Real-Parameter Black-Box Optimization Benchmarking 2010: Noisy Functions Definitions

Nikolaus Hansen[*], Steffen Finck[†], Raymond Ros[‡] and Anne Auger[§]

INRIA research report RR-6869, compiled March 24, 2012

Contents

0 Introduction **3**
0.1 Symbols and Definitions . . . 3
0.2 General Setup . . . 3
0.3 Noise Models . . . 4

1 Functions with moderate noise **6**
1.1 Sphere . . . 6
1.1.1 101 Sphere with moderate gaussian noise . . . 6
1.1.2 102 Sphere with moderate uniform noise . . . 6
1.1.3 103 Sphere with moderate seldom cauchy noise . . . 6
1.2 Rosenbrock . . . 6
1.2.1 104 Rosenbrock with moderate gaussian noise . . . 6
1.2.2 105 Rosenbrock with moderate uniform noise . . . 6
1.2.3 106 Rosenbrock with moderate seldom cauchy noise . . . 6

2 Functions with severe noise **7**
2.1 Sphere . . . 7
2.1.1 107 Sphere with gaussian noise . . . 7
2.1.2 108 Sphere with uniform noise . . . 7
2.1.3 109 Sphere with seldom cauchy noise . . . 7
2.2 Rosenbrock . . . 7
2.2.1 110 Rosenbrock with gaussian noise . . . 7
2.2.2 111 Rosenbrock with uniform noise . . . 7
2.2.3 112 Rosenbrock with seldom cauchy noise . . . 7
2.3 Step ellipsoid . . . 8
2.3.1 113 Step ellipsoid with gaussian noise . . . 8
2.3.2 114 Step ellipsoid with uniform noise . . . 8
2.3.3 115 Step ellipsoid with seldom cauchy noise . . . 8
2.4 Ellipsoid . . . 8
2.4.1 116 Ellipsoid with gaussian noise . . . 8

[*] NH is with the TAO Team of INRIA Saclay–Île-de-France at the LRI, Université-Paris Sud, 91405 Orsay cedex, France

[†] SF is with the Research Center PPE, University of Applied Science Vorarlberg, Hochschulstrasse 1, 6850 Dornbirn, Austria

[‡] RR is with the TAO Team of INRIA Saclay–Île-de-France at the LRI, Université-Paris Sud, 91405 Orsay cedex, France

[§] AA is with the TAO Team of INRIA Saclay–Île-de-France at the LRI, Université-Paris Sud, 91405 Orsay cedex, France

2.4.2 117 Ellipsoid with uniform noise 8
2.4.3 118 Ellipsoid with seldom cauchy noise 8
2.5 Different Powers 9
2.5.1 119 Different Powers with gaussian noise 9
2.5.2 120 Different Powers with uniform noise 9
2.5.3 121 Different Powers with seldom cauchy noise 9

3 Highly multi-modal functions with severe noise **9**
3.1 Schaffer's F7 9
3.1.1 122 Schaffer's F7 with gaussian noise 9
3.1.2 123 Schaffer's F7 with uniform noise 9
3.1.3 124 Schaffer's F7 with seldom cauchy noise 9
3.2 Composite Griewank-Rosenbrock 10
3.2.1 125 Composite Griewank-Rosenbrock with gaussian noise 10
3.2.2 126 Composite Griewank-Rosenbrock with uniform noise 10
3.2.3 127 Composite Griewank-Rosenbrock with seldom cauchy noise 10
3.3 Gallagher's Gaussian Peaks, globally rotated 10
3.3.1 128 Gallagher's Gaussian Peaks 101-me with gaussian noise 10
3.3.2 129 Gallagher's Gaussian Peaks 101-me with uniform noise 11
3.3.3 130 Gallagher's Gaussian Peaks 101-me with seldom cauchy noise 11

A Utility Functions **12**

0 Introduction

This document is based on the BBOB 2009 noisy function definition document [?]. In the following the benchmark functions with noise are defined (for a graphical presentation see [?]). For a general motivation of the choice of functions please refer to [?].

