that the three vectors have a joint multi-gaussian distribution<sup>1</sup>

*Comments.*

- We could have included the third random function  $S(x)$  in  $Y(x)$  which looks more general. However for some of the applications  $S(x)$  will be given by indirect information such as seismics, and thus it is interesting to identify it explicitly.
- To simplify the notation we will designate the constraints by the symbol  $C$  when there is no risk of confusion. In the cases where we want to specify one particular facies we will write  $C_i$  and if we need to distinguish which random function it is related to, we will add an upper index (e.g.,  $C^Z$  to specify that it is related to  $Z$ ).

Having defined the multi-gaussian random functions, we now define the functions that may be nonlinear. Let  $\Phi$ ,  $\psi$  and  $\eta$  be three functions from  $\mathbb{R}^n$  onto  $E \in \mathbb{R}^n$ ; from  $\mathbb{R}^m$  onto  $F \in \mathbb{R}^m$  and from  $\mathbb{R}^p$  onto  $H \in \mathbb{R}^p$ . We assume that these functions are surjective but not necessarily bijective, that is, for any  $t$  in the arrival space we can find *at least* one point in the initial space whose image is  $t$ .

Our objective is to simulate  $n_F$  facies  $F_i$   $i = 1, \dots, n_F$ , defined by truncating  $\Phi(Z)$  at the target points subject to inequality constraints on the three variables at different points. We are now going to express this problem in three different ways.

*Problem I.* To simulate facies  $F_i$  defined by  $\phi(Z(x)) \in C^\phi(x)$ , subject to the following constraints:

$$\begin{aligned} \phi(Z(x_Z)) &\in C^\phi(x_Z) \\ \psi(Y(x_Y)) &\in C^\psi(x_Y) \\ \eta(S(x_S)) &= S^\eta(x_S). \end{aligned} \tag{I}$$

*Problem II.* The model can be made more general by rewriting the constraints using another function  $\Theta$  and a multi-gaussian  $B$ :

---

<sup>1</sup>At this level of generality no hypothesis has to be made concerning the stationarity or the correlation structure of the gaussians. They can be stationary or non stationary. Later on for practical reasons – mainly the inference of the covariances from the data but also because of the strong vertical non-stationarity of facies in sedimentary deposits linked to sequence stratigraphy – we will choose to work with second order stationary gaussians and possibly variable proportions. But from a theoretical point of view there is no need to specify these properties.

$$\Theta(\mathbf{B}) = \begin{bmatrix} \phi(Z(x_{Z,1})) \\ \dots \\ \phi(Z(x_{Z,m_Z})) \\ \psi(Y(x_{Y,1})) \\ \dots \\ \psi(Y(x_{Y,m_Y})) \\ \eta(S(x_{S,1})) \\ \dots \\ \eta(S(x_{S,m_S})) \end{bmatrix}$$

While this makes it easier to write the simulation problem, it is formally equivalent to problem I. So the problem is to simulate  $F_i(x)$  defined by

$$F_i = \{x \in \mathbb{R}^3; \phi(Z(x)) \in \mathcal{C}_i^\phi(x)\} \text{ given } \Theta(\mathbf{B}) \in \mathcal{C}^\Theta. \quad (\text{II})$$

Note that the  $\mathcal{C}^\Theta$  are quite complex now because they are embedded in a higher dimension space. Although this problem looks complicated at first glance it is much simpler than it appears. As we can solve for the functions  $\psi$  and  $\eta$  we can also solve for the function  $\Theta$ . So

$$\Theta(\mathbf{B}) \in \mathcal{C}^\Theta \Leftrightarrow \mathbf{B} \in \mathcal{C} \triangleq \Theta^{-1}(\mathcal{C}),$$

where

$$\Theta^{-1}(\mathcal{C}) = \{u \in \mathbb{R}^{nm_Y + mm_Z + pm_S}; \Theta(u) \in \mathcal{C}\}.$$

Similarly

$$\phi(Z(x)) \in \mathcal{C}_i^\phi(x) \Leftrightarrow Z(x) \in \mathcal{C}_i(x),$$

With  $\mathcal{C}_i \triangleq \{u \in \mathbb{R}^n; \phi(u) \in \mathcal{C}_i^\phi(x)\}$ . Finally we see that Problem II reduces to a simpler one which we call Problem IIIa.

*Problem III.* In this case the problem is to simulate<sup>2</sup>

$$Z(x)|\mathbf{B} \in \mathcal{C}. \quad (\text{IIIa})$$

If need be, this problem can be rewritten in terms of  $Z$ ,  $Y$  and  $S$

$$[Z(x)|Z(x_Z) \in \mathcal{C}(x_Z), Y(x_Y) \in \mathcal{C}(x_Y), S(x_S) \in \mathcal{C}(x_S)]. \quad (\text{IIIb})$$

---

<sup>2</sup>Here  $\mathcal{C}$  refers to all the constraints ie at points  $x_Z$ ,  $x_Y$  and  $x_S$ .

Note that if  $\eta$  is not injective, the point constraints for  $S(x_S)$  are now transformed into constraints that  $S(x_S)$  belongs to a set.

Problems IIIa and IIIb are far easier to solve than Problem I, because we can now use all the nice properties of gaussian random functions for simulation. The variograms will also be far simpler to model than in the initial form. The only potential drawback is that the constraints might have complex shapes and might be more difficult to compute. Now we show how to solve Problem IIIb.

Let  $g(z, z_Z, z_Y, z_S)$  be the density of  $(Z(x), Z(x_Z), Y(x_Y), S(x_S))$ . Note that each term is itself a vector. For example  $Z$  has  $n$  components and  $x$  lies on a grid. Whatever the distribution<sup>3</sup> of  $Z$ ,  $Y$  and  $S$ , the conditioned vector that we have to simulate has the following density (up to a normation factor that we are in general unable to compute):

$$g(z, z_Z, z_Y, z_S) 1_{z_Z \in \mathcal{C}_Z} 1_{z_Y \in \mathcal{C}_Y} 1_{z_S \in \mathcal{C}_S}. \quad (2.1)$$

Using Bayes law we can rewrite it as:

$$g(z|z_Z, z_Y, z_S) h(z_Z, z_Y, z_S) 1_{z_Z \in \mathcal{C}_Z} 1_{z_Y \in \mathcal{C}_Y} 1_{z_S \in \mathcal{C}_S}, \quad (2.2)$$

where  $h$  is the density of  $z_Z, z_Y, z_S$ . Alternatively it can be written as

$$g(z|z_Z, z_Y, z_S) h(z_Z|z_Y, z_S) 1_{z_Z \in \mathcal{C}_Z} v(z_Y|z_S) 1_{z_Y \in \mathcal{C}_Y} w(z_S) 1_{z_S \in \mathcal{C}_S}. \quad (2.3)$$

Equations (2.2) and (2.3) show that we can disconnect the simulation at data points from the one at grid nodes. More precisely as soon as we have a conditional simulation at the sample locations, we can simulate  $Z$  on the grid given the values simulated at constraints points. Going further, from (2.3) we can simulate  $S$  at its own data points then  $Y$  at its data locations given the  $S(x_S)$  and so on. This is independent of any hypothesis (except the existence of a density for all the variables). In the gaussian case we can go even further: (2.2) and (2.3) are saying that once we have simulated the values of the various gaussians (truncated or not) at the constraint points, we can use the well known gaussian conditional simulation methods using the preceding values as if they were known conditioning values. So now the only problem remaining is to simulate

$$h(z_Z, z_Y, z_S) 1_{z_Z \in \mathcal{C}_Z} 1_{z_Y \in \mathcal{C}_Y} 1_{z_S \in \mathcal{C}_S} \quad (2.4)$$

or in the second version

$$h(z_Z|z_Y, z_S) 1_{z_Z \in \mathcal{C}_Z} v(z_Y|z_S) 1_{z_Y \in \mathcal{C}_Y} w(z_S) 1_{z_S \in \mathcal{C}_S}. \quad (2.5)$$

For sake of completeness Annex 1 at the end of this chapter gives the proof that the conditional distributions obtained from a gaussian vector given a sub-vector are

---

<sup>3</sup>We need only assume that it has a density.

also gaussian (with the simple co-kriging as its mean and the variance of residuals as its variance).

In the multi-gaussian case using this result we see that all the three densities ( $h$ ,  $v$  and  $w$ ) are standardized multi-gaussians.<sup>4</sup> The conditional densities are also gaussian so we only need covariance matrix of the initial multi-gaussian to fully characterize these distributions. It is interesting to note that they can be computed quite easily (by a simple inversion) from covariance between facies, or categorical variables, or between one of these and the seismic data. The formulas for the two common cases are given below.

