Real-Parameter Black-Box Optimization Benchmarking 2010: Presentation of the Noiseless Functions

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

Report 2009/20, Research Center PPE, re-compiled February 18, 2019

Contents

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

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

‡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

Errata

Errata (2019) are marked with colored text.

0 Introduction

This document is based on the BBOB 2009 function document [1]. In the following, 24 noise-free real-parameter single-objective benchmark functions are presented.[1] Our intention behind the selection of benchmark functions was to evaluate the performance of algorithms with regard to typical difficulties which we believe occur in continuous domain search. We hope that the function collection reflects, at least to a certain extend and with a few exceptions, a more difficult portion of the problem distribution that will be seen in practice (easy functions are evidently of lesser interest).

We prefer benchmark functions that are comprehensible such that algorithm behaviours can be understood in the topological context. In this way, a desired search behaviour can be pictured and deficiencies of algorithms can be profoundly analysed. Last but not least, this can eventually lead to a systematic improvement of algorithms.

All benchmark functions are scalable with the dimension. Most functions have no specific value of their optimal solution (they are randomly shifted in x-space). All functions have an artificially chosen optimal function value (they are randomly shifted in f-space). Consequently, for each function different *instances* can be generated: for each instance the randomly chosen values are drawn anew[2]. Apart from the first subgroup, the benchmarks are non-separable. Other specific properties are discussed in the appendix.

0.1 General Setup

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

Location of the optimal $\mathbf{x}^{\text{opt}}$ and of $f_{\text{opt}} = f(\mathbf{x}^{\text{opt}})$ All functions have their global optimum in $[-5, 5]^D$. The majority of functions has the global optimum in $[-4, 4]^D$ and for many of them $\mathbf{x}^{\text{opt}}$ is drawn uniformly from this compact. The value for f_{opt} is drawn from a Cauchy distributed random variable, with zero median and with roughly 50% of the values between -100 and 100. The value is rounded after two decimal places and set to ± 1000 if its absolute value exceeds 1000.

[1] For our experimental setup see [4, 5] and for our performance assessment methodology see [3].

[2] The implementation provides an instance ID as input, such that a set of uniquely specified instances can be explicitly chosen.

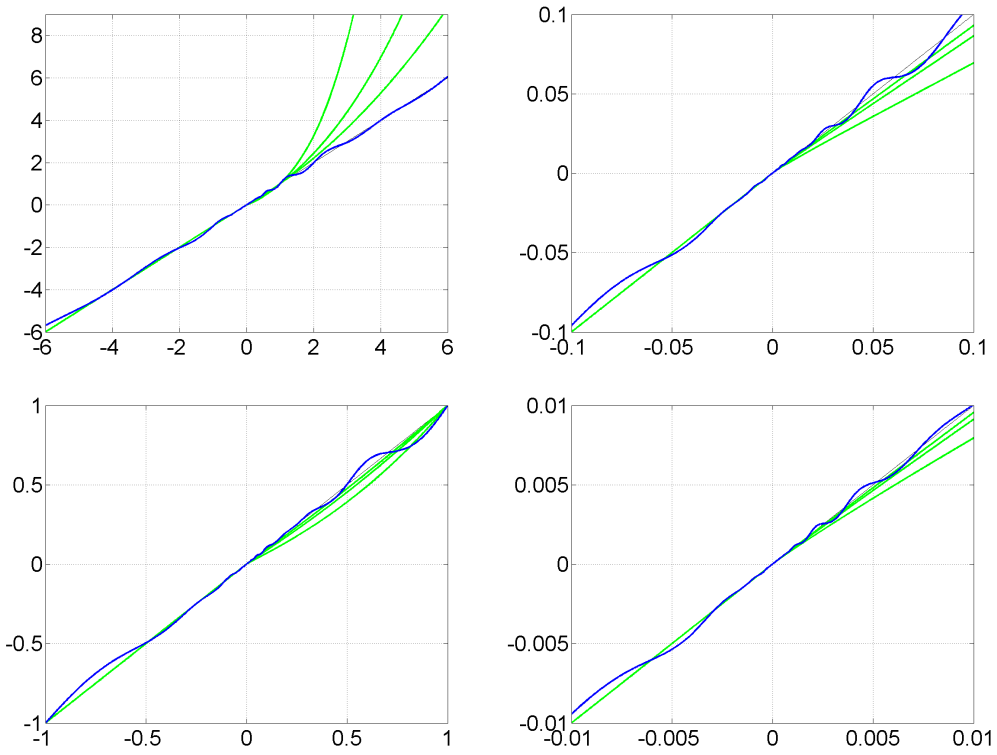


Figure 1: T_{osz} (blue) and D-th coordinate of T_{asy} for $\beta = 0.1, 0.2, 0.5$ (green)

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 $\mathbf{z}^{\text{opt}} = \mathbf{0}$, if not stated otherwise.

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

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 $T_{\text{osz}} : \mathcal{R}^n \to \mathcal{R}^n$, $n \in \{1, D\}$, and $T_{\text{asy}} : \mathcal{R}^D \to \mathcal{R}^D$ are defined coordinate-wise (see below). 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.

0.2 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, $\otimes : \mathcal{R}^D \times \mathcal{R}^D \to \mathcal{R}^D, (\mathbf{x}, \mathbf{y}) \mapsto \text{diag}(\mathbf{x}) \times \mathbf{y} = (x_i \times y_i)_{i=1,\ldots,D}$

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

$[.]$ denotes the nearest integer value

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

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

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

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

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

Q, R orthogonal (rotation) matrices. For one function in one dimension a different realization for respectively **Q** and **R** is used for each instantiation of the function. Orthogonal matrices are generated from standard normally distributed entries by Gram-Schmidt orthonormalization. Columns and rows of an orthogonal matrix form an orthonormal basis.

R see **Q**

$T^{\beta}_{\mathrm{asy}} : \mathcal{R}^D \to \mathcal{R}^D$, $x_i \mapsto \begin{cases} x_i^{1+\beta\frac{i-1}{D-1}\sqrt{x_i}} & \text{if } x_i > 0 \\ x_i & \text{otherwise} \end{cases}$, for $i = 1, \ldots, D$. See Figure 1.

$T_{\mathrm{osz}} : \mathcal{R}^n \to \mathcal{R}^n$, for any positive integer n ($n = 1$ and $n = D$ are used in the following), 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 $\hat{x} = \begin{cases} \log(|x|) & \text{if } x \neq 0 \\ 0 & \text{otherwise} \end{cases}$, $\mathrm{sign}(x) = \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{otherwise} \end{cases}$, $c_1 = \begin{cases} 10 & \text{if } x > 0 \\ 5.5 & \text{otherwise} \end{cases}$ and $c_2 = \begin{cases} 7.9 & \text{if } x > 0 \\ 3.1 & \text{otherwise} \end{cases}$. See Figure 1.

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

0.3 Figures

The benchmark function definitions in the next section are accompanied with a number of figures.

1. a (3-D) surface plot, where $D = 2$
2. a contour plot, where $D = 2$
3. two projected contour plots, where $D = 20$. Plotted are, starting from the optimum $\mathbf{x}^{\mathrm{opt}}$, first versus second variable (left) and first versus fifth variable (right).
4. sections (f versus x) through the global optimum along the first variable x_1, the second variable x_2, and the all-ones vector. The sections for different dimensions appear
   (a) in a non-log plot (above), where the maximum f-value is normalized to one for each single graph.
   (b) in a semi-log plot (middle row)
   (c) in a log-log plot (below) starting close to the global optimum along x_1, $-x_1$, x_2, $-x_2$, $\mathbf{1}$, and $-\mathbf{1}$.

1 Separable functions

1.1 Sphere Function

$$f_1(\mathbf{x}) = \|\mathbf{z}\|^2 + f_{\text{opt}} \tag{1}$$

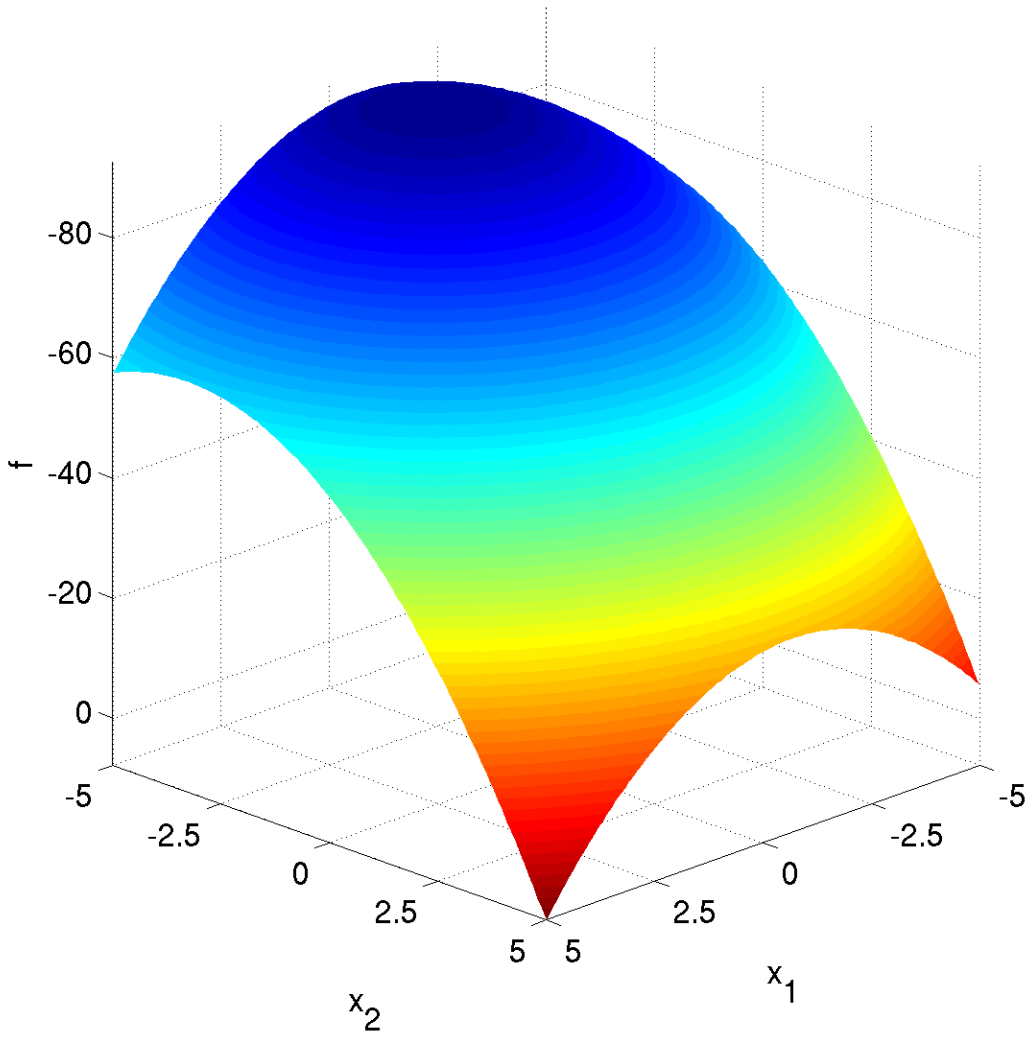
- $\mathbf{z} = \mathbf{x} - \mathbf{x}^{\text{opt}}$

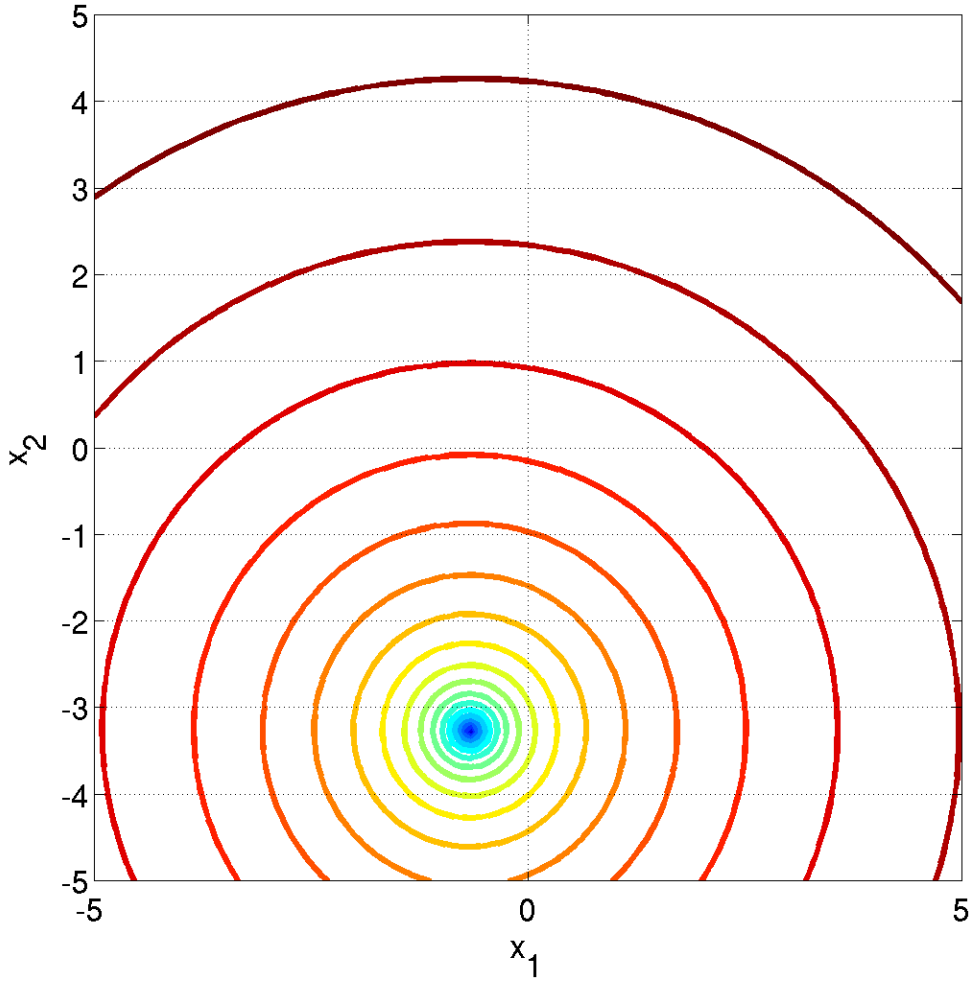
Properties Presumably the most easy continuous domain search problem, given the volume of the searched solution is small (i.e. where pure monte-carlo random search is too expensive).

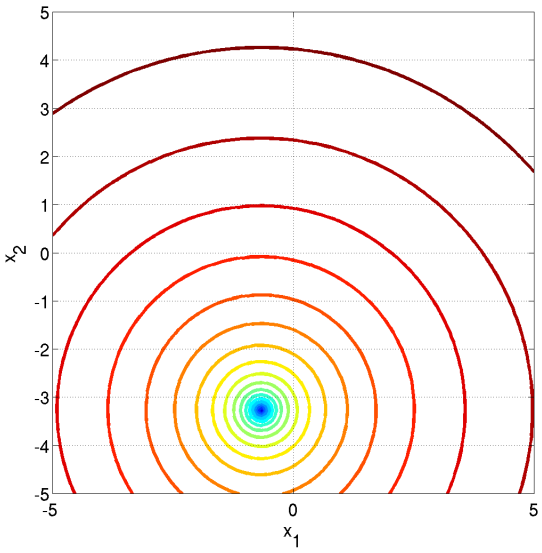
- unimodal
- highly symmetric, in particular rotationally invariant, scale invariant

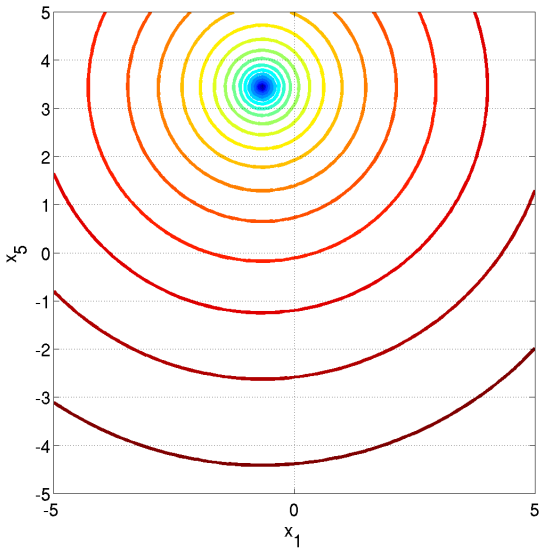
Information gained from this function:

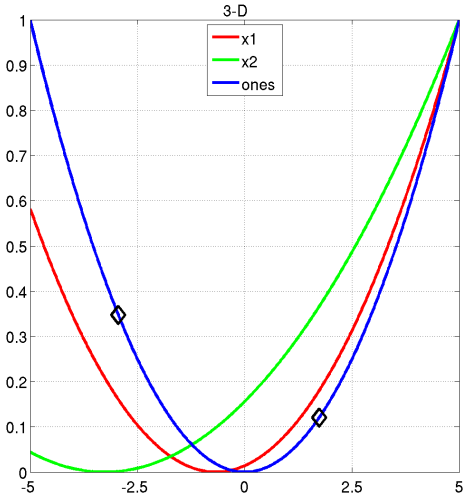
- What is the optimal convergence rate of an algorithm?

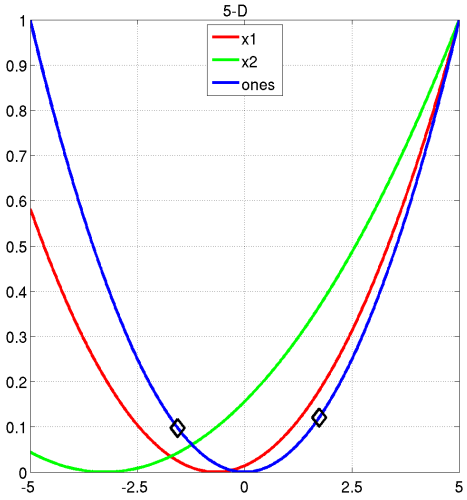


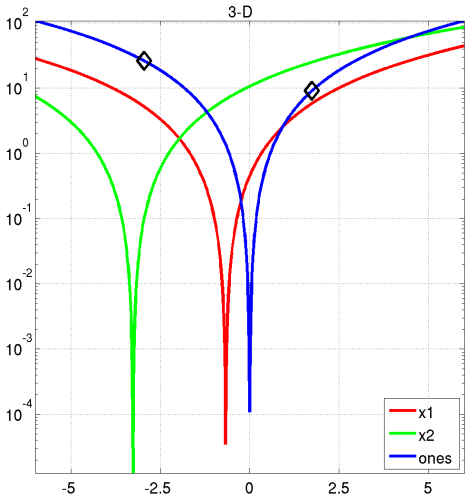

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

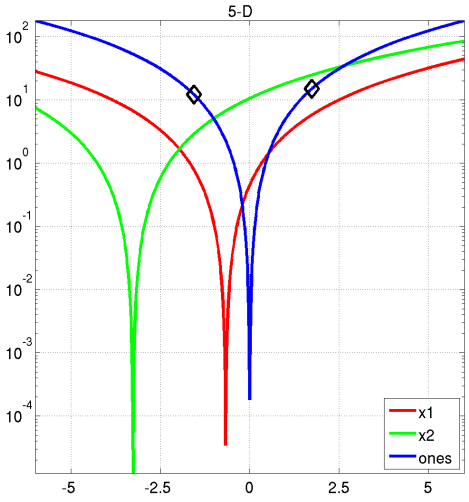

x2
x1

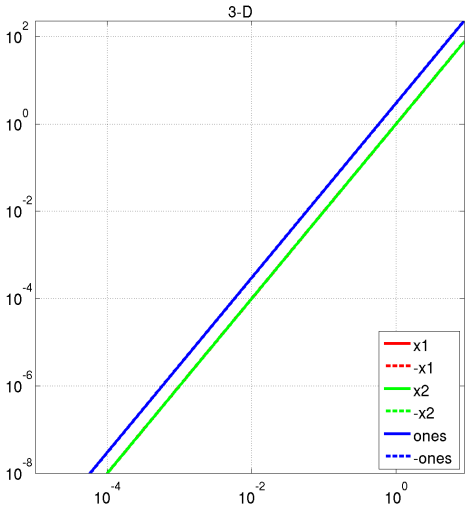

x5
x1

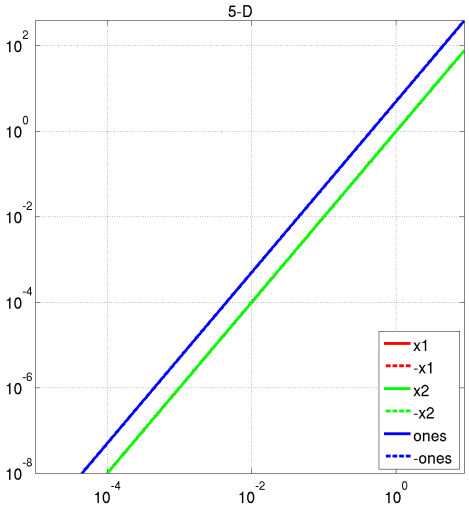

3-D
x1
x2
ones

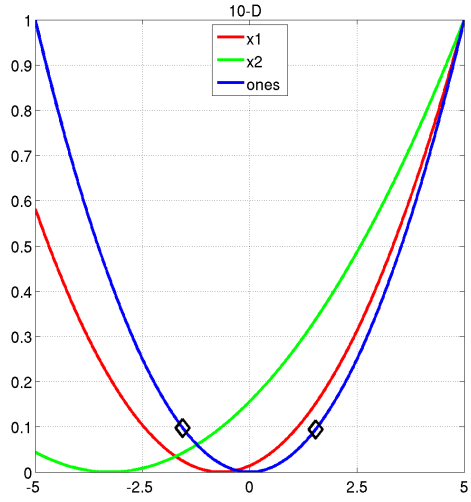

5-D
x1
x2
ones

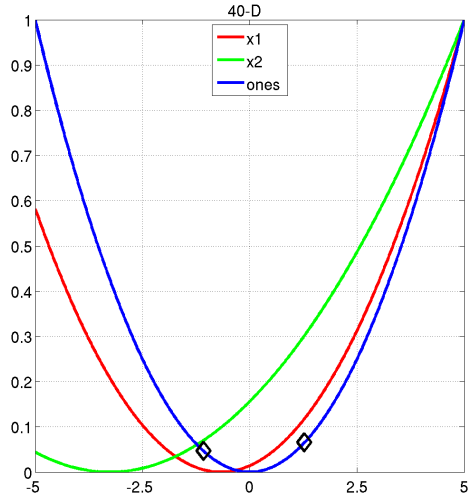

3-D
x1
x2
ones

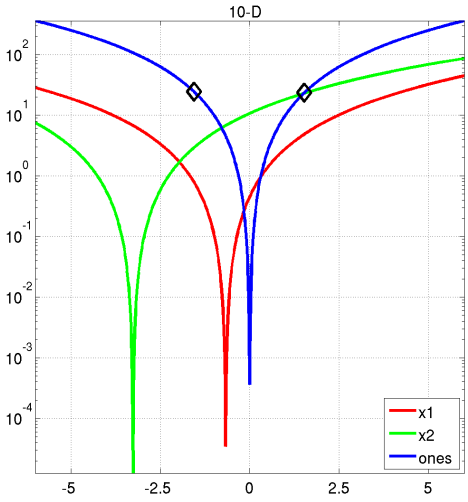

5-D
x1
x2
ones

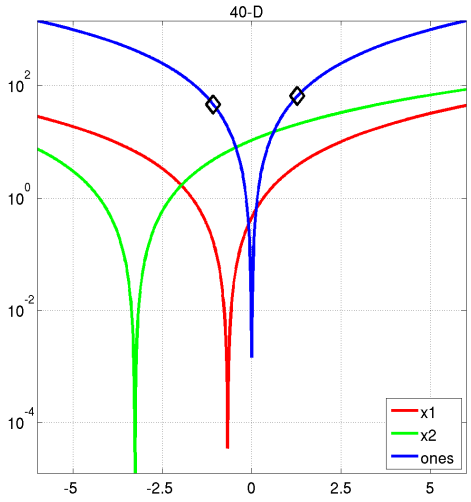

3-D
x1
-x1
x2
-x2
ones
-ones

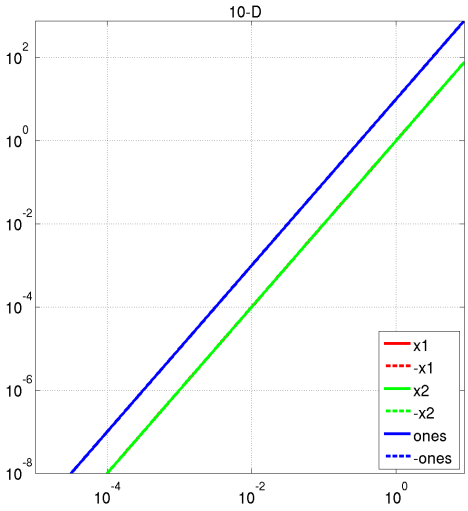

5-D
x1
-x1
x2
-x2
ones
-ones

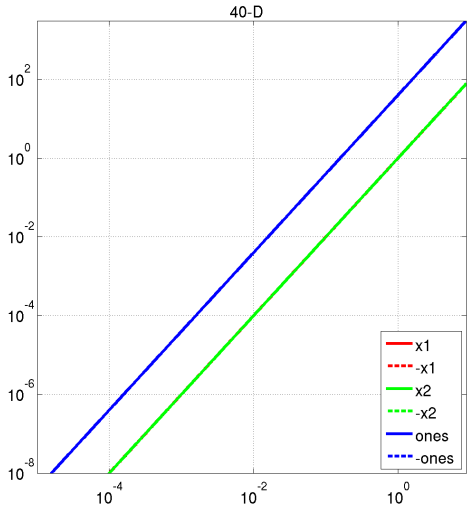

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

1.2 Ellipsoidal Function

$$f_2(\mathbf{x}) = \sum_{i=1}^{D} 10^{6\frac{i-1}{D-1}} z_i^2 + f_{\text{opt}} \tag{2}$$