0.1 Symbols and Definitions

Used symbols and definitions of, e.g., auxiliary functions are given in the following. Vectors are typeset in bold and refer to column vectors.

$\otimes$ indicates element-wise multiplication of two D-dimensional vectors, $\otimes : \mathcal{R}^D \times \mathcal{R}^D \to \mathcal{R}^D, (\mathbf{x}, \mathbf{y}) \mapsto \mathrm{diag}(\mathbf{x}) \times \mathbf{y} = (x_i \times y_i)_{i=1,\dots,D}$

$\|.\|$ denotes the Euclidean norm, $\|\mathbf{x}\|^2 = \sum_i x_i^2$.

$[.]$ denotes the nearest integer value

$\mathbf{0}$ $= (0, \dots, 0)^{\mathrm{T}}$ all zero vector

$\mathbf{1}$ $= (1, \dots, 1)^{\mathrm{T}}$ all one vector

Λ^α is a diagonal matrix in D dimensions with the ith diagonal element as $\lambda_{ii} = \alpha^{\frac{1}{2}\frac{i-1}{D-1}}$

f_{pen} : $\mathcal{R}^D \to \mathcal{R}$, $\mathbf{x} \mapsto 100 \sum_{i=1}^{D} \max(0, |x_i| - 5)^2$

$\mathbf{1}_-^+$ a D-dimensional vector with entries of -1 or 1 both drawn equal probability.

Q, R orthogonal (rotation) matrices. For each function instance in each dimension a single realization for respectively $\mathbf{Q}$ and $\mathbf{R}$ is used. Rotation matrices are generated from standard normally distributed entries by Gram-Schmidt orthogonalization. Columns and rows of a rotation matrix form an orthogonal basis.

R see **Q**

T_{asy}^{β} : $\mathcal{R}^D \to \mathcal{R}^D$, $x_i \mapsto \begin{cases} x_i^{1+\beta \frac{i-1}{D-1}\sqrt{x_i}} & \text{if } x_i > 0 \\ x_i & \text{otherwise} \end{cases}$, for $i = 1, \dots, D$.

T_{osz} : $\mathcal{R}^n \to \mathcal{R}^n$, for any positive integer n, maps element-wise

$$x \mapsto \mathrm{sign}(x) \exp\left(\hat{x} + 0.049\left(\sin(c_1 \hat{x}) + \sin(c_2 \hat{x})\right)\right)$$

with $\hat{x} = \begin{cases} \log(|x|) & \text{if } x \neq 0 \\ \in \mathcal{R} & \text{otherwise} \end{cases}$, $\mathrm{sign}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{otherwise} \end{cases}$, $c_1 = \begin{cases} 10 & \text{if } x > 0 \\ 5.5 & \text{otherwise} \end{cases}$ and $c_2 = \begin{cases} 7.9 & \text{if } x > 0 \\ 3.1 & \text{otherwise} \end{cases}$

$\mathbf{x}^{\mathrm{opt}}$ optimal solution vector, such that $f(\mathbf{x}^{\mathrm{opt}})$ is minimal.

0.2 General Setup

Search Space All functions are defined and can be evaluated over $\mathcal{R}^D$, while the actual search domain is given as $[-5, 5]^D$.