## Non Centered Covariance Between Two Indicators

$$\begin{aligned}
 \text{Cov}(1_{F_i}(x), 1_{F_j}(y)) &= P(x \in F_i, y \in F_j) \\
 &= E[1_{F_i}(x) 1_{F_j}(y)] \\
 &= E[1_{Z(x) \in C_i(x)} 1_{Z(y) \in C_j(y)}] \\
 &= \int_{C_i(x)} \int_{C_j(y)} g_{Z(x), Z(y)}(u, v) \, du dv,
 \end{aligned} \tag{2.6}$$

where  $g_{Z(x), Z(y)}(u, v)$  is the  $2n$ -variate gaussian density of the vector  $(Z(x), Z(y))$ , so each integral is over  $\mathbb{R}^n$ .

For the sake of completeness we write the other non-centred covariances which are very similar

$$P(x \in F_i^Y, y \in F_j^Y) = \int_{C_i(x)} \int_{C_j(y)} g_{Y(x), Y(y)}(u, v) \, du dv, \tag{2.7}$$

where  $F^Y$  is the categorical variable defined by  $F_i^Y = \{x \in \mathbb{R}^d; Y(x) \in C_i(x)\}$ . Note that each integral is now over  $\mathbb{R}^m$ .

$$P(x \in F_i, y \in F_j^Y) = \int_{C_i^Z(x)} \int_{C_j^Y(y)} g_{Z(x), Y(y)}(u, v) \, du dv. \tag{2.8}$$

Note that  $g_{Z(x), Y(y)}(u, v)$  is the density of the  $n + m$  vector  $(Z(x), Y(y))$  so the dimension of the first integral is  $n$  and the second is  $m$ .

---

<sup>4</sup>But the conditional versions are no longer standardized. See Annex 1.

## Centered Covariance Between One Indicator and One Gaussian

$$\begin{aligned}
 \text{Cov}(F_i(x), G(y)) &= E[1_{F_i(x)} G(y)] \\
 &= E[1_{Z(x) \in \mathcal{C}_i(x)} G(y)] \\
 &= \int_{\mathcal{C}_i(x)} du \int_{\mathbb{R}} s g_{Z(x), G(y)}(u, s) ds.
 \end{aligned} \tag{2.9}$$

The dimension of the first integral is 1 and  $n$  for the second one.

Before studying the conditioning process we present some examples to illustrate the flexibility of the method and the variety of shapes it can produce. Geological realism will be postponed until the next chapter and the case studies.

### Examples

In this section we present several examples of cases where the functions  $\phi$ ,  $\psi$  and  $\eta$  are linear, then several non-linear cases including one where  $\phi$  is quadratic, which gives rise to a gamma distribution. For the linear cases we start out with the simplest case where  $n = 1$  (which corresponds to the well-known truncated gaussian case). Then we consider several cases with  $n = 2$ , with different types of sample data.

#### *Linear Case with One Gaussian ( $n = 1$ )*

In this case there is no  $Y$  or  $S$  and  $n = 1$ . This is usually referred to as the truncated gaussian method. The problem is to simulate  $Z(x)|Z(x_Z) \in \mathcal{C}(x_Z)$ . Each simulation of  $Z$  conditioned by  $Z(x_Z)$  has only to be truncated in order to define the facies  $F_i$  according to the definition:

$$F_i = \{x \in \mathbb{R}^2; Z(x) \in \mathcal{C}_i(x)\}.$$

In this case  $\mathcal{C}_i(x)$  is generally an interval, and as was mentioned earlier the  $\mathcal{C}_i(x)$  for  $i = 1, \dots, n_F$  are a partition of  $\mathbb{R}$  for every  $x$ . As this interval depends on  $x$ , even if we choose  $Z(x)$  to be stationary, the spatial distribution of the facies can be non-stationary. The non centered cross-covariance of facies is easy to relate to the covariance of the gaussian and to the constraints. In this case (2.9) simplifies to two integrals over  $\mathbb{R}$ .

$$\begin{aligned}
 E[1_{F_i(x)} 1_{F_j(y)}] &= E[1_{Z(x) \in \mathcal{C}_i(x)} 1_{Z(y) \in \mathcal{C}_j(y)}] \\
 &= \int_{\mathcal{C}_i(x)} du \int_{\mathcal{C}_j(y)} dv g_{Z(x), Z(y)}(u, v) du,
 \end{aligned} \tag{2.10}$$

where

$$g_{Z(x), Z(y)}(u, v) = \frac{1}{(2\pi)^{n/2} \det C_{x,y}} \exp \left( -\frac{1}{2} (u^t, v^t) C_{x,y}^{-1} \begin{pmatrix} u \\ v \end{pmatrix} \right)$$

and  $C_{x,y}$  is the covariance  $Z(x)$  and  $Z(y)$ . See Chap. 5 for more information, and in particular how to determine the covariances and the thresholds (domains  $C_i$ ), from facies data points. Equation (2.2) becomes

$$g(z|z_Z)h(z_Z)1_{z_Z \in \mathcal{C}(x_Z)}. \quad (2.11)$$

Here  $g(z|z_Z)$  is the multi-variate gaussian density of dimension  $N$  (the number of target points). The mean at each target point is the simple kriging of the  $Z(x)$  using the  $Z(x_Z)$ , and its covariance is that of the residuals. The second and third terms give the facies information at data points,  $h$  is also a multi-variate gaussian density with mean 0 and variance 1. These two terms correspond to the density of a truncated gaussian vector to be simulated. So in order to simulate  $Z(x)$  we first simulate the  $Z(x_Z)$ , that is, the truncated gaussian values at data points. Then for each simulation we simulate the  $Z(x)$  at target points according to (2.11), using the preceding simulated values as conditioning points. That is, we are left with the simulation of a gaussian random function whose conditional distribution is  $g(z|z_Z)$ . The three examples shown in Fig. 2.1 were obtained by truncating the same gaussian random function – one with an anisotropic gaussian variogram with a parameter 100 in the horizontal direction and 50 in the vertical direction. Only the threshold intervals differ. The grey facies corresponds to the interval  $[1, \infty[$  (Fig. 2.1a) while it corresponds to the interval  $[0.58, \sqrt{6}[$  (Fig. 2.1b). Both have approximately 30% of grey facies. The interval  $[-0.1, 0.1[$  for the grey facies is symmetric around zero (Fig. 2.1c). It has been included to help in understanding the case with two gaussians and a nonlinear transformation shown in Fig. 2.2.

### ***Non Linear Case with One Gaussian: A Gamma Process***

Let  $\phi(t) = t^2$  so the facies are defined by:

$$F_i = \{x \in \mathbb{R}^3; Z^2(x) \in C_i(x)\}.$$

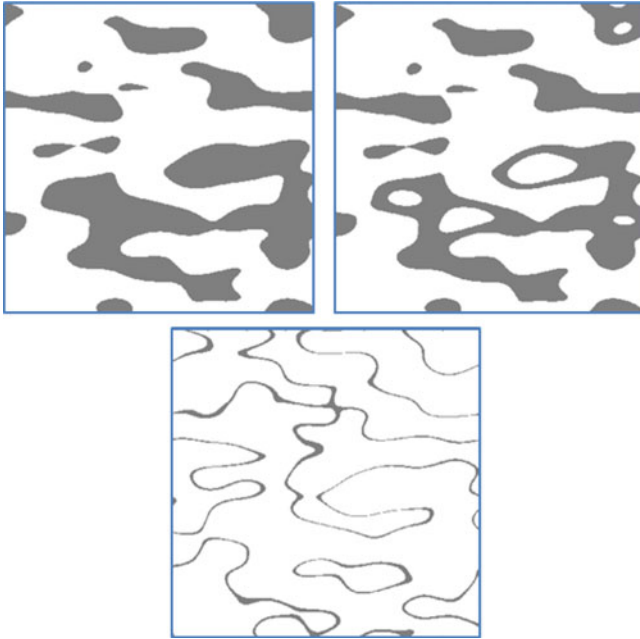
For each value  $t$  in  $C_i(x) \in \mathbb{R}_+$  we have two roots  $\pm \sqrt{t}$ , so

$$F_i = \left\{ x \in \mathbb{R}^2; Z(x) \in \pm \sqrt{C_i(x)} \right\}.$$

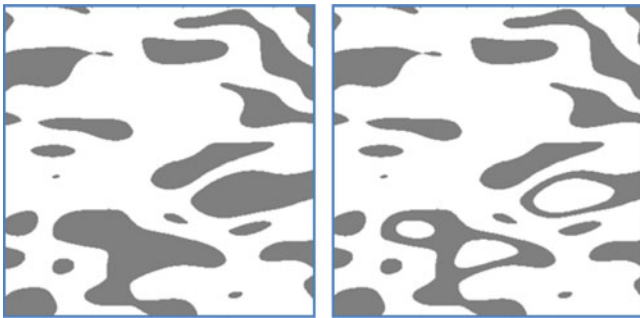
For example, if  $C_i(x)$  is the interval  $[a, b[$  we get:

