

## Chapter 3

# Unconstrained Global Optimization

### 3.1 A Randomized Curve of Steepest Descent

After a short discussion of local optimization, we investigate the following unconstrained global minimization problem:

$$\underset{\mathbf{x}}{\text{globmin}}\{f(\mathbf{x})\}, \quad f: \mathbb{R}^n \rightarrow \mathbb{R}, \quad n \in \mathbb{N}, \quad f \in C^2(\mathbb{R}^n, \mathbb{R}).$$

In other words, we have to compute a point  $\mathbf{x}_{\text{gl}} \in \mathbb{R}^n$  such that

$$f(\mathbf{x}) \geq f(\mathbf{x}_{\text{gl}}) \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

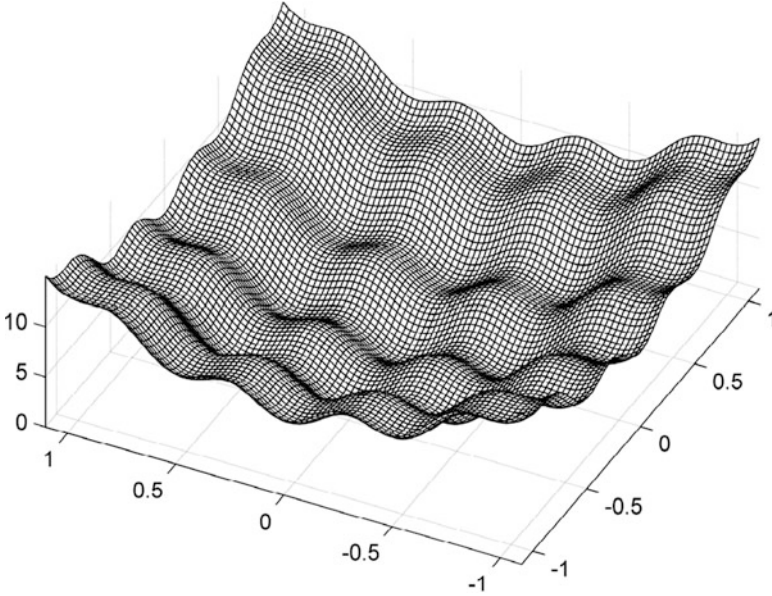
It is not reasonable to develop methods for solving unconstrained global minimization problems neglecting the existence of powerful methods solving unconstrained local minimization problems. Therefore, we are only interested in computing a suitable starting point close enough to a global minimum point in order to apply a local minimizing procedure.

Let the objective function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  be such that

$$\int_{\mathbb{R}^n} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x} < \infty$$

for some  $\epsilon > 0$ . In this case, the function

$$\lambda_f: \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \frac{\exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right)}{\int_{\mathbb{R}^n} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x}}$$



**Fig. 3.1** Objective function  $f$

can be interpreted as a Lebesgue density function of an  $n$ -dimensional real-valued random variable  $\mathbf{X}_f$ . We illustrate the main relationship between  $f$  and  $\lambda_f$  in the following example.

*Example 3.1.* The objective function

$$f : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x, y)^\top \mapsto 6x^2 + 6y^2 - \cos(12x) - \cos(12y) + 2$$

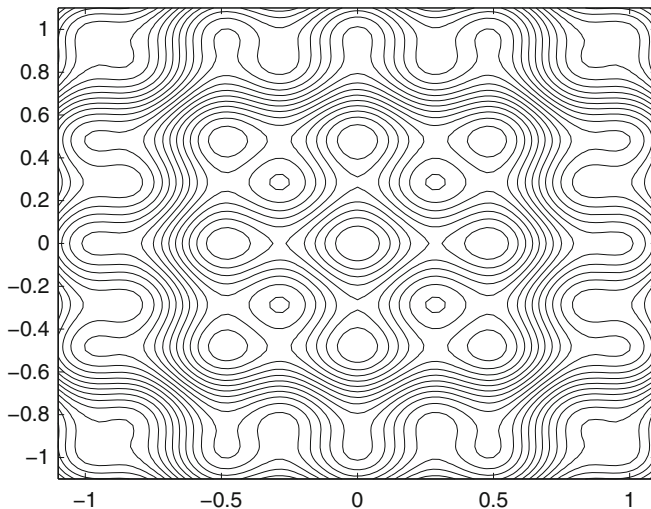
has 25 isolated minimum points with the unique global minimum point at the origin (Figs. 3.1 and 3.2).

Figure 3.3 shows the corresponding Lebesgue density function  $\lambda_f$ .

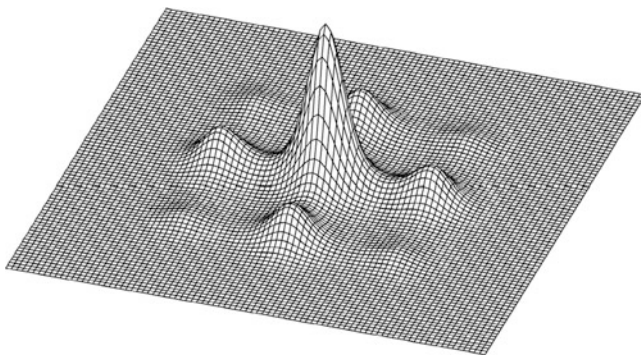
The smaller the function value of  $f$ , the larger the function value of  $\lambda_f$  and finally the likelihood of  $\mathbf{X}_f$ . If we would be able to generate realizations of the random variable  $\mathbf{X}_f$ , we would be able to generate suitable starting points for local minimization procedures to compute a global minimum point of  $f$ . Unfortunately, there is no easy way to compute realizations of  $\mathbf{X}_f$  in general. Therefore, we will use a dynamical system representing a suitable randomization of the curve of steepest descent in order to compute realizations of  $\mathbf{X}_f$ . For that, we have to introduce  $n$ -dimensional Brownian Motion.

Let

$$\Omega := C^0([0, \infty), \mathbb{R}^n)$$



**Fig. 3.2** Contour lines of  $f$



**Fig. 3.3** Lebesgue density function  $\lambda_f$

be the set of all continuous functions  $\omega : [0, \infty) \rightarrow \mathbb{R}^n$  (for  $t = 0$ , continuity from the right). We define a metric  $d_\Omega$  on  $\Omega$  by

$$d_\Omega : \Omega \times \Omega \rightarrow \mathbb{R}, \quad (\omega_1, \omega_2) \mapsto \sum_{m=1}^{\infty} \frac{1}{2^m} \min \left\{ \max_{0 \leq t \leq m} \|\omega_1(t) - \omega_2(t)\|_2, 1 \right\}.$$

With  $\mathcal{B}(\Omega)$ , we denote the smallest  $\sigma$ -field containing all open subsets of  $\Omega$  with respect to the topology defined by the metric  $d_\Omega$ . Using the Borel  $\sigma$ -field  $\mathcal{B}(\mathbb{R}^n)$  which is defined by all open subsets of  $\mathbb{R}^n$  in the topology given by the Euclidean norm, we obtain  $\mathcal{B}(\Omega)$ - $\mathcal{B}(\mathbb{R}^n)$ -measurable functions:

$$\mathbf{B}_t : \Omega \rightarrow \mathbb{R}^n, \quad \omega \mapsto \omega(t), \quad t \in [0, \infty).$$

There is a uniquely defined probability measure  $W$  on  $\mathcal{B}(\Omega)$  (the so-called Wiener measure) fulfilling the following conditions, where  $\mathcal{N}(\mathbf{e}, \mathbf{\Sigma})$  denotes the  $n$ -dimensional Gaussian distribution with expectation  $\mathbf{e}$  and covariance matrix  $\mathbf{\Sigma}$  (see, e.g., [KarShr08]):

- (i)  $W(\{\omega \in \Omega; \mathbf{B}_0(\omega) = \mathbf{0}\}) = 1$ .
- (ii) For all  $0 \leq t_0 < t_1 < \dots < t_k, k \in \mathbb{N}$ , the random variables

$$\mathbf{B}_{t_0}, \mathbf{B}_{t_1} - \mathbf{B}_{t_0}, \dots, \mathbf{B}_{t_k} - \mathbf{B}_{t_{k-1}}$$

are stochastically independent.

- (iii) For every  $0 \leq s < t$ , the random variable  $\mathbf{B}_t - \mathbf{B}_s$  is  $\mathcal{N}(\mathbf{0}, (t-s)\mathbf{I}_n)$  Gaussian distributed.

The stochastic process  $\{\mathbf{B}_t\}_{t \in [0, \infty)}$  is called  $n$ -dimensional Brownian Motion.

Based on the Wiener space, we will now investigate a stochastic process  $\{\mathbf{X}_t\}_{t \in [0, \infty)}$ ,

$$\mathbf{X}_t : \Omega \rightarrow \mathbb{R}^n, \quad t \in [0, \infty),$$

given by

$$\mathbf{X}_t(\omega) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau(\omega)) d\tau + \epsilon (\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)), \quad \omega \in \Omega,$$

where  $\mathbf{x}_0 \in \mathbb{R}^n$  and  $\epsilon > 0$  are fixed. This process can be interpreted as a randomized curve of steepest descent, where the curve of steepest descent in integral form is given by

$$\mathbf{x}(t) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{x}(\tau)) d\tau, \quad t \in [0, \infty),$$

as noted above. The balance between steepest descent represented by

$$\mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau(\omega)) d\tau$$

and a purely random search using Gaussian distributed random variables

$$\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)$$

is controlled by the parameter  $\epsilon$ .

In Sect. 2.1, we investigated the curve of steepest descent under weak assumptions on the objective function. In order to analyze the randomized curve of steepest descent, it is necessary again to formulate weak assumptions on the objective function:

**Assumption 3.2** *There exists a real number  $\epsilon > 0$  such that*

$$\mathbf{x}^\top \nabla f(\mathbf{x}) \geq \frac{1 + n\epsilon^2}{2} \max\{1, \|\nabla f(\mathbf{x})\|_2\}$$

for all  $\mathbf{x} \in \{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 > \rho\}$  for some  $\rho \in \mathbb{R}$ ,  $\rho > 0$ .

This assumption describes a special behavior of  $f$  outside a ball with radius  $\rho$ , for which only the existence is postulated. Starting at the origin, the objective function  $f$  has to increase sufficiently fast along each straight line outside the mentioned ball. Therefore, each function  $f \in C^2(\mathbb{R}^n, \mathbb{R})$  fulfilling Assumption 3.2 has at least one global minimum point  $\mathbf{x}_g$  within the ball  $\{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 \leq \rho\}$ . The converse is not true as the  $\sin$ -function shows. The fact that only the existence of some possibly very large  $\rho > 0$  is postulated ensures the weakness of this assumption.

In the theory of partial differential equations (see, e.g., [RenRog04]),  $\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is called a *coercive function* if

$$\lim_{\|\mathbf{x}\|_2 \rightarrow \infty} \frac{\mathbf{x}^\top \mathbf{g}(\mathbf{x})}{\|\mathbf{x}\|_2} = \infty.$$

The following alternative formulation of Assumption 3.2:

*There exists a real number  $\epsilon > 0$  such that*

$$\frac{\mathbf{x}^\top \nabla f(\mathbf{x})}{\|\mathbf{x}\|_2} \geq \frac{1 + n\epsilon^2}{2} \max \left\{ \frac{1}{\|\mathbf{x}\|_2}, \frac{\|\nabla f(\mathbf{x})\|_2}{\|\mathbf{x}\|_2} \right\}$$

for all  $\mathbf{x} \in \{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 > \rho\}$  for some  $\rho \in \mathbb{R}$ ,  $\rho > 0$

shows the relation of these two conditions on  $\mathbf{g}$  and  $\nabla f$ .