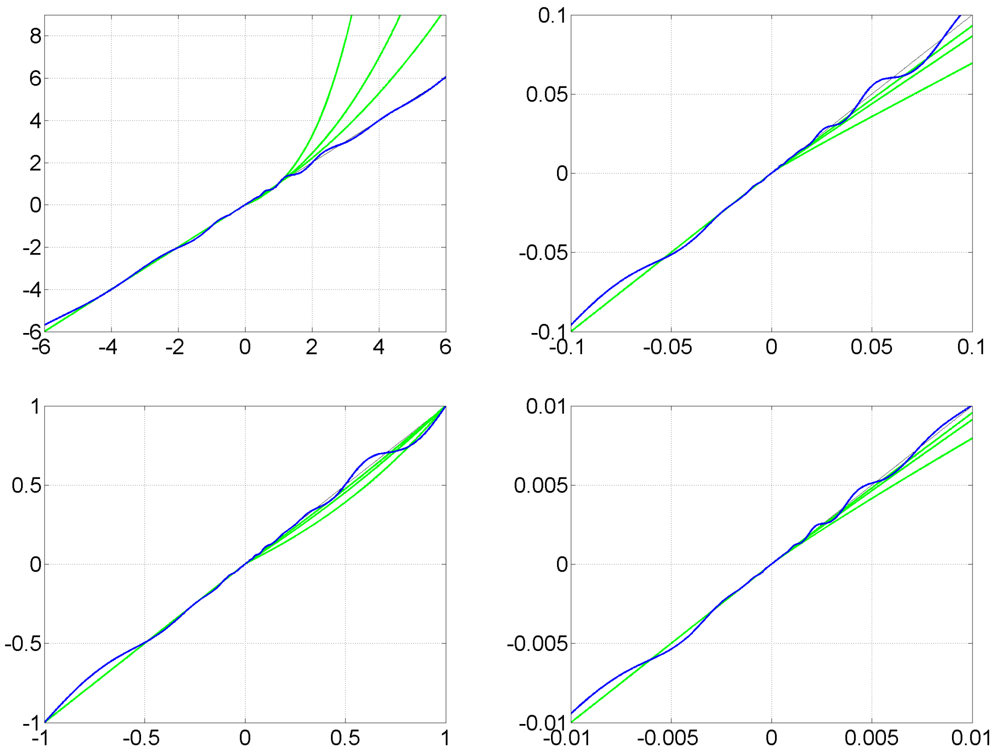

Figure 1: T_{osz} (blue) and D-th coordinate of T_{asy} for $\beta = 0.1, 0.2, 0.5$ (green)

Location of the optimal $\mathbf{x}^{\mathrm{opt}}$ and of $f_{\mathrm{opt}} = f(\mathbf{x}^{\mathrm{opt}})$ All functions have their global optimum in $[-5, 5]^D$. The majority of functions has the global optimum in $[-4, 4]^D$. The value for f_{opt} is drawn from a cauchy distributed random variable, with roughly 50% of the values between -100 and 100. The value is rounded after two decimal places and the maximum and minimum are set to 1000 and -1000 respectively. In the function definitions a transformed variable vector $\mathbf{z}$ is often used instead of the argument $\mathbf{x}$. The vector $\mathbf{z}$ has its optimum in $\mathbf{z}^{\mathrm{opt}} = \mathbf{0}$, if not stated otherwise.

Boundary Handling On all functions a penalty boundary handling is applied as given with f_{pen} (see section 0.1).

Linear Transformations Linear transformations of the search space are applied to derive non-separable functions from separable ones and to control the conditioning of the function.

Non-Linear Transformations and Symmetry Breaking In order to make relatively simple, but well understood functions less regular, on some functions non-linear transformations are applied in x- or f-space. Both transformations $T_{\mathrm{osz}} : \mathcal{R}^n \to \mathcal{R}^n$, $n \in \{1, D\}$, and $T_{\mathrm{asy}} : \mathcal{R}^D \to \mathcal{R}^D$ are defined coordinate-wise (see Section 0.1). They are smooth and have, coordinate-wise, a strictly positive derivative. They are shown in Figure 1. T_{osz} is oscillating about the identity, where the oscillation is scale invariant w.r.t. the origin. T_{asy} is the identity for negative values. When T_{asy} is applied, a portion of $1/2^D$ of the search space remains untransformed.