$$F_i = \left\{ x \in \mathbb{R}^2; Z(x) \in ]-b, -a] \cup [a, b[ \right\}.$$



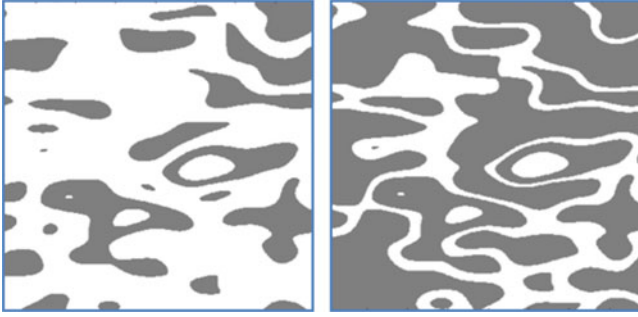


**Fig. 2.1** Three examples obtained by truncating the same gaussian. The grey facies corresponds to the interval  $[1, \infty[$  (top left), to  $[0.58, \sqrt{6}[$  (top right) and to  $[-0.1, 0.1[$  (bottom)



**Fig. 2.2** Truncated gamma function. The underlying gaussian is the same as Figure 2.1 and the proportion of grey facies is very similar too. The truncation interval is  $[1, \infty[$  (left) and  $[1, 6[$  (right)

Each facies is now defined by the union of two intervals instead of one, but except for that, we are back to the first example: linear case with one gaussian. In particular the covariances are easily expressed using (2.10) (after replacing  $\mathcal{C}_i(\mathbf{x})$  by  $-\sqrt{\mathcal{C}_i(\mathbf{x})} \cup \sqrt{\mathcal{C}_i(\mathbf{x})}$ ), the simple and cross covariances for the facies can be expressed as a sum of four covariances corresponding to the linear case with one gaussian. Two simulations of a gamma truncated process are shown in Fig. 2.2.



**Fig. 2.3** The grey corresponds to  $\phi(G) \leq 0$  with  $t_0 = -1.08$  (right) and  $t_0 = -0.1$  (left). In both cases,  $t_1 = 1$  and  $t_2 = 3$

The underlying gaussian is the same as in Fig. 2.1. The grey facies have been given the same proportion as in top examples. The truncation on the gamma random function is  $[1, \infty[$  on the left and  $[1, 6[$  on the right. Comparing these to Fig. 2.1a, b the obvious difference is that the shapes are less continuous. Looking more carefully we see grey facies at the bottom left of Fig. 2.3 that was not present on Fig. 2.1. They correspond to low values of the initial gaussian.

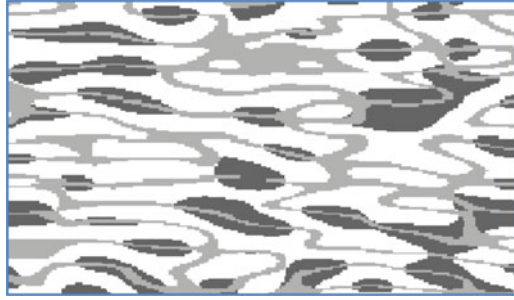
In many cases it might be interesting to use the gaussian which defines the truncation, as an auxiliary correlated variable to simulate values in the facies (e.g., porosities or permeabilities) in order to have more continuity on the edges of the facies. This was first done by Freulon et al. (1990) for the univariate case. By doing this, we could, for example, generate a bimodal permeability distribution within one lithofacies. One population of permeabilities would correspond to high values of the underlying gaussian, and the other to low values.

### ***Nonlinear Case with One Gaussian and a Third Order Polynomial***

Consider the case of a third order polynomial with three real roots  $t_0, t_1, t_2$ :

$$\phi(t) = (t - t_0)(t - t_1)(t - t_2).$$

The truncation now depends on the roots of the polynomial. In Fig. 2.3 we use the same underlying gaussian  $G$  as in Figs. 2.1 and 2.2. The grey facies corresponds to values of  $G$  such that  $\phi(G) \leq 0$ , that is, to values lower than  $t_0$  or in the interval  $[t_1, t_2[$ . For Fig. 2.3a we set  $t_0 = -1.08$ ,  $t_1 = 1$  and  $t_2 = 3$ , which gives approximately the same proportion of grey as in Figs. 2.1a, b and 2.2a. For Fig. 2.3b we changed  $t_0$  to 0.1, while keeping the other two parameters the same. Continuing the reasoning used for the gamma case, we find that the non centered simple and cross covariances between facies are the sum of nine non centred simple and cross covariances of a classical truncated gaussian. On Fig. 2.3a the grey facies is less continuous than on Fig. 2.2a, which itself was less continuous than on Fig. 2.1b.



**Fig. 2.4** Three facies obtained using one Gaussian and its derivative. The grey facies could represent meandering streams; the black one could be crevasse splays (in oil reservoirs) or washout in coal fields. Reproduced with permission

If we want to simulate petrophysical values on the grey facies using the underlying gaussian defining it as a correlated variable, we would have a different bivariate distribution than in the gamma case.

These two very simple examples show that non linear transforms of  $Z(x)$  are a convenient way to generate truncation rules that are more complex than intervals. They can provide much more freedom when fitting complex variograms of facies indicators and are useful because it is easier to understand the properties of the resulting facies by looking at the transform itself than the resulting rule on  $Z(x)$ . This can also help to simulate multimodal distributions for petrophysical variables.

### ***Linear Case with Two Gaussians ( $n = 2$ )***

The only constraints are on  $Z(x) = (Z_1(x), Z_2(x))$ .

#### **Classical Plurigaussian Case**

The problem is to simulate  $Z(x)|Z(x_z) \in \mathcal{C}(x_z)$ . For each  $x_z$ ,  $\mathcal{C}(x_z)$  is a domain in  $\mathbb{R}^2$  ( $n = 2$ ). In most practical applications  $\mathcal{C}(x_z) = I_1(x_z) \times I_2(x_z)$ , where  $I_1$  and  $I_2$  are intervals in  $\mathbb{R}$ . So as soon as we have simulated  $Z(x)$  the facies at location  $x$  is given by:

$$F_i = \{x; Z_1(x) \in I_1^i(x) \text{ \& } Z_2(x) \in I_2^i(x)\}.$$

Rewriting (2.2) in terms of the two components we get:

$$g(z_1, z_2 | z_{1Z}, z_{2Z}) h(z_{1Z}, z_{2Z}) 1_{z_{1Z} \in I_1(x_z)} 1_{z_{2Z} \in I_2(x_z)} \quad (2.12)$$

where  $g(z_1, z_2 | z_{1Z}, z_{2Z})$  is the multi-variate gaussian distribution. The dimension of both  $z_1$  and  $z_2$  is that of the grid to be simulated. For every grid node  $x$ , the mean of

the distributions is the simple co-kriging of the  $Z(x)$  by the  $Z(x_z)$ , and its covariance is that of residuals. Here the co-kriging of  $Z(x)$  is the simple co-kriging of each component using the two components at data points. When  $Z_1$  is independent of  $Z_2$  this simple co-kriging simplifies to two simple krigings.

The other three terms in the product is the distribution of the truncated gaussian vector to be simulated (up to a factor). So as in the previous case we start by simulating the two truncated gaussian random functions at data points using the preceding distribution. Then we use the simulated values as conditioning points for simulating the two gaussians on the grid according to (2.12). (See Chap. 7 for details on how (2.6) is written for this case).

Up to now, we have not specified the covariance and cross covariance of  $Z_1$  and  $Z_2$ . There is no specific constraint, provided they constitute a valid model for coregionalisation. One way to ensure this would be for example to use the linear model for coregionalisation, but there are plenty of other possibilities, e.g., by specifying a functional relationship between the two gaussians. For example  $Z_2$  could be a convolution of  $Z_1$  (or a convolution of a variable correlated to  $Z_1$ ). Below we present a case where  $Z_2$  is a partial derivative of  $Z_1$

In Chap. 7 we discuss the link between the co-regionalisation structure of the  $Y$ s and that of the facies for the case of two gaussians, in more detail.