If Assumption 3.2 is not fulfilled, one can try to use an auxiliary objective function  $\tilde{f}$  of the following type:

$$\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto f(\mathbf{x}) + \left( P \left( \|\mathbf{x}\|_2^2 - c \right) \right)^m, \quad m \in \mathbb{N}, m \geq 3, c \in \mathbb{R}, c > 0,$$

where

$$P : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \begin{cases} x & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}.$$

For  $\tilde{f}$ , we obtain:

- $\tilde{f} \in C^2(\mathbb{R}^n, \mathbb{R})$ .
- $\tilde{f}(x) = f(x)$  for all  $x \in \{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2^2 \leq c\}$ .
- $\tilde{f}(x) > f(x)$  for all  $x \in \{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2^2 > c\}$ .

The practicability of  $\bar{f}$  instead of  $f$  is shown by

$$\mathbf{x}^\top \nabla \bar{f}(\mathbf{x}) = \mathbf{x}^\top \nabla f(\mathbf{x}) + 2m \left( P \left( \|\mathbf{x}\|_2^2 - c \right) \right)^{m-1} \|\mathbf{x}\|_2^2.$$

In the following theorem, we study properties of the randomized curve of steepest descent.

**Theorem 3.3** *Consider*

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad n \in \mathbb{N}, \quad f \in C^2(\mathbb{R}^n, \mathbb{R})$$

and let the following Assumption 3.2 be fulfilled:

There exists a real number  $\epsilon > 0$  such that

$$\mathbf{x}^\top \nabla f(\mathbf{x}) \geq \frac{1 + n\epsilon^2}{2} \max\{1, \|\nabla f(\mathbf{x})\|_2\}$$

for all  $\mathbf{x} \in \{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 > \rho\}$  for some  $\rho \in \mathbb{R}$ ,  $\rho > 0$ ,

then we obtain:

- (i) Using the Wiener space  $(\Omega, \mathcal{B}(\Omega), W)$  and  $\epsilon > 0$  from Assumption 3.2, the integral equation

$$\mathbf{X}(t, \omega) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}(\tau, \omega)) d\tau + \epsilon (\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)), \quad t \in [0, \infty), \quad \omega \in \Omega$$

has a unique solution  $\mathbf{X} : [0, \infty) \times \Omega \rightarrow \mathbb{R}^n$  for each  $\mathbf{x}_0 \in \mathbb{R}^n$ .

- (ii) For each  $t \in [0, \infty)$ , the mapping

$$\mathbf{X}_t : \Omega \rightarrow \mathbb{R}^n, \quad \omega \mapsto \mathbf{X}(t, \omega)$$

is an  $n$ -dimensional random variable (therefore,  $\mathcal{B}(\Omega) - \mathcal{B}(\mathbb{R}^n)$  measurable) and its probability distribution is given by a Lebesgue density function

$$p_t : \mathbb{R}^n \rightarrow \mathbb{R}$$

with

$$\lim_{t \rightarrow \infty} p_t(\mathbf{x}) = \frac{\exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right)}{\int_{\mathbb{R}^n} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x}} \quad \text{for all } \mathbf{x} \in \mathbb{R}^n.$$

The pointwise convergence of the Lebesgue density functions  $p_t$  to

$$\lambda_f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \frac{\exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right)}{\int_{\mathbb{R}^n} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x}}$$

can be interpreted as follows:

The randomized curve of steepest descent can be interpreted as a machine learning scheme for the computation of pseudorandom vectors according to the probability distribution given by  $\lambda_f$  using evaluations of  $\nabla f$  (or higher derivatives) and using the computation of standard Gaussian distributed pseudorandom numbers as inputs. This will be done by the numerical approximation of a path of the randomized curve of steepest descent.

For lucidity of the proof of Theorem 3.3, we first prove two lemmata.

**Lemma 3.4** *Let*

$$\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

*be a globally Lipschitz-continuous function with Lipschitz constant  $L > 0$ , and let*

$$\mathbf{B} : [0, \infty) \rightarrow \mathbb{R}^n$$

*be a continuous function; then the integral equation*

$$\mathbf{x}(t) = \mathbf{x}_0 - \int_0^t \mathbf{g}(\mathbf{x}(\tau)) d\tau + \mathbf{B}(t), \quad t \in [0, \infty),$$

*has a unique solution*

$$\mathbf{x} : [0, \infty) \rightarrow \mathbb{R}^n$$

*for each  $\mathbf{x}_0 \in \mathbb{R}^n$ .*

*Proof.* Since  $\mathbf{g}$  is globally Lipschitz-continuous with Lipschitz constant  $L > 0$ , we prove existence and uniqueness of a solution  $\mathbf{x} : [0, \infty) \rightarrow \mathbb{R}^n$  of the considered integral equation using the fixed point theorem of Banach. Let  $C^0([0, T], \mathbb{R}^n)$  be the set of all continuous functions  $\mathbf{u} : [0, T] \rightarrow \mathbb{R}^n$ , and

$$K : C^0([0, T], \mathbb{R}^n) \rightarrow C^0([0, T], \mathbb{R}^n),$$

$$K(\mathbf{u})(t) = \mathbf{x}_0 - \int_0^t \mathbf{g}(\mathbf{u}(\tau)) d\tau + \mathbf{B}(t), \quad t \in [0, T].$$

Obviously, each solution  $\mathbf{x}_T$  of

$$\mathbf{z}(t) = \mathbf{x}_0 - \int_0^t \mathbf{g}(\mathbf{z}(\tau)) d\tau + \mathbf{B}(t), \quad t \in [0, T]$$

is a fixed point of  $K$  and vice versa. With

$$d : C^0([0, T], \mathbb{R}^n) \times C^0([0, T], \mathbb{R}^n) \rightarrow \mathbb{R}, \quad (\mathbf{u}, \mathbf{v}) \mapsto \max_{t \in [0, T]} (\|\mathbf{u}(t) - \mathbf{v}(t)\|_2 e^{-2Lt}),$$

$(C^0([0, T], \mathbb{R}^n), d)$  is a complete metric space as mentioned before.

Again, the calculation

$$\begin{aligned} \|K(\mathbf{u})(t) - K(\mathbf{v})(t)\|_2 e^{-2Lt} &= \left\| \int_0^t (\mathbf{g}(\mathbf{v}(\tau)) - \mathbf{g}(\mathbf{u}(\tau))) d\tau \right\|_2 e^{-2Lt} \\ &\leq \int_0^t \|\mathbf{g}(\mathbf{v}(\tau)) - \mathbf{g}(\mathbf{u}(\tau))\|_2 d\tau \cdot e^{-2Lt} \\ &= \int_0^t \|\mathbf{g}(\mathbf{v}(\tau)) - \mathbf{g}(\mathbf{u}(\tau))\|_2 e^{-2L\tau} e^{2L\tau} d\tau \cdot e^{-2Lt} \\ &\leq L \int_0^t \|\mathbf{v}(\tau) - \mathbf{u}(\tau)\|_2 e^{-2L\tau} e^{2L\tau} d\tau \cdot e^{-2Lt} \\ &\leq L \cdot d(\mathbf{u}, \mathbf{v}) \int_0^t e^{2L\tau} d\tau \cdot e^{-2Lt} \\ &= L \cdot d(\mathbf{u}, \mathbf{v}) \frac{1}{2L} (e^{2Lt} - 1) e^{-2Lt} \\ &\leq \frac{L}{2L} d(\mathbf{u}, \mathbf{v}) = \frac{1}{2} d(\mathbf{u}, \mathbf{v}), \quad t \in [0, T] \end{aligned}$$

shows that

$$d(K(\mathbf{u}), K(\mathbf{v})) \leq \frac{1}{2} d(\mathbf{u}, \mathbf{v}),$$

and that the fixed point theorem of Banach is applicable. Consequently, we have found a unique solution  $\mathbf{x}_T : [0, T] \rightarrow \mathbb{R}^n$  of

$$\mathbf{z}(t) = \mathbf{x}_0 - \int_0^t \mathbf{g}(\mathbf{z}(\tau)) d\tau + \mathbf{B}(t), \quad t \in [0, T]$$

for all  $T > 0$ , and this gives us a unique solution  $\mathbf{x} : [0, \infty) \rightarrow \mathbb{R}^n$  of

$$\mathbf{x}(t) = \mathbf{x}_0 - \int_0^t \mathbf{g}(\mathbf{x}(\tau)) d\tau + \mathbf{B}(t), \quad t \in [0, \infty),$$

**q.e.d.**



**Lemma 3.5** *Let  $f \in C^2(\mathbb{R}^n, \mathbb{R})$  and  $\epsilon > 0$  such that Assumption 3.2 is fulfilled, then*

$$\int_{\mathbb{R}^n} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x} < \infty.$$

*Proof.* For every  $\mathbf{y} \in \mathbb{R}^n$ ,  $\|\mathbf{y}\|_2 \neq \mathbf{0}$ , we obtain for  $\gamma > \rho$

$$\nabla f\left(\gamma \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\right)^\top \gamma \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \geq \frac{1 + n\epsilon^2}{2}$$

and therefore

$$\frac{d}{d\gamma} f\left(\gamma \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\right) \geq \frac{1 + n\epsilon^2}{2\gamma}.$$

Integration over  $[\rho, \xi]$ ,  $\xi > \rho$ , with respect to  $\gamma$ , leads to

$$f\left(\xi \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\right) \geq \frac{1 + n\epsilon^2}{2} \ln(\xi) - \frac{1 + n\epsilon^2}{2} \ln(\rho) + f\left(\rho \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\right).$$

Let

$$c := \min_{\|\mathbf{y}\|_2 \neq 0} \left\{ f\left(\rho \frac{\mathbf{y}}{\|\mathbf{y}\|_2}\right) \right\}.$$

For every  $\mathbf{x} \in \{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 > \rho\}$ , there exists a unique  $\xi > \rho$  and a unique  $\mathbf{y} \in \mathbb{R}^n$  with

$$\mathbf{x} = \xi \frac{\mathbf{y}}{\|\mathbf{y}\|_2}.$$

We obtain:

$$f(\mathbf{x}) \geq \frac{1 + n\epsilon^2}{2} \ln(\|\mathbf{x}\|_2) + c - \frac{1 + n\epsilon^2}{2} \ln(\rho)$$

for all  $\mathbf{x} \in \{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 > \rho\}$ . This inequality is equivalent to

$$\exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) \leq c_1 \|\mathbf{x}\|_2^{-n-\frac{1}{\epsilon^2}}$$

with

$$c_1 = \exp\left(\frac{1 + n\epsilon^2}{\epsilon^2} \ln(\rho) - \frac{2c}{\epsilon^2}\right).$$

Integration leads to

$$\begin{aligned}
 \int_{\mathbb{R}^n} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x} &= \int_{\{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 \leq \rho\}} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x} \\
 &\quad + \int_{\{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 > \rho\}} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x} \\
 &\leq \int_{\{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 \leq \rho\}} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x} \\
 &\quad + \int_{\{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 > \rho\}} c_1 \|\mathbf{x}\|_2^{-n-\frac{1}{\epsilon^2}} d\mathbf{x} \\
 &< \infty
 \end{aligned}$$

using  $n$ -dimensional polar coordinates for the computation of

$$\int_{\{\mathbf{z} \in \mathbb{R}^n; \|\mathbf{z}\|_2 > \rho\}} c_1 \|\mathbf{x}\|_2^{-n-\frac{1}{\epsilon^2}} d\mathbf{x}.$$

**q.e.d.**

*Proof of Theorem 3.3.* Using

$$\mathbf{g} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \mathbf{x} \mapsto \begin{cases} \nabla f(\mathbf{x}) & \text{if } \|\mathbf{x} - \mathbf{x}_0\|_2 \leq r \\ \nabla f\left(\mathbf{x}_0 + \frac{r(\mathbf{x} - \mathbf{x}_0)}{\|\mathbf{x} - \mathbf{x}_0\|_2}\right) & \text{if } \|\mathbf{x} - \mathbf{x}_0\|_2 > r \end{cases}, \quad r > 0,$$

we consider the integral equation

$$\mathbf{Z}(t, \omega) = \mathbf{x}_0 - \int_0^t \mathbf{g}(\mathbf{Z}(\tau, \omega)) d\tau + \epsilon (\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)), \quad t \in [0, \infty), \quad \omega \in \Omega.$$

Since  $\mathbf{g}$  is globally Lipschitz-continuous with Lipschitz constant  $L > 0$ , and since each path of a Brownian Motion is continuous, the application of Lemma 3.4 for each  $\omega \in \Omega$  shows the existence and uniqueness of a solution

$$\mathbf{Z} : [0, \infty) \times \Omega \rightarrow \mathbb{R}^n$$

of the above integral equation.