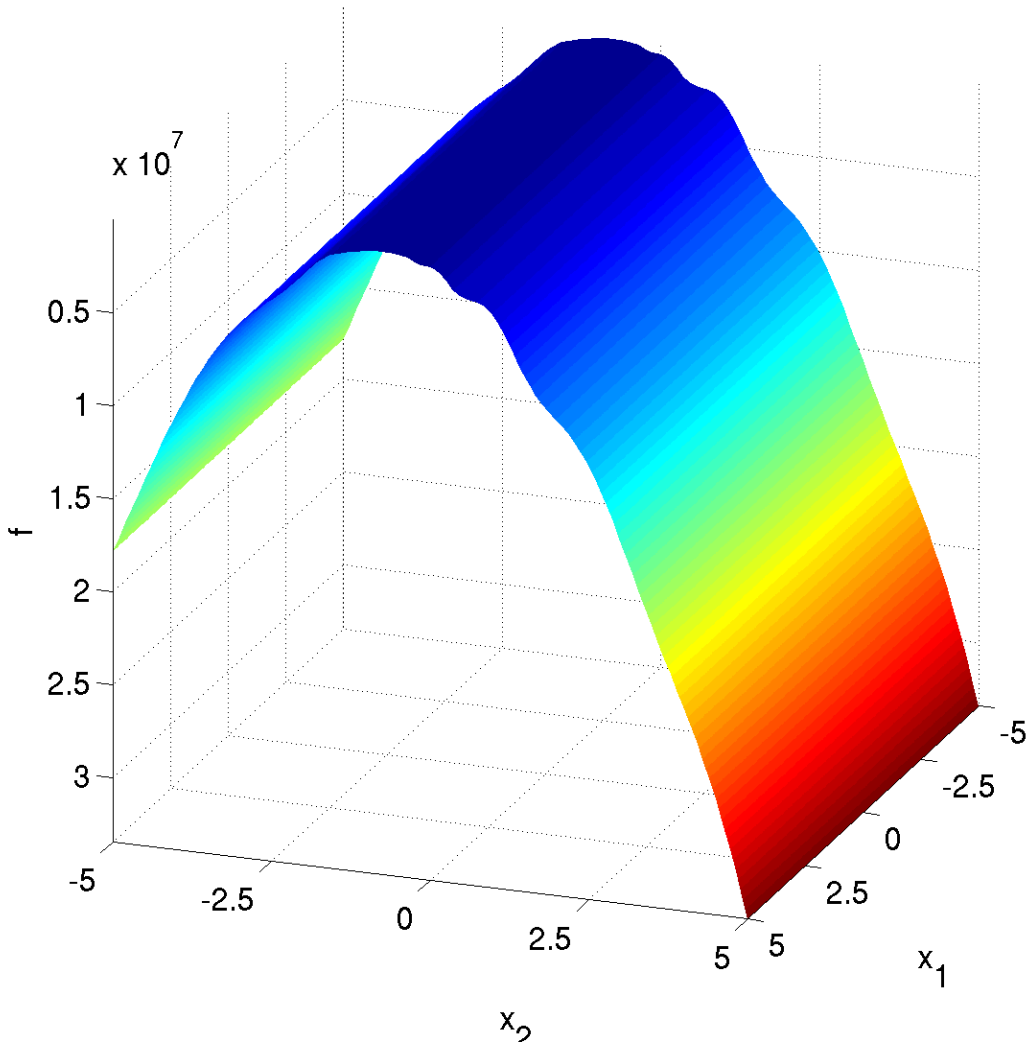
- $\mathbf{z} = T_{\text{osz}}(\mathbf{x} - \mathbf{x}^{\text{opt}})$

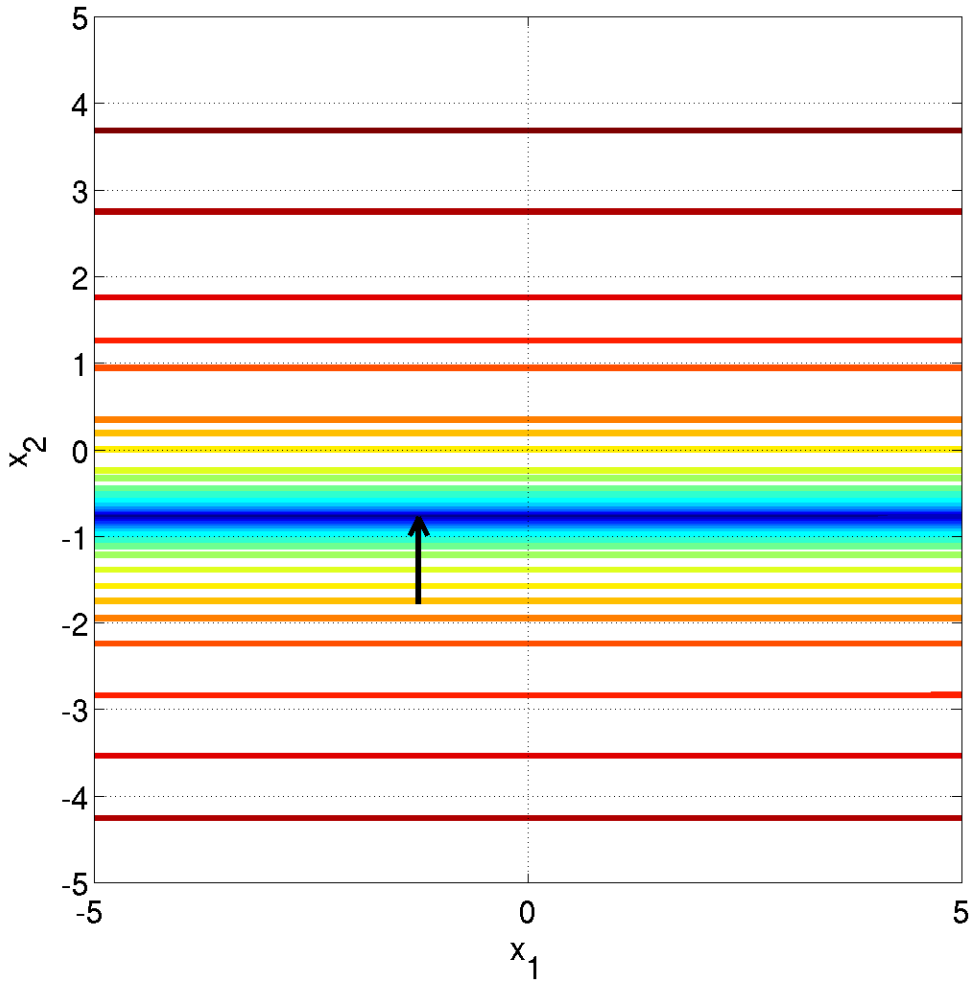
Properties Globally quadratic and ill-conditioned function with smooth local irregularities.

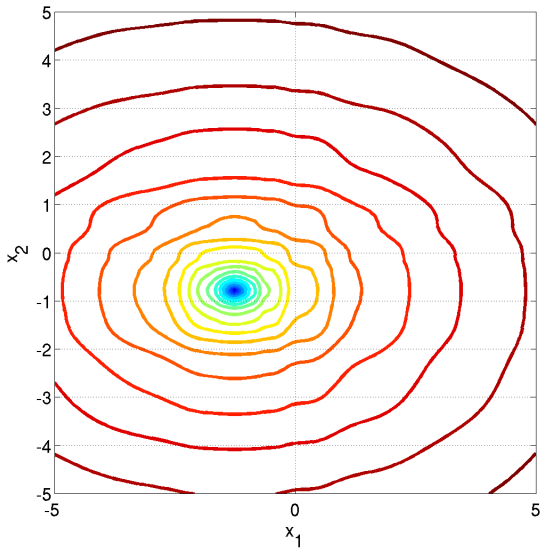
- unimodal
- conditioning is about 10^6

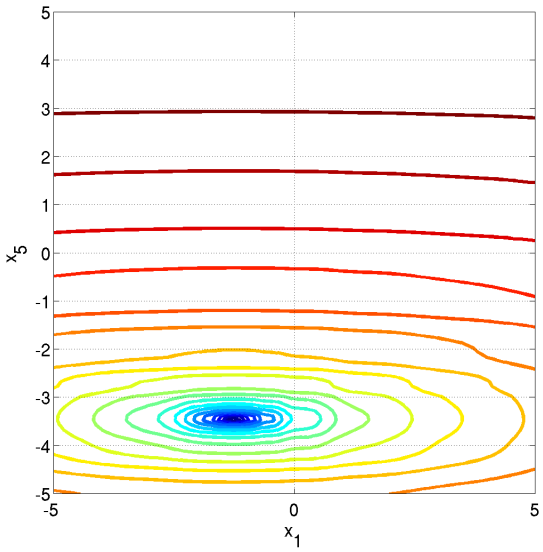
Information gained from this function:

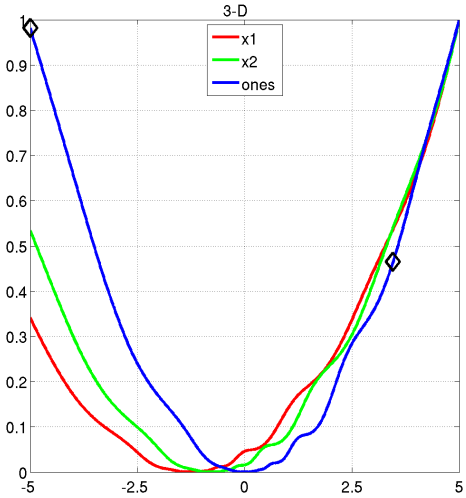
- In comparison to f10: Is separability exploited?

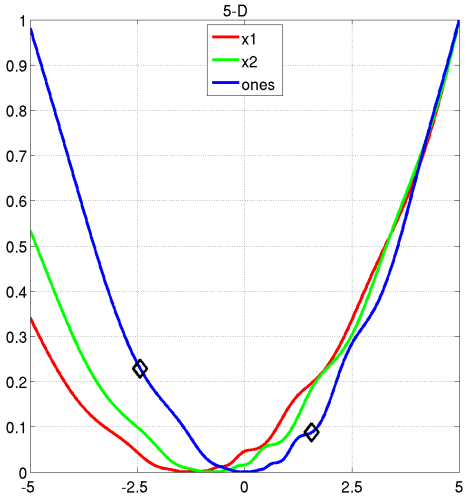


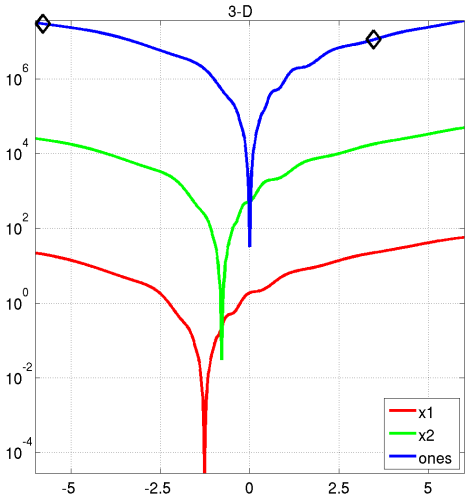

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

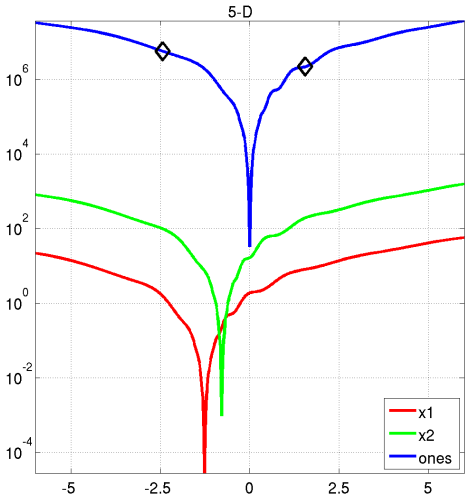

x2
x1

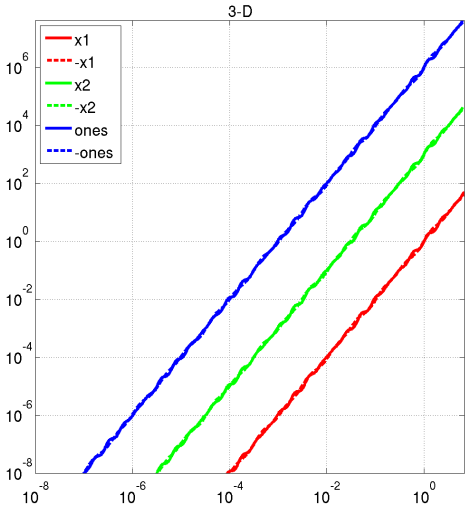

x5
x1

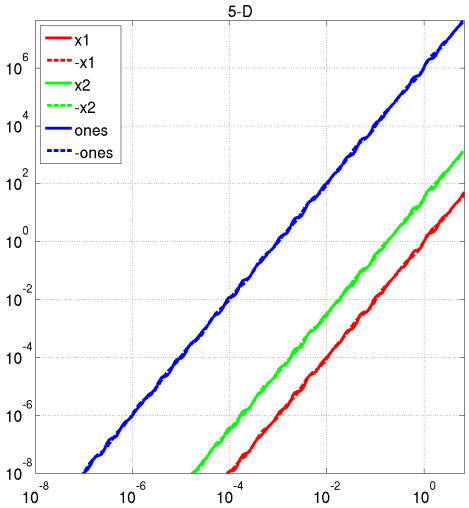

3-D
x1
x2
ones

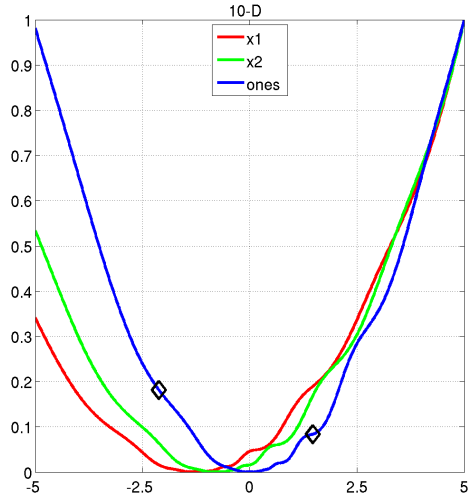

5-D
x1
x2
ones

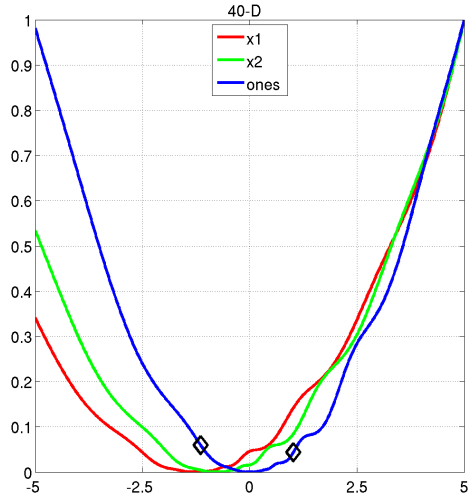

3-D
x1
x2
ones

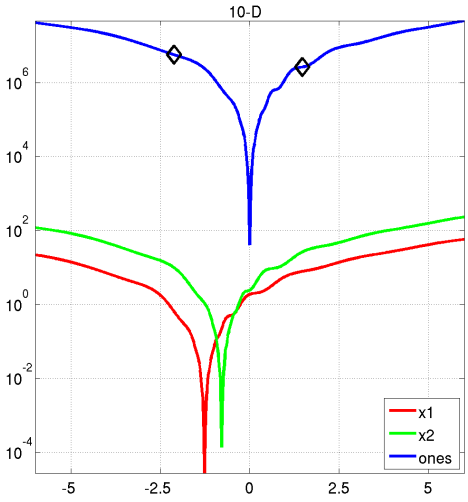

5-D
x1
x2
ones

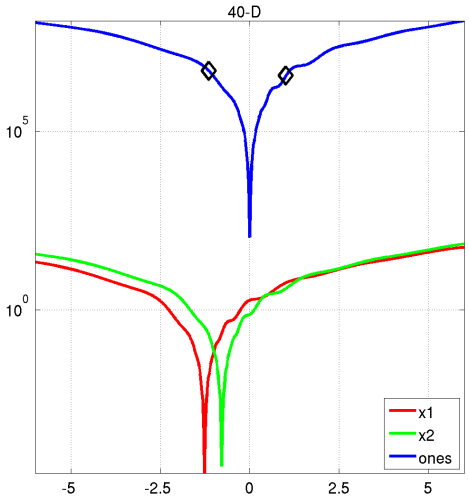

3-D
x1
-x1
x2
-x2
ones
-ones

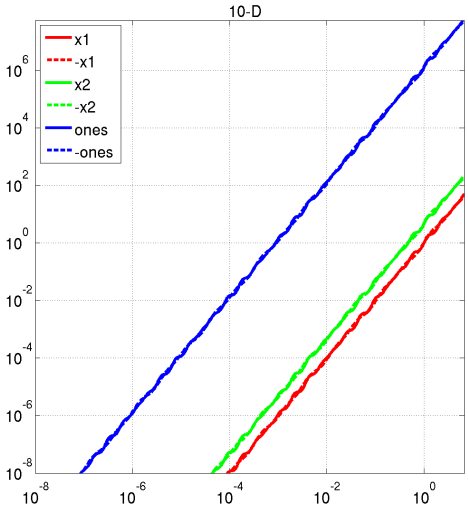

5-D
x1
-x1
x2
-x2
ones
-ones

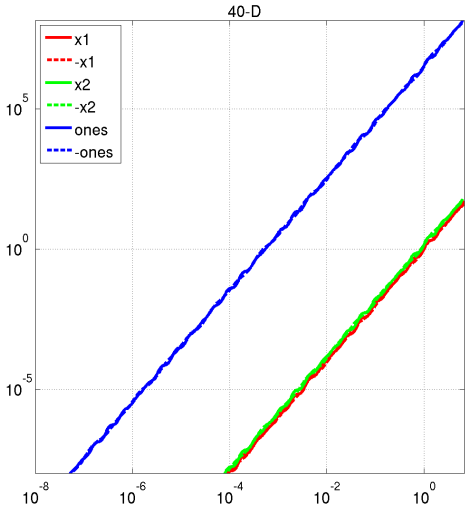

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

1.3 Rastrigin Function

$$f_3(\mathbf{x}) = 10\left(D - \sum_{i=1}^{D}\cos(2\pi z_i)\right) + \|\mathbf{z}\|^2 + f_{\text{opt}} \tag{3}$$

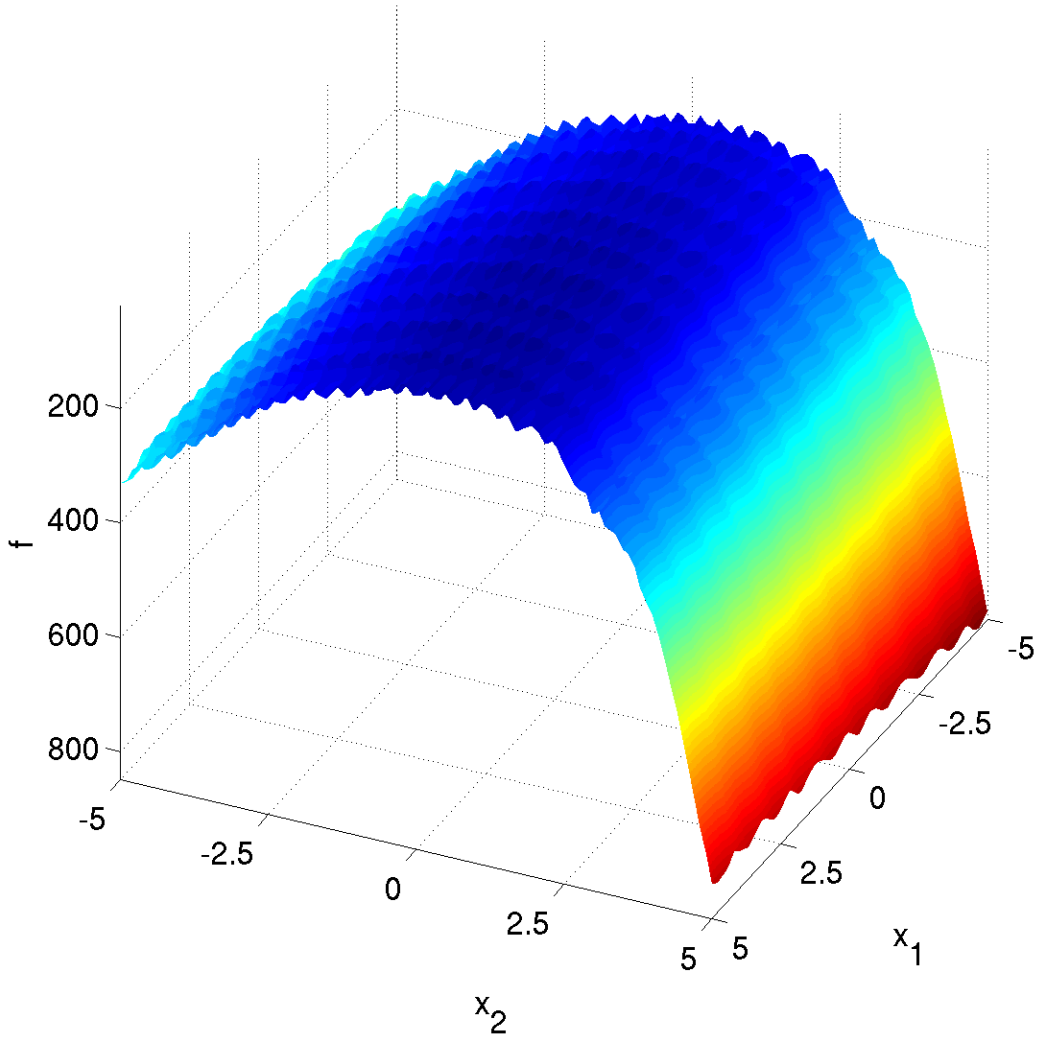
- $\mathbf{z} = \Lambda^{10} T_{\text{asy}}^{0.2}(T_{\text{osz}}(\mathbf{x} - \mathbf{x}^{\text{opt}}))$