## Derivative Based Plurigaussian Simulations

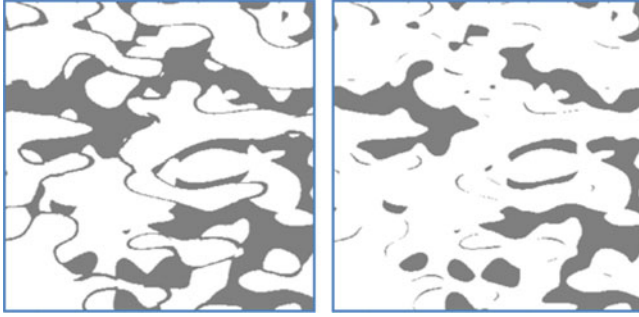
Here we present an example where the second gaussian is the partial derivative of the first one (Armstrong and Galli 1999). In order for the derivative of  $Z_1(x)$  to exist, the covariance of  $Z_1(x)$  must be twice differentiable. More general models based on other derivatives can also be used for  $Z_2$ . For example, the derivative of another random function  $Z_3$  that is possibly correlated with  $Z_1$  could be included:

$$Z_2(x, y) = a \frac{\partial}{\partial x} Z_1(x, y) + b \frac{\partial}{\partial y} Z_3(x, y).$$

The method is exactly the same as in the general case presented above. The only differences will be in the covariance of  $Z_2(x)$  and the cross covariance with  $Z_1(x)$ . If we assume that the processes are stationary and if we let  $h$  be the vector of differences, then we have:

$$\begin{aligned} C_{Z_2}(h) &= -a^2 \frac{\partial^2}{\partial h_1^2} C_{Z_1}(h) - b^2 \frac{\partial^2}{\partial h_2^2} C_{Z_3}(h) + ab \frac{\partial^2}{\partial h_1 \partial h_2} C_{Z_1, Z_3}(h) \\ C_{Z_1, Z_2}(h) &= a \frac{\partial}{\partial h_1} C_{Z_1}(h). \end{aligned}$$

Figure 2.4 taken from Armstrong and Galli (1999) presents an example with three facies. The grey facies represents meandering streams or rivers; the black one



**Fig. 2.5** Truncating a non-linear function of two independent gaussians, with  $I_1 = [-0.2, 2]$ ,  $I_2 = [-0.1, 5]$  on the *right*, and  $I_1 = [-0.5, 0.5]$ ,  $I_2 = [0.1, 5]$  on the *left*

could be crevasse splays in an oil reservoir or washouts in a coal measure. This sort of geometry is difficult to reproduce without using this type of model.

### ***Linear Case with Two Gaussians and Two Categorical Constraints***

There are situations where the dataset is composed of two types of categorical data at the same sample locations or at different locations. In some cases these could be related to two facies types (initial facies and diagenetic facies for instance see Pontiggia et al. 2010), whereas in others, they could be related to the same facies (for instance, different interpretations of logs). In that case we would not want to include this information in the truncation rule, because only the correlation matters.

Let the two gaussians for the facies be  $Z_1(x)$  and  $Z_2(x)$ ; let  $Y_1(x)$  and  $Y_2(x)$  be the two gaussians corresponding to categorical constraints. The simulation method is the same as before, except that an extra step has to be included: first simulate  $Z(x_Z)$  and then simulate  $Z(x)$  with the right conditional distribution which involves the  $Z(x_Y)$  and the  $Z(x_Z)$ . For simplicity we assume that various  $\mathcal{C}$  are the products of intervals. Equation (2.2) becomes:

$$\begin{aligned}
 &g(z_1, z_2 | z_{1Z}, z_{2Z}, z_{1Y}, z_{2Y}) \\
 &\quad \times h(z_{1Z}, z_{2Z} | z_{1Y}, z_{2Y}) \mathbf{1}_{z_{1Z} \in I_{1Z}^1(x_Z)} \mathbf{1}_{z_{2Z} \in I_{2Z}^2(x_Z)} \\
 &\quad \times v(z_{1Y}, z_{2Y}) \mathbf{1}_{z_{1Y} \in I_{1Y}^1(x_Y)} \mathbf{1}_{z_{2Y} \in I_{2Y}^2(x_Y)}.
 \end{aligned} \tag{2.13}$$

The conditional distribution of the  $Z(x_Z)$  given the  $Y(x_Y)$ , is a truncated gaussian whose mean is the cokriging of each component of the  $Z$  by the known values of the  $Y$  and the covariance matrix is that of the corresponding residuals. So the simulation of  $Z$  at points  $x_Z$  is carried out using this distribution. The next step is to simulate the two components at grid nodes using their conditional distribution which is also gaussian.

From a theoretical point of view, there is little difference between this and the case without constraints on  $Y$ , but from a practical point of view we must have a consistent coregionalisation model for four random functions  $(Z_1, Z_2, Y_1, Y_2)$  which can be complicated to infer. Furthermore the cokriging starts to become quite involved. Note that from (2.6)–(2.8) we can relate facies and categorical data to the various gaussian. So provided we have enough data, the coregionalisation structure of the gaussian can be inferred from that of the facies and the second categorical variable.

### ***Non Linear Case with Two Gaussians***

Continuing the preceding discussion we show that the non linear transforms can be more complicated because both components can involve one or two gaussians. So the truncation rules will be far more complex, as in the simple example below. The conclusions found for one gaussian hold here too, except that the simple and cross variograms of facies are more complex.

#### **Case with Two Facies**

The first gaussian is the same as in the previous examples with an anisotropic gaussian variogram with  $a_X = 100$ ,  $a_Y = 50$ , while the second is isotropic with a gaussian variogram with  $a = 50$ . To simplify further we let the two gaussians be independent

$$F_1 = \left\{ \begin{array}{l} Z_1(x) \in I^1 \\ Z_1(x)Z_2(x) \in I^2 \end{array} \right\}.$$

This case is interesting when the origin belongs to  $I^1$ . At points where  $Z_1$  equals 0 no facies  $F_1$  can be present unless  $I^2$  also contains 0. So we are able to generate two types of grey facies, a massive one and a thin one. In the case on the left the thin ones are connected to the massive ones, while on the right they are isolated (Fig. 2.5).

### ***Linear Case with One Gaussian $Z(x)$ and Seismic Constraints $S(x)$***

Suppose that we want to simulate three facies obtained by truncating  $Z(x)$  given seismic information  $S(x)$ . For simplicity we assume that this seismic information is a vertical convolution: (in 3D  $x$  is denoted by  $(x_1, x_2, x_3)$ )

$$S(x) = \int_c^d w(x_3 - y) G(x_1, x_2, y) dy. \quad (2.14)$$

Here  $G(x)$  is a standard normal distribution  $N(0,1)$  that is correlated to  $Z(x)$ . The function  $w$  has been normalized so that  $S(x)$  is also  $N(0,1)$ . In this very simple case with conditioning data, (2.5) simplifies (up to a constant factor) to:

$$h(z|z_S)1_{z \in C_Z}.$$

Here  $z_S$  stands for the values of seismics at sample locations  $S(x_S)$ , and  $h$  is the gaussian distribution of  $Z$  given  $S$ .

The mean of this gaussian conditional distribution is just the simple cokriging of  $Z$  knowing the seismics at data points; its covariance is the covariance of the simple kriging residual. So provided we know the covariance between  $Z$  and  $S$  we can carry out all the computation required. In order to compute this covariance, we use the classic convolution formula, but for that we need the covariance between  $F_i$  and  $G$ , as shown below.

$$\begin{aligned} C_{F_i, S}(x + h, x) &= E(1_{F_i}(x + h)S(x)) \\ &= \int_c^d w(x_3 + h_3 - y) \text{Cov}(1_{Z(x+h) \in C_i(x+h)}, G(x_1, x_2, y)) dy. \end{aligned}$$

By applying (2.13) or by direct computation, we obtain:

$$\begin{aligned} \text{Cov}(1_{Z(x+h) \in C_i(x+h)}, G(x_1, x_2, y)) &= E(1_{Z(x+h) \in C_i(x+h)} G(x_1, x_2, y)) \\ &= \int_{\mathbb{R}} \text{sg}(s) \left[ N\left(\frac{b_i - \lambda(x + h, x)s}{\sigma(x + h, x)}\right) - N\left(\frac{a_i - \lambda(x + h, x)s}{\sigma(x + h, x)}\right) \right] ds, \end{aligned} \quad (2.15)$$