Now, we have to consider the connection between the function  $\mathbf{Z}$  and the integral equation

$$\mathbf{X}(t, \omega) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}(\tau, \omega)) d\tau + \epsilon (\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)), \quad t \in [0, \infty), \quad \omega \in \Omega.$$

Therefore, we introduce the functions

$$s_r : \Omega \rightarrow \mathbb{R} \cup \{\infty\},$$

$$\omega \mapsto \begin{cases} \inf \{t \geq 0; \|\mathbf{Z}(t, \omega) - \mathbf{x}_0\|_2 \geq r\} & \text{if } \{t \geq 0; \|\mathbf{Z}(t, \omega) - \mathbf{x}_0\|_2 \geq r\} \neq \emptyset \\ \infty & \text{if } \{t \geq 0; \|\mathbf{Z}(t, \omega) - \mathbf{x}_0\|_2 \geq r\} = \emptyset \end{cases}$$

for every  $r > 0$ . Using  $s_r$ , it is obvious that the set of functions

$$\{\mathbf{Z}|_{s_r} : [0, s_r(\omega)) \rightarrow \mathbb{R}^n, t \mapsto \mathbf{Z}(t, \omega); \quad \omega \in \Omega\}$$

defines the unique solution of

$$\mathbf{X}(t, \omega) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}(\tau, \omega)) d\tau + \epsilon (\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)), \quad t \in [0, s_r(\omega)), \quad \omega \in \Omega.$$

Hence, we have to show that

$$\lim_{r \rightarrow \infty} s_r(\omega) = \infty \quad \text{for all } \omega \in \Omega.$$

For every  $\omega \in \Omega$ , we obtain a monotonically increasing function

$$s_\omega : [0, \infty) \rightarrow \mathbb{R} \cup \{\infty\}, \quad r \mapsto s_r(\omega).$$

Now, we assume the existence of  $\hat{\omega} \in \Omega$  such that

$$\lim_{r \rightarrow \infty} s_{\hat{\omega}}(r) = \lim_{r \rightarrow \infty} s_r(\hat{\omega}) = s < \infty.$$

In order to lead this assumption to a contradiction, we consider the function

$$k : (0, s) \rightarrow \mathbb{R}, \quad t \mapsto \frac{d}{dt} \left( \frac{1}{2} \|\mathbf{Z}(t, \hat{\omega}) - \epsilon (\mathbf{B}_t(\hat{\omega}) - \mathbf{B}_0(\hat{\omega})) + \epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_0(\hat{\omega}))\|_2^2 \right).$$

Choosing  $\bar{t} \in [0, s)$  such that

- $\|\mathbf{Z}(\bar{t}, \hat{\omega})\|_2 > \rho$  ( $\rho$  from Assumption 3.2),
- $\|\epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_t(\hat{\omega}))\|_2 < \frac{1+n\epsilon^2}{4}$  for all  $t \in [\bar{t}, s)$ ,

we obtain for all  $t \in [\bar{t}, s)$  with  $\|\mathbf{Z}(t, \hat{\omega})\|_2 > \rho$

$$\begin{aligned}
k(t) &= -(\mathbf{Z}(t, \hat{\omega}) + \epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_t(\hat{\omega})))^\top \nabla f(\mathbf{Z}(t, \hat{\omega})) \\
&= -\mathbf{Z}(t, \hat{\omega})^\top \nabla f(\mathbf{Z}(t, \hat{\omega})) - \epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_t(\hat{\omega}))^\top \nabla f(\mathbf{Z}(t, \hat{\omega})) \\
&\leq -\frac{1+n\epsilon^2}{2} \max\{1, \|\nabla f(\mathbf{Z}(t, \hat{\omega}))\|_2\} + \frac{1+n\epsilon^2}{4} \|\nabla f(\mathbf{Z}(t, \hat{\omega}))\|_2 \\
&= -\frac{1+n\epsilon^2}{4} (\max\{2, 2\|\nabla f(\mathbf{Z}(t, \hat{\omega}))\|_2\} - \|\nabla f(\mathbf{Z}(t, \hat{\omega}))\|_2) \\
&= -\frac{1+n\epsilon^2}{4} \max\{2 - \|\nabla f(\mathbf{Z}(t, \hat{\omega}))\|_2, \|\nabla f(\mathbf{Z}(t, \hat{\omega}))\|_2\} \\
&\leq -\frac{1+n\epsilon^2}{4} < 0.
\end{aligned}$$

Consequently, for all  $t \in [\bar{t}, s)$  holds:

$$\begin{aligned}
\|\mathbf{Z}(t, \hat{\omega})\|_2 &= \|\mathbf{Z}(t, \hat{\omega}) + \epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_t(\hat{\omega})) - \epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_t(\hat{\omega}))\|_2 \\
&\leq \|\mathbf{Z}(t, \hat{\omega}) + \epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_t(\hat{\omega}))\|_2 + \|\epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_t(\hat{\omega}))\|_2 \\
&= \|\mathbf{Z}(t, \hat{\omega}) - \epsilon (\mathbf{B}_t(\hat{\omega}) - \mathbf{B}_0(\hat{\omega})) + \epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_0(\hat{\omega}))\|_2 \\
&\quad + \|\epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_t(\hat{\omega}))\|_2 \\
&\leq \|\mathbf{Z}(\bar{t}, \hat{\omega})\|_2 + \max_{\bar{t} \leq t \leq s} \{\|\epsilon (\mathbf{B}_s(\hat{\omega}) - \mathbf{B}_0(\hat{\omega})) - \epsilon (\mathbf{B}_t(\hat{\omega}) - \mathbf{B}_0(\hat{\omega}))\|_2\} \\
&\quad + \frac{1+n\epsilon^2}{4}.
\end{aligned}$$

This is a contradiction to

$$\lim_{r \rightarrow \infty} \|\mathbf{Z}(s_r(\hat{\omega}), \hat{\omega}) - \mathbf{x}_0\|_2 = \infty$$

and the proof of the first part of the theorem is finished.

Now, we choose  $t \in (0, \infty)$ ,  $m \in \mathbb{N}$ , and  $t_j := j \frac{t}{m}$ ,  $j = 0, \dots, m$ .

Since

$$\mathbf{X}_t : \Omega \rightarrow \mathbb{R}^n, \quad \omega \mapsto \mathbf{X}_t(\omega)$$

is the limit of a fixed point iteration

$$\mathbf{X}_t(\omega) = \lim_{k \rightarrow \infty} \mathbf{X}_t^k(\omega)$$

with

- $\mathbf{X}_t^0 : \Omega \rightarrow \mathbb{R}^n, \quad \omega \mapsto \mathbf{x}_0,$
- $\mathbf{X}_t^k : \Omega \rightarrow \mathbb{R}^n, \quad \omega \mapsto \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau^{k-1}(\omega)) d\tau + \epsilon (\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)),$

and since

$$\int_0^t \nabla f(\mathbf{X}_\tau^{k-1}(\omega)) d\tau = \lim_{m \rightarrow \infty} \sum_{j=1}^m \nabla f(\mathbf{X}_{t_{j-1}}^{k-1}(\omega)) (t_j - t_{j-1}),$$

each function  $\mathbf{X}_t$  is  $\mathcal{B}(\Omega) - \mathcal{B}(\mathbb{R}^n)$  measurable. Existence and properties of the Lebesgue density functions  $p_t$  are proven by Lemma 3.5 in combination with the analysis of the Cauchy problem for parabolic equations (see, e.g. [Fried06], Chap. 6, Sect. 4). **q.e.d.**

From Theorem 3.3, we know that choosing any starting point  $\mathbf{x}_0 \in \mathbb{R}^n$ , the numerical computation of a path

$$\mathbf{X}_{\tilde{\omega}} : [0, \infty) \rightarrow \mathbb{R}^n, \quad t \mapsto \mathbf{X}_t(\tilde{\omega})$$

with

$$\mathbf{X}_t(\tilde{\omega}) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau(\tilde{\omega})) d\tau + \epsilon (\mathbf{B}_t(\tilde{\omega}) - \mathbf{B}_0(\tilde{\omega})), \quad t \in [0, \infty)$$

leads to a realization of  $\mathbf{X}_f$  with Lebesgue density function

$$\lambda_f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \frac{\exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right)}{\int_{\mathbb{R}^n} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x}}.$$

Furthermore, we obtain from the stability theory of stochastic differential equations (see [Has80], Chap. 3, Theorem 7.1):

**Theorem 3.6** *Let*

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad n \in \mathbb{N}, \quad f \in C^2(\mathbb{R}^n, \mathbb{R}),$$

*and let Assumption 3.2 be fulfilled. Choose any  $r > 0$  and let  $\mathbf{x}_{gl}$  be any global minimum point of  $f$ . Using*

- *The Wiener space  $(\Omega, \mathcal{B}(\Omega), W)$ ,*
- *Any  $\epsilon > 0$  from Assumption 3.2,*
- *The integral equation*

$$\mathbf{X}(t, \omega) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}(\tau, \omega)) d\tau + \epsilon (\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)), \quad t \in [0, \infty), \quad \omega \in \Omega$$

*with unique solution  $\mathbf{X}$  according to Theorem 3.3,*

- *The stopping time*

$$st : \Omega \rightarrow \mathbb{R} \cup \{\infty\},$$

$$\omega \mapsto \begin{cases} \inf \{t \geq 0; \|\mathbf{X}(t, \omega) - \mathbf{x}_{gl}\|_2 \leq r\} & \text{if } \{t \geq 0; \|\mathbf{X}(t, \omega) - \mathbf{x}_{gl}\|_2 \leq r\} \neq \emptyset \\ \infty & \text{if } \{t \geq 0; \|\mathbf{X}(t, \omega) - \mathbf{x}_{gl}\|_2 \leq r\} = \emptyset \end{cases},$$

we obtain the following results:

- (i)  $W(\{\omega \in \Omega; st(\omega) < \infty\}) = 1$ .
- (ii) For the expectation of  $st$  holds

$$\mathbb{E}(st) < \infty.$$

Theorem 3.6 shows that almost all paths of the stochastic process  $\{\mathbf{X}_t\}_{t \in [0, \infty)}$  generated by

$$\mathbf{X}_t(\omega) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau(\omega)) d\tau + \epsilon(\mathbf{B}_t(\omega) - \mathbf{B}_0(\omega)), \quad t \in [0, \infty), \quad \omega \in \Omega$$

approximate each global minimum point of  $f$  with arbitrary accuracy in finite time.