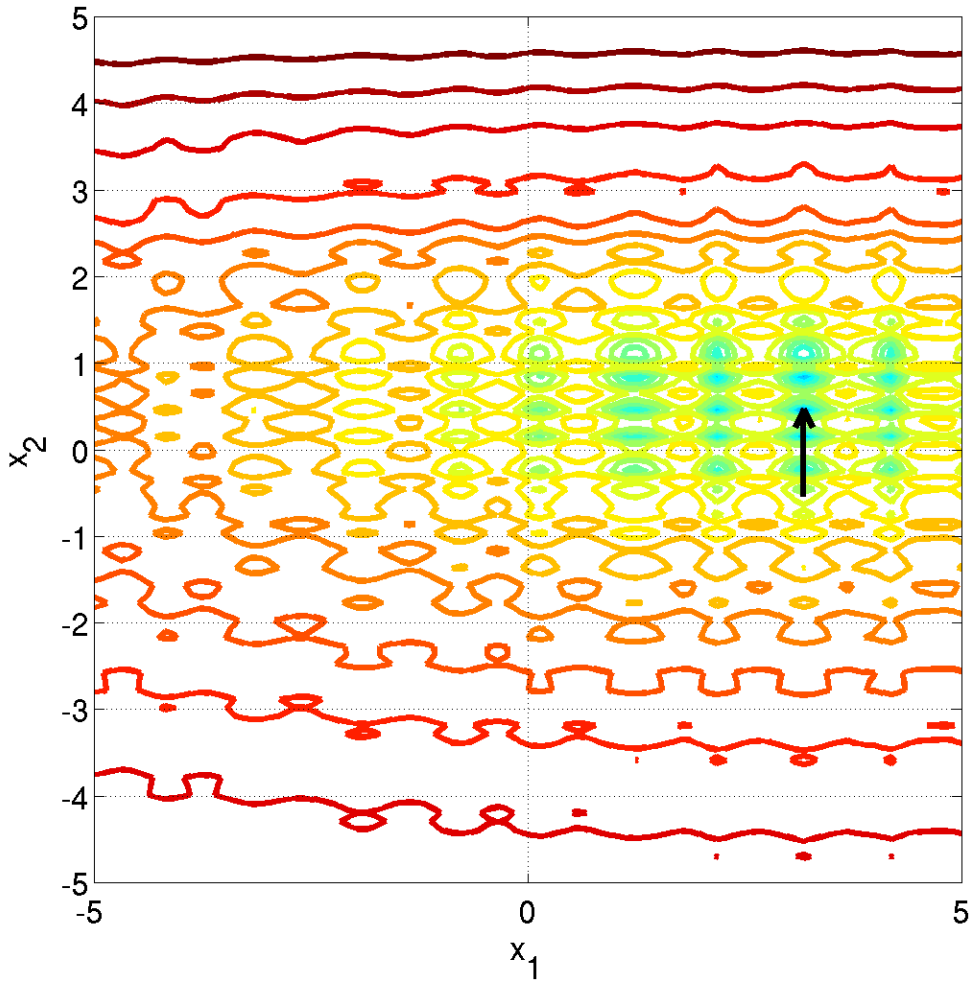
Properties Highly multimodal function with a comparatively regular structure for the placement of the optima. The transformations T_{asy} and T_{osz} alleviate the symmetry and regularity of the original Rastrigin function

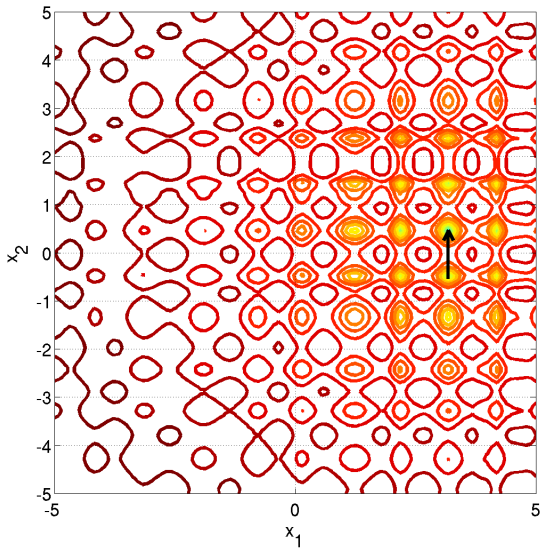
- roughly 10^D local optima
- conditioning is about 10

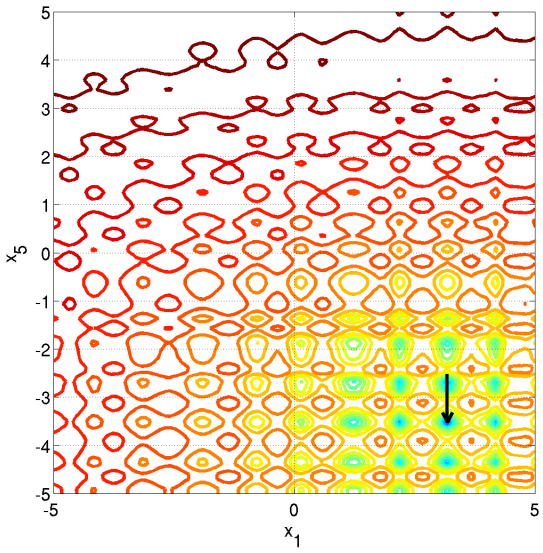
Information gained from this function:

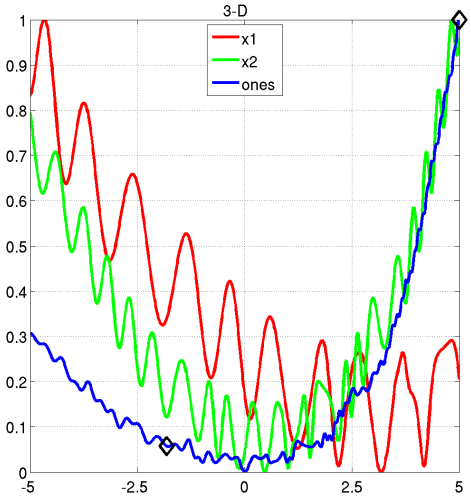
- In comparison to f15: Is separability exploited?

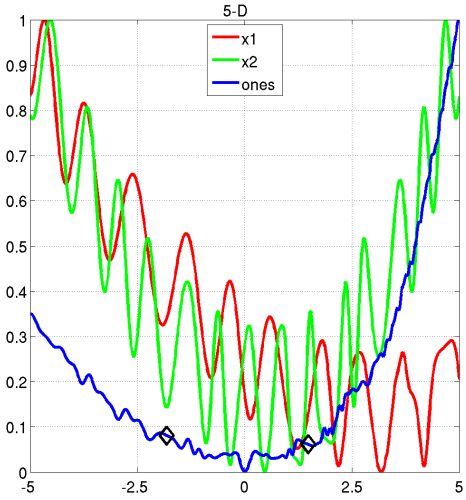


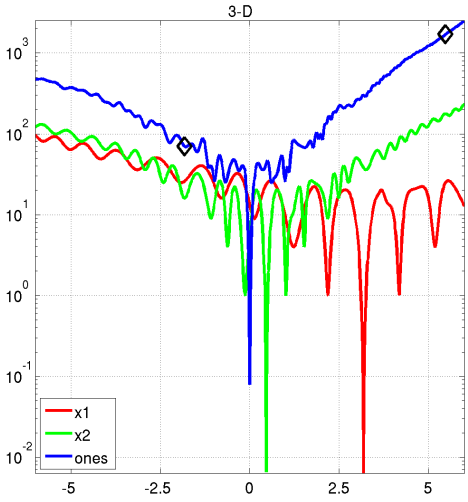

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

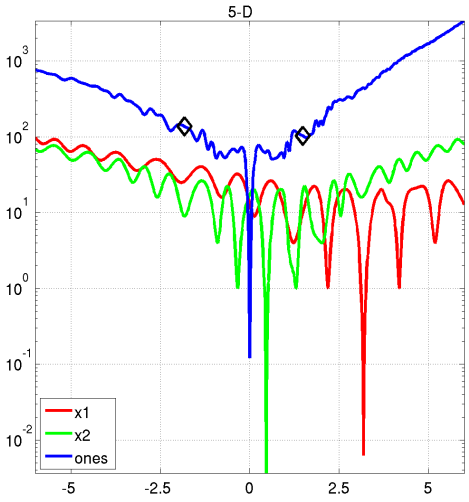

x2
x1

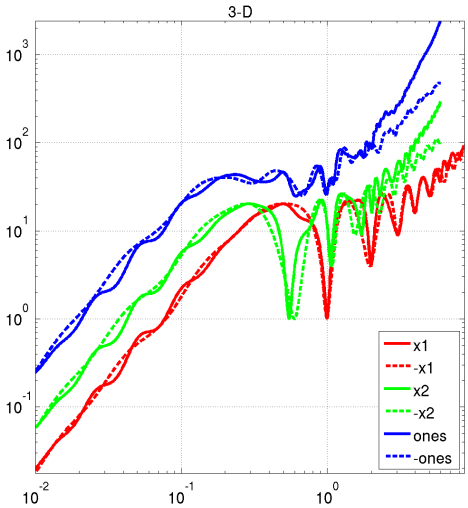

x5
x1

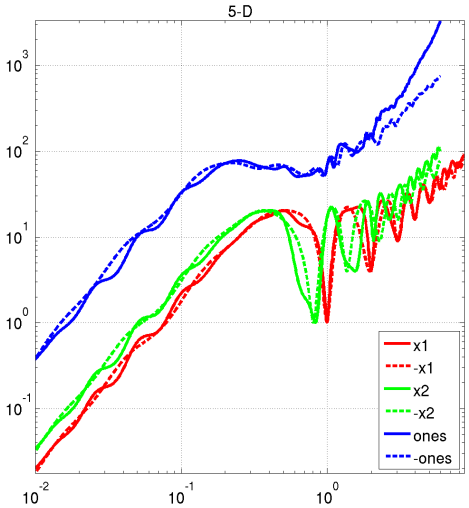

3-D
x1
x2
ones

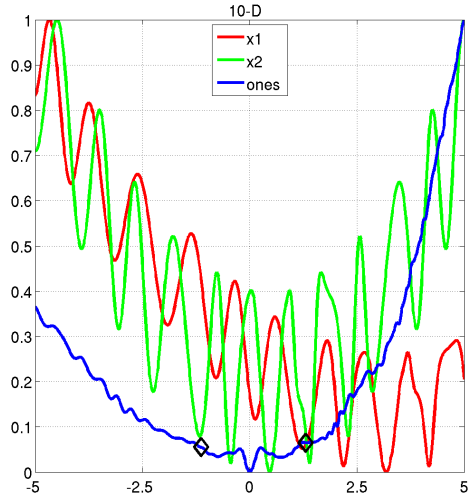

5-D
x1
x2
ones

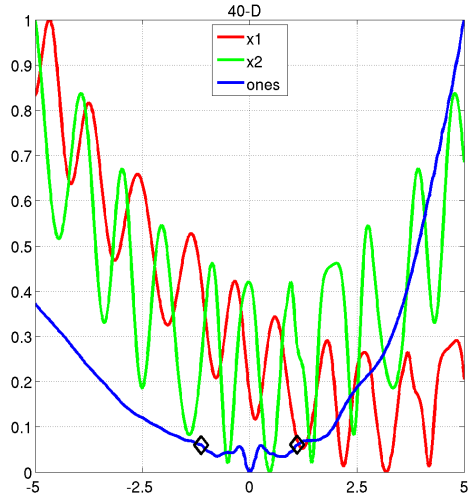

3-D
x1
x2
ones

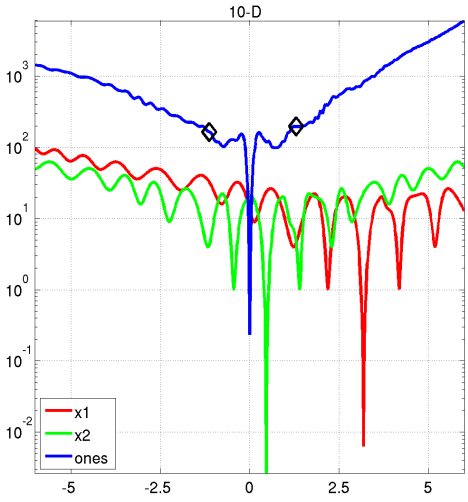

5-D
x1
x2
ones

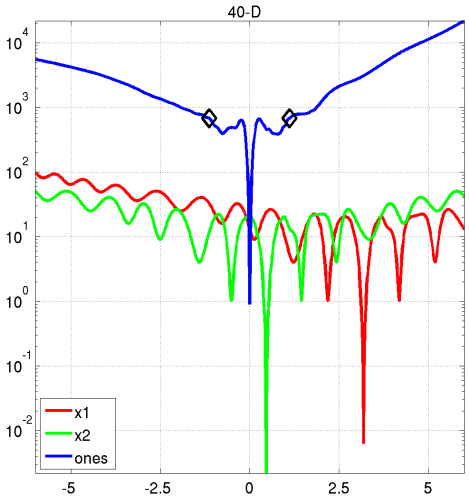

3-D
x1
-x1
x2
-x2
ones
-ones

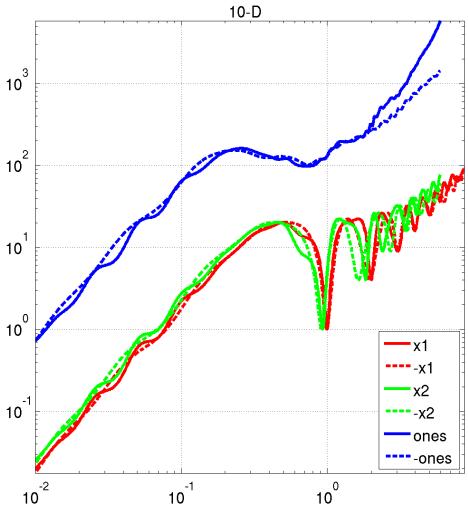

5-D
x1
-x1
x2
-x2
ones
-ones

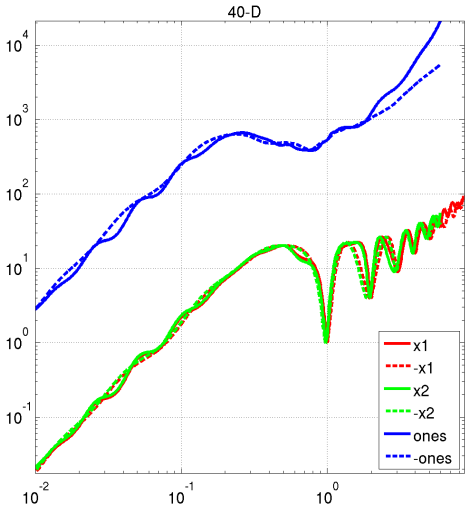

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

1.4 Büche-Rastrigin Function

$$f_4(\mathbf{x}) = 10\left(D - \sum_{i=1}^{D}\cos(2\pi z_i)\right) + \sum_{i=1}^{D} z_i^2 + 100\, f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{4}$$

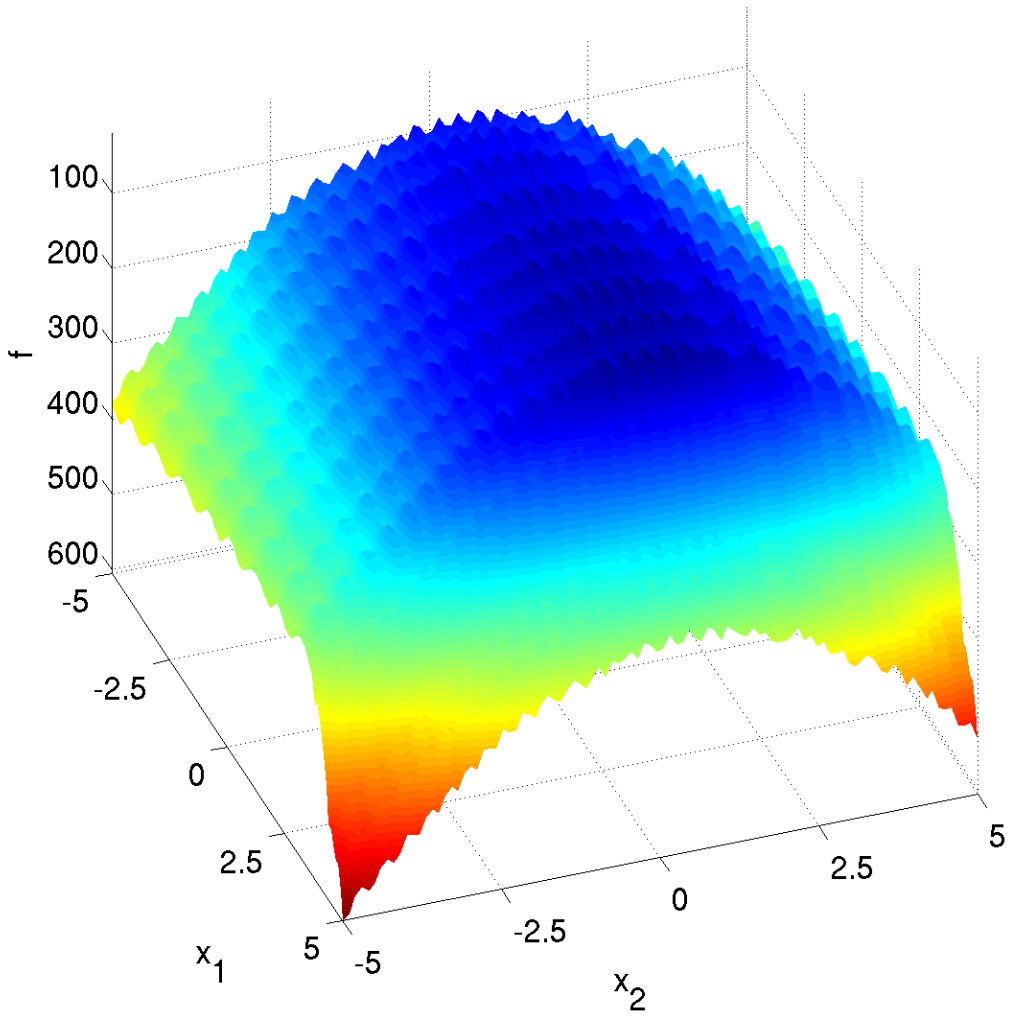
- $z_i = s_i\, T_{\text{osz}}(x_i - x_i^{\text{opt}}) \quad \text{for } i = 1 \ldots D$
- $s_i = \begin{cases} 10 \times 10^{\frac{1}{2}\frac{i-1}{D-1}} & \text{if } z_i > 0 \text{ and } i = 1, 3, 5, \ldots \\ 10^{\frac{1}{2}\frac{i-1}{D-1}} & \text{otherwise} \end{cases} \quad \text{for } i = 1, \ldots, D$

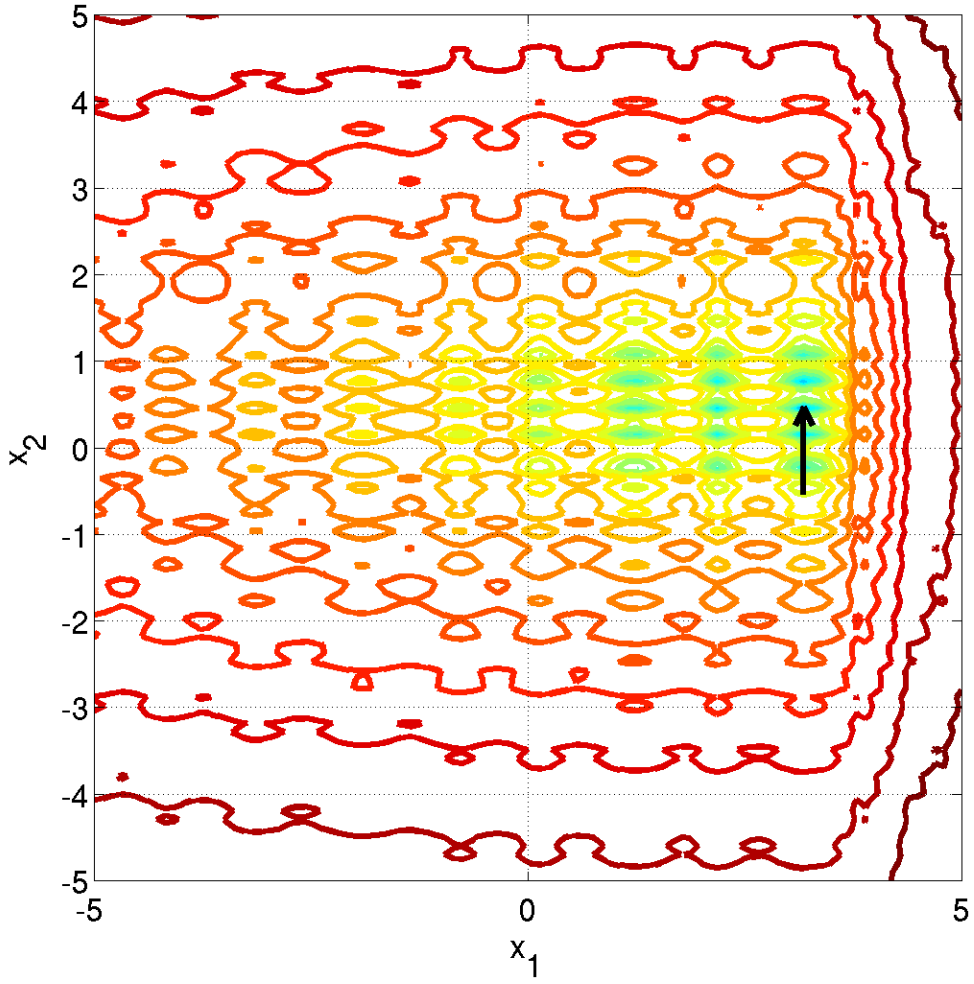
Properties Highly multimodal function with a structured but highly asymmetric placement of the optima. Constructed as a deceptive function for symmetrically distributed search operators.

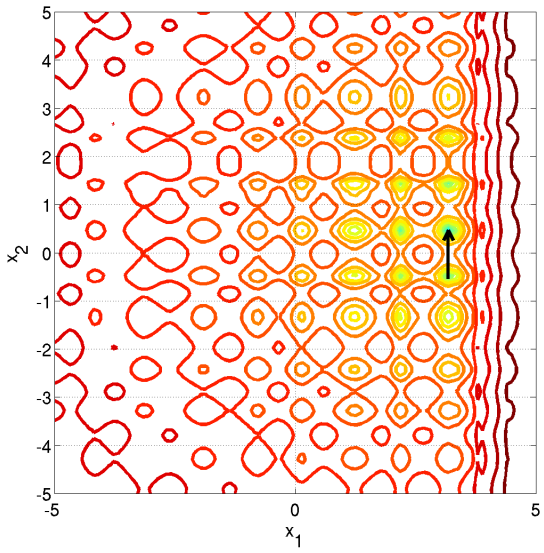
- roughly 10^D local optima, conditioning is about 10, skew factor is about 10 in x-space and 100 in f-space

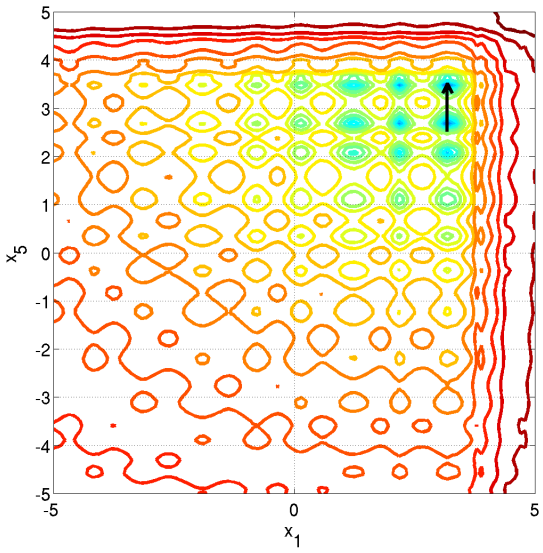
Information gained from this function:

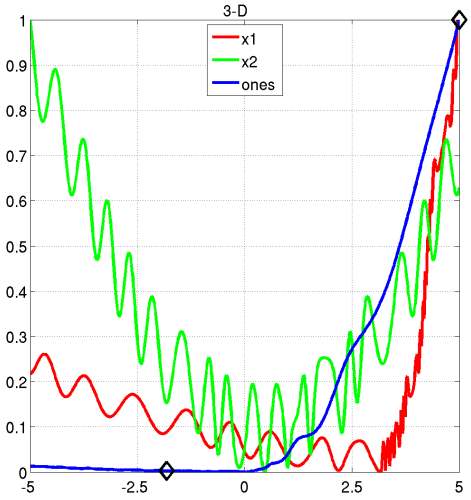
- In comparison to f3: What is the effect of asymmetry?