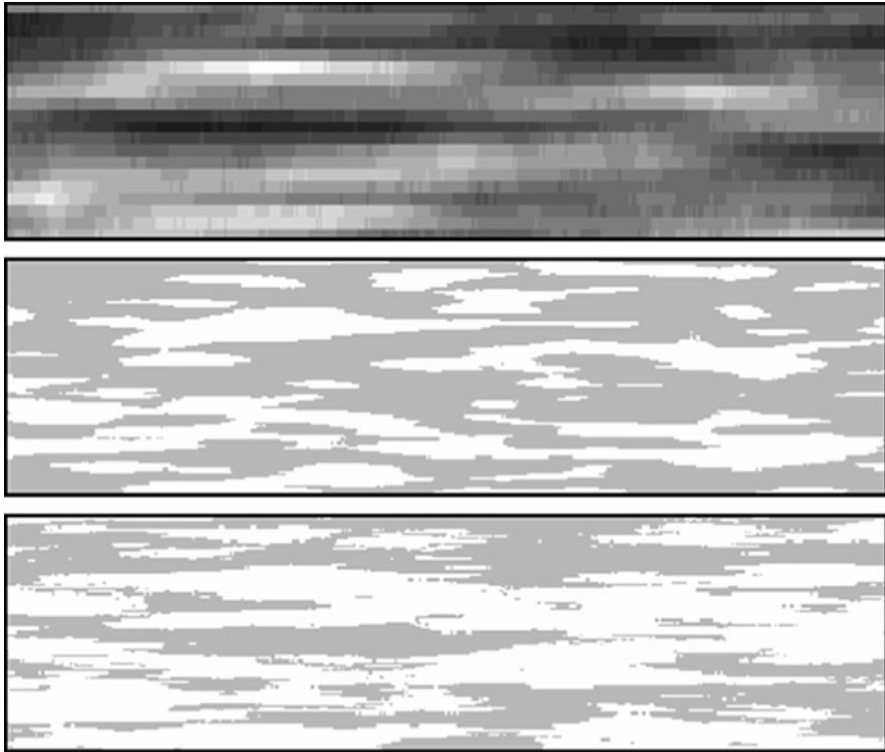
where

$$\begin{aligned} \lambda(x + h, x) &= C_{Z, G}(x + h, x) \\ \sigma(x + h, x) &= \sqrt{1 - C_{Z, G}^2(x + h, x)} \end{aligned}$$

and  $C_{Z, G}$  is the cross-covariance between  $Z$  and  $G$  and  $C_i(x + h) = ]a_i, b_i]$ .

Assuming the stationarity of the cross-covariance between  $Y$  and  $S$  we can estimate the covariance between  $Z$  and  $G$  simply by inverting equations (2.15) if we have enough information to compute an experimental cross-covariance (or cross variogram) between the facies and the seismic. Then we use the convolution formula to obtain the required covariances,  $C_S$  and  $C_Z$ .

Figure 2.6 shows a synthetic example where seismics (top) were used when simulating the facies. Two different correlation coefficients were compared: a weaker one ( $\rho = 0.4$ ) above and a stronger one ( $\rho = 0.8$ ) below.



**Fig. 2.6** Synthetic example where the seismics (above) were taken into account when simulating the facies. The middle panel corresponds to a correlation of 0.4, compared to 0.8 (below)

We have seen that the gaussian  $Z$  could be used as a driving factor to simulate continuous variables such as grades, porosities or permeabilities in the facies. In some cases it might be interesting to simulate the additional categorical variable  $Y(x)$ . In other cases, a shift could be introduced into the correlation between the gaussian and the facies. For example, Langlais et al. (2008) used a deterministic shift in a case study on a roll-front uranium deposit, while Emery (2007a, b) used a stochastic one.

## Simulating a Truncated Gaussian

There are many direct methods available for simulating gaussian random functions. The simplest are by matrix factorization, then by simulating the residuals, or by the sequential gaussian simulation where each point is simulated in turn. The problem is more difficult for truncated gaussian functions. To the best of our knowledge, no



direct method exists for truncated gaussians random functions so we have to resort to MCMC.

To start we note that if  $Z = (Z_1, Z_2)$  is a multivariate gaussian vector with  $n$  components that has been partitioned into  $n_1, n_2$  then the conditional distribution of the vector  $Z_1$  given  $Z_2$  is gaussian. Its mean is the simple co-kriging of  $Z_1$  by  $Z_2$  and its covariance matrix is the covariance of residuals. So if all of the values but one are known, then we only have to simulate a truncated random variable within a known set. Many algorithms are available for doing this. The simplest based on inversion is not the most efficient. Others based on acceptation & rejection algorithms using suitable functions (Devroye 1986; Geweke 1991; Robert 1995) have a high acceptance rate. Chopin (2011) proposed an extremely fast algorithm when only one sample has to be drawn. A variant of it exists for bivariate truncated Gaussians for infinite intervals.

This suggests that the Gibbs sampler might be a good choice because it chooses one component after another and draws values according to the conditional distribution given all the others (that is, from a univariate truncated gaussian). The Gibbs sampler can be described as follows:

Given a truncated multivariate gaussian vector  $Z$  to be simulated in a set  $\mathcal{C}$  we first partition the vector  $Z$  into  $(Z_1, Z_2)$  with  $n_1$  and  $n_2$  components respectively. Then the algorithm is:

Choose an initial vector  $Z^0 = (Z_1^0, Z_2^0)$  satisfying the constraints.

Iterate

1. Draw  $Z_1^{n+1}$  following a truncated gaussian distribution in the set  $\mathcal{C}$  given the values of  $Z_2^n$
2. Draw  $Z_2^{n+1}$  following a truncated gaussian distribution in the set  $\mathcal{C}$  given the values of  $Z_1^{n+1}$

Two points in the method have to be noted:

1. The question of when to stop the iterations is treated in Chap. 7
2. The question of how to draw the two sub-vectors can be solved either by Gibbs sampling the sub-vector or in the case where  $n_1$  (or  $n_2$ ) equals 1 by drawing an univariate truncated gaussian using one of the algorithms mentioned below

The case where  $n_2 = 1$  was first treated by Freulon (1994) in order to simulate conditionally truncated univariate gaussians with interval constraints, and by Le Loc'h and Galli (1997) for the bivariate truncated plurigaussian with interval constraints. But it might be worthwhile to consider different blocking cases for practical reasons, for example taking  $Z_1$  as the vector  $Z(x_Z)$  and  $Z_2$  as  $Y(x_Y)$ .

Other ways to proceed could be to follow (2.5) and simulate  $Y$  first under its constraints and then  $Z$ . But as is well known (Robert and Sahu 1997; Galli and Gao 2001) the convergence rate depends strongly on the blocking and also the ordering of the components. A study of the best ordering and blocking can be made for the gaussian case and used for the plurigaussian simulation. A practical example with four conditioning samples is treated in Chap. 7.

## Convergence of the Gibbs Sampler

### *The Untruncated Case*

The Gibbs sampler is known to converge in the untruncated gaussian case. See Barone and Frigessi (1990), Amit and Grenader (1991), Roberts and Sahu (1997) or Galli and Gao (2001). In fact, the Gibbs Sampler corresponds to Gauss-Seidel iterations on the inverse of the covariance matrix. This can be proved by expressing the inverse of the covariance matrix  $C^{-1}$  as  $D(I-L-U)$  where  $D$  is the block diagonal part of  $C^{-1}$ ,  $L$  and  $U$  are lower and upper block triangular matrices and  $I$  is the identity matrix. Then the iteration matrix of the Gibbs sampler can be written as  $A = (I-L)^{-1}U$ . So the rate of convergence depends on its spectral radius  $\rho(A)$ . A simple proof based on relating these iterations to linear fixed point iterations is given in Galli and Gao (2001). The iteration is:

$$Y^{(n+1)} = AY^{(n)} + (I-L)^{-1}R^{(n+1)},$$

where  $R^{(n+1)} = L_R V^{(n+1)}$ . Here  $L_R$  is the lower triangular part of the Cholesky decomposition of  $D^{-1}$  and  $V^{(n+1)}$  is a vector containing independent  $N(0,1)$  components. Consequently the covariance matrix of  $R^{(n+1)}$  is  $D^{-1}$ . The equation can be written in a more compact form.

$$Y^{(n+1)} = AY^{(n)} + U^{(n+1)},$$

where the covariance matrix of  $U^{(n+1)}$  is

$$\Sigma = (I-L)^{-1}D^{-1}(I-L)^{-t}.$$

### Blocking Factor

Galli and Gao (2001) also demonstrated the importance of the block size used in the Gibbs sampler (i.e., in the matrices  $D$ ,  $L$  and  $U$ ) for geostatistical applications. The most natural choice for  $D$  is simply the diagonal terms of  $C^{-1}$ . This corresponds to blocks of size  $1 \times 1$ . Other choices are possible and these can significantly increase the speed of convergence.