## 3.2 Concepts of Numerical Analysis

As pointed out above, the solution of

$$\mathbf{X}_t(\tilde{\omega}) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau(\tilde{\omega})) d\tau + \epsilon(\mathbf{B}_t(\tilde{\omega}) - \mathbf{B}_0(\tilde{\omega}))$$

can be interpreted as a path of a randomized curve of steepest descent, where the curve of steepest descent in integral form is given by

$$\mathbf{x}(t) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{x}(\tau)) d\tau, \quad t \in [0, \infty).$$

The balance between steepest descent represented by

$$\mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau(\tilde{\omega})) d\tau$$

and a purely random search using Gaussian distributed random variables represented by

$$\mathbf{B}_t(\tilde{\omega}) - \mathbf{B}_0(\tilde{\omega})$$

is controlled by the parameter  $\epsilon$ .

The optimal choice of  $\epsilon > 0$  (such that Assumption 3.2 is fulfilled) depends on the scaling of the objective function and has to be guided by the following considerations:

- If a long time is spent on the chosen path close to (local) minimum points of the objective function, then local minimization dominates and  $\epsilon$  has to be increased.
- If minimum points of the objective function play no significant role along the path, then purely random search dominates and  $\epsilon$  has to be decreased.

One can try to replace the parameter  $\epsilon$  with a function

$$\epsilon : [0, \infty) \rightarrow \mathbb{R}, \quad t \mapsto \epsilon(t)$$

such that

1.  $\epsilon(t) > 0$  for all  $t \geq 0$ ,
2.  $\lim_{t \rightarrow \infty} \epsilon(t) = 0$ ,

hoping that the Lebesgue density function

$$\lambda_f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \frac{\exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right)}{\int_{\mathbb{R}^n} \exp\left(-\frac{2f(\mathbf{x})}{\epsilon^2}\right) d\mathbf{x}}$$

will converge pointwise to the point measure on all global minimum points (cooling strategy). But this strategy will only work if

- $\epsilon(0) > C$ , where  $C$  is a constant depending on  $f$ .
- $\lim_{t \rightarrow \infty} \epsilon^2(t) \ln(t) > 0$ .

If one of these conditions is not fulfilled, we find only local minimum points in general (as shown in [GemHwa86]).

Since almost all paths of the stochastic process  $\{\mathbf{X}_t\}_{t \in [0, \infty)}$  are continuous but nowhere differentiable, classical strategies of step size control in numerical analysis are not applicable. Therefore, a step size control based on following principle is used:

Using a computed approximation  $\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega})$  of a path at  $\bar{t}$ , an approximation  $\mathbf{x}_{\text{app}}(\bar{t} + h, \tilde{\omega})$  is computed twice, in one step with step size  $h$  and in two steps with step sizes  $\frac{h}{2}$ . If both approximations are close enough to each other, the step is accepted; otherwise, the approach must be repeated with  $\frac{h}{2}$  instead of  $h$ . This step size control is based on the assumption that the exact solution is sufficiently close to the computed approximations, if these approximations are close enough to each other.

### 3.3 A Semi-implicit Euler Method

In this section, we investigate a semi-implicit Euler method for the numerical solution of

$$\mathbf{X}_t(\tilde{\omega}) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau(\tilde{\omega})) d\tau + \epsilon (\mathbf{B}_t(\tilde{\omega}) - \mathbf{B}_0(\tilde{\omega})), \quad t \in [0, \infty).$$

The implicit Euler method based on an approximation  $\mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega})$  of  $\mathbf{X}_{\tilde{t}}(\tilde{\omega})$  leads to a system

$$\mathbf{x}_{\text{app}}(\tilde{t} + h, \tilde{\omega}) = \mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega}) - h \nabla f(\mathbf{x}_{\text{app}}(\tilde{t} + h, \tilde{\omega})) + \epsilon (\mathbf{B}_{\tilde{t}+h}(\tilde{\omega}) - \mathbf{B}_{\tilde{t}}(\tilde{\omega}))$$

or equivalently

$$\mathbf{x}_{\text{app}}(\tilde{t} + h, \tilde{\omega}) - \mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega}) + h \nabla f(\mathbf{x}_{\text{app}}(\tilde{t} + h, \tilde{\omega})) - \epsilon (\mathbf{B}_{\tilde{t}+h}(\tilde{\omega}) - \mathbf{B}_{\tilde{t}}(\tilde{\omega})) = \mathbf{0}$$

of nonlinear equations in general. We consider the linearization of

$$\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n, \quad \mathbf{z} \mapsto \mathbf{z} - \mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega}) + h \nabla f(\mathbf{z}) - \epsilon (\mathbf{B}_{\tilde{t}+h}(\tilde{\omega}) - \mathbf{B}_{\tilde{t}}(\tilde{\omega})),$$

at  $\mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega})$ , which is given by

$$\begin{aligned} \mathbf{LF} : \mathbb{R}^n &\rightarrow \mathbb{R}^n, \\ \mathbf{z} &\mapsto h \nabla f(\mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega})) + (\mathbf{I}_n + h \nabla^2 f(\mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega}))) (\mathbf{z} - \mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega})) \\ &\quad - \epsilon (\mathbf{B}_{\tilde{t}+h}(\tilde{\omega}) - \mathbf{B}_{\tilde{t}}(\tilde{\omega})). \end{aligned}$$

Solving  $\mathbf{LF} = \mathbf{0}$  instead of  $\mathbf{F} = \mathbf{0}$  leads to

$$\begin{aligned} \mathbf{x}_{\text{app}}(\tilde{t} + h, \tilde{\omega}) &= \mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega}) \\ &\quad - \left( \frac{1}{h} \mathbf{I}_n + \nabla^2 f(\mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega})) \right)^{-1} \left( \nabla f(\mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega})) - \frac{\epsilon}{h} (\mathbf{B}_{\tilde{t}+h}(\tilde{\omega}) - \mathbf{B}_{\tilde{t}}(\tilde{\omega})) \right) \end{aligned}$$

for small enough  $h > 0$  (at least such that  $(\frac{1}{h} \mathbf{I}_n + \nabla^2 f(\mathbf{x}_{\text{app}}(\tilde{t}, \tilde{\omega})))$  is positive definite).

Since  $(\mathbf{B}_{\tilde{t}+h} - \mathbf{B}_{\tilde{t}})$  is a  $\mathcal{N}(\mathbf{0}, h \mathbf{I}_n)$  Gaussian distributed random variable, the numerical evaluation of  $\frac{\epsilon}{h} (\mathbf{B}_{\tilde{t}+h}(\tilde{\omega}) - \mathbf{B}_{\tilde{t}}(\tilde{\omega}))$  can be done by algorithmic generation of  $n$  stochastically independent  $\mathcal{N}(0, 1)$  Gaussian distributed pseudorandom numbers  $p_1, \dots, p_n \in \mathbb{R}$ . Hence, the vector  $\frac{\epsilon}{h} (\mathbf{B}_{\tilde{t}+h}(\tilde{\omega}) - \mathbf{B}_{\tilde{t}}(\tilde{\omega}))$  can be realized by

$$\frac{\epsilon}{h} (\mathbf{B}_{\tilde{t}+h}(\tilde{\omega}) - \mathbf{B}_{\tilde{t}}(\tilde{\omega})) = \frac{\epsilon}{\sqrt{h}} \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}.$$



Assume that we have computed  $\mathbf{x}_{\text{app}}(h, \tilde{\omega})$  by a step with fixed step size  $h$  based on the starting point  $\mathbf{x}_0 \in \mathbb{R}^n$  as described above:

$$\mathbf{x}_{\text{app}}(h, \tilde{\omega}) = \mathbf{x}_0 - \left( \frac{1}{h} \mathbf{I}_n + \nabla^2 f(\mathbf{x}_0) \right)^{-1} \left( \nabla f(\mathbf{x}_0) - \frac{\epsilon}{\sqrt{h}} \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \right).$$

Prima facie, the choice of  $\tilde{\omega} \in \Omega$  is left to the computer by computing  $p_1, \dots, p_n \in \mathbb{R}$ . But let us consider the set

$$\Omega_h := \{\omega \in \Omega; \mathbf{X}_h(\omega) = \mathbf{X}_h(\tilde{\omega})\}.$$

Apparently, the vector  $\mathbf{x}_{\text{app}}(h, \tilde{\omega})$  is not only an approximation of  $\mathbf{X}_h(\tilde{\omega})$  but also of  $\mathbf{X}_h(\omega)$  for all  $\omega \in \Omega_h$ . Consequently, the computation of  $p_1, \dots, p_n \in \mathbb{R}$  does not cause a determination of  $\tilde{\omega} \in \Omega$  but only a reduction of  $\Omega$  to  $\Omega_h$ .

Assume that we would like to compute  $\mathbf{x}_{\text{app}}(2h, \tilde{\omega})$  by a step with fixed step size  $h$  based on the point  $\mathbf{x}_{\text{app}}(h, \tilde{\omega})$ . Therefore, we have to compute again  $n$  stochastically independent  $\mathcal{N}(0, 1)$  Gaussian distributed pseudorandom numbers  $q_1, \dots, q_n \in \mathbb{R}$  and

$$\begin{aligned} \mathbf{x}_{\text{app}}(2h, \tilde{\omega}) &= \mathbf{x}_{\text{app}}(h, \tilde{\omega}) \\ &- \left( \frac{1}{h} \mathbf{I}_n + \nabla^2 f(\mathbf{x}_{\text{app}}(h, \tilde{\omega})) \right)^{-1} \left( \nabla f(\mathbf{x}_{\text{app}}(h, \tilde{\omega})) - \frac{\epsilon}{\sqrt{h}} \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} \right). \end{aligned}$$

Since the random variables  $(\mathbf{B}_{2h} - \mathbf{B}_h)$  and  $(\mathbf{B}_h - \mathbf{B}_0)$  are stochastically independent, it is allowed to compute  $q_1, \dots, q_n$  independent of  $p_1, \dots, p_n$ . The vector  $\mathbf{x}_{\text{app}}(2h, \tilde{\omega})$  is a numerical approximation of  $\mathbf{X}_{2h}(\omega)$  for all  $\omega \in \Omega_{2h}$  with

$$\Omega_{2h} := \{\omega \in \Omega; \mathbf{X}_h(\omega) = \mathbf{X}_h(\tilde{\omega}) \text{ and } \mathbf{X}_{2h}(\omega) = \mathbf{X}_{2h}(\tilde{\omega})\}.$$