0.3 Noise Models

In this benchmarking suite three different noise models are used. The first two, $f_{\mathrm{GN}}\left(\right)$ and $f_{\mathrm{UN}}\left(\right)$, are multiplicative noise models while the third model, $f_{\mathrm{CN}}\left(\right)$, is an additive noise model. All noise models are applied to a function value f under the assumption that $f \geq 0$. All noise models reveal *stochastic dominance* between any two solutions and are therefore utility-free (see Appendix A).

Gaussian Noise The Gaussian noise model is scale invariant and defined as

$$f_{\mathrm{GN}}\left(f, \beta\right) = f \times \exp(\beta\, \mathcal{N}(0,1)) \ . \tag{1}$$

The noise strength is controlled with β. The distribution of the noise is log-normal, thus no negative noise values can be sampled. For the benchmark functions with moderate noise $\beta = 0.01$, otherwise $\beta = 1$. For small values of β this noise model resembles $f \times (1 + \beta\, \mathcal{N}(0,1))$.

Uniform Noise The uniform noise model is introduced as a more severe noise model than the Gaussian and is defined as

$$f_{\mathrm{UN}}\left(f, \alpha, \beta\right) = f \times \mathcal{U}(0,1)^{\beta} \max\left(1, \left(\frac{10^9}{f+\epsilon}\right)^{\alpha\, \mathcal{U}(0,1)}\right) \ . \tag{2}$$

The noise model uses two random factors. The first factor is in the interval $[0, 1]$, uniformly distributed for $\beta = 1$. The second factor, $\max(\dots)$, is ≥ 1. The parameters α and β control the noise strength. For moderate noise $\alpha = 0.01\,(0.49 + 1/D)$ and $\beta = 0.01$, otherwise $\alpha = 0.49 + 1/D$ and $\beta = 1$. Furthermore, ϵ is set to 10^{-99} in order to prevent division by zero.

The uniform noise model is not scale invariant. Due to the last factor in Eq. (2) the noise strength increases with decreasing (positive) value of f. Therefore the noise strength becomes more severe when approaching the optimum.

Cauchy Noise The Cauchy noise model represents a different type of noise with two important aspects. First, only a comparatively small percentage of function values is disturbed by noise. Second, the noise distribution is comparatively "weird". Large outliers occur once in a while, and because they stem from a continuous distribution they cannot be easily diagnosed. The Cauchy noise model is defined as

$$f_{\mathrm{CN}}\left(f, \alpha, p\right) = f + \alpha \max\left(0, 1000 + \mathbb{I}_{\{\mathcal{U}(0,1)<p\}} \frac{\mathcal{N}(0,1)}{|\mathcal{N}(0,1)| + \epsilon}\right) \ , \tag{3}$$

where α defines the noise strength and p determines the frequency of the noise disturbance. In the moderate noise case $\alpha = 0.01$ and $p = 0.05$, otherwise $\alpha = 1$ and $p = 0.2$. The summand of 1000 was chosen to sample positive and negative "outliers" (as the function value is cut from below, see next paragraph) and ϵ is set to 10^{-199}.

Final Function Value In order to achieve a convenient testing for the target function value, in all noise models 1.01×10^{-8} is added to the function value and, if the input argument f is smaller than 10^{-8}, the undisturbed f is returned.

$$f_{\mathrm{XX}}(f, \dots) \leftarrow \begin{cases} f_{\mathrm{XX}}(f, \dots) + 1.01 \times 10^{-8} & \text{if } f \geq 10^{-8} \\ f & \text{otherwise} \end{cases} \tag{4}$$

1 Functions with moderate noise

1.1 Sphere

$$f_{\text{sphere}}(\mathbf{x}) = \|\mathbf{z}\|^2$$

- $\mathbf{z} = \mathbf{x} - \mathbf{x}^{\text{opt}}$

Properties Presumably the most easy continuous domain search problem, given the volume of the searched solution is small (i.e. where pure monte-carlo random search is too expensive).