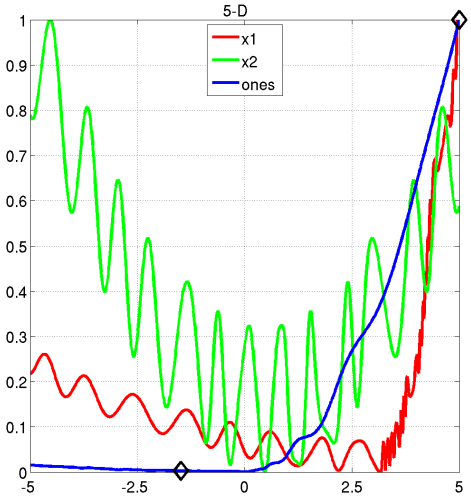


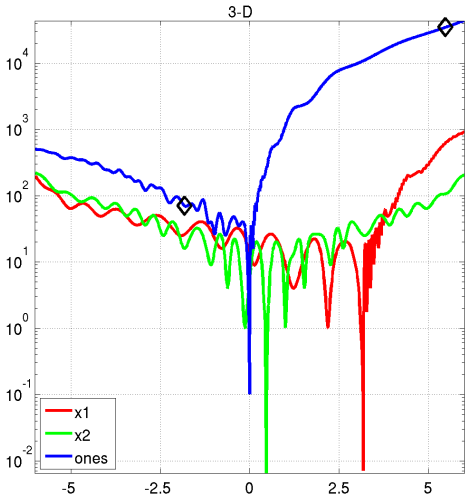

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

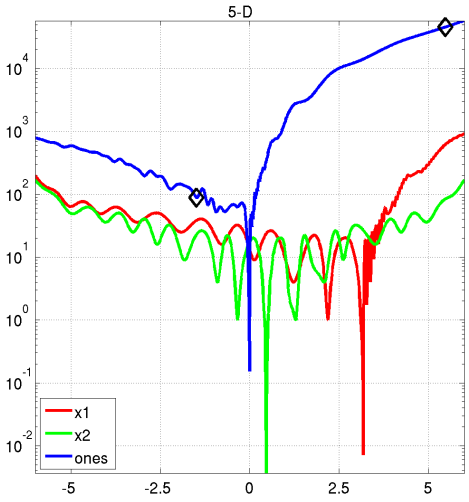

x2
x1

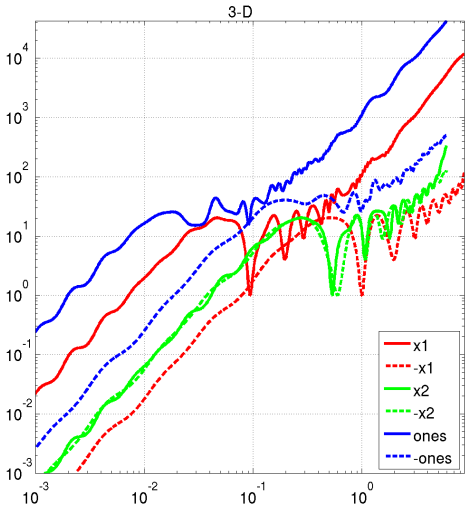

x5
x1

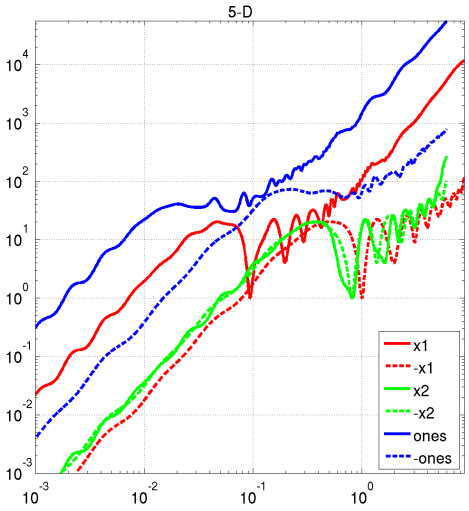

3-D
x1
x2
ones

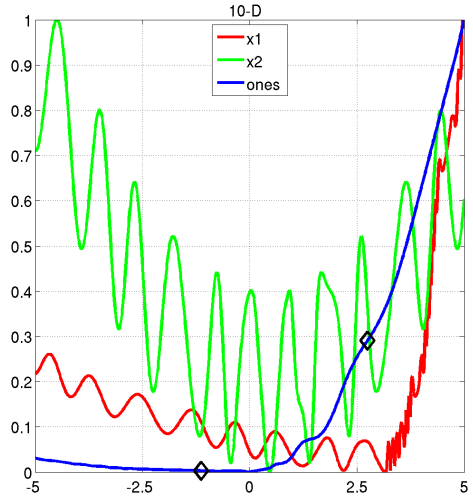

5-D
x1
x2
ones

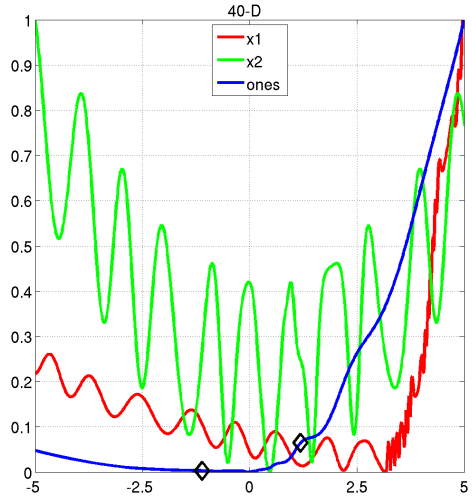

3-D
x1
x2
ones

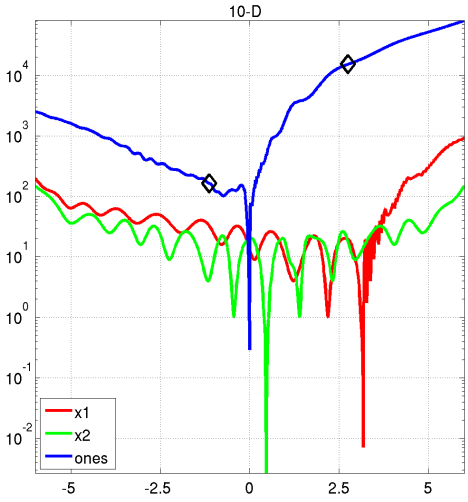

5-D
x1
x2
ones

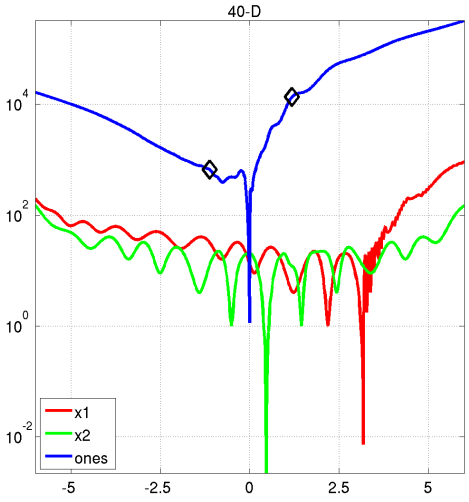

3-D
x1
-x1
x2
-x2
ones
-ones

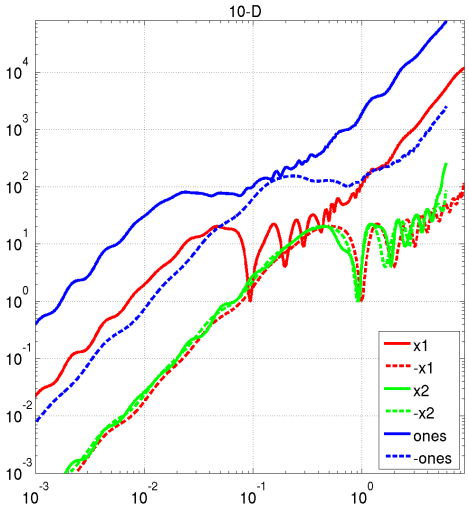

5-D
x1
-x1
x2
-x2
ones
-ones

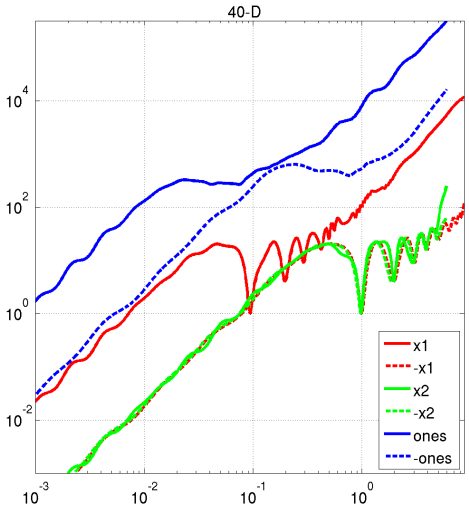

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

1.5 Linear Slope

$$f_5(\mathbf{x}) = \sum_{i=1}^{D} 5\,|s_i| - s_i z_i + f_{\text{opt}} \tag{5}$$

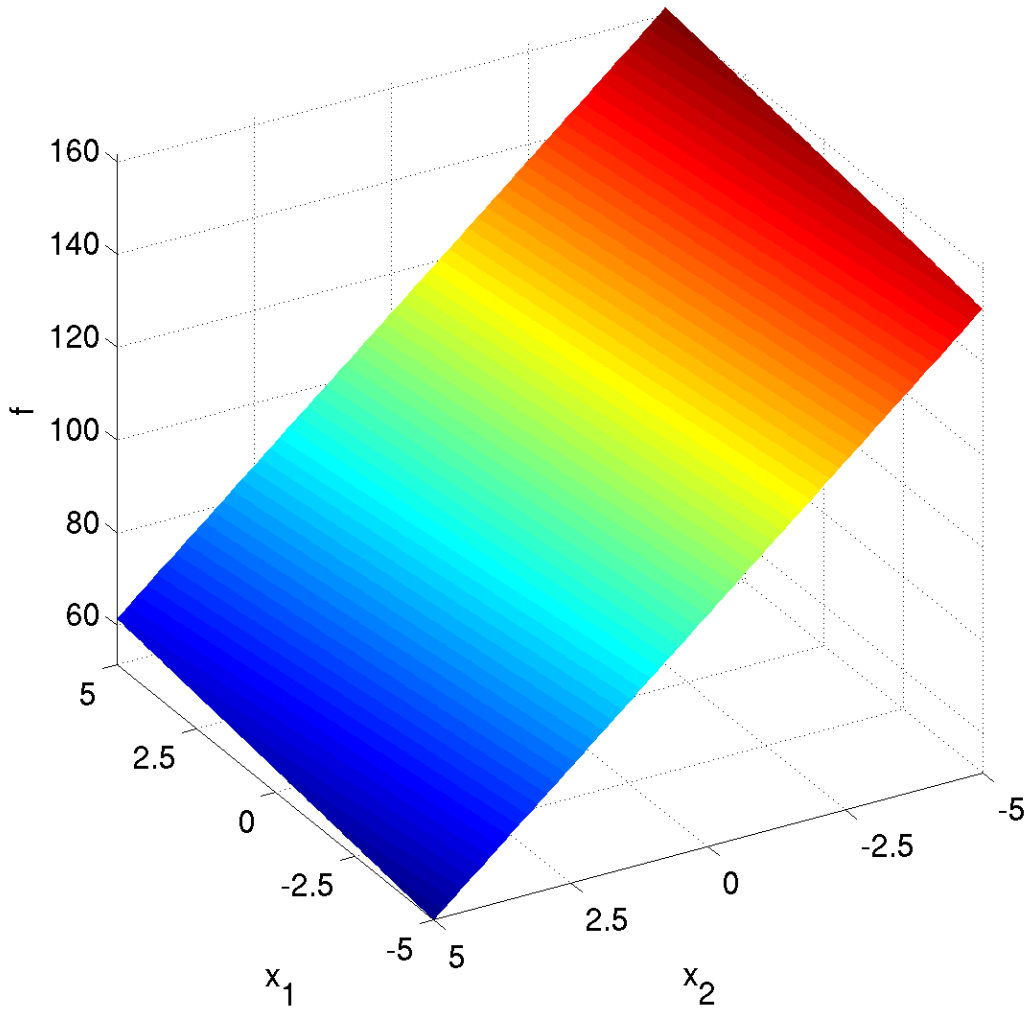
- $z_i = x_i$ if $x_i^{\text{opt}} x_i < 5^2$ and $z_i = x_i^{\text{opt}}$ otherwise, for $i = 1, \ldots, D$. That is, if x_i exceeds x_i^{opt} it will mapped back into the domain and the function appears to be constant in this direction.
- $s_i = \text{sign}\left(x_i^{\text{opt}}\right) 10^{\frac{i-1}{D-1}}$ for $i = 1, \ldots, D$.
- $\mathbf{x}^{\text{opt}} = \mathbf{z}^{\text{opt}} = 5 \times \mathbf{1}_-^+$

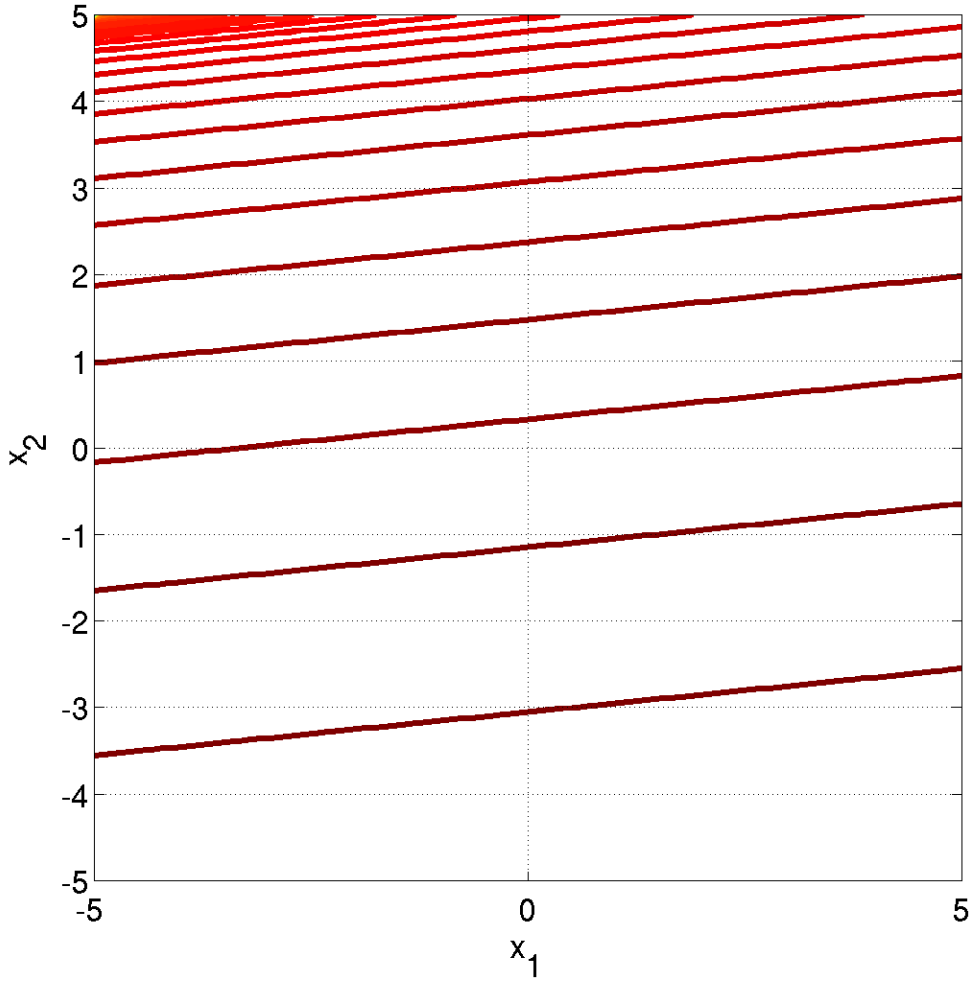
Properties Purely linear function testing whether the search can go outside the initial convex hull of solutions right into the domain boundary.

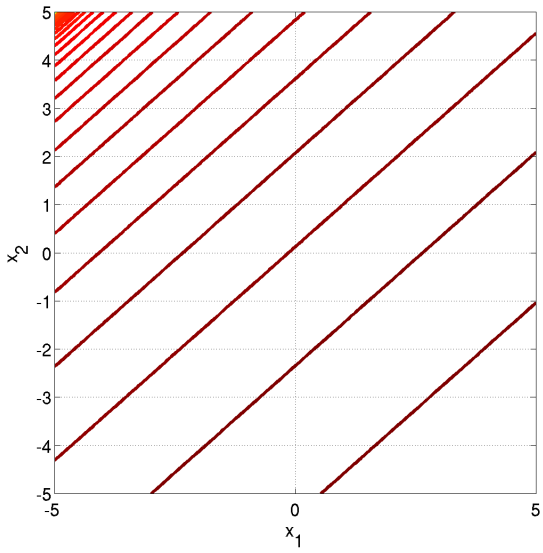
- $\mathbf{x}^{\text{opt}}$ is on the domain boundary

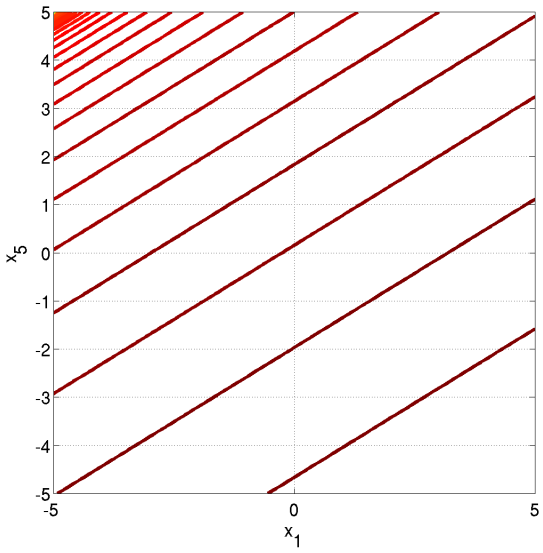
Information gained from this function:

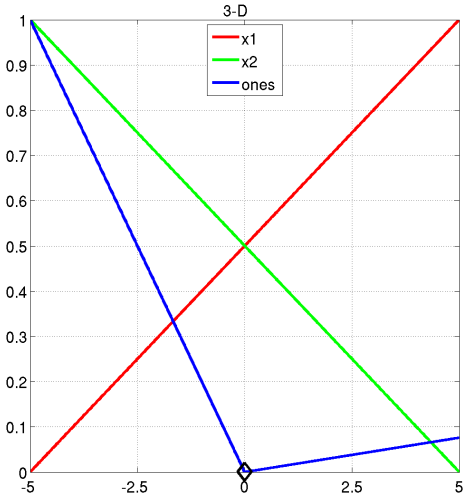
- Can the search go outside the initial convex hull of solutions into the domain boundary? Can the step size be increased accordingly?