To illustrate its impact, we compare the spectral radius for the  $1 \times 1$  blocking size with that for the  $2 \times 2$  case for the four sample example considered earlier in the chapter. In that case the covariance matrix is of the form:

$$C = \begin{bmatrix} 1 & a & a^2 & a^3 \\ a & 1 & a & a^2 \\ a^2 & a & 1 & a \\ a^3 & a^2 & a & 1 \end{bmatrix} \text{ and } C^{-1} = \frac{1}{(a^2-1)} \begin{bmatrix} 1 & a & 0 & 0 \\ a & -(a^2+1) & a & 0 \\ 0 & a & -(a^2+1) & a \\ 0 & 0 & a & 1 \end{bmatrix}.$$

In this case, it is not difficult to calculate the various matrices and hence find the spectral radii for the two block sizes and also if the points are arranged in a different order.

### Block Size of 1 (Order: 1, 2, 3, 4)

$$D = \frac{1}{(a^2 - 1)} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -(a^2 + 1) & 0 & 0 \\ 0 & 0 & -(a^2 + 1) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A = (I - L)^{-1}U = \begin{bmatrix} 0 & b & 0 & 0 \\ 0 & b^2 & b & 0 \\ 0 & b^3 & b^2 & b \\ 0 & b^4 & b^3 & b^2 \end{bmatrix},$$

where

$$b = \frac{a}{(a^2 + 1)}$$

and hence the spectral radius  $\rho(A)$  is the largest root of the equation:

$$\lambda^2 [\lambda^2 - 3b^2 \lambda + b^4] = 0.$$

That is, it is

$$\rho(A) = b^2 \frac{3 + \sqrt{5}}{2} < 1 \quad \text{since} \quad \max(b^2) = \max\left(\frac{a^2}{(a^2 + 1)^2}\right) = \frac{1}{4}.$$

### Block Size of 2 (Points 1 & 2 and 3 & 4)

$$D = \frac{1}{(a^2 - 1)} \begin{bmatrix} 1 & a & 0 & 0 \\ a & -(a^2 + 1) & 0 & 0 \\ 0 & 0 & -(a^2 + 1) & a \\ 0 & 0 & a & 1 \end{bmatrix},$$

$$A = \begin{bmatrix} 0 & 0 & a^2 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & a^2 & 0 \\ 0 & 0 & a^3 & 0 \end{bmatrix}$$

and hence the spectral radius  $\rho(A)$  is the largest root of the equation:

$$\lambda^3 [a^2 - \lambda] = 0 \Rightarrow \rho(A) = \sqrt{a}.$$

### Block Size of 2 (Points 1&3 and 2&4)

$$D = \frac{1}{(a^2 - 1)} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -(a^2 + 1) & 0 & 0 \\ 0 & 0 & -(a^2 + 1) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$A = \begin{bmatrix} 0 & 0 & b & 0 \\ 0 & 0 & b & b \\ 0 & 0 & 2b^2 & b^2 \\ 0 & 0 & b^2 & b^2 \end{bmatrix}$$

and hence the spectral radius  $\rho(A)$  is the largest root of the equation:

$$\lambda^2 [\lambda^2 - 3b^2 \lambda + b^4] = 0.$$

So it is the same as in the first case.

### *Convergence in the Truncated Gaussian Case*

In his thesis Freulon (1992) considered two iterative ways of simulating truncated gaussian random functions. The first was based on work in the field of statistical mechanics by Metropolis et al. (1953) while the second is the Gibbs sampler (see Geman and Geman 1984). Freulon (pp 58–65) proved that both procedures converge to a truncated gaussian distribution, for the total variation distance, but his demonstration of their ergodicity only works for the case where the constraint is a compact (i.e., a closed and bounded interval). Secondly it does not provide any way to compute the rate of convergence because it is the minimum of  $k(x, y)$  on  $D \times D$  which is difficult to evaluate because of the denominator. In order to obtain more general results, we need to work in a more general framework, and so we introduce Markov chains.

### Markov Chains

Here we give an intuitive introduction to Markov chains. For more precise and more complete descriptions see Numelin (1984), Meyn and Tweedie (1993) or Gikhman

and Skorokhod (1969). A Markov chain is a chain  $X^{(0)}, X^{(1)}, \dots, X^{(n)} \dots$  with the property that for  $n > 1$ :

$$E(X^{(n)} | X^{(n-1)}, \dots, X^{(1)}, X^{(0)}) = E(X^{(n)} | X^{(n-1)}).$$

So a chain is fully determined as soon as we know the transition probabilities from one state to another. Here the possible states are not countable so they are characterised by an integral kernel  $k_n(x, y)$ . If the kernel does not depend on the position in the chain, ie  $k_n(x, y) = k(x, y)$  the chain is said to be homogeneous. We will only consider this case.

## ***Two Integral Operators***

We can define two integral operators from this kernel. If  $f(x)$  is a density then the operator  $g(y) = Kf(y)$  as defined below is a density:

$$g(y) = \int_S k(x, y) f(x) dx.$$

Furthermore, if  $f$  is the density of the  $(n - 1)^{\text{st}}$  iterate then  $g(y)$  is the density of the  $n^{\text{th}}$  one. Let  $h(x)$  be the conditional expectation of  $t(X^{(n)})$  given  $X^{(n-1)}$ . Then

$$h(x) = E(t(X^{(n)}) | X^{(n-1)} = x) = \int_S k(x, y) t(y) dy.$$

The two operators can be expressed in a more symmetrical way by letting

$$\tilde{k}(y, x) = k(x, y).$$

In that case

$$K_F t(y) = \int_S \tilde{k}(x, y) t(x) dx.$$

Let  $K^{(n)}$  be the iterated operator, that is,

$$K^{(n)}(f) = K(K^{(n-1)}(f)).$$

Similarly let  $K_F^{(n)}$  be

$$K_F^{(n)}(t) = K_F(K_F^{(n-1)}(t)).$$

Finally let  $k^{(n)}(x, y)$  be the corresponding kernel. The operator  $K^{(n)}$  expresses the relationship between the density of  $X^{(p)}$  and that of  $X^{(p+n)}$ . In contrast  $K_F^{(n)}$  gives the conditional expectation of a function of  $X^{(p+n)}$  given  $X^{(p)}$

$$E(t(X^{(n+p)})|X^{(p)}).$$

Note that for a homogeneous chain the joint distribution of the iterates  $X^{(n+p)}$  and  $X^{(n)}$  is:

$$g^{(n+p)}(x, y) = k^{(n)}(x, y)f^{(p)}(x) = g_y^{(n)}(x)f^{(n+p)}(y).$$

Here  $f^{(p)}$  and  $f^{(n+p)}$  stand for the marginal densities obtained by running the chain. The last equality holds because of Bayes Theorem.

If we let  $k^*(y, x)$  be the transition kernel corresponding to reverse passage from  $y$  to  $x$ , then  $g_y^{(n)}(x)$  corresponds to  $k^{*(n)}(y, x)$ . So we get:

$$g^{(n+p)}(x, y) = k^{(n)}(x, y)f^{(p)}(x) = k^{*(n)}(y, x)f^{(n+p)}(y). \quad (2.16)$$

### ***Stationary Distribution***

We say that a chain has a stationary distribution if the influence of the initial point  $X^{(0)}$  disappears after some time. More formally a chain has a stationary distribution if  $P(X^{(n)} \in A | X^{(0)})$  tends toward  $\pi(A)$  for any measurable set  $A$ , whatever the initial state  $X^{(0)}$ . In our case this is equivalent to  $k^{(n)}(x, y) \rightarrow \pi(y)$ . The chain is said to have an invariant distribution  $\pi$  if there exists a distribution  $\pi(x)$  such that

$$\pi(y) = \int_S k(x, y) \pi(x) dx.$$

Clearly a stationary distribution must be an invariant distribution. Assuming that  $\pi(x)$  is an invariant distribution, and applying (2.16) with  $p = 1$ ,  $n = 1$ , and  $f(x) = \pi(x)$  we get:

$$g^{(2)}(x, y) = k(x, y) \pi(x) = k^*(y, x) \pi(y). \quad (2.17)$$

### ***Irreducible Chain***

A Markov chain is said to be  $\psi$  irreducible if for each  $x$  in the support of a measure  $\psi$ , and for each measurable set  $A$  such that  $\psi(A) > 0$ , there exists an integer  $p \geq 1$  such that

$$P(X^{(p)} \in A | X^{(0)} = x) > 0.$$

A chain is said to be periodic if there exists an integer  $d \geq 2$  and a sequence of nonempty disjoint sets  $D_1, \dots, D_{d-1}$   $i = 1, \dots, (d-1)$  such that

$$\forall i = 1, \dots, d-1 \quad \text{and} \quad \forall x \in D_i, \quad K(x, D_j) = 1,$$

where

$$j = (i + 1) \bmod d.$$

A chain is aperiodic if it is not periodic. Note that if a chain is irreducible with  $p = 1$ , then it is aperiodic.