Since the function

$$\mathbf{X}_t(\tilde{\omega}) : [0, \infty) \rightarrow \mathbb{R}^n$$

is continuous but nowhere differentiable for almost all  $\tilde{\omega} \in \Omega$ , classical strategies of step size control in numerical analysis are not applicable. Therefore, we present a step size control based on the principle introduced in the last section. Beginning with a starting value  $h_{\max}$  for  $h$  such that

$$\left( \frac{1}{h_{\max}} \mathbf{I}_n + \nabla^2 f(\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega})) \right)$$

is positive definite, we compute

$$\begin{aligned} \tilde{\mathbf{x}}\left(\bar{t} + \frac{h_{\max}}{2}\right) &:= \mathbf{x}_{\text{app}}\left(\bar{t} + \frac{h_{\max}}{2}, \tilde{\omega}\right) = \mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega}) \\ &\quad - \left(\frac{1}{\frac{h_{\max}}{2}} \mathbf{I}_n + \nabla^2 f(\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega}))\right)^{-1} \left(\nabla f(\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega})) - \frac{\epsilon}{\sqrt{\frac{h_{\max}}{2}}} \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}\right), \end{aligned}$$

$$\begin{aligned} \tilde{\mathbf{x}}^1(\bar{t} + h_{\max}) &:= \mathbf{x}_{\text{app}}^1(\bar{t} + h_{\max}, \tilde{\omega}) = \tilde{\mathbf{x}}\left(\bar{t} + \frac{h_{\max}}{2}\right) \\ &\quad - \left(\frac{1}{\frac{h_{\max}}{2}} \mathbf{I}_n + \nabla^2 f\left(\tilde{\mathbf{x}}\left(\bar{t} + \frac{h_{\max}}{2}\right)\right)\right)^{-1} \left(\nabla f\left(\tilde{\mathbf{x}}\left(\bar{t} + \frac{h_{\max}}{2}\right)\right) - \frac{\epsilon}{\sqrt{\frac{h_{\max}}{2}}} \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix}\right), \end{aligned}$$

and

$$\begin{aligned} \tilde{\mathbf{x}}^2(\bar{t} + h_{\max}) &:= \mathbf{x}_{\text{app}}^2(\bar{t} + h_{\max}, \tilde{\omega}) = \mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega}) \\ &\quad - \left(\frac{1}{h_{\max}} \mathbf{I}_n + \nabla^2 f(\tilde{\mathbf{x}}(\bar{t}))\right)^{-1} \left(\nabla f(\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega})) - \frac{\epsilon}{\sqrt{2h_{\max}}} \begin{pmatrix} p_1 + q_1 \\ \vdots \\ p_n + q_n \end{pmatrix}\right), \end{aligned}$$

The vectors  $\tilde{\mathbf{x}}^1(\bar{t} + h_{\max})$  and  $\tilde{\mathbf{x}}^2(\bar{t} + h_{\max})$  represent two numerical approximations of  $\mathbf{X}_{\bar{t}+h_{\max}}(\tilde{\omega})$ . The approximation  $\tilde{\mathbf{x}}^1(\bar{t} + h_{\max})$  is computed by two  $\frac{h_{\max}}{2}$  steps based on  $\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega})$  and using the random numbers  $p_1, \dots, p_n$  for the first step and  $q_1, \dots, q_n$  for the second step.

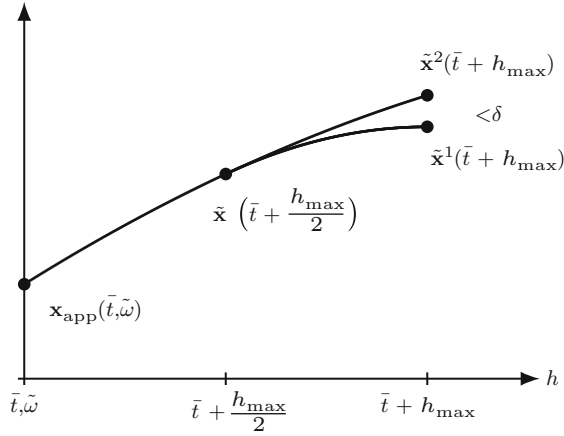
The approximation  $\tilde{\mathbf{x}}^2(\bar{t} + h_{\max})$  is computed by one  $h_{\max}$  step based on  $\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega})$  and using the random numbers  $p_1, \dots, p_n$  and  $q_1, \dots, q_n$ . This is necessary because we have to compute different approximations of the same path. Chosen any  $\delta > 0$ , we accept  $\tilde{\mathbf{x}}^1(\bar{t} + h_{\max})$  as a numerical approximation of  $\mathbf{X}_{\bar{t}+h_{\max}}(\tilde{\omega})$  if

$$\|\tilde{\mathbf{x}}^1(\bar{t} + h_{\max}) - \tilde{\mathbf{x}}^2(\bar{t} + h_{\max})\|_2 < \delta.$$

Otherwise, we have to repeat the step size procedure with  $h = \frac{h_{\max}}{2}$  (Fig. 3.4).

The following algorithm describes a semi-implicit Euler method for the computation of

$$\mathbf{X}_{\bullet}(\tilde{\omega}) : [0, \infty) \rightarrow \mathbb{R}^n, \quad t \mapsto \mathbf{X}_t(\tilde{\omega}).$$



**Fig. 3.4** Step size control

### Step 0: (Initialization)

Choose  $\mathbf{x}_0 \in \mathbb{R}^n$  and  $\epsilon, \delta > 0$ ,  
 Choose  $\text{maxit} \in \mathbb{N}$ ,  
 $j := 0$ ,  
 goto step 1.

In step 0, the starting point  $\mathbf{x}_0$ , the parameter  $\epsilon$  according to assumption 3.2, the parameter  $\delta > 0$  according to the step size control, and the maximal number of iterations have to be determined by the user.

### Step 1: (Derivatives)

$h := 1$ ,  
 compute  $\nabla f(\mathbf{x}_j), \nabla^2 f(\mathbf{x}_j)$ ,  
 goto step 2.

The initial value  $h_{\max}$  of the step size is chosen equal to 1.

### Step 2: (Pseudorandom Numbers)

Compute  $2n$  stochastically independent  $\mathcal{N}(0, 1)$  Gaussian distributed pseudorandom numbers  $p_1, \dots, p_n, q_1, \dots, q_n \in \mathbb{R}$ ,  
 goto step 3.

In this step, the choice of the path is determined successively by the computer.

### Step 3: (Cholesky Decomposition)

If  $(\frac{1}{h} \mathbf{I}_n + \nabla^2 f(\mathbf{x}_j)) \in \mathbb{R}^{n,n}$  is positive definite,

then

compute  $\mathbf{L} \in \mathbb{R}^{n,n}$  such that:

$$\mathbf{L}\mathbf{L}^\top = \left(\frac{1}{h}\mathbf{I}_n + \nabla^2 f(\mathbf{x}_j)\right) \quad (\text{Cholesky}),$$

goto step 4.

else

$$h := \frac{h}{2},$$

goto step 3.

#### Step 4: (Computation of $\mathbf{x}_{j+1}^2$ by one step with step size $h$ )

Compute  $\mathbf{x}_{j+1}^2$  by solving

$$\mathbf{L}\mathbf{L}^\top \mathbf{x}_{j+1}^2 = \left( \nabla f(\mathbf{x}_j) - \frac{\epsilon}{\sqrt{2}h} \begin{pmatrix} p_1 + q_1 \\ \vdots \\ p_n + q_n \end{pmatrix} \right),$$

$$\mathbf{x}_{j+1}^2 := \mathbf{x}_j - \mathbf{x}_{j+1}^2,$$

goto step 5.

$\mathbf{x}_{j+1}^2$  is computed by a step with starting point  $\mathbf{x}_j$  using the step size  $h$ .

#### Step 5: (Cholesky Decomposition)

Compute  $\mathbf{L} \in \mathbb{R}^{n,n}$  such that:

$$\mathbf{L}\mathbf{L}^\top = \left(\frac{2}{h}\mathbf{I}_n + \nabla^2 f(\mathbf{x}_j)\right),$$

goto step 6.

The matrix  $\left(\frac{2}{h}\mathbf{I}_n + \nabla^2 f(\mathbf{x}_j)\right)$  is obviously positive definite.

#### Step 6: (Computation of $\mathbf{x}_{\frac{h}{2}}$ )

Compute  $\mathbf{x}_{\frac{h}{2}}$  by solving

$$\mathbf{L}\mathbf{L}^\top \mathbf{x}_{\frac{h}{2}} = \left( \nabla f(\mathbf{x}_j) - \frac{\epsilon}{\sqrt{\frac{h}{2}}} \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \right),$$

$$\mathbf{x}_{\frac{h}{2}} := \mathbf{x}_j - \mathbf{x}_{\frac{h}{2}},$$

goto step 7.

$\mathbf{x}_{\frac{h}{2}}$  is computed by a step with starting point  $\mathbf{x}_j$  using the step size  $\frac{h}{2}$ .

#### Step 7: (Derivatives)

Compute  $\nabla f(\mathbf{x}_{\frac{h}{2}}), \nabla^2 f(\mathbf{x}_{\frac{h}{2}}),$

goto step 8.

#### Step 8: (Cholesky Decomposition)

If  $\left(\frac{2}{h}\mathbf{I}_n + \nabla^2 f(\mathbf{x}_{\frac{h}{2}})\right) \in \mathbb{R}^{n,n}$  is positive definite,

then

compute  $\mathbf{L} \in \mathbb{R}^{n,n}$  such that:

$$\mathbf{L}\mathbf{L}^\top = \left( \frac{2}{h} \mathbf{I}_n + \nabla^2 f \left( \mathbf{x}_{\frac{h}{2}} \right) \right) \quad (\text{Cholesky}),$$

goto step 9.

else

$$h := \frac{h}{2},$$

goto step 3.

**Step 9: (Computation of  $\mathbf{x}_{j+1}^1$  by two steps with step size  $\frac{h}{2}$ )**

Compute  $\mathbf{x}_{j+1}^1$  by solving

$$\mathbf{L}\mathbf{L}^\top \mathbf{x}_{j+1}^1 = \left( \nabla f \left( \mathbf{x}_{\frac{h}{2}} \right) - \frac{\epsilon}{\sqrt{\frac{h}{2}}} \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix} \right),$$

$$\mathbf{x}_{j+1}^1 := \mathbf{x}_{\frac{h}{2}} - \mathbf{x}_{j+1}^1,$$

goto step 10.

$\mathbf{x}_{j+1}^1$  is computed by a step with starting point  $\mathbf{x}_{\frac{h}{2}}$  using the step size  $\frac{h}{2}$ .

**Step 10: (Acceptance condition)**

If  $\|\mathbf{x}_{j+1}^1 - \mathbf{x}_{j+1}^2\|_2 < \delta$ ,

then

$$\mathbf{x}_{j+1} := \mathbf{x}_{j+1}^1,$$

print  $(j + 1, \mathbf{x}_{j+1}, f(\mathbf{x}_{j+1}))$ ,

goto step 11.

else

$$h := \frac{h}{2},$$

goto step 3.

**Step 11: (Termination condition)**

If  $j + 1 < \text{maxit}$ ,

then

$$j := j + 1,$$

goto step 1.

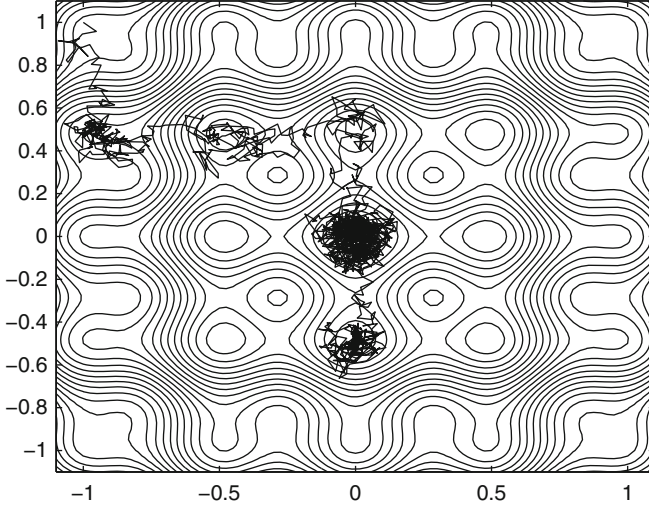
else

STOP.

The point

$$\mathbf{x}_s \in \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\text{maxit}}\}$$

with the smallest function value is chosen as a starting point for a local minimization procedure.



**Fig. 3.5** Path of the randomized curve of steepest descent and contour lines, Problem 1,  $\epsilon = 1$ , 1,500 points

Now, we apply this algorithm to eight global minimization problems. The first problem serves as an example for the visualization of the numerical results.