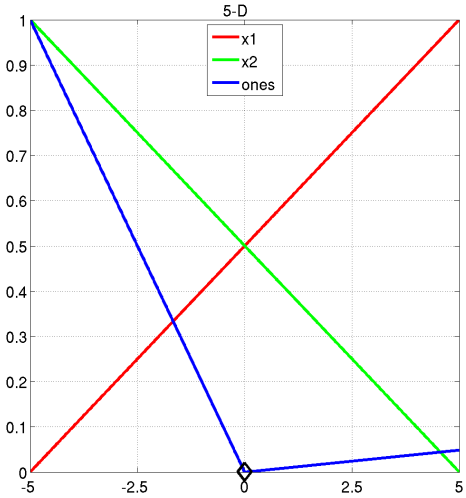


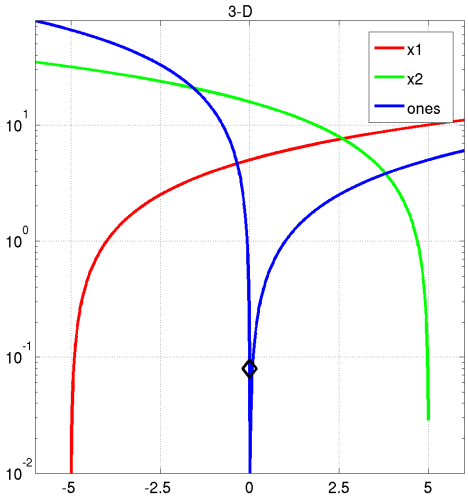

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

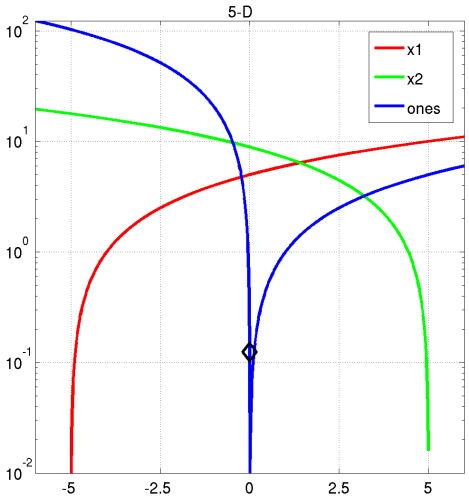

x2
x1

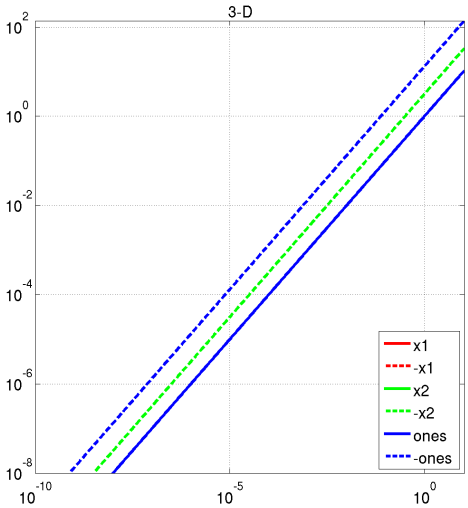

x5
x1

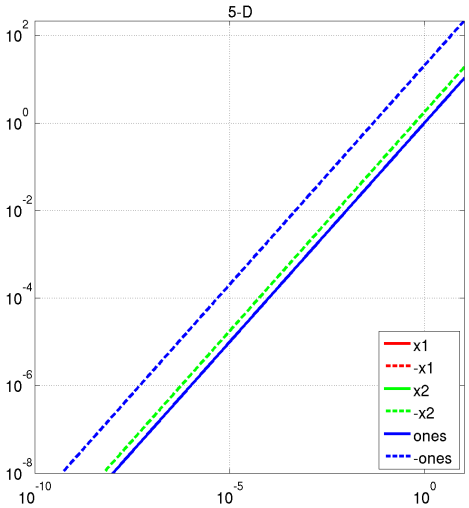

3-D
x1
x2
ones

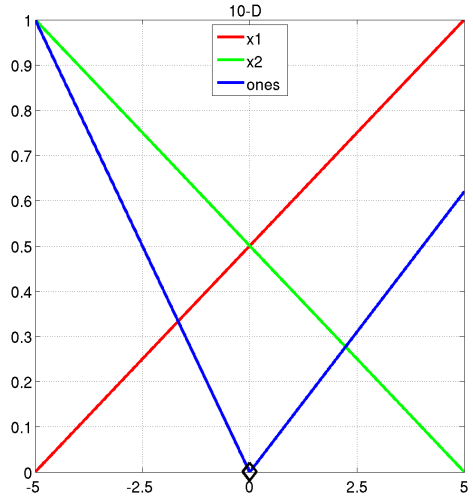

5-D
x1
x2
ones

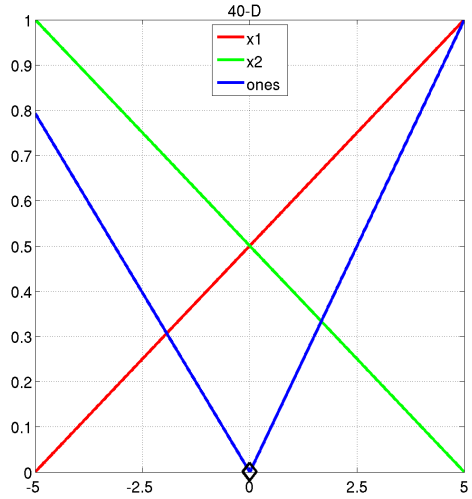

3-D
x1
x2
ones

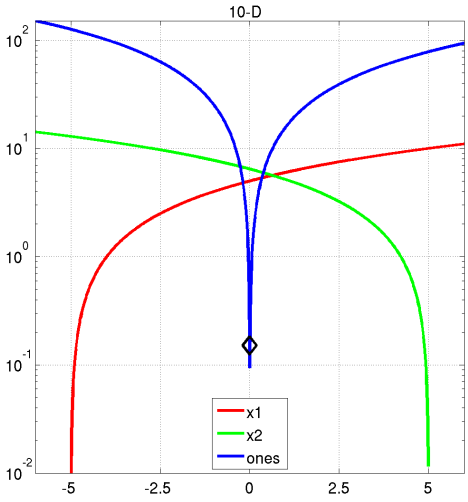

5-D
x1
x2
ones

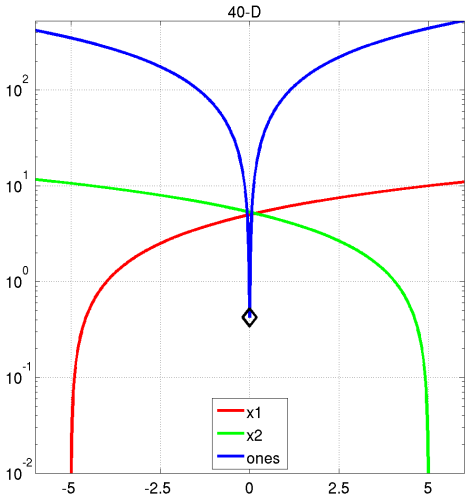

3-D
x1
-x1
x2
-x2
ones
-ones

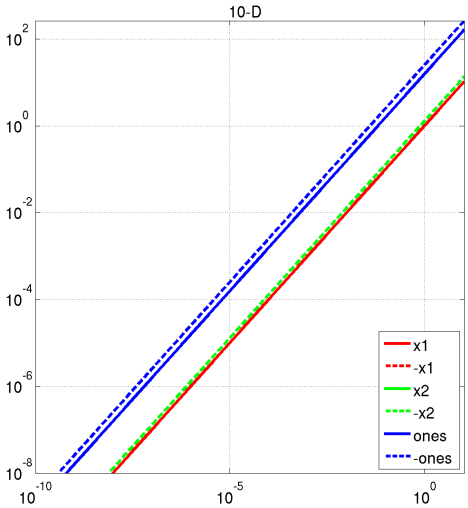

5-D
x1
-x1
x2
-x2
ones
-ones

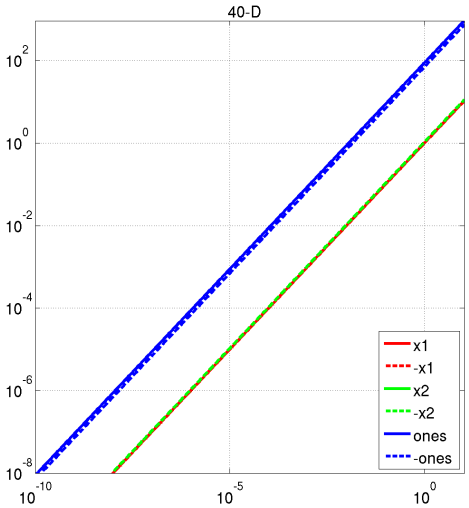

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

2 Functions with low or moderate conditioning

2.6 Attractive Sector Function

$$f_6(\mathbf{x}) = T_{\text{osz}}\left(\sum_{i=1}^{D}(s_i z_i)^2\right)^{0.9} + f_{\text{opt}} \tag{6}$$

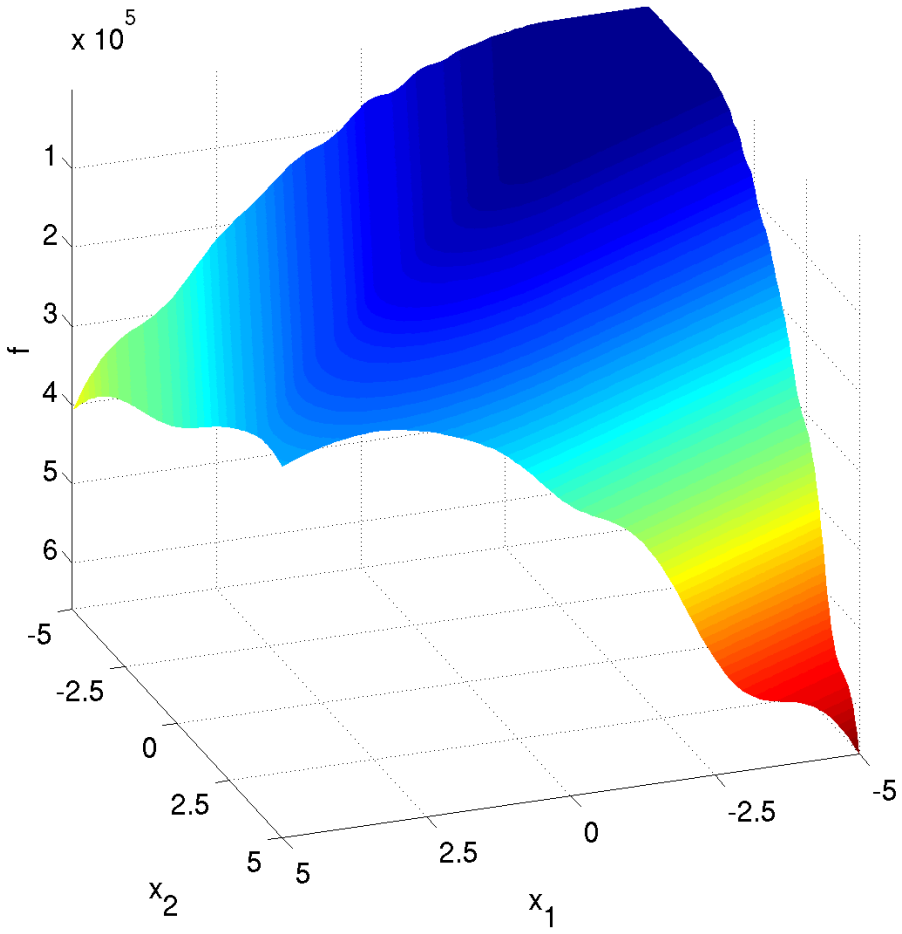
- $\mathbf{z} = \mathbf{Q}\Lambda^{10}\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}})$
- $s_i = \begin{cases} 10^2 & \text{if } z_i \times x_i^{\text{opt}} > 0 \\ 1 & \text{otherwise} \end{cases}$

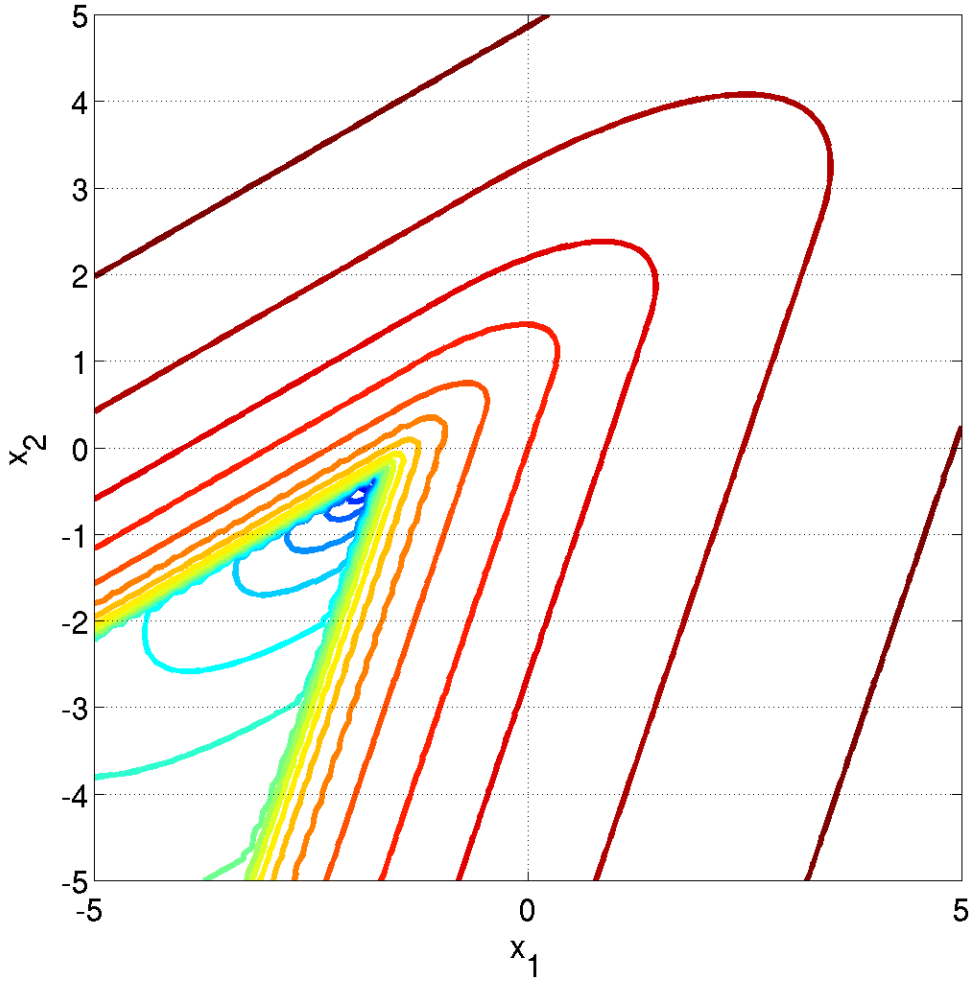
Properties Highly asymmetric function, where only one "hypercone" (with angular base area) with a volume of roughly $1/2^D$ yields low function values. The optimum is located at the tip of this cone.

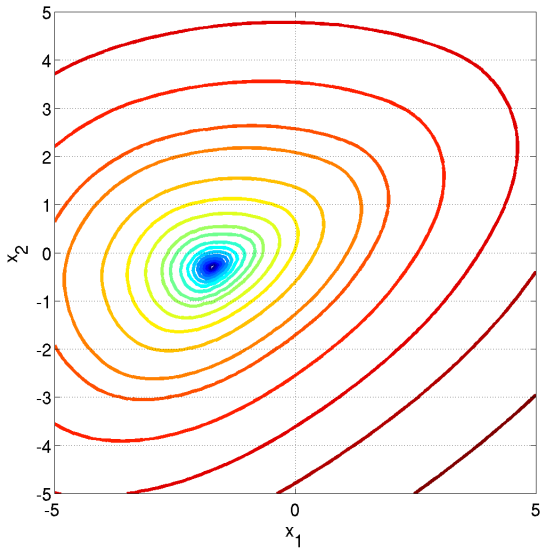
- unimodal

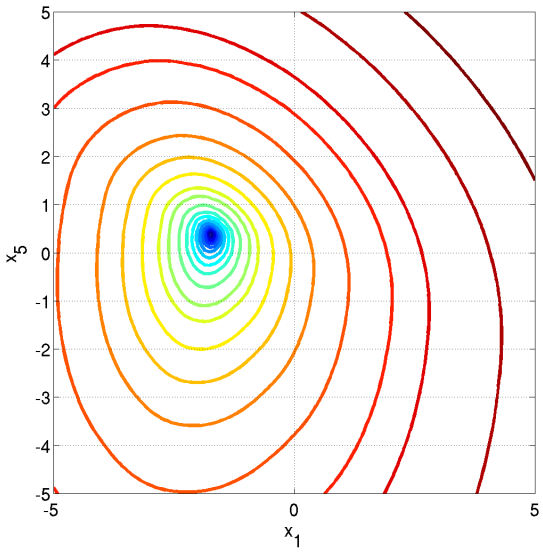
Information gained from this function:

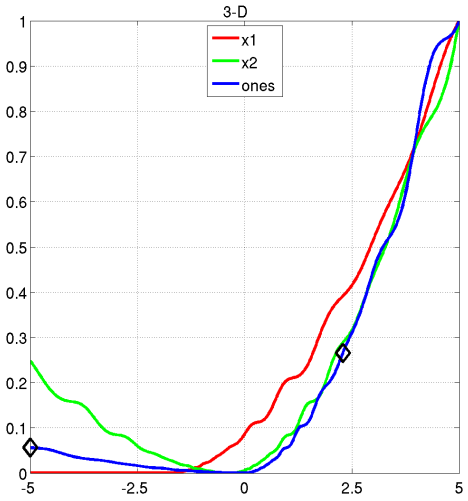
- In comparison to f1: What is the effect of a highly asymmetric landscape?

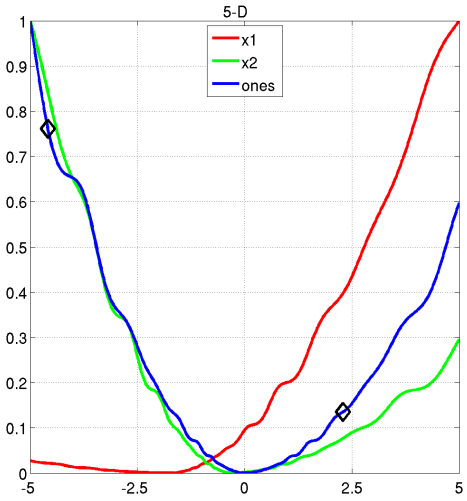


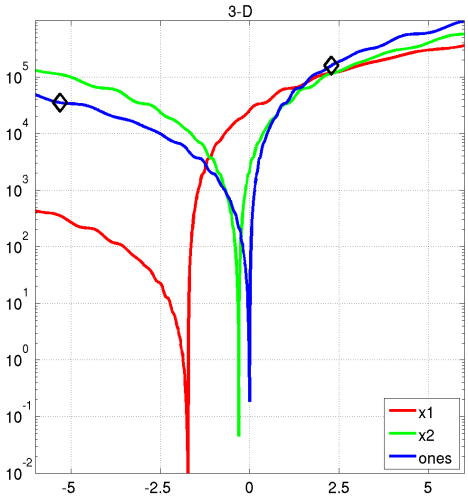

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

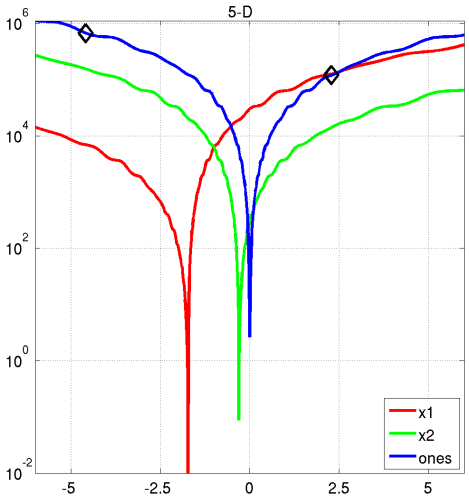

x2
x1

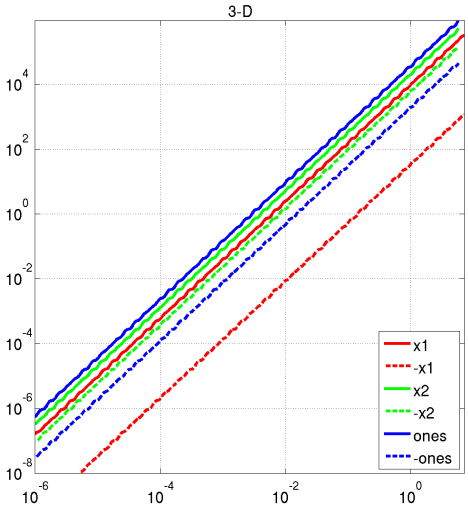

x5
x1

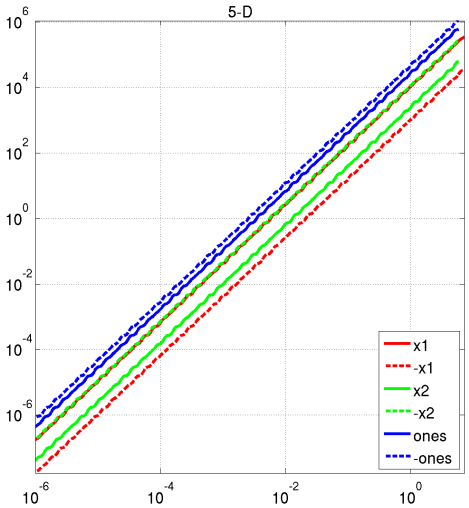

3-D
x1
x2
ones

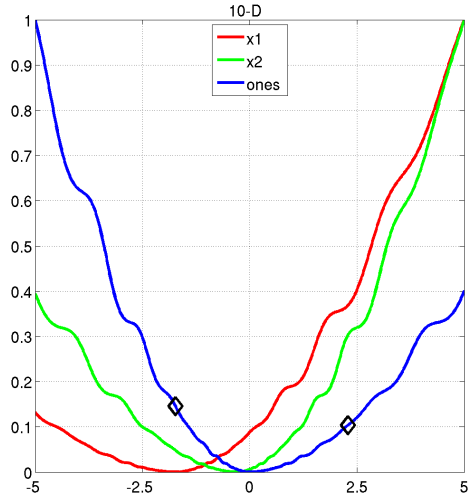

5-D
x1
x2
ones

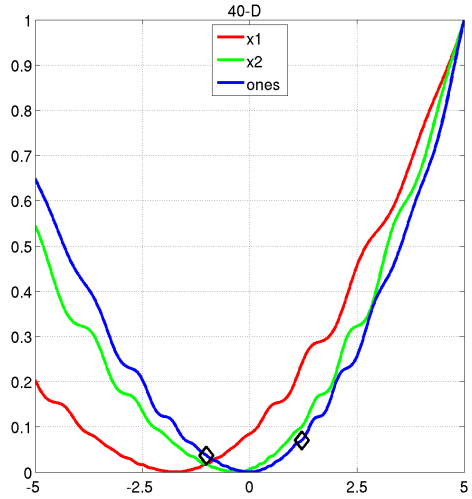

3-D
x1
x2
ones

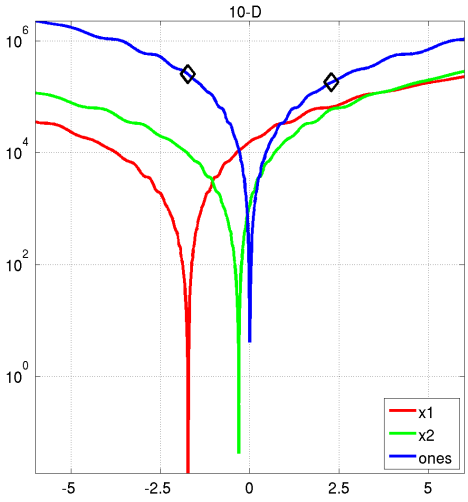

5-D
x1
x2
ones

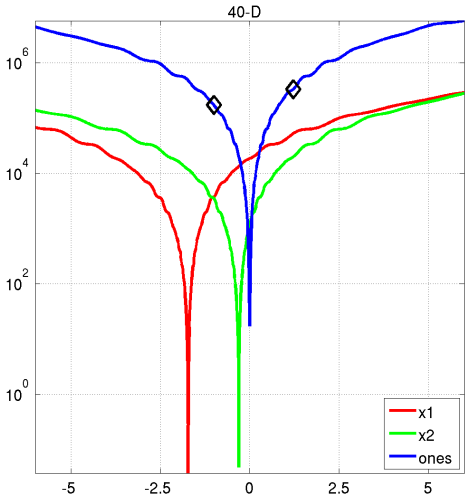

3-D
x1
-x1
x2
-x2
ones
-ones

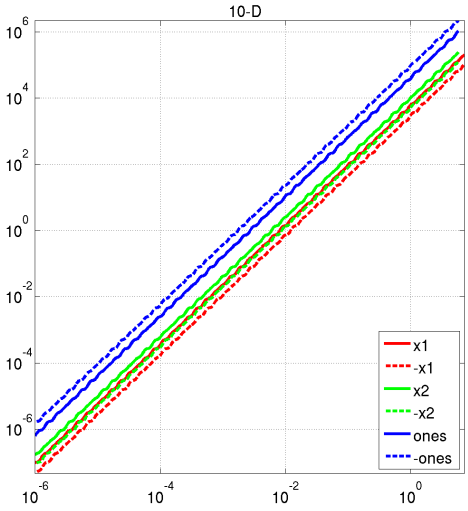

5-D
x1
-x1
x2
-x2
ones
-ones

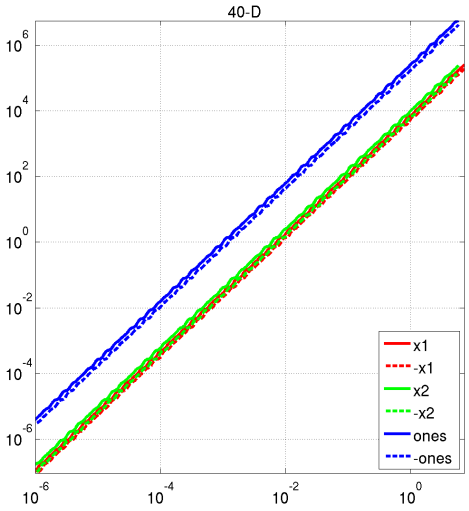

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

2.7 Step Ellipsoidal Function

$$f_7(\mathbf{x}) = 0.1 \max\left(|\hat{z}_1|/10^4, \sum_{i=1}^{D} 10^{2\frac{i-1}{D-1}} z_i^2\right) + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{7}$$