- unimodal
- highly symmetric, in particular rotationally invariant

1.1.1 101 Sphere with moderate gaussian noise

$$f_{101}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{sphere}}(\mathbf{x}), 0.01\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{101}$$

1.1.2 102 Sphere with moderate uniform noise

$$f_{102}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{sphere}}(\mathbf{x}), 0.01\left(0.49 + \frac{1}{D}\right), 0.01\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{102}$$

1.1.3 103 Sphere with moderate seldom cauchy noise

$$f_{103}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{sphere}}(\mathbf{x}), 0.01, 0.05\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{103}$$

1.2 Rosenbrock

$$f_{\text{rosenbrock}}(\mathbf{x}) = \sum_{i=1}^{D-1} 100\left(z_i^2 - z_{i+1}\right)^2 + (z_i - 1)^2$$

- $\mathbf{z} = \max\left(1, \frac{\sqrt{D}}{8}\right)(\mathbf{x} - \mathbf{x}^{\text{opt}}) + 1$
- $\mathbf{z}^{\text{opt}} = \mathbf{1}$

Properties So-called banana function due to its 2-D contour lines as a bent ridge (or valley). In the beginning, the prominent first term of the function definition attracts to the point $\mathbf{z} = \mathbf{0}$. Then, a long bending valley needs to be followed to reach the global optimum. The ridge changes its orientation $D - 1$ times.

- in larger dimensions the function has a local optimum with an attraction volume of about 25%

1.2.1 104 Rosenbrock with moderate gaussian noise

$$f_{104}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{rosenbrock}}(\mathbf{x}), 0.01\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{104}$$

1.2.2 105 Rosenbrock with moderate uniform noise

$$f_{105}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{rosenbrock}}(\mathbf{x}), 0.01\left(0.49 + \frac{1}{D}\right), 0.01\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{105}$$

1.2.3 106 Rosenbrock with moderate seldom cauchy noise

$$f_{106}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{rosenbrock}}(\mathbf{x}), 0.01, 0.05\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{106}$$

2 Functions with severe noise

2.1 Sphere

$$f_{\text{sphere}}(\mathbf{x}) = \|\mathbf{z}\|^2$$

- $\mathbf{z} = \mathbf{x} - \mathbf{x}^{\text{opt}}$

Properties Presumably the most easy continuous domain search problem, given the volume of the searched solution is small (i.e. where pure monte-carlo random search is too expensive).

- unimodal
- highly symmetric, in particular rotationally invariant

2.1.1 107 Sphere with gaussian noise

$$f_{107}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{sphere}}(\mathbf{x}), 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{107}$$

2.1.2 108 Sphere with uniform noise

$$f_{108}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{sphere}}(\mathbf{x}), 0.49 + \frac{1}{D}, 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{108}$$

2.1.3 109 Sphere with seldom cauchy noise

$$f_{109}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{sphere}}(\mathbf{x}), 1, 0.2\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{109}$$

2.2 Rosenbrock

$$f_{\text{rosenbrock}}(\mathbf{x}) = \sum_{i=1}^{D-1} 100\left(z_i^2 - z_{i+1}\right)^2 + (z_i - 1)^2$$

- $\mathbf{z} = \max\left(1, \frac{\sqrt{D}}{8}\right)(\mathbf{x} - \mathbf{x}^{\text{opt}}) + 1$
- $\mathbf{z}^{\text{opt}} = \mathbf{1}$

Properties So-called banana function due to its 2-D contour lines as a bent ridge (or valley). In the beginning, the prominent first term of the function definition attracts to the point $\mathbf{z} = \mathbf{0}$. Then, a long bending valley needs to be followed to reach the global optimum. The ridge changes its orientation $D - 1$ times.