### ***Distance Metric and Convergence***

To define convergence we first have to define a distance criterion. One of the most commonly used distances is total variation (TV). The fact that a measure  $\mu_n$  converges toward a measure  $\mu$  for this distance, implies that for any bounded measurable function  $f$  we have convergence of the expectation. That is,

$$\|\mu_n - \mu\|_{TV} \rightarrow 0 \quad \text{implies that} \quad \int (\mu_n - \mu)(dx)f(x) \rightarrow 0$$

for any measurable bounded function  $f$ . This can be written as

$$E_n(f) \rightarrow E(f).$$

We have the following convergence result for Markov chains.

**Theorem 1 (Tierney 1994, 1996).** *If a chain is an irreducible aperiodic Markov chain with transition kernel  $k$  and an invariant distribution  $\pi$  then:*

$$\left\| k_F^{(n)}(x, \cdot) - \pi(\cdot) \right\|_{TV} \rightarrow 0.$$

*Although this result is interesting, its usefulness is limited because of its generality. Secondly it does not give the rate of convergence. Stronger results can be obtained in  $L^2$  type spaces, provided one makes some additional assumptions about the compactness of the integral operators defined previously. Here we will restrict ourselves to the case where the state space  $S$  is compact. For the truncated gaussian case this excludes unbounded intervals. From a practical point of view, it means adding two additional facies with infinitesimal proportions which are not seen in the data in order to bound the interval by a large value.*

For that case the previous result provides a method to estimate the speed of convergence in practice, at least for reasonably small spaces. In the case of truncated gaussian this means a reasonable number of constraints. The idea is to estimate the transition operator while running the chain as was proposed by Lantuéjoul (2002a, b). But here the chain has a continuous state space, so it is not possible to consider all possible transitions.

Given a finite partition of  $S$ , that is, a family of disjoint sets  $\Delta_i$  such that:

$$S = \bigcup_{i=1}^n \Delta_i,$$

we consider the transition matrix:

$$P_{i,j}^\Delta = \frac{1}{|\Delta_i|} \int_{\Delta_i} \int_{\Delta_j} k(x,y) \, dx \, dy.$$

This can be estimated reasonably by running the chain for a long enough time. Then we can compute the eigenvalues of  $P_{i,j}^\Delta$ . So the question is now: How close are these eigenvalues to those of  $K$ ? Our assumptions are

- The state space  $S$  is compact.
- The kernel  $k$  is continuous
- The assumptions of Theorem 1 hold.

The first and second assumptions ensure that  $K$  will map  $C(S)$  to  $C(S)$ . This is easy to verify because for continuous functions on a compact, there is no problem inverting the limit and the integral. Then using a result in Yosida (1978, p 277) we can show that this new operator  $K$  considered from  $C(S)$  to  $C(S)$  is compact. The convergence result of Theorem 1 ensures that the eigenvalues of  $K$  are lower than or equal to 1, and that it has one and only one eigenvalue 1, which is associated to the eigenvector  $\pi$ . This is easy to show because the set of bounded measurable functions on  $S$  includes continuous functions. Consequently

$$\sup_{f \in C[S]; \|f\| \leq 1} \left| \int_S (k^{(n)}(x,y) - \pi(y)) f(y) \, dy \right| \leq \|k^{(n)} - \pi\|_{TV}.$$

If there were a distribution  $g \neq \pi$  such that  $K_F(g) = cg$  with  $c \geq 1$ , then  $K^{(n)}_F(g)$  would equal  $c^n g$  and the integral

$$\int_S (c^n g(y) - \pi(y)) \, dy$$

would not converge to 0. This, in turn, implies that  $\|k^{(n)} - \pi\|_{TV}$  would not converge. Finally if we define

$$P_\Delta(f)(t) = \frac{1}{|\Delta_i|} \int_{\Delta_i} f(u) \, du \quad \text{for } t \text{ in } \Delta_i.$$



We see that when

$$|\Delta| \rightarrow 0, P_\Delta f(t) \rightarrow f(t).$$

Finally it can be shown that this corresponds to a Galerkin approximation. See Chatelin (1983, pp 170–174). That guarantees the convergence of  $P^\Delta$  toward  $k$  as  $|\Delta|$  tends toward 0. Interestingly, the rate of convergence depends on the derivability of  $k$  (see exercise 4.39 p 191 in Chatelin (1983)). Furthermore the convergence of eigenvalues is also ensured and bounds are known.

## The Gibbs Sampler

Let  $\pi(x)$  be the density of the vector  $(X_1, \dots, X_n)$ , and let  $\pi(x_i | x_{i-}, x_{i+})$  be the conditional density of  $X_i$  given  $(X_{i-}, X_{i+})$  where

$$X_{i-} = (X_1, \dots, X_{i-1}) \quad \text{and} \quad X_{i+} = (X_{i+1}, \dots, X_n).$$

By convention

$$X_{0-} = \emptyset \quad \text{and} \quad X_{n+} = \emptyset.$$

The Gibbs sampler modifies all the coordinates of the vector, one after another, using the newly simulated one and the former ones. Note that  $X_i$  may be either a single component or a block of components. In the latter case we should speak of the block Gibbs sampler, but for simplicity we will call it a Gibbs sampler even if we are using a blocking factor.

The transition kernel for one iteration of the Gibbs sampler from  $x$  to  $y$  is:

$$k(x, y) = \prod_{i=1}^n \pi(y_i | y_{i-}, x_{i+}).$$

The positivity condition is said to hold if the support of  $\pi$  is the product of the supports of its marginal distributions. This turns out to be true in our case provided the constraints are products of unions of intervals.

Using this positivity condition we can establish the Hammersley Clifford decomposition (Besag 1974). It is easy to show that

$$\begin{aligned} \pi(x) &= \pi(x_1 | x_{1-}, x_{1+}) \pi(x_{1-}, x_{1+}) \\ &= \frac{\pi(x_1 | x_{1-}, x_{1+})}{\pi(y_1 | x_{1-}, x_{1+})} \pi(y_1, x_{1-}, x_{1+}). \end{aligned}$$

By repeating this operation on  $\pi(y_1, x_{1-}, x_{1+})$  we obtain:

$$\pi(x) = \frac{\prod_{i=1}^n \pi(x_i | y_{i-}, x_{i+})}{\prod_{i=1}^n \pi(y_i | y_{i-}, x_{i+})} \pi(y) \quad (\text{Hammersley Clifford})$$

The denominator is  $k(x, y)$  and the numerator is the kernel  $k^*(y, x)$  corresponding to one iteration of the Gibbs sampler from  $y$  to  $x$  in the reverse order. So we can rewrite the Hammersley Clifford decomposition in the following way:

$$\pi(x)k(x, y) = \pi(y)k^*(y, x). \quad (2.18)$$

Integrating this gives

$$\begin{aligned} \int \pi(x)k(x, y)dx &= \int \pi(y)k^*(y, x)dx \\ &= \pi(y) \int k^*(y, x)dx \\ &= \pi(y). \end{aligned}$$

Consequently  $\pi$  is an invariant distribution of the Markov chain with transition kernel  $k(x, y)$  and also of the Markov chain with transition kernel  $k^*(y, x)$ . Equation (2.18) could be obtained directly from (2.16) without the positivity assumption but assuming that  $\pi$  is the stationary distribution of the chain.

### ***Truncated Gaussian Case***

In the truncated gaussian case

$$\pi(x) = \kappa g(x)1_D(x),$$

where  $\kappa$  is the normalisation constant and  $1_D(x)$  is the indicator of constraints. We will only consider constraints which are products of unions of intervals. So the positivity condition holds because the conditional distributions  $\pi(x_i | x_{i-}, x_{i+})$  are truncated gaussians whose domain is the union of intervals corresponding to the constraint for the  $i^{\text{th}}$  component. Let the inverse of the covariance matrix of  $X$  be

$$C^{-1} = D(I - L - U),$$

where  $D$  is the block diagonal part of the inverse of  $C$ , and  $U$  and  $L$  are the upper and lower triangular parts.

Using the results for the non truncated gaussian case, it is easy to show that the transition kernel is:

$$k(x, y) = k(x)e^{-\frac{1}{2}(y-Ax)^T \Sigma (y-Ax)} 1_D(y)1_D(x),$$

where  $k(x)$  is a normation factor whose value depends on  $x$  and

$$A = (I - L)^{-1}U \quad \text{and} \quad \Sigma = (I - L)^t D (I - L).$$

Similarly,

$$k^*(y, x) = k^*(y) e^{-\frac{1}{2}(x - \tilde{A}y)^t \tilde{\Sigma}(x - \tilde{A}y)} 1_D(y) 1_D(x),$$

where

$$\tilde{A} = (I - U)^{-1}L \quad \text{and} \quad \tilde{\Sigma} = (I - U)^t D (I - U).$$

It is obvious that the conditions for Theorem 1 hold so the Gibbs sampler converges toward our candidate distribution  $\pi$ , for the total variation distance.