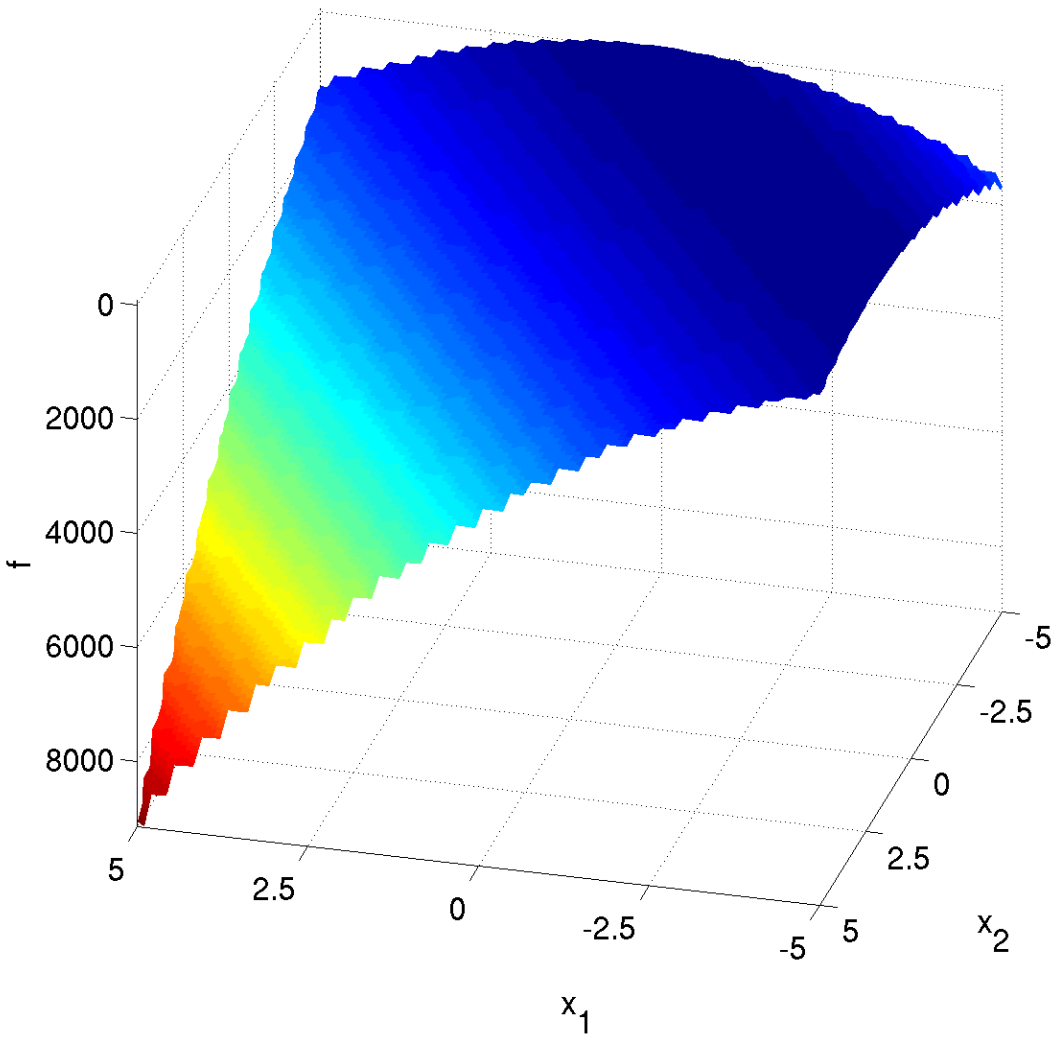
- $\hat{\mathbf{z}} = \Lambda^{10}\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}})$
- $\tilde{z}_i = \begin{cases} \lfloor 0.5 + \hat{z}_i \rfloor & \text{if } [\hat{z}_i \longrightarrow] \ |\hat{z}_i| > 0.5 \\ \lfloor 0.5 + 10\,\hat{z}_i \rfloor / 10 & \text{otherwise} \end{cases}$ for $i = 1, \ldots,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

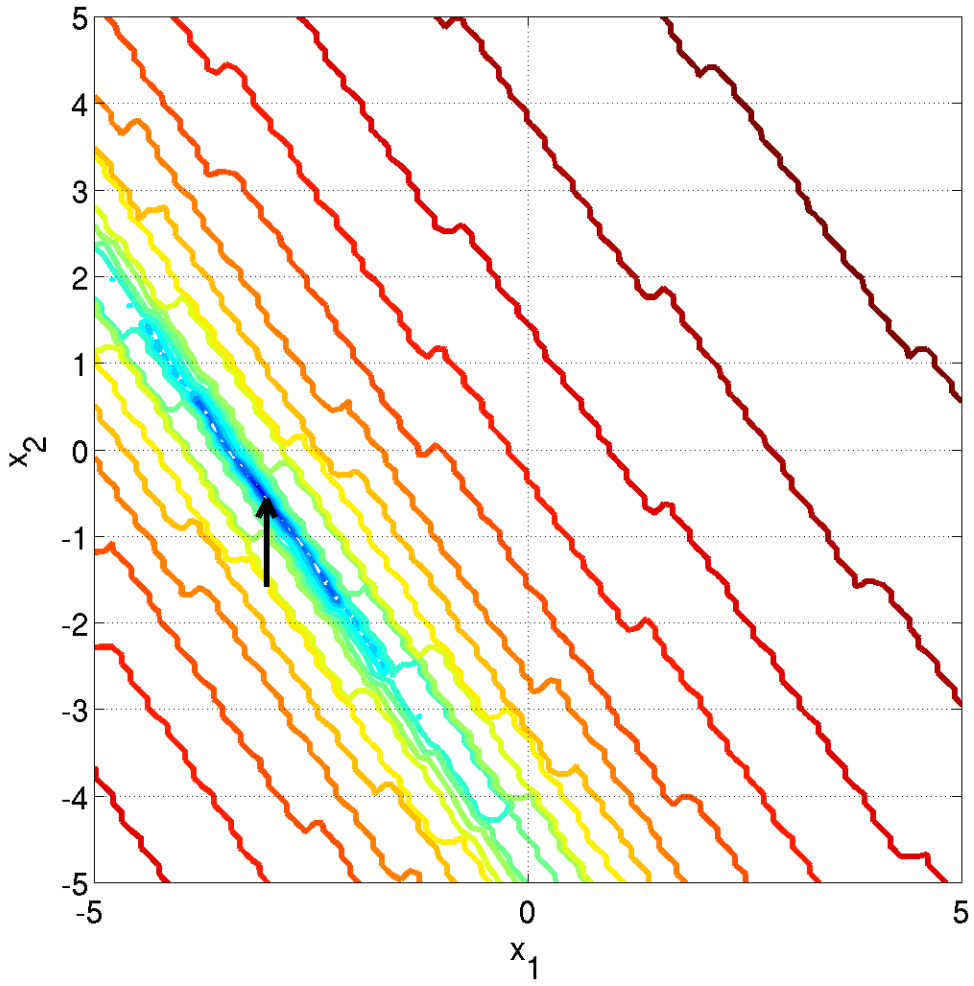
Properties The function consists of many plateaus of different sizes. Apart from a small area close to the global optimum, the gradient is zero almost everywhere.

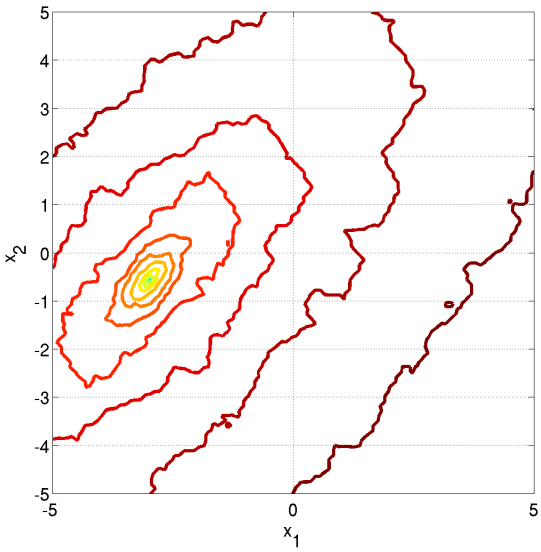
- unimodal, non-separable, conditioning is about 100

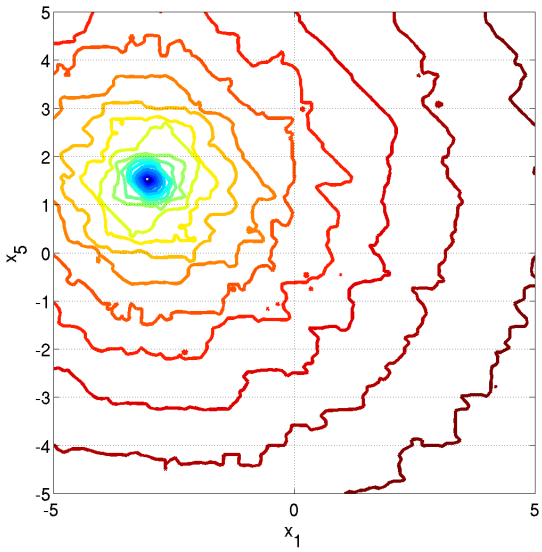
Information gained from this function:

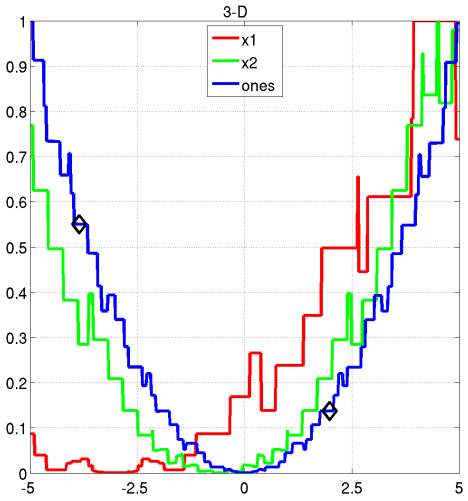
- Does the search get stuck on plateaus?

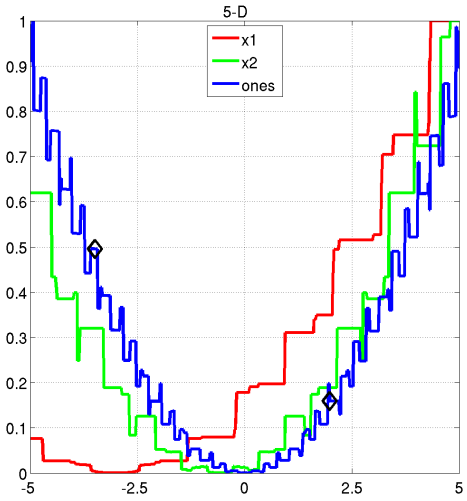


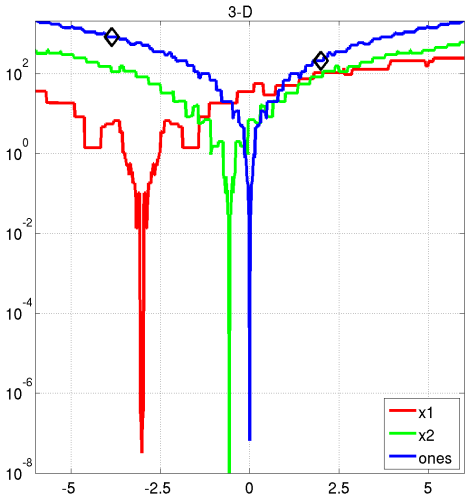

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

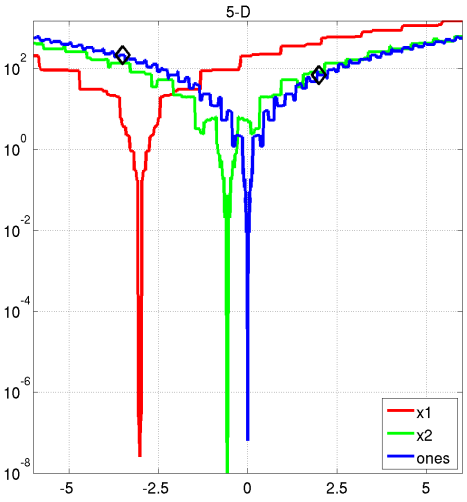

x2
x1

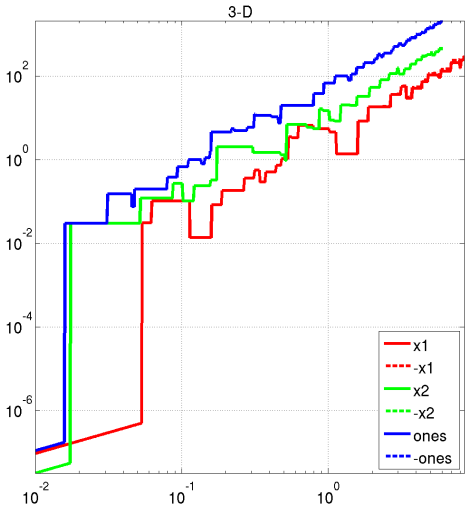

x5
x1

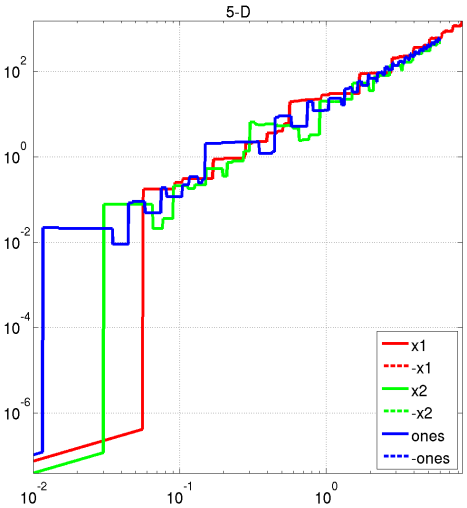

3-D
x1
x2
ones

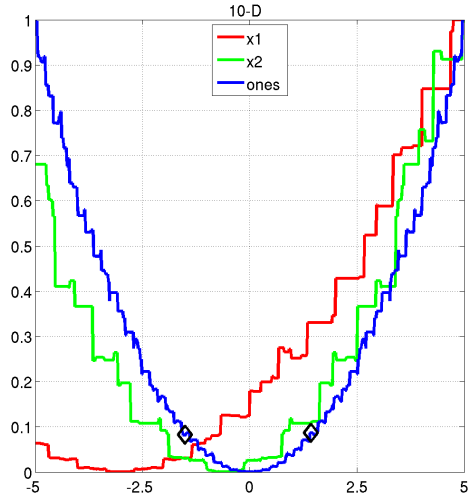

5-D
x1
x2
ones

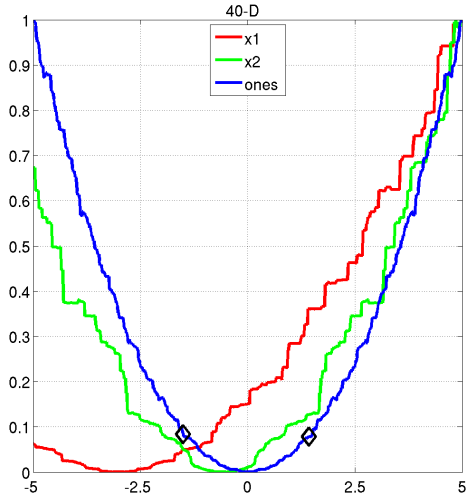

3-D
x1
x2
ones

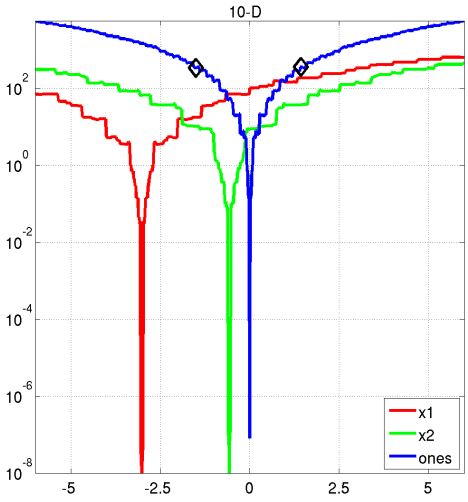

5-D
x1
x2
ones

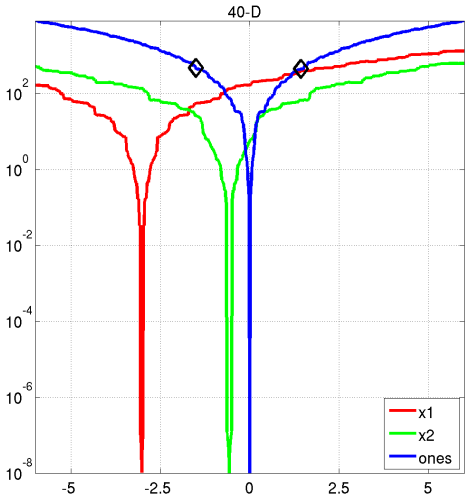

3-D
x1
-x1
x2
-x2
ones
-ones

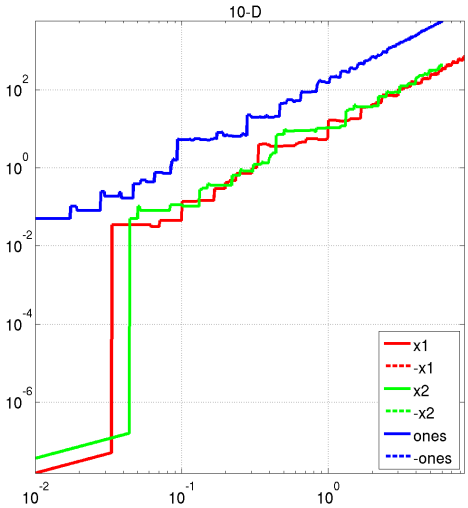

5-D
x1
-x1
x2
-x2
ones
-ones

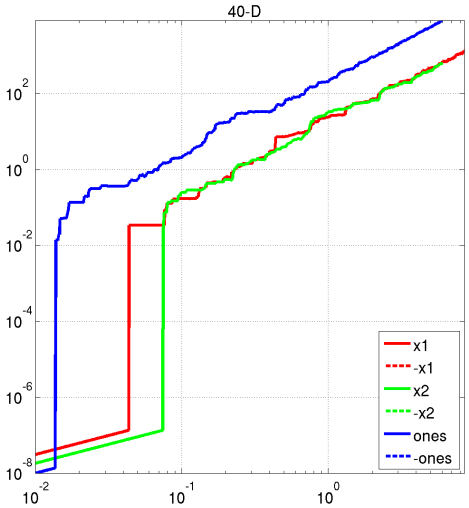

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

2.8 Rosenbrock Function, original

$$f_8(\mathbf{x}) = \sum_{i=1}^{D-1} \left(100 \left(z_i^2 - z_{i+1}\right)^2 + (z_i - 1)^2\right) + f_{\text{opt}} \tag{8}$$

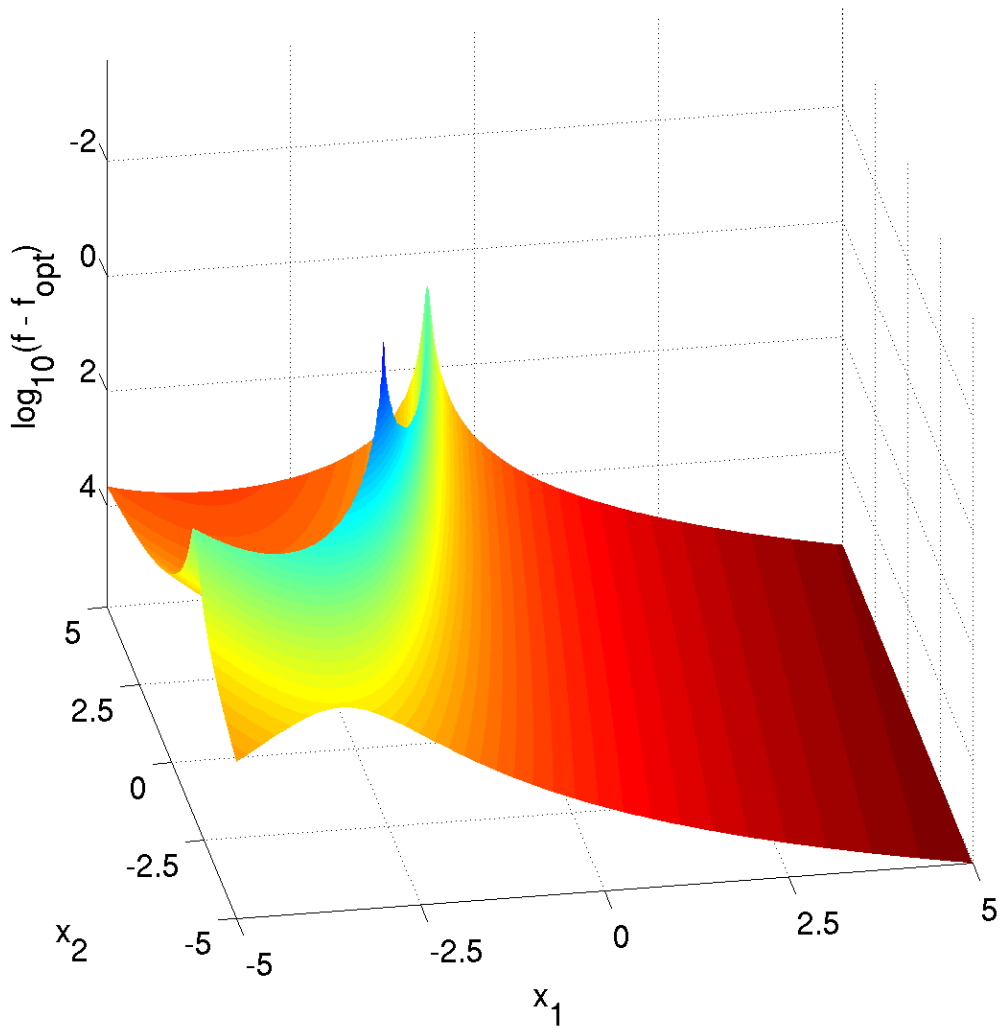
- $\mathbf{z} = \max\left(1, \frac{\sqrt{D}}{8}\right)(\mathbf{x} - \mathbf{x}^{\text{opt}}) + 1$
- $\mathbf{z}^{\text{opt}} = \mathbf{1}$

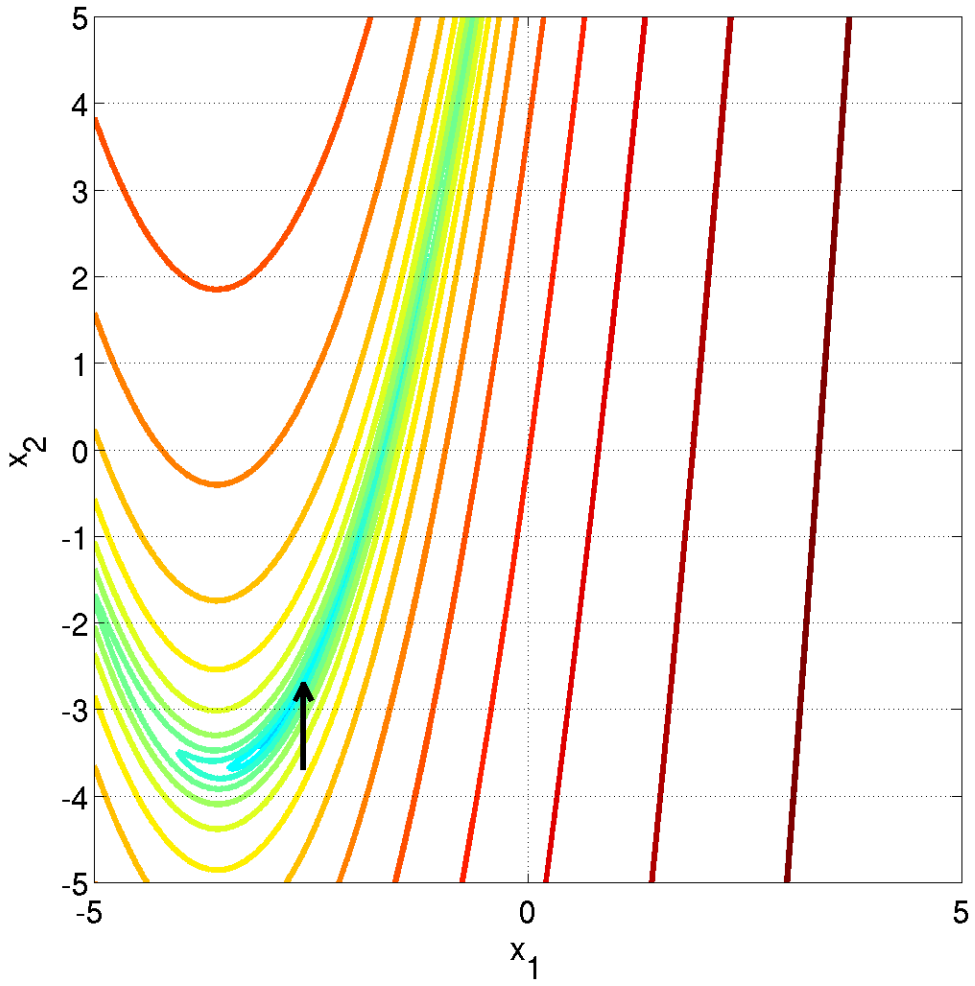
Properties So-called banana function due to its 2-D contour lines as a bent ridge (or valley) [8]. In the beginning, the prominent first term of the function definition attracts to the point $\mathbf{z} = \mathbf{0}$. Then, a long bending valley needs to be followed to reach the global optimum. The ridge changes its orientation $D - 1$ times. Exceptionally, here $\mathbf{x}^{\text{opt}} \in [-3, 3]^D$.