- a local optimum with an attraction volume of about 25%

2.2.1 110 Rosenbrock with gaussian noise

$$f_{110}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{rosenbrock}}(\mathbf{x}), 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{110}$$

2.2.2 111 Rosenbrock with uniform noise

$$f_{111}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{rosenbrock}}(\mathbf{x}), 0.49 + \frac{1}{D}, 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{111}$$

2.2.3 112 Rosenbrock with seldom cauchy noise

$$f_{112}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{rosenbrock}}(\mathbf{x}), 1, 0.2\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{112}$$

2.3 Step ellipsoid

$$f_{\text{step}}(\mathbf{x}) = 0.1 \max\left(|\hat{z}_1|/10^4,\, \sum_{i=1}^{D} 10^{2\frac{i-1}{D-1}} z_i^2\right)$$

- $\hat{\mathbf{z}} = \Lambda^{10}\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}})$
- $\tilde{z}_i = \begin{cases} \lfloor 0.5 + \hat{z}_i \rfloor & \text{if } \hat{z}_i > 0.5 \\ \lfloor 0.5 + 10\,\hat{z}_i \rfloor / 10 & \text{otherwise} \end{cases}$ for $i = 1, \ldots, D$,
  denotes the rounding procedure in order to produce the plateaus.
- $\mathbf{z} = \mathbf{Q}\tilde{\mathbf{z}}$

Properties The function consists of many plateaus of different sizes. Apart from a small area close to the global optimum, the gradient is zero almost everywhere.

- condition number is about 100

2.3.1 113 Step ellipsoid with gaussian noise

$$f_{113}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{step}}(\mathbf{x}), 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{113}$$

2.3.2 114 Step ellipsoid with uniform noise

$$f_{114}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{step}}(\mathbf{x}), 0.49 + \frac{1}{D}, 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{114}$$

2.3.3 115 Step ellipsoid with seldom cauchy noise

$$f_{115}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{step}}(\mathbf{x}), 1, 0.2\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{115}$$

2.4 Ellipsoid

$$f_{\text{ellipsoid}}(\mathbf{x}) = \sum_{i=1}^{D} 10^{4\frac{i-1}{D-1}} z_i^2$$

- $\mathbf{z} = T_{\text{osz}}(\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}}))$

Properties Globally quadratic ill-conditioned function with smooth local irregularities.

- condition number is 10^4

2.4.1 116 Ellipsoid with gaussian noise

$$f_{116}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{ellipsoid}}(\mathbf{x}), 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{116}$$

2.4.2 117 Ellipsoid with uniform noise

$$f_{117}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{ellipsoid}}(\mathbf{x}), 0.49 + \frac{1}{D}, 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{117}$$

2.4.3 118 Ellipsoid with seldom cauchy noise

$$f_{118}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{ellipsoid}}(\mathbf{x}), 1, 0.2\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{118}$$

2.5 Different Powers

$$f_{\text{diffpowers}}(\mathbf{x}) = \sqrt{\sum_{i=1}^{D} |z_i|^{2+4\frac{i-1}{D-1}}}$$

- $\mathbf{z} = \mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}})$

Properties Due to the different exponents the sensitivies of the z_i-variables become more and more different when approaching the optimum.

2.5.1 119 Different Powers with gaussian noise

$$f_{119}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{diffpowers}}(\mathbf{x}), 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{119}$$

2.5.2 120 Different Powers with uniform noise

$$f_{120}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{diffpowers}}(\mathbf{x}), 0.49 + \frac{1}{D}, 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{120}$$

2.5.3 121 Different Powers with seldom cauchy noise

$$f_{121}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{diffpowers}}(\mathbf{x}), 1, 0.2\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{121}$$

3 Highly multi-modal functions with severe noise

3.1 Schaffer's F7

$$f_{\text{schaffer}}(\mathbf{x}) = \left(\frac{1}{D-1}\sum_{i=1}^{D-1} \sqrt{s_i} + \sqrt{s_i}\sin^2\left(50\, s_i^{1/5}\right)\right)^2$$

- $\mathbf{z} = \Lambda^{10}\mathbf{Q}\, T_{\text{asy}}^{0.5}(\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}}))$
- $s_i = \sqrt{z_i^2 + z_{i+1}^2}$ for $i = 1, \ldots, D$

Properties A highly multimodal function where frequency and amplitude of the modulation vary.