**Problem 1.**  $n = 2$  (Example 3.1)

$$\underset{\mathbf{x}}{\text{globmin}} \left\{ f : \mathbb{R}^2 \rightarrow \mathbb{R}, \mathbf{x} \mapsto 6x_1^2 + 6x_2^2 - \cos 12x_1 - \cos 12x_2 + 2 \right\},$$

This problem has 25 isolated minimum points within the box  $[-1, 1] \times [-1, 1]$  with six different function values. Starting at  $(-1, 1)^\top$  (very close to a local minimum point with the largest function value), Figs. 3.5–3.7 show a typical behavior of a path of the randomized curve of steepest descent (numerically approximated using the semi-implicit Euler method) in combination with

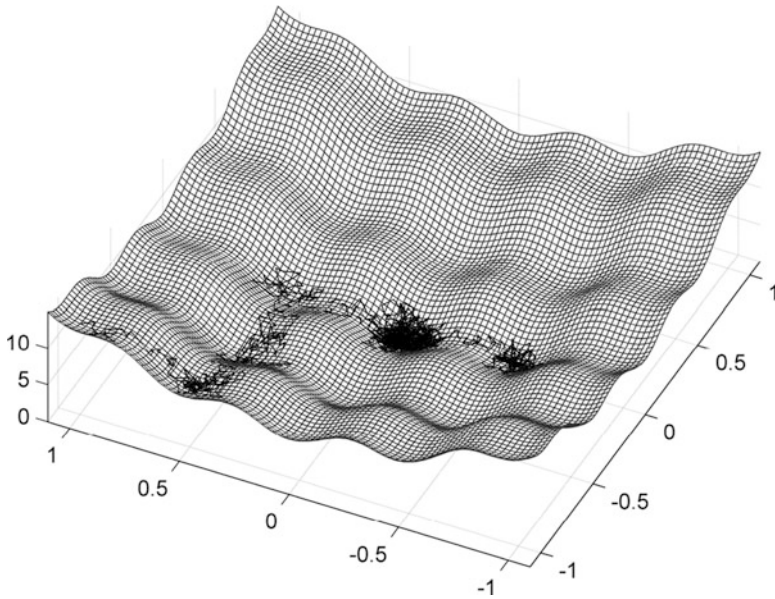
- Contour lines of the objective function
- Graph of the objective function
- Graph of the appropriate Lebesgue density function

Figure 3.8 shows a path with an  $\epsilon$  too large.

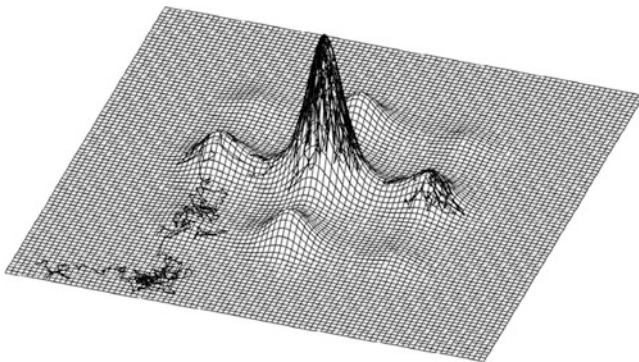
This looks like purely random search using Gaussian distributed pseudorandom vectors without minimization part.

Now, we consider an  $\epsilon$  too small (Figs. 3.9 and 3.10).

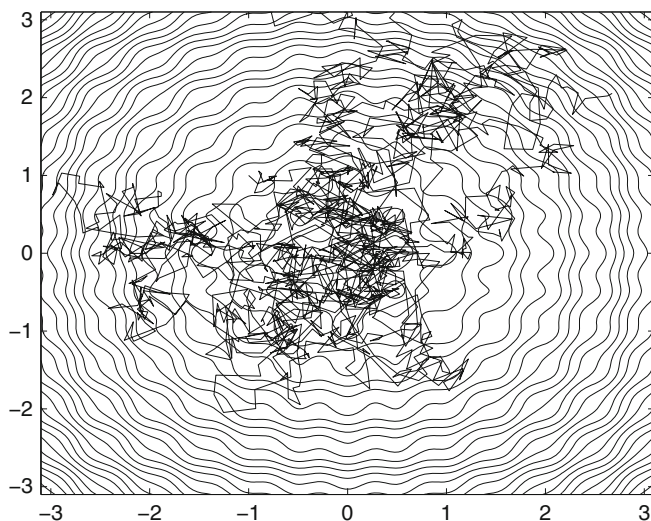
Figures 3.8, 3.9, and 3.10 show the importance of the possibility to match the free parameter  $\epsilon$  to the global minimization problem.



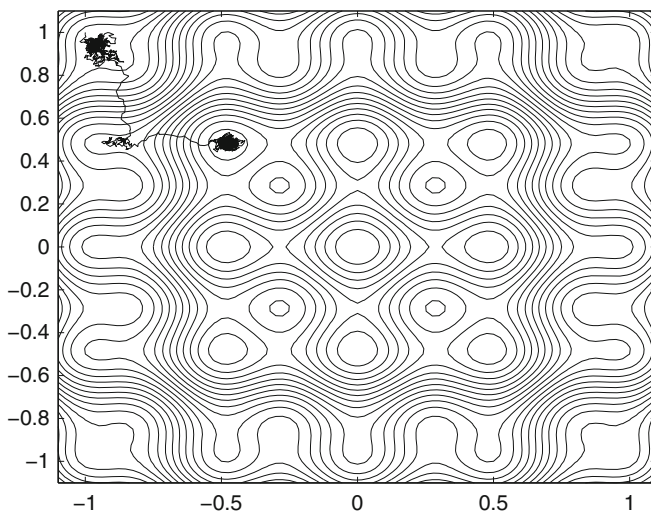
**Fig. 3.6** Path of the randomized curve of steepest descent and objective function, Problem 1,  $\epsilon = 1$ , 1,500 points



**Fig. 3.7** Path of the randomized curve of steepest descent and Lebesgue density, Problem 1,  $\epsilon = 1$ , 1,500 points

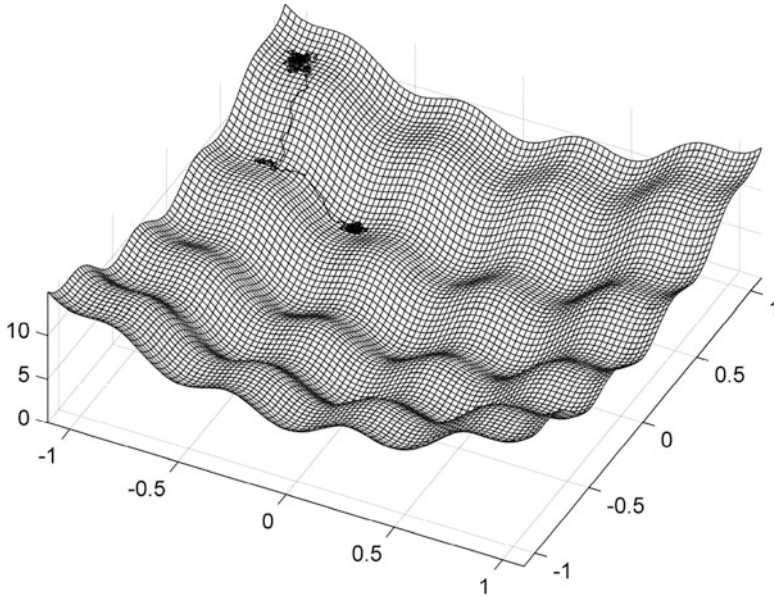


**Fig. 3.8** Path of the randomized curve of steepest descent and contour lines, Problem 1,  $\epsilon = 5$ , 1,500 points



**Fig. 3.9** Path of the randomized curve of steepest descent and contour lines, Problem 1,  $\epsilon = 0.3$ , 1,500 points





**Fig. 3.10** Path of the randomized curve of steepest descent and objective function, Problem 1,  $\epsilon = 0.3$ , 1,500 points

**Problem 2.**  $n = 90$

$$f : \mathbb{R}^{90} \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto 9.5x_1^2 - \cos(19x_1) + 792 \sum_{i=3}^{90} (3x_i + \sin^2(x_{i-1}))^2$$

$$+ \sqrt{3 + 12x_2^2 - 2\cos(12x_2) - 12x_2^2 \cos(12x_2) + 36x_2^4 - \sin^2(12x_2)}$$

This objective function has 35 isolated minimum points with function values less than 17. The unique global minimum point is given by  $\mathbf{x}_{\text{gl}} = \mathbf{0}$  with  $f(\mathbf{x}_{\text{gl}}) = 0$ .

$$\epsilon = 5, \quad \delta = 0.1.$$

Chosen number of iterations:  $\text{maxit} = 500$ .

$\mathbf{x}_0$  is chosen as a local minimum point with the largest function value.

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

**Results:**  $f(\mathbf{x}_s) = 0.1$ .

Number of gradient evaluations: 1,246.

Number of Hessian evaluations: 1,246.

Local minimization leads to the global minimum point.

**Problem 3.**  $n = 100$

$$f : \mathbb{R}^{100} \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto 6x_{100}^2 - \cos(12x_{100}) + 980 \sum_{i=1}^{97} (x_i - x_{i+2}^2)^2 \\ + 15x_{99}^2 - 2.5 \cos(12x_{99}) + 18x_{98}^2 - 3 \cos(12x_{98}) + 6.5$$

This objective function has 125 isolated minimum points with function values less than 21. The unique global minimum point is given by  $\mathbf{x}_{\text{gl}} = \mathbf{0}$  with  $f(\mathbf{x}_{\text{gl}}) = 0$ .

$$\epsilon = 3.5, \quad \delta = 0.1.$$

Chosen number of iterations:  $\text{maxit} = 3,000$ .

$\mathbf{x}_0$  is chosen as a local minimum point with the largest function value.

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

**Results:**  $f(\mathbf{x}_s) = 0.27$ .

Number of gradient evaluations: 6,160.

Number of Hessian evaluations: 6,160.

Local minimization leads to the global minimum point.