Then the kernels are continuous in the support of  $\pi$ , and if  $D$  is compact the two operators  $K$  and  $K_F$  corresponding to the truncated gaussian case define compacts operators from  $C(D)$  to  $C(D)$ . So we can apply the approximation method described earlier to estimate the eigenvalues, and to estimate the speed of convergence experimentally. It is also interesting to note that provided the constraints are the product of compact intervals, the minorisation condition holds.

**Definition.** We say that the kernel  $k(x, y)$  satisfies the minorisation condition if there exists a density  $\psi$ , and a parameter  $0 < \rho < 1$  such that

$$\forall x \in D \forall y \in D \quad k(x, y) \geq \rho \psi(y).$$

**Theorem 2.** If the constraints  $D$  are compact the kernel  $k(x, y)$  of the transition operator satisfies the minorisation condition.

*Proof.* From above

$$k(x, y) = k(x) e^{-\frac{1}{2}(y - Ax)^t \Sigma (y - Ax)} 1_D(y) 1_D(x).$$

As the proof depends on the properties of convex sets, we start out by reviewing some basic facts about them. A point  $v$  in a convex is an extremum if it cannot be written as a convex combination of other points in the convex. In a polyhedral convex such as  $y - A(D)$ , the extremal points are just the vertices of the polyhedra.

Because  $D$  is compact and  $k(x)$  is continuous, it is bounded below by a parameter  $c$  which is greater than 0 because  $k(x) \approx P(D - Ax)$ , and its minimum is reached at at least one point  $x_0$  and  $P(D - Ax_0) > 0$ . In order to minimise the exponential term we have to maximise the exponent.

Let  $u = y - Ax$ . This is equivalent to maximising

$$\|u\|_K^2 = u^t \Sigma u$$

subject to the constraints

$$u \in y - A(D).$$

As the function is continuous and as the set  $y - A(D)$  is compact, the minimum is attained. Because  $\Sigma \geq 0$ , the function is strictly convex.

As is well known, the maximum of a convex function on a convex is reached at extremal points of the convex. See for example Rockafellar (1970). Now each extremal point of  $y - A(D)$  is of the form  $y - A(x_i)$  where  $x_i$  is an extremum in  $D$ . There are  $2^n$  extremal points  $x_i$  in  $D$ , they are the vertices of the hyper-rectangle. Let  $T_i$  be the points  $y$  whose corresponding extremal point is  $x_i$ . By definition  $D$  is the union of the  $T_i$  and we can write:

$$k(x, y) > c \sum_i 1_{T_i}(y) e^{-\frac{1}{2}(y - Ax_i)^t \Sigma (y - Ax_i)}.$$

If we let  $\tilde{f}(y)$  denote

$$\tilde{f}(y) = c \sum_i 1_{T_i}(y) e^{-\frac{1}{2}(y - Ax_i)^t \Sigma (y - Ax_i)},$$

we find that

$$1 = \int_D k(x, y) dy > \int_D \tilde{f}(y) dy > 0.$$

If we let  $\rho$  be the integral of  $\tilde{f}$  over  $D$  we can define the density  $f(y) = \tilde{f}(y)/\rho$ . We find that:

$$\forall y \in D, \forall x \in D \quad k(x, y) > \rho f(y) \quad \text{with} \quad \rho < 1.$$

This minorisation condition ensures the convergence of the Gibbs sampler in  $L^1$ . See for example, Roberts and Polson (1994).

## Summary

The Gibbs sampler is a very powerful simulation method. It is not difficult to establish its convergence in the truncated gaussian case. However it is far more difficult to find its rate of convergence experimentally, even in the case of compact constraints which considerably simplify the problem. For this case we proposed a finite rank operator which approximates the transition operator (which is compact).

## Annex 1: Conditional Distribution of a Gaussian

### *Basic Properties of a Gaussian Vector*

Let  $Z$  be an arbitrary gaussian vector of dimension  $p + n$ . Assume its covariance matrix  $\Sigma$  has an inverse  $\Sigma^{-1}$ . Its density is:

$$g(w_1, w_2, \dots, w_{p+n}) = \frac{1}{(2\pi)^{(p+n)/2} |\Sigma|} \exp \left\{ -\frac{1}{2} (w - m)^t \Sigma^{-1} (w - m) \right\},$$

where  $w$  and  $m$  are column vectors with  $m = E(Z)$  and where  $|\Sigma|$  is the determinant of the covariance matrix.

We split the vector  $Z$  into two parts  $X$  and  $Y$ , of dimension  $p$  and  $n$  respectively. These correspond to the  $p$  data points and the  $n$  target points to be simulated. Partitioning the covariance matrix in the same way we obtain:

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^t & \Sigma_{YY} \end{pmatrix},$$

where  $\Sigma_{XX}$  and  $\Sigma_{YY}$  are the covariance matrices of  $X$  and  $Y$ , respectively, and  $\Sigma_{XY}$  is the cross covariance matrix between  $X$  and  $Y$ . Let  $\Lambda$  be the vector defined by

$$\Lambda = \Sigma_{XX}^{-1} \Sigma_{XY}.$$

It is easy to see that its  $i^{\text{th}}$  column consists of the weights for the minimum variance linear estimate of the  $i^{\text{th}}$  element of  $Y$  given the vector  $X$  as data. Let  $R$  be the covariance matrix of the residuals when estimating  $Y$  given  $X$ ; that is,  $R = \text{cov}(Y - \Lambda^t X)$ . It is well known<sup>5</sup> that the inverse of the covariance matrix  $\Sigma^{-1}$  is

$$\Sigma^{-1} = \begin{bmatrix} \Sigma_{XX}^{-1} + \Lambda R^{-1} \Lambda^t & -\Lambda R^{-1} \\ -R^{-1} \Lambda^t & R^{-1} \end{bmatrix}.$$

If we let  $u$  and  $v$  be vectors containing the first  $p$  and the remaining  $n$  components of  $w$  then

$$(u^t v^t) \Sigma^{-1} \begin{pmatrix} u \\ v \end{pmatrix} = (v - \Lambda^t u)^t R^{-1} (v - \Lambda^t u) + u^t \Sigma_{XX}^{-1} u.$$

Now we use this decomposition to express the gaussian density given earlier as the product:

---

<sup>5</sup>See for example Galli and Gao (2001).

$$g(u,v) = k g_u(v) g(u),$$

where the normation constant  $k$  is just

$$k = \frac{|\Sigma_{XX}| |R|}{|\Sigma|},$$

$$g(u) = \frac{1}{(2\pi)^{p/2} |\Sigma_{XX}|} \exp \left\{ -\frac{1}{2} u^t \Sigma_{XX}^{-1} u \right\}$$

and

$$g_u(v) = \frac{1}{(2\pi)^{n/2} |R|} \exp \left\{ -\frac{1}{2} (v - \Lambda^t u)^t R^{-1} (v - \Lambda^t u) \right\}.$$

Both are normal distributions; the first has zero mean and covariance  $\Sigma_{XX}$ , the second has mean  $\Lambda^t u$  and covariance  $R$ . As their integrals are equal to 1, the normation factor  $k$  turns out 1, which was not obvious initially. The first one  $g(u)$  is the distribution of the vector  $X$ . The second is the conditional density of  $Y$  given  $X$ . It is still gaussian but its mean now equals the minimum variance estimate of  $Y$  given  $X$ , and its covariance matrix is  $R$ , the covariance matrix of the residuals. Some typical examples are:

$$Z = (Z(x_Z), Z(x)) \Rightarrow X = Z(x_Z), Y = Z(x)$$

$$Z = (Z(x_Z), Z(x), Y(x_Y)) \Rightarrow X = (Z(x_Z), Y(x_Y))$$

$$Z = (Z(x_Z), Z(x), Y(x_Y), S(x_S)) \Rightarrow X = (Z(x_Z), Y(x_Y), S(x_S)).$$

Except when working with univariate truncated gaussians, the matrix  $\Sigma$  is the simple co-kriging matrix. The  $\Lambda$  are the co-kriging weights,  $u$  stands for the data points so  $\Lambda^t u$  is the co-kriging estimate at locations  $x$  and  $R$  is the matrix of the covariance of the co-kriging residuals. See Chap. 7 for more detail.

Plurigaussian Simulations in Geosciences

Armstrong, M.; Galli, A.; Beucher, H.; Loc'h, G.; Renard,  
D.; Doligez, B.; Eschard, R.; Geffroy, F.

2011, X, 176 p., Hardcover

ISBN: 978-3-642-19606-5