- conditioning is low

3.1.1 122 Schaffer's F7 with gaussian noise

$$f_{122}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{schaffer}}(\mathbf{x}), 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{122}$$

3.1.2 123 Schaffer's F7 with uniform noise

$$f_{123}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{schaffer}}(\mathbf{x}), 0.49 + \frac{1}{D}, 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{123}$$

3.1.3 124 Schaffer's F7 with seldom cauchy noise

$$f_{124}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{schaffer}}(\mathbf{x}), 1, 0.2\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{124}$$

3.2 Composite Griewank-Rosenbrock

$$f_{\text{f8f2}}(\mathbf{x}) = \frac{1}{D-1}\sum_{i=1}^{D-1}\left(\frac{s_i}{4000} - \cos(s_i)\right) + 1$$

- $\mathbf{z} = \max\left(1, \frac{\sqrt{D}}{8}\right)\mathbf{R}\mathbf{x} + 0.5$
- $s_i = 100\,(z_i^2 - z_{i+1})^2 + (z_i - 1)^2$ for $i = 1, \ldots, D$
- $\mathbf{z}^{\text{opt}} = \mathbf{1}$

Properties Resembling the Rosenbrock function in a highly multimodal way.

3.2.1 125 Composite Griewank-Rosenbrock with gaussian noise

$$f_{125}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{f8f2}}(\mathbf{x}), 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{125}$$

3.2.2 126 Composite Griewank-Rosenbrock with uniform noise

$$f_{126}(\mathbf{x}) = f_{\text{UN}}\left(f_{\text{f8f2}}(\mathbf{x}), 0.49 + \frac{1}{D}, 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{126}$$

3.2.3 127 Composite Griewank-Rosenbrock with seldom cauchy noise

$$f_{127}(\mathbf{x}) = f_{\text{CN}}\left(f_{\text{f8f2}}(\mathbf{x}), 1, 0.2\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{127}$$

3.3 Gallagher's Gaussian Peaks, globally rotated

$$f_{\text{gallagher}}(\mathbf{x}) = T_{\text{osz}}\left(10 - \max_{i=1}^{101} w_i \exp\left(-\frac{1}{2D}\left(\mathbf{x} - \mathbf{y}_i\right)^{\text{T}}\mathbf{R}^{\text{T}}\mathbf{C}_i\mathbf{R}(\mathbf{x} - \mathbf{y}_i)\right)\right)^2$$

- $w_i = \begin{cases} 1.1 + 8 \times \dfrac{i-2}{99} & \text{for } i = 2, \ldots, 101 \\ 10 & \text{for } i = 1 \end{cases}$, three optima have a value larger than 9
- $\mathbf{C}_i = \Lambda^{\alpha_i}/\alpha_i^{1/4}$ where Λ^{α_i} is defined as usual (see Section 0.1), but with randomly permuted diagonal elements. For $i = 2, \ldots, 101$, α_i is drawn uniformly randomly from the set $\left\{1000^{2\frac{j}{99}} \mid j = 0, \ldots, 99\right\}$ without replacement, and $\alpha_i = 1000$ for $i = 1$.
- the local optima $\mathbf{y}_i$ are uniformly drawn from the domain $[-4.9, 4.9]^D$ for $i = 2, \ldots, 101$ and $\mathbf{y}_1 \in [-4, 4]^D$. The global optimum is at $\mathbf{x}^{\text{opt}} = \mathbf{y}_1$.

Properties The function consists of 101 optima with position and height being unrelated and randomly chosen.