Now, we consider test problems from linear complementarity theory (see [Cottle.etal92]), which plays a very important role in game theory (computation of Nash equilibrium points (cf. [Owen68] and [Schäfer08])), and free boundary value problems (see [Crank84] and [Has.etal05]):

Given  $\mathbf{c} \in \mathbb{R}^n$  and  $\mathbf{C} \in \mathbb{R}^{n,n}$ , find any  $\mathbf{x} \in \mathbb{R}^n$  such that:

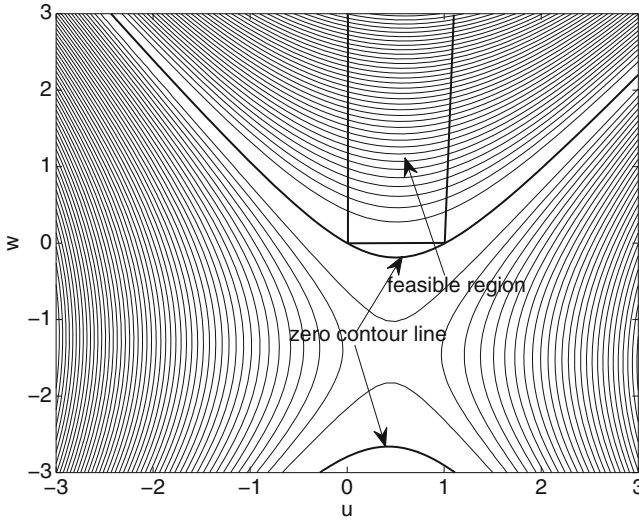
$$\begin{aligned} (\mathbf{c} + \mathbf{C}\mathbf{x})^\top \mathbf{x} &= 0 \\ x_i &\geq 0 \quad i = 1, \dots, n \\ (\mathbf{c} + \mathbf{C}\mathbf{x})_i &\geq 0 \quad i = 1, \dots, n. \end{aligned} \tag{LCP}$$

Using

$$P : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \begin{cases} x & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases},$$

the first idea to solve (LCP) may consist in the investigation of unconstrained global optimization problems of the following type:

$$\text{globmin}_{\mathbf{x}} \left\{ \mathbf{c}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{C} \mathbf{x} + \mu \left( \sum_{i=1}^n (P(-x_i))^4 + \sum_{i=1}^n (P(-(\mathbf{c} + \mathbf{C}\mathbf{x})_i))^4 \right) \right\},$$



**Fig. 3.11** Contour lines of  $f$ , LCP with two solutions,  $n = 2$

where  $\mu > 0$ . Unfortunately, the objective function

$$g : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \mathbf{c}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{C} \mathbf{x} + \mu \left( \sum_{i=1}^n (P(-x_i))^4 + \sum_{i=1}^n (P(-(\mathbf{c} + \mathbf{C}\mathbf{x})_i))^4 \right)$$

is not bounded from below in general. Therefore, we use the following objective function (again with  $\mu > 0$ ):

$$f : \mathbb{R}^n \rightarrow \mathbb{R},$$

$$\mathbf{x} \mapsto \sqrt{1 + (\mathbf{c}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{C} \mathbf{x})^2} - 1 + \mu \left( \sum_{i=1}^n (P(-x_i))^4 + \sum_{i=1}^n (P(-(\mathbf{c} + \mathbf{C}\mathbf{x})_i))^4 \right),$$

with

- $f(\mathbf{x}) \geq 0$  for all  $\mathbf{x} \in \mathbb{R}^n$ .
- $\mathbf{x}^*$  solves (LCP) iff  $f(\mathbf{x}^*) = 0$ .

*Example 3.7.* We consider two examples with  $n = 2$  visualized in Figs. 3.11–3.14:

The method of Best and Ritter [BesRit88] is chosen as a local minimization procedure of the quadratic problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{locmin}} \{ & \mathbf{c}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{C} \mathbf{x}; \quad x_i \geq 0 \quad i = 1, \dots, n \\ & (\mathbf{c} + \mathbf{C}\mathbf{x})_i \geq 0 \quad i = 1, \dots, n \}. \end{aligned}$$

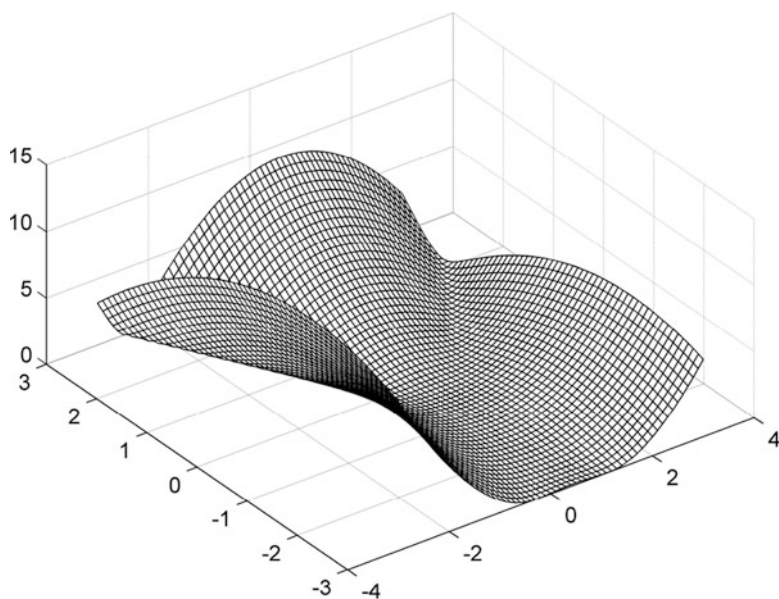


Fig. 3.12 Objective function  $f$ , LCP with two solutions,  $n = 2$

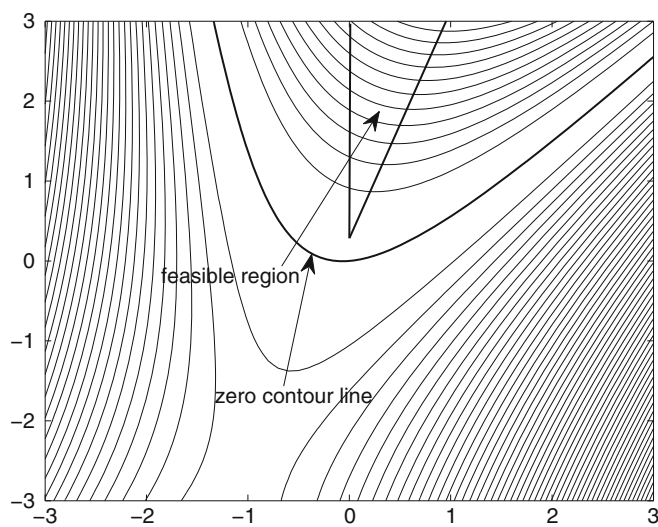
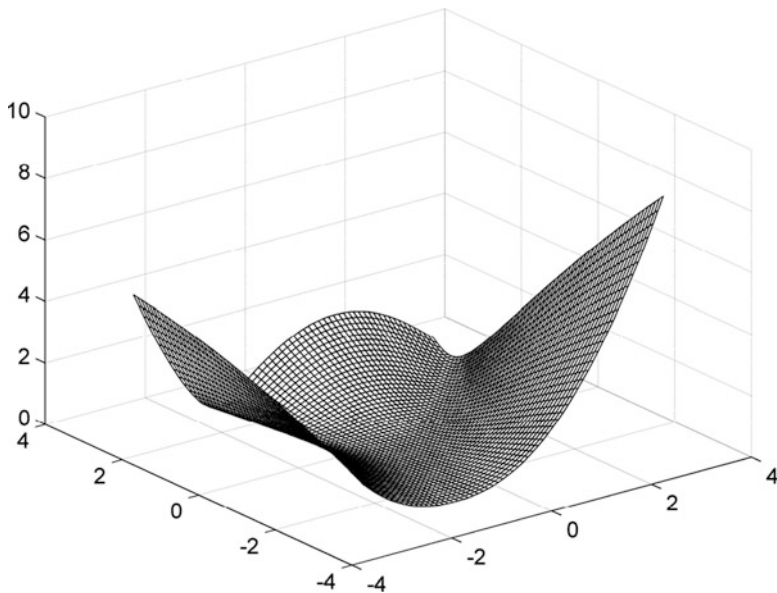


Fig. 3.13 Contour lines of  $f$ , LCP with no solution,  $n = 2$



**Fig. 3.14** Objective function  $f$ , LCP with no solution,  $n = 2$

The following numerical results can be found in [Schä95] (with  $\mu = 1,000$ ).

**Problem 4.**  $n = 30$

$x_{2k}^* > 0$  is chosen randomly,  $k = 1, \dots, 15$ .

$x_{2k-1}^* = 0$ ,  $k = 1, \dots, 15$ .

$\mathbf{C} \in \mathbb{R}^{30,30}$  is computed by pseudorandom numbers.

$\mathbf{c}_{2k} = -(\mathbf{C}\mathbf{x}^*)_{2k}$ ,  $k = 1, \dots, 15$ .

$\mathbf{c}_{2k-1}$  is chosen randomly such that  $\mathbf{c}_{2k-1} \geq -(\mathbf{C}\mathbf{x}^*)_{2k-1}$ ,  $k = 1, \dots, 15$ .

$\mathbf{x}_0$  is chosen far away from  $\mathbf{x}^*$ ,  $f(\mathbf{x}_0) = 2.8 \cdot 10^{10}$ .

$\epsilon = 2$ ,  $\delta = 0.1$ .

Chosen number of iterations:  $\text{maxit} = 1,000$ .

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

$\mathbf{x}_{gl}$  denotes the computed global minimum point.

**Results:**  $f(\mathbf{x}_s) = 20.7$ ,  $f(\mathbf{x}_{gl}) = 6.7 \cdot 10^{-5}$ .

**Problem 5.**  $n = 40$ 

$x_{2k}^* > 0$  is chosen randomly,  $k = 1, \dots, 20$ .

$x_{2k-1}^* = 0, k = 1, \dots, 20$ .

$\mathbf{C} \in \mathbb{R}^{40,40}$  is computed by pseudorandom numbers.

$\mathbf{c}_{2k} = -(\mathbf{C}\mathbf{x}^*)_{2k}, k = 1, \dots, 20$ .

$\mathbf{c}_{2k-1}$  is chosen randomly such that  $\mathbf{c}_{2k-1} \geq -(\mathbf{C}\mathbf{x}^*)_{2k-1}, k = 1, \dots, 20$ .

$\mathbf{x}_0$  is chosen far away from  $\mathbf{x}^*, f(\mathbf{x}_0) = 2.6 \cdot 10^{10}$ .

$\epsilon = 2, \quad \delta = 0.1$ .

Chosen number of iterations: maxit = 1,000.

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

$\mathbf{x}_{gl}$  denotes the computed global minimum point.

**Results:**  $f(\mathbf{x}_s) = 31.6, f(\mathbf{x}_{gl}) = 4.7 \cdot 10^{-5}$ .

**Problem 6.**  $n = 50$ 

$x_{2k}^* > 0$  is chosen randomly,  $k = 1, \dots, 25$ .

$x_{2k-1}^* = 0, k = 1, \dots, 25$ .

$\mathbf{C} \in \mathbb{R}^{50,50}$  is computed by pseudorandom numbers.

$\mathbf{c}_{2k} = -(\mathbf{C}\mathbf{x}^*)_{2k}, k = 1, \dots, 25$ .

$\mathbf{c}_{2k-1}$  is chosen randomly such that  $\mathbf{c}_{2k-1} \geq -(\mathbf{C}\mathbf{x}^*)_{2k-1}, k = 1, \dots, 25$ .

$\mathbf{x}_0$  is chosen far away from  $\mathbf{x}^*, f(\mathbf{x}_0) = 2.3 \cdot 10^{10}$ .

$\epsilon = 2, \quad \delta = 0.1$ .

Chosen number of iterations: maxit = 1,000.

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

$\mathbf{x}_{gl}$  denotes the computed global minimum point.

**Results:**  $f(\mathbf{x}_s) = 40.8, f(\mathbf{x}_{gl}) = 8.9 \cdot 10^{-6}$ .

**Problem 7.**  $n = 60$ 

$x_{2k}^* > 0$  is chosen randomly,  $k = 1, \dots, 30$ .

$x_{2k-1}^* = 0, k = 1, \dots, 30$ .

$\mathbf{C} \in \mathbb{R}^{60,60}$  is computed by pseudorandom numbers.

$$\mathbf{c}_{2k} = -(\mathbf{C}\mathbf{x}^*)_{2k}, k = 1, \dots, 30.$$

$\mathbf{c}_{2k-1}$  is chosen randomly such that  $\mathbf{c}_{2k-1} \geq -(\mathbf{C}\mathbf{x}^*)_{2k-1}, k = 1, \dots, 30.$

$\mathbf{x}_0$  is chosen far away from  $\mathbf{x}^*$ ,  $f(\mathbf{x}_0) = 3.8 \cdot 10^{10}.$

$$\epsilon = 2, \quad \delta = 0.1.$$

Chosen number of iterations: maxit = 1,000.