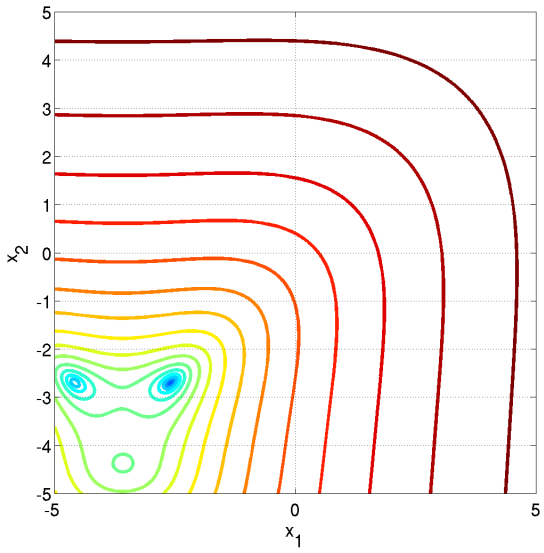
- tri-band dependency structure, in larger dimensions the function has a local optimum with an attraction volume of about 25%

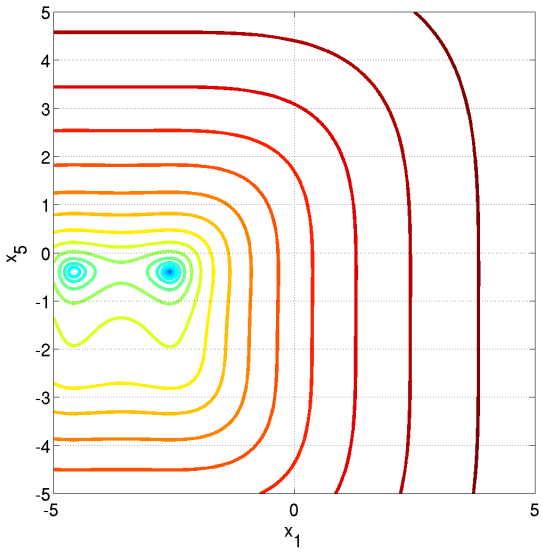
Information gained from this function:

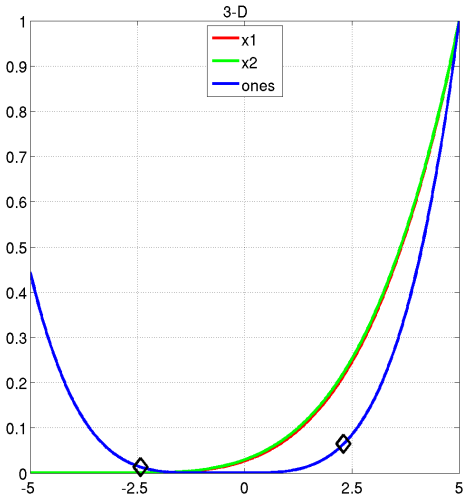
- Can the search follow a long path with $D - 1$ changes in the direction?

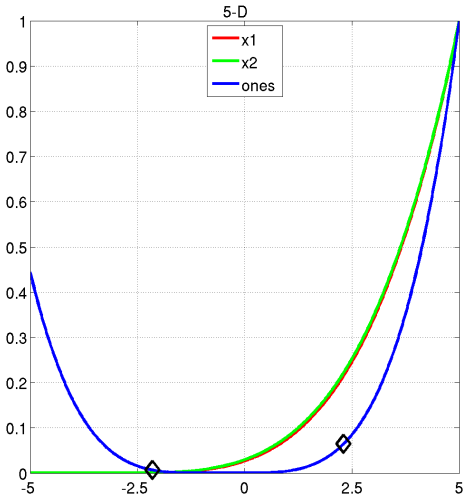


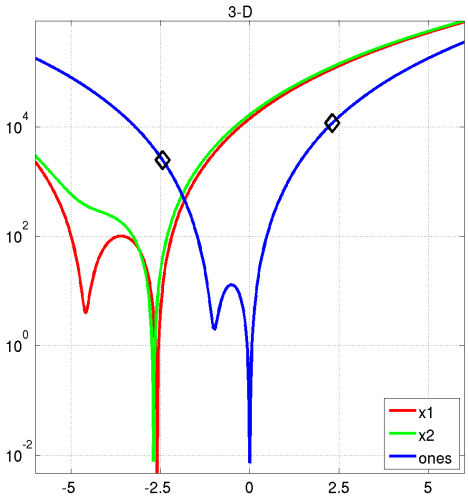

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

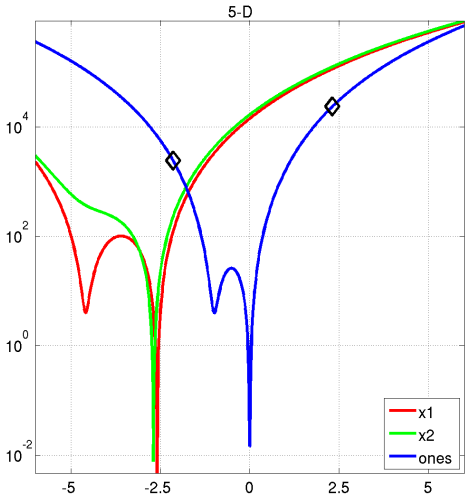

x2
x1

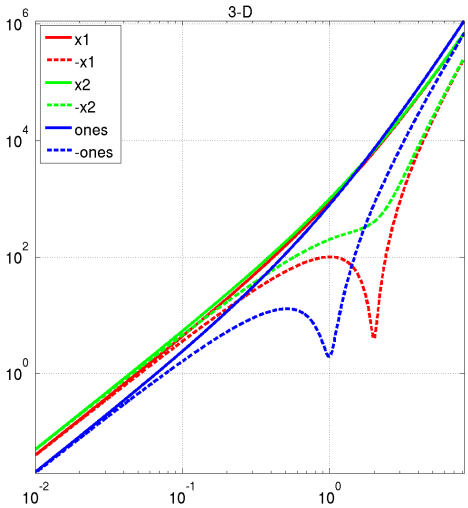

x5
x1

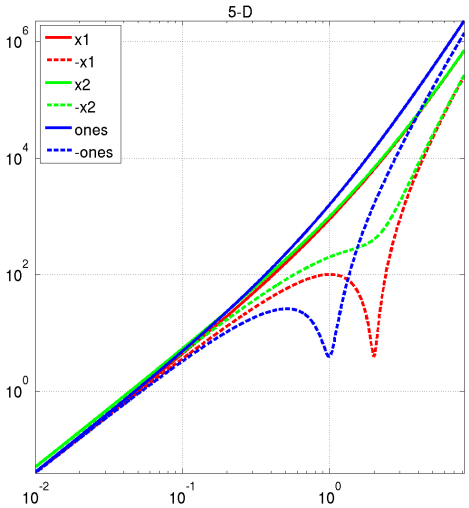

3-D
x1
x2
ones

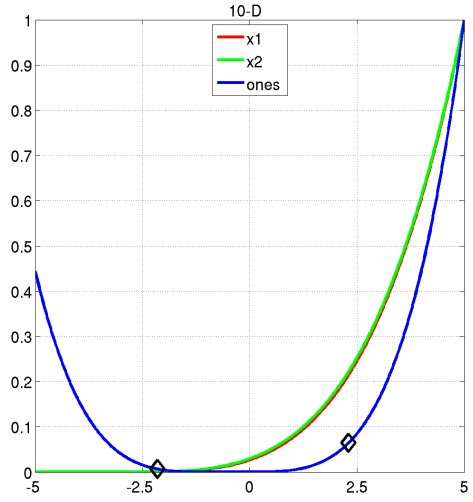

5-D
x1
x2
ones

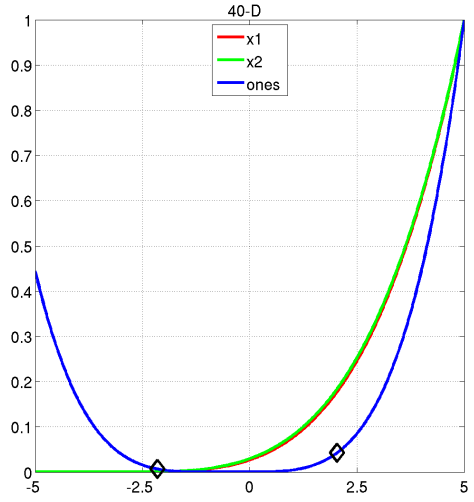

3-D
x1
x2
ones

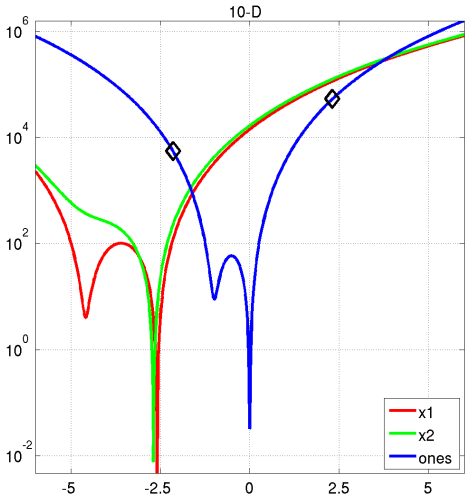

5-D
x1
x2
ones

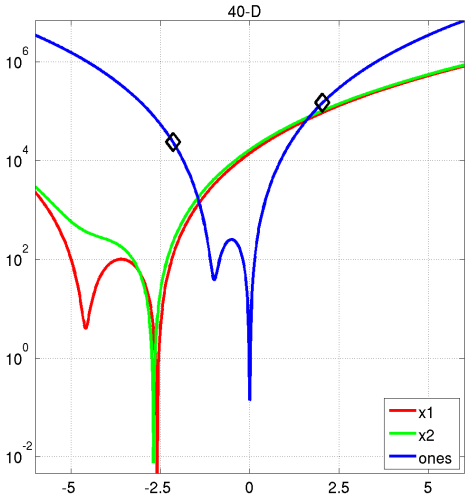

3-D
x1
-x1
x2
-x2
ones
-ones

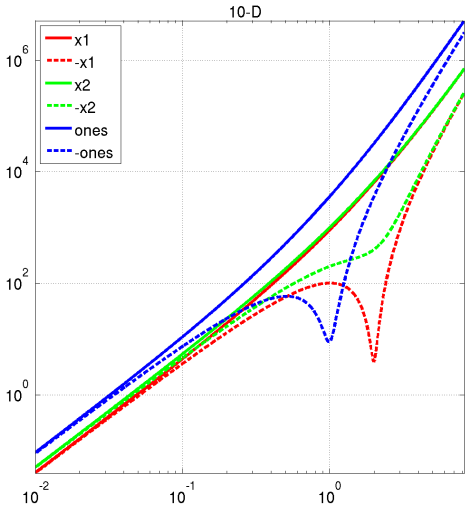

5-D
x1
-x1
x2
-x2
ones
-ones

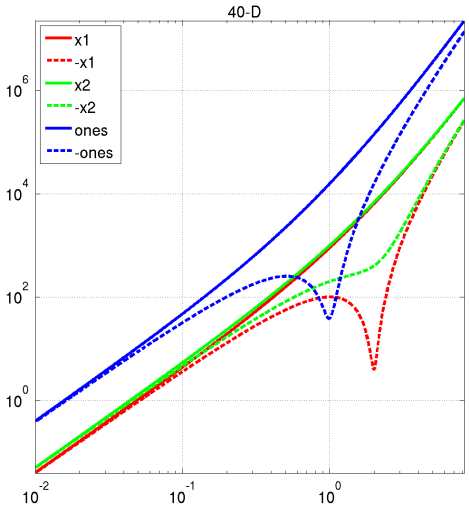

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

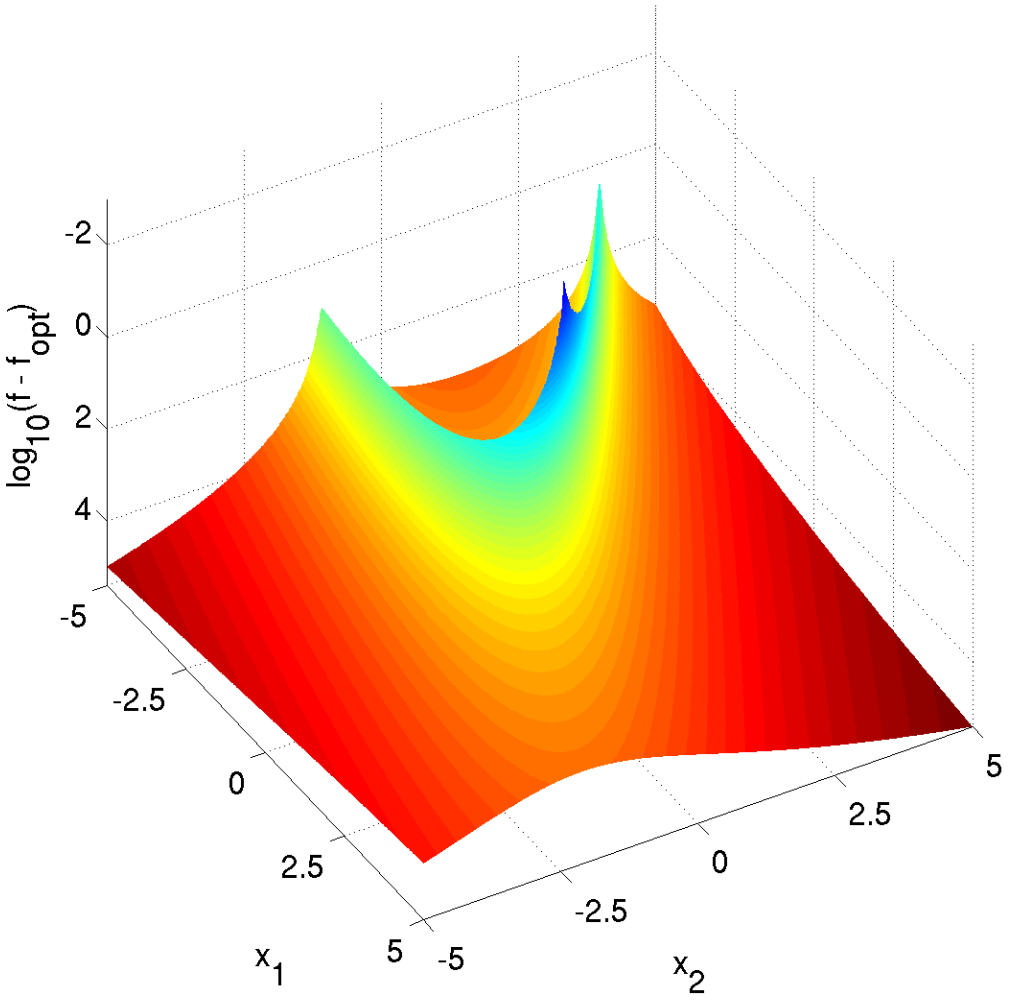
2.9 Rosenbrock Function, rotated

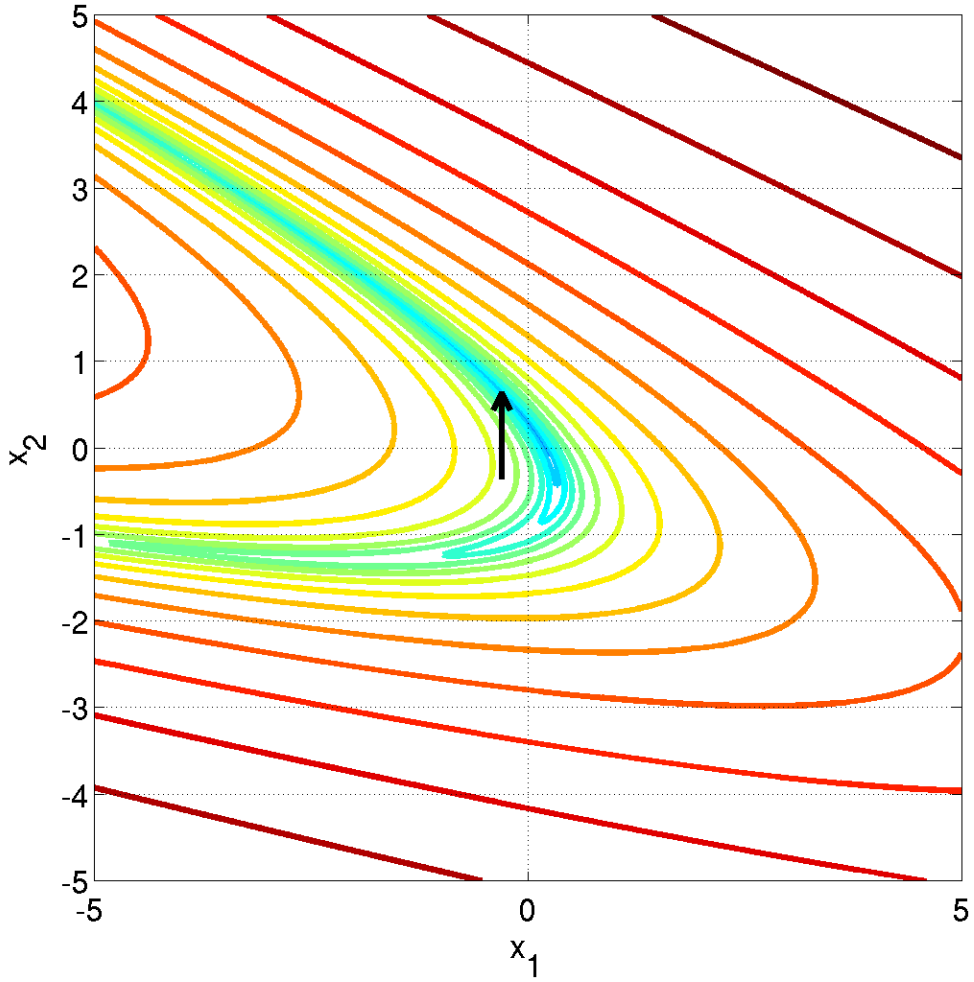
$$f_9(\mathbf{x}) = \sum_{i=1}^{D-1} \left(100 \left(z_i^2 - z_{i+1}\right)^2 + (z_i - 1)^2\right) + f_{\text{opt}} \tag{9}$$

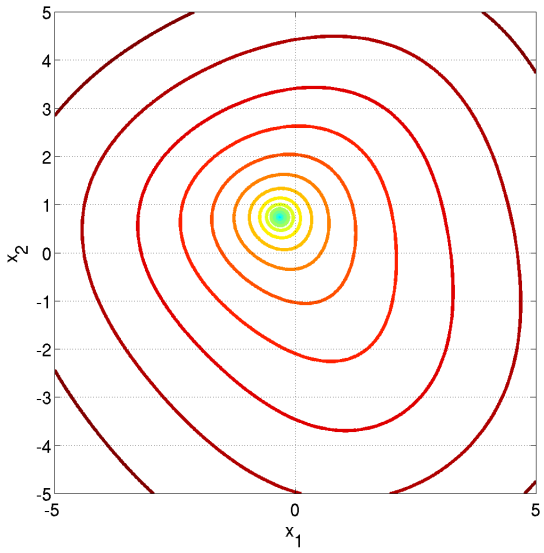
- $\mathbf{z} = \max\left(1, \frac{\sqrt{D}}{8}\right)\mathbf{R}\mathbf{x} + \mathbf{1}/2$
- $\mathbf{z}^{\text{opt}} = \mathbf{1}$

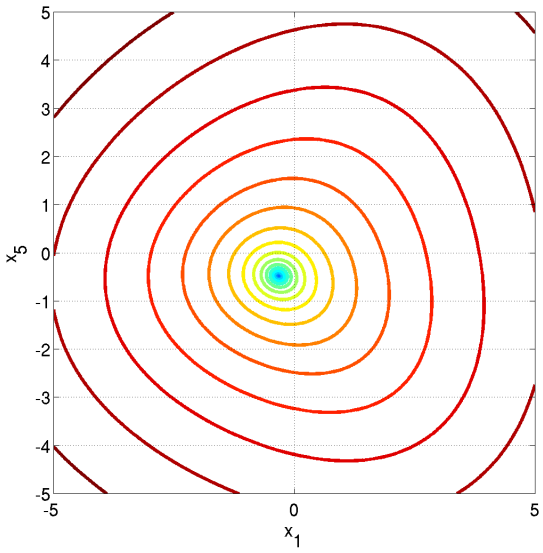
Properties rotated version of the previously defined Rosenbrock function.
Information gained from this function:

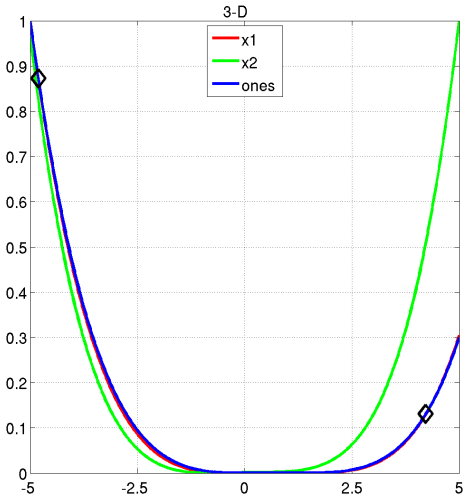
- In comparison to f8: Can the search follow a long path with $D-1$ changes in the direction without exploiting partial separability?