- condition number around the global optimum is about 30
- same overall rotation matrix

3.3.1 128 Gallagher's Gaussian Peaks 101-me with gaussian noise

$$f_{128}(\mathbf{x}) = f_{\text{GN}}\left(f_{\text{gallagher}}(\mathbf{x}), 1\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{128}$$

3.3.2 129 Gallagher's Gaussian Peaks 101-me with uniform noise

$$f_{129}(\mathbf{x}) = f_{\mathrm{UN}}\left(f_{\mathrm{gallagher}}(\mathbf{x}), 0.49 + \frac{1}{D}, 1\right) + f_{\mathrm{pen}}(\mathbf{x}) + f_{\mathrm{opt}} \quad (129)$$

3.3.3 130 Gallagher's Gaussian Peaks 101-me with seldom cauchy noise

$$f_{130}(\mathbf{x}) = f_{\mathrm{CN}}\left(f_{\mathrm{gallagher}}(\mathbf{x}), 1, 0.2\right) + f_{\mathrm{pen}}(\mathbf{x}) + f_{\mathrm{opt}} \quad (130)$$

Acknowledgments

The authors would like to thank Hans-Georg Beyer for his constructive comments.
Steffen Finck was supported by the Austrian Science Fund (FWF) under grant P19069-N18.

APPENDIX

A Utility Functions

The objective when optimizing a noisy function is not immediately obvious. Function values are noisy by definition and the ordering of solutions is not unique, but depends on the realized random values. For this reason, often a utility is defined, which is a *deterministic* function over the search space. Let $F : \mathcal{R}^D \to \mathcal{R}$ be the noisy function, i.e. for each $\mathbf{x} \in \mathcal{R}^D$, $F(\mathbf{x})$ is a random variable with values in $\mathcal{R}$. Then a utility maps each $F(\mathbf{x})$ to a real value and therefore, equivalently maps each $\mathbf{x}$ to this value. A typical utility is $\mathbf{x} \mapsto E(F(\mathbf{x}))$, where E denotes the expected value (the expectation is taken over $F(\mathbf{x})$ for a single given $\mathbf{x}$). The objective can now be stated in that the utility shall be minimized. Most search procedures, like evolutionary algorithms, do not optimize for the expected value. Therefore, evaluating their performance with respect to $E(F(\mathbf{x}))$ could render benchmarking result rather meaningless.

In the given testbed, we have chosen to implement **utility-free** noisy functions. We call them utility-free, because any utility that is compliant with the minimization task will eventually lead to the same result: the ordering between solutions will remain the same for all compliant utility functions and in particular the global optimum will be the same. This property is provided by *stochastic dominance* between the F-distributions of any two solutions, $\mathbf{x}$ and $\mathbf{y}$. For the cumulative distribution functions P of $F(\mathbf{x})$ and $F(\mathbf{y})$ holds either for all $z \in \mathcal{R}$ is $P_{F(\mathbf{x})}(z) < P_{F(\mathbf{y})}(z)$,[1] or for all $z \in \mathcal{R}$ is $P_{F(\mathbf{x})}(z) > P_{F(\mathbf{y})}(z)$, or $F(\mathbf{x})$ and $F(\mathbf{y})$ have the same distribution. Some compliant utility functions are given.

- $\mathbf{x} \mapsto E(F(\mathbf{x}))$, the expected value of $F(\mathbf{x})$.
- $\mathbf{x} \mapsto E(g(F(\mathbf{x})))$, for any stricly increasing $g : \mathcal{R} \to \mathcal{R}$, is a generalization of the previous case.
- any distribution percentile of $F(\mathbf{x})$.
- let $P_p(F(\mathbf{x}))$ be the p%-percentile of the distribution of $F(\mathbf{x})$, then $\int_{p=0}^{100} w(p) P_p(F(\mathbf{x}))$ is another possible utility, for any non-trivial $w : p \mapsto w(p) \geq 0$. When w is dirac, the previous case is recovered.

The advantage of a utility-free testbed is that any minimization algorithm, independent of its explicitely or implicitely build-in utility, can be reasonably evaluated.

[1] $P_{F(\mathbf{x})}(z)$ denotes the probability that $F(\mathbf{x}) \leq z$.