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

$\mathbf{x}_{gl}$  denotes the computed global minimum point.

**Results:**  $f(\mathbf{x}_s) = 54.2, f(\mathbf{x}_{gl}) = 2.7 \cdot 10^{-5}.$

**Problem 8.**  $n = 70$

$x_{2k}^* > 0$  is chosen randomly,  $k = 1, \dots, 35.$

$$x_{2k-1}^* = 0, k = 1, \dots, 35.$$

$\mathbf{C} \in \mathbb{R}^{70,70}$  is computed by pseudorandom numbers.

$$\mathbf{c}_{2k} = -(\mathbf{C}\mathbf{x}^*)_{2k}, k = 1, \dots, 35.$$

$\mathbf{c}_{2k-1}$  is chosen randomly such that  $\mathbf{c}_{2k-1} \geq -(\mathbf{C}\mathbf{x}^*)_{2k-1}, k = 1, \dots, 35.$

$\mathbf{x}_0$  is chosen far away from  $\mathbf{x}^*$ ,  $f(\mathbf{x}_0) = 4.0 \cdot 10^{10}.$

$$\epsilon = 2, \quad \delta = 0.1.$$

Chosen number of iterations: maxit = 1,000.

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

$\mathbf{x}_{gl}$  denotes the computed global minimum point.

**Results:**  $f(\mathbf{x}_s) = 66.4, f(\mathbf{x}_{gl}) = 1.2 \cdot 10^{-5}.$

A collection of linear complementarity problems discussed in the literature is published in the *Handbook of Test Problems in Local and Global Optimization* ([Flo.etal99], Chap. 10). The maximal dimension of these problems is  $n = 16.$

It is important to see that the efficiency of the semi-implicit Euler method apparently does not depend on the dimension of the problem. This is a very atypical property of a global optimization procedure.

On the other hand, the proposed approach may not work for unconstrained global minimization problems with a huge number of local minimum points as the following example shows.

*Example 3.8.* Considering

$$g_n : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \sum_{i=1}^n (4x_i^2 - \cos(8x_i) + 1),$$

Assumption 3.2 is fulfilled for all  $\epsilon > 0$  and we know that each function  $g_n$  has  $3^n$  isolated minimum points and a unique global minimum point at  $\mathbf{x} = \mathbf{0}$  (cf. Example 1.1).

The region of attraction to the unique global minimum point is approximately given by  $[-0.4, 0.4]^n$ .

Choosing  $\epsilon = \sqrt{2}$  and using the probability distribution given by the Lebesgue density function

$$\lambda_{g_n} : \mathbb{R}^n \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto \frac{\exp(-g_n(\mathbf{x}))}{\int_{\mathbb{R}^n} \exp(-g_n(\mathbf{x})) d\mathbf{x}},$$

the probability for the region of attraction of the global minimum point is

$$p_{g_n} := \frac{\int_{[-0.4, 0.4]^n} \exp(-g_n(\mathbf{x})) d\mathbf{x}}{\int_{\mathbb{R}^n} \exp(-g_n(\mathbf{x})) d\mathbf{x}} \approx 0.8^n.$$

This probability decreases exponentially with the exponential growth of the number of isolated local minimum points.

### 3.4 An Euler Method with Gradient Approximations

In this section, we investigate an Euler method for the numerical solution of

$$\mathbf{X}_t(\tilde{\omega}) = \mathbf{x}_0 - \int_0^t \nabla f(\mathbf{X}_\tau(\tilde{\omega})) d\tau + \epsilon (\mathbf{B}_t(\tilde{\omega}) - \mathbf{B}_0(\tilde{\omega})), \quad t \in [0, \infty).$$

This Euler method is based on an approximation  $\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega})$  of  $\mathbf{X}_{\bar{t}}(\tilde{\omega})$  and is given by

$$\mathbf{x}_{\text{app}}(\bar{t} + h, \tilde{\omega}) = \mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega}) - h \nabla f(\mathbf{x}_{\text{app}}(\bar{t}, \tilde{\omega})) + \epsilon (\mathbf{B}_{\bar{t}+h}(\tilde{\omega}) - \mathbf{B}_{\bar{t}}(\tilde{\omega}))$$

approximating the gradient by centered differences

$$\mathbf{D}f(\mathbf{x}(t, \tilde{\omega})) = \begin{pmatrix} \frac{f(\mathbf{x}(t, \tilde{\omega})_1 + \gamma, \mathbf{x}(t, \tilde{\omega})_2, \dots, \mathbf{x}(t, \tilde{\omega})_n) - f(\mathbf{x}(t, \tilde{\omega})_1 - \gamma, \mathbf{x}(t, \tilde{\omega})_2, \dots, \mathbf{x}(t, \tilde{\omega})_n)}{2\gamma} \\ \vdots \\ \frac{f(\mathbf{x}(t, \tilde{\omega})_1, \dots, \mathbf{x}(t, \tilde{\omega})_{n-1}, \mathbf{x}(t, \tilde{\omega})_n + \gamma) - f(\mathbf{x}(t, \tilde{\omega})_1, \dots, \mathbf{x}(t, \tilde{\omega})_{n-1}, \mathbf{x}(t, \tilde{\omega})_n - \gamma)}{2\gamma} \end{pmatrix}.$$

We arrive at the following algorithm, which uses only function values.



**Step 0: (Initialization)**

Choose  $\mathbf{x}_0 \in \mathbb{R}^n$  and  $\epsilon, \delta, \gamma > 0$ ,  
 Choose  $\text{maxit} \in \mathbb{N}$ ,  
 $j := 0$ ,  
 goto step 1.

In step 0, the starting point  $\mathbf{x}_0$ , the parameter  $\epsilon$  according to Assumption 3.2, the parameter  $\delta > 0$  according to the step size control, the parameter  $\gamma > 0$  according to the approximation of the gradient information, and the maximal number of iterations have to be determined by the user.

**Step 1: (Gradient Approximation)**

$h := 1$ ,  
 compute  $\mathbf{D}f(\mathbf{x}_j)$   
 goto step 2.

The initial value  $h_{\max}$  of the step size is chosen equal to 1.

**Step 2: (Pseudorandom Numbers)**

Compute  $2n$  stochastically independent  $\mathcal{N}(0, 1)$  Gaussian distributed pseudorandom numbers  $p_1, \dots, p_n, q_1, \dots, q_n \in \mathbb{R}$ ,  
 goto step 3.

In this step, the choice of the path is determined successively by the computer.

**Step 3: (Computation of  $\mathbf{x}_{j+1}^2$  by one step with step size  $h$ )**

$$\mathbf{x}_{j+1}^2 := \mathbf{x}_j - h\mathbf{D}f(\mathbf{x}_j) + \epsilon\sqrt{\frac{h}{2}} \begin{pmatrix} p_1 + q_1 \\ \vdots \\ p_n + q_n \end{pmatrix}.$$

goto step 4.

$\mathbf{x}_{j+1}^2$  is computed by a step with starting point  $\mathbf{x}_j$  using the step size  $h$ .

**Step 4: (Computation of  $\mathbf{x}_{\frac{h}{2}}$ )**

$$\mathbf{x}_{\frac{h}{2}} := \mathbf{x}_j - \frac{h}{2}\mathbf{D}f(\mathbf{x}_j) + \epsilon\sqrt{\frac{h}{2}} \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}.$$

goto step 5.

$\mathbf{x}_{\frac{h}{2}}$  is computed by a step with starting point  $\mathbf{x}_j$  using the step size  $\frac{h}{2}$ .

**Step 5: (Gradient Approximation)**

compute  $\mathbf{D}f(\mathbf{x}_{\frac{h}{2}})$   
goto step 6.

**Step 6: (Computation of  $\mathbf{x}_{j+1}^1$  by two steps with step size  $\frac{h}{2}$ )**

$$\mathbf{x}_{j+1}^1 := \mathbf{x}_{\frac{h}{2}} - \frac{h}{2} \mathbf{D}f(\mathbf{x}_{\frac{h}{2}}) + \epsilon \sqrt{\frac{h}{2}} \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix}.$$

goto step 7.

$\mathbf{x}_{j+1}^1$  is computed by a step with starting point  $\mathbf{x}_{\frac{h}{2}}$  using the step size  $\frac{h}{2}$ .

**Step 7: (Acceptance condition)**

If  $\|\mathbf{x}_{j+1}^1 - \mathbf{x}_{j+1}^2\|_2 < \delta$ ,  
then  
     $\mathbf{x}_{j+1} := \mathbf{x}_{j+1}^1$ ,  
    print  $(j + 1, \mathbf{x}_{j+1}, f(\mathbf{x}_{j+1}))$ ,  
    goto step 8.  
else  
     $h := \frac{h}{2}$ ,  
    goto step 3.

**Step 7: (Termination condition)**

If  $j + 1 < \text{maxit}$ ,  
then  
     $j := j + 1$ ,  
    goto step 1.  
else  
    STOP.

The point

$$\mathbf{x}_s \in \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{\text{maxit}}\}$$

with the smallest function value is chosen as a starting point for a local minimization procedure. The following numerical results are summarized in [Bar97].

**Problem 9.**  $n = 80$ 

$$f : \mathbb{R}^{80} \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto 2 + 12x_{80}^2 - 2 \cos(12x_{80}) + 720 \sum_{i=1}^{79} (x_i - \sin(\cos(x_{i+1}) - 1))^2.$$

This objective function has five isolated minimum points with function values less than 16. The unique global minimum point is given by  $\mathbf{x}_{\text{gl}} = \mathbf{0}$  with  $f(\mathbf{x}_{\text{gl}}) = 0$ .

$$\epsilon = 0.5, \quad \delta = 0.5, \quad \gamma = 10^{-6}.$$

Chosen number of iterations: maxit = 300.

$$\mathbf{x}_0 = (10, \dots, 10)^\top, \quad f(\mathbf{x}_0) = 1.1 \cdot 10^4.$$

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

$$\textbf{Results: } f(\mathbf{x}_s) = 3.8565, \quad \frac{\|\mathbf{x}_s\|_2}{80} = 0.024.$$

Number of function evaluations: 144,781.

Local minimization leads to the global minimum point.

**Problem 10.**  $n = 70$ 

$$f : \mathbb{R}^{70} \rightarrow \mathbb{R}, \quad \mathbf{x} \mapsto 1000 \sum_{i=2}^{70} (x_i - \ln(x_{i-1}^2 + 1))^2 - 1 + \\ + \sqrt{3 + 19x_1^2 - 2 \cos(19x_1) - 19x_1^2 \cos(19x_1) + 90.25x_1^4 - \sin^2(19x_1)}$$

This objective function has 7 isolated minimum points with function values less than 12. The unique global minimum point is given by  $\mathbf{x}_{\text{gl}} = \mathbf{0}$  with  $f(\mathbf{x}_{\text{gl}}) = 0$ .

$$\epsilon = 3.5, \quad \delta = 1.0, \quad \gamma = 10^{-6}.$$

Chosen number of iterations: maxit = 1,000.

$$\mathbf{x}_0 = (10, \dots, 10)^\top, \quad f(\mathbf{x}_0) = 2 \cdot 10^6.$$

$\mathbf{x}_s$  denotes the computed starting point for local minimization.

$$\textbf{Results: } f(\mathbf{x}_s) = 302.7, \quad \frac{\|\mathbf{x}_s\|_2}{70} = 0.00811.$$

Number of function evaluations: 422,681.

Local minimization leads to the global minimum point.



<http://www.springer.com/978-1-4614-3926-4>

Global Optimization

A Stochastic Approach

Schäffler, S.

2012, XVI, 148 p., Hardcover

ISBN: 978-1-4614-3926-4