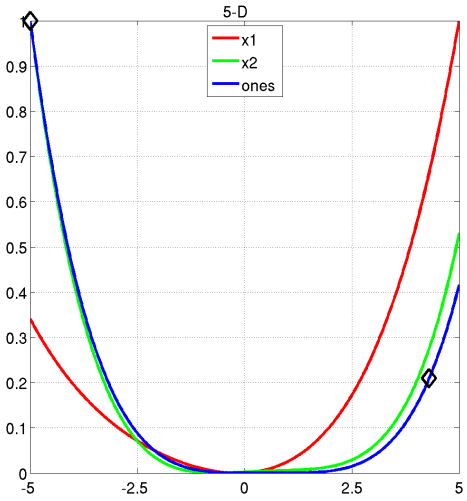


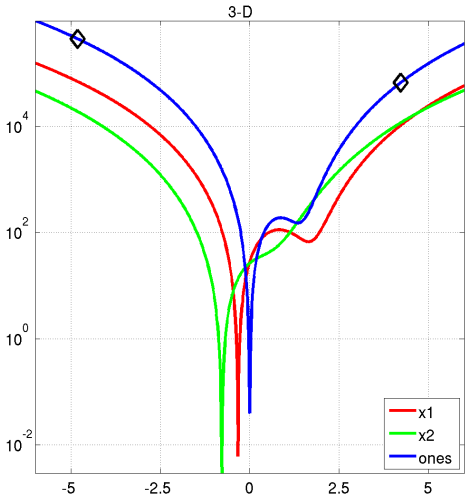

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

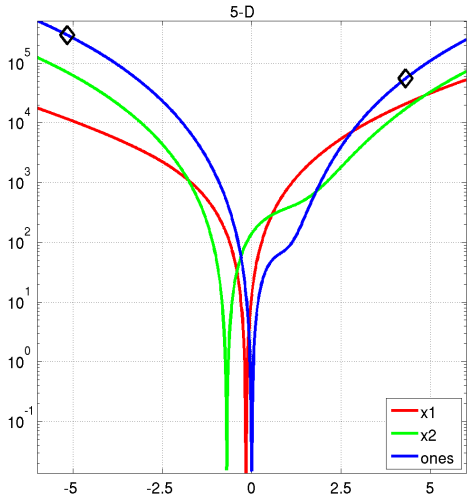

x2
x1

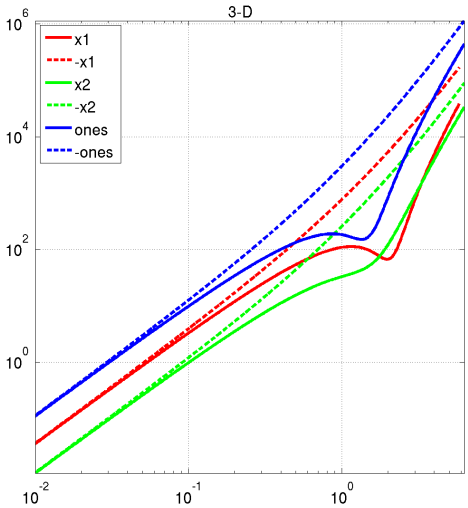

x5
x1

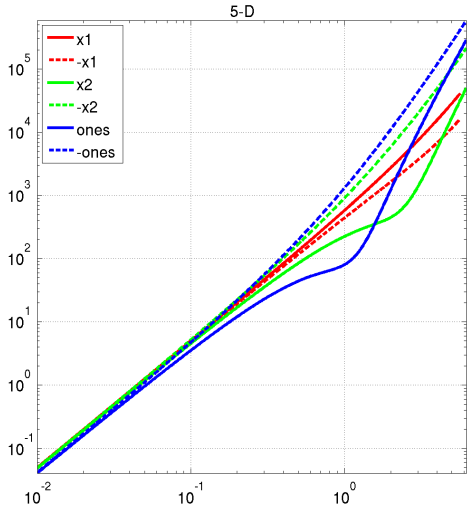

3-D
x1
x2
ones

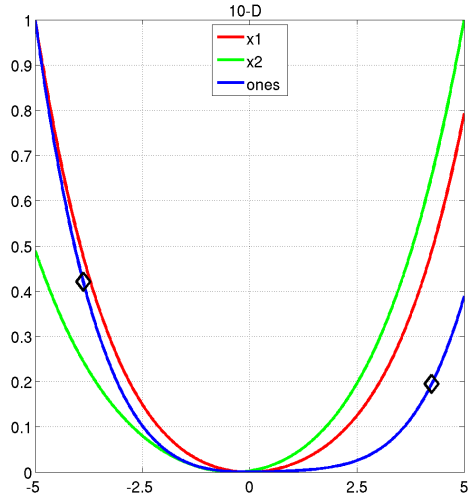

5-D
x1
x2
ones

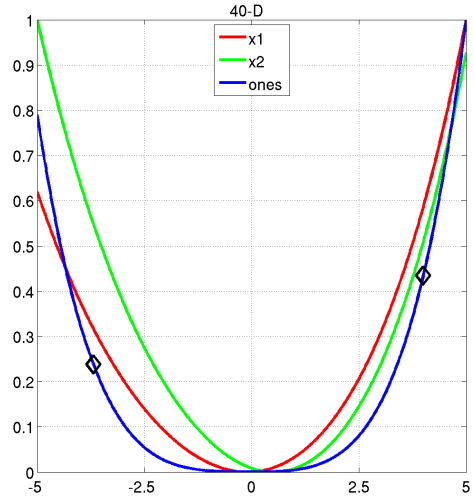

3-D
x1
x2
ones

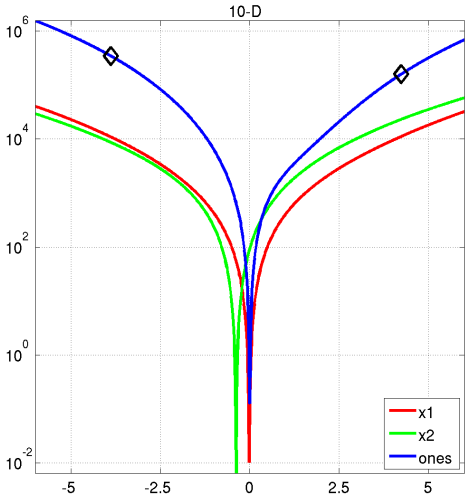

5-D
x1
x2
ones

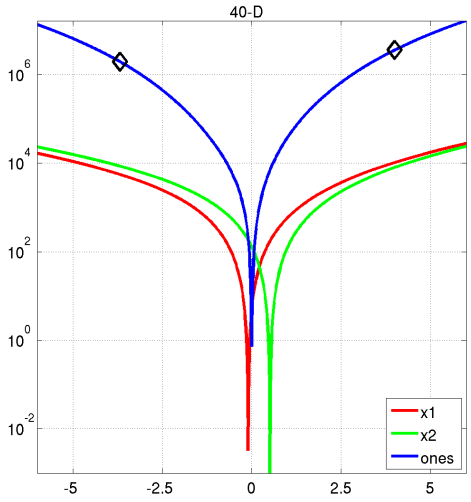

3-D
x1
-x1
x2
-x2
ones
-ones

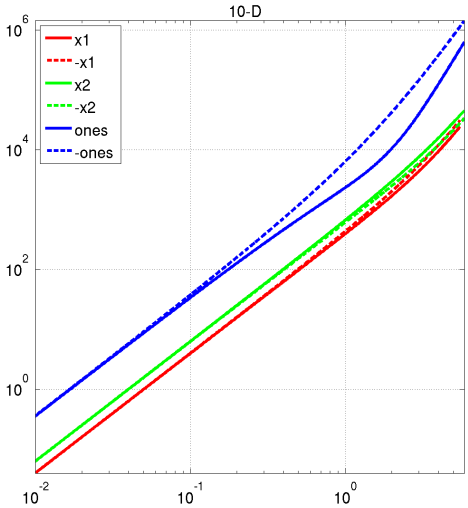

5-D
x1
-x1
x2
-x2
ones
-ones

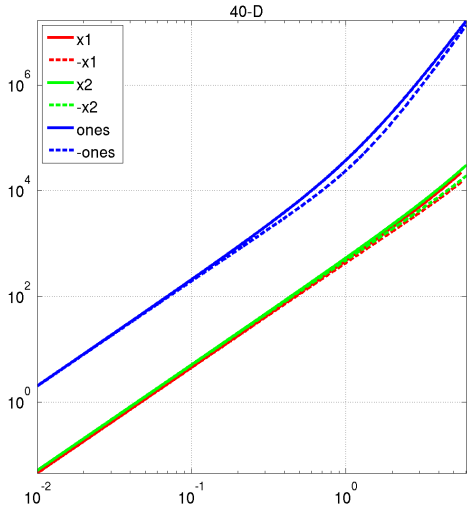

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

3 Functions with high conditioning and unimodal

3.10 Ellipsoidal Function

$$f_{10}(\mathbf{x}) = \sum_{i=1}^{D} 10^{6\frac{i-1}{D-1}} z_i^2 + f_{\text{opt}} \tag{10}$$

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

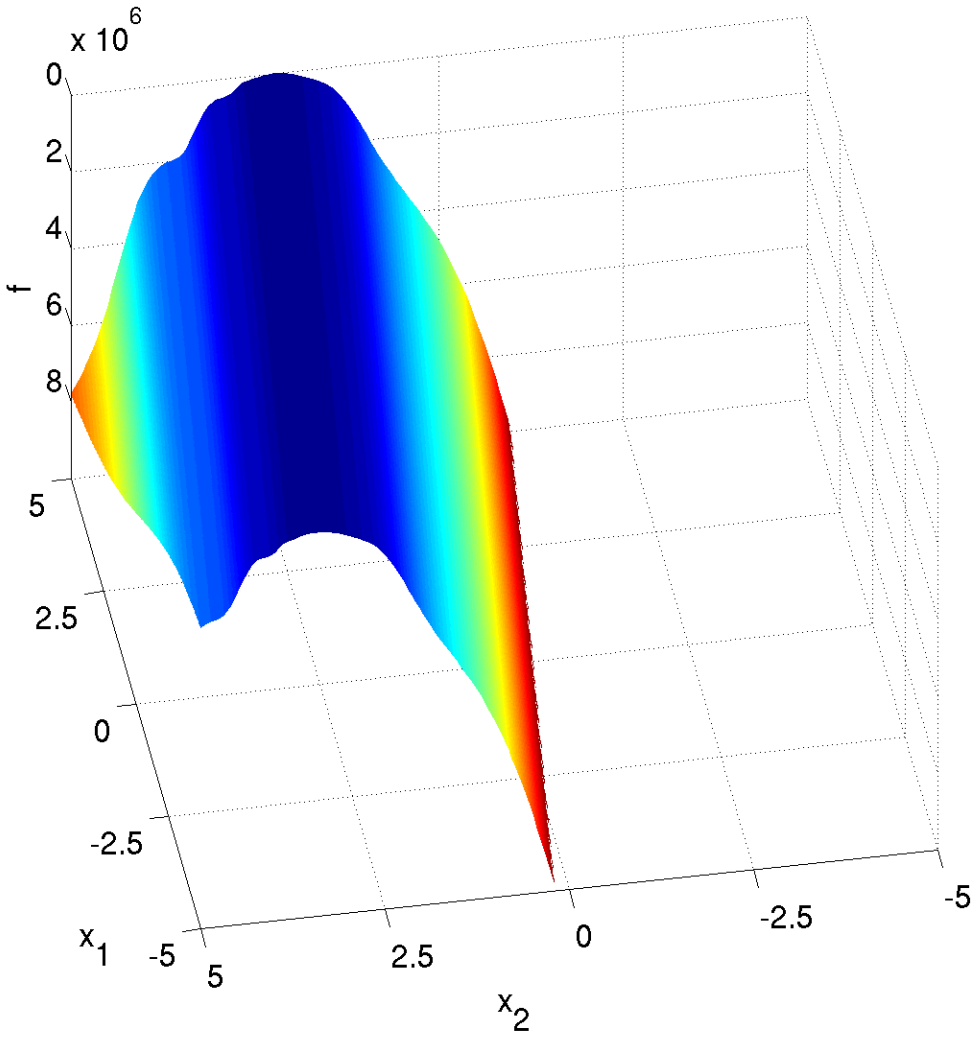
Properties Globally quadratic ill-conditioned function with smooth local irregularities, non-separable counterpart to f_2.

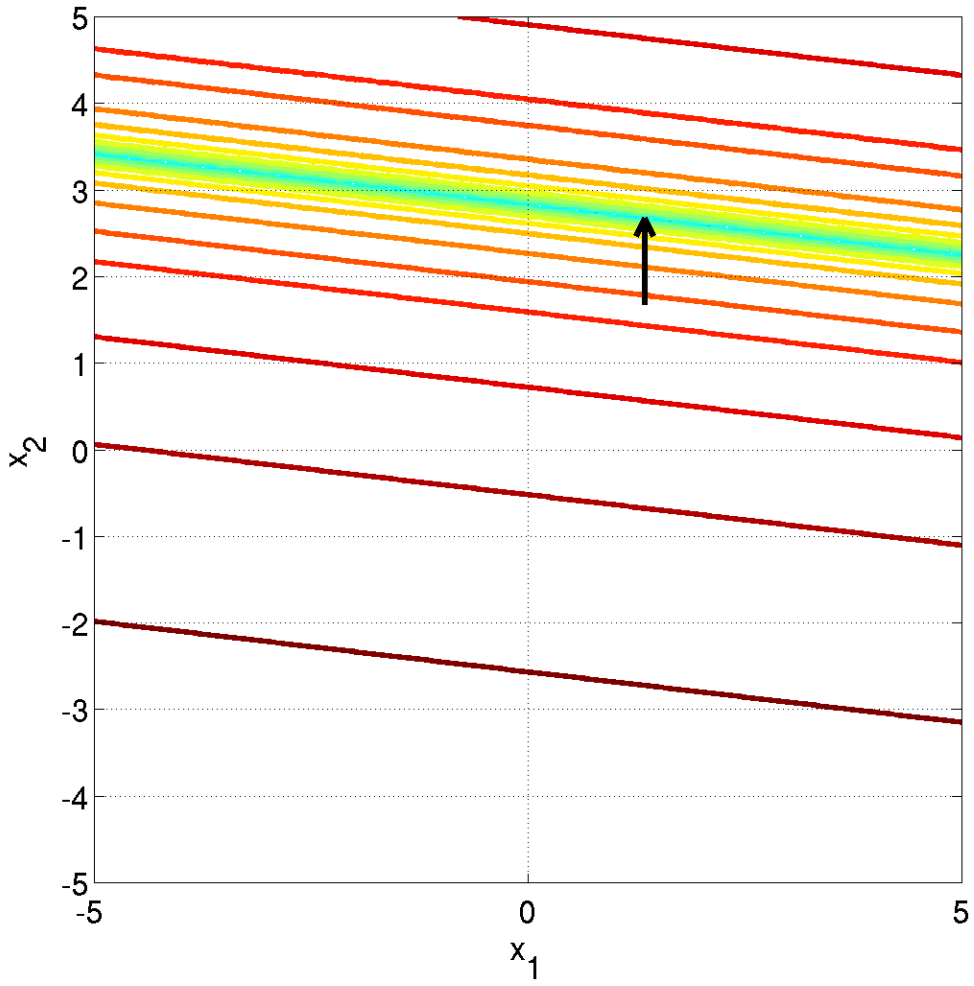
- unimodal, conditioning is 10^6

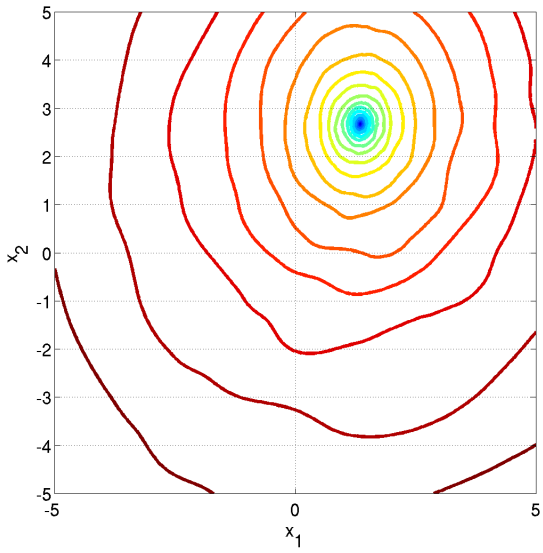
Information gained from this function:

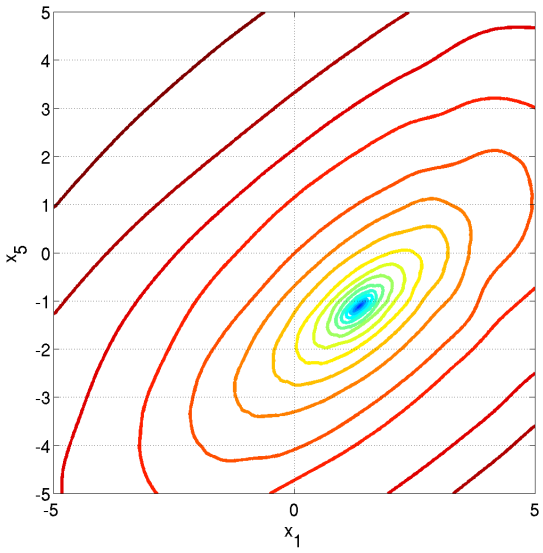
- In comparison to f2: What is the effect of rotation (non-separability)?

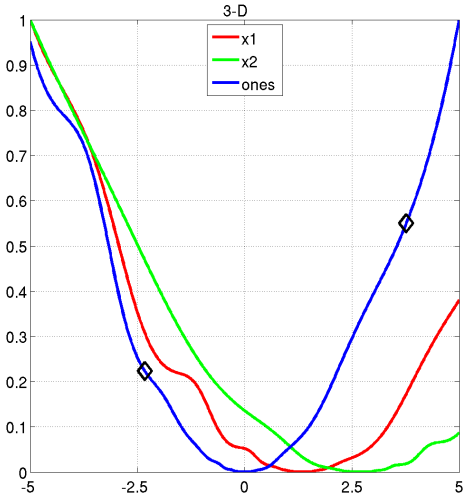
Note: The 3d plot shows only a part of the complete function in the vicinity of the optimum.

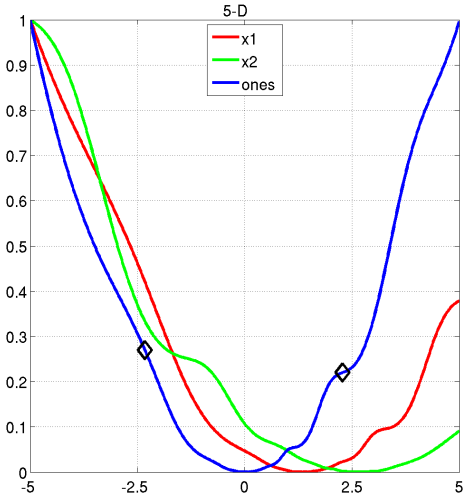


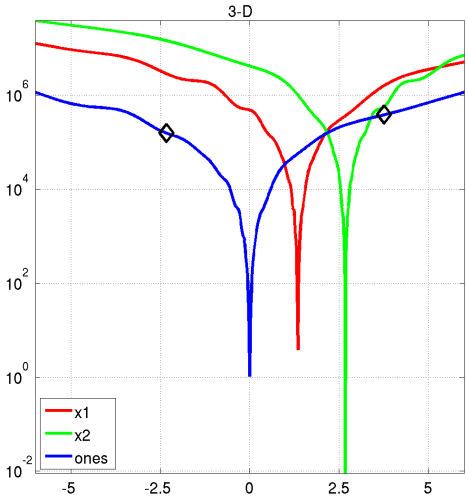

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

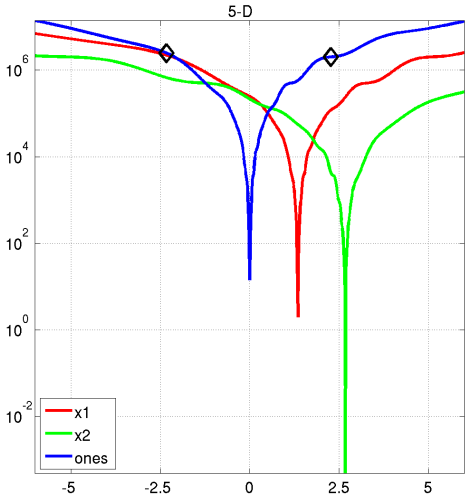

x2
x1

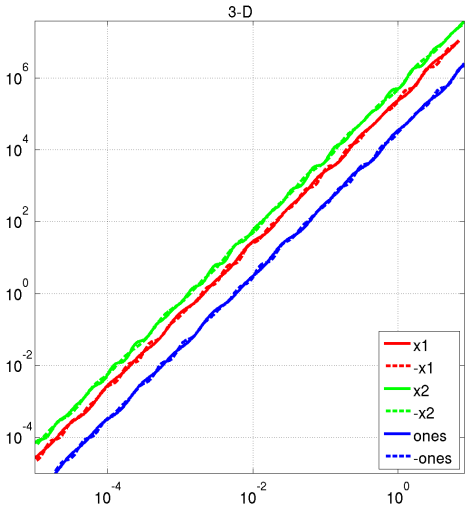

x5
x1

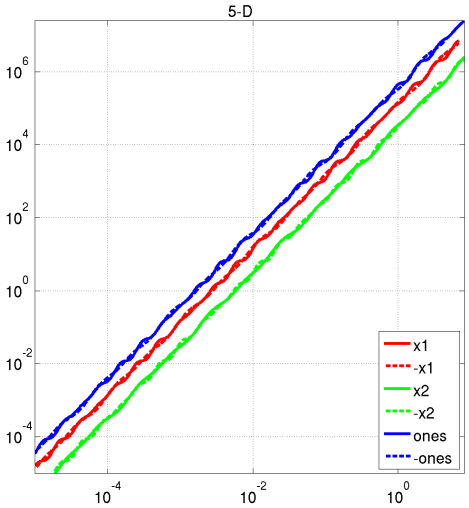

3-D
x1
x2
ones

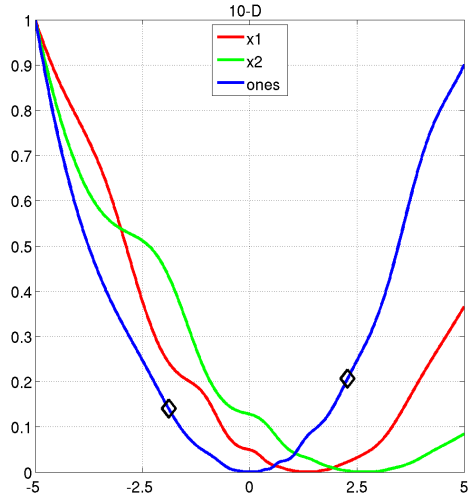

5-D
x1
x2
ones

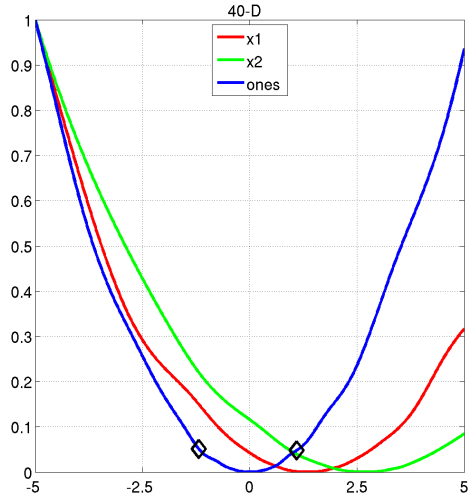

3-D
x1
x2
ones

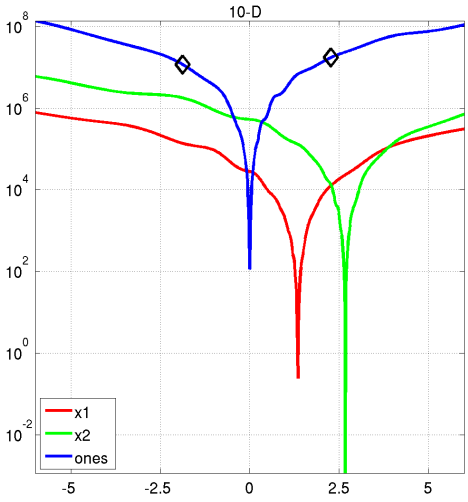

5-D
x1
x2
ones

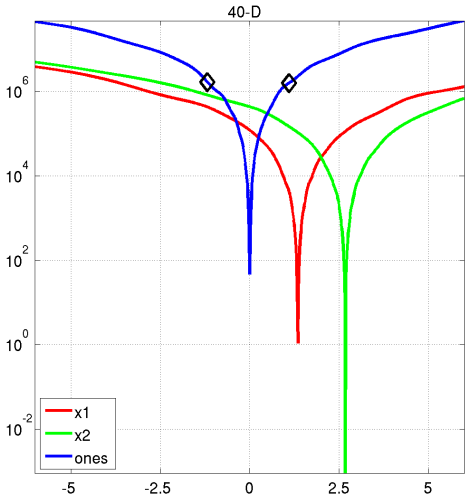

3-D
x1
-x1
x2
-x2
ones
-ones

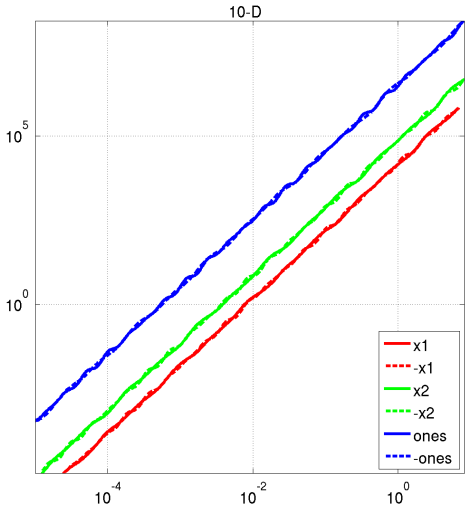

5-D
x1
-x1
x2
-x2
ones
-ones

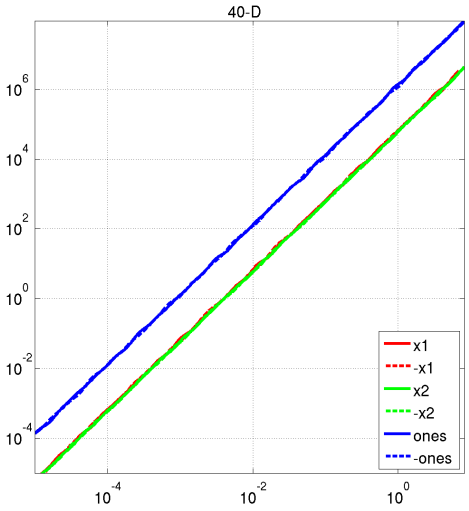

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

3.11 Discus Function

$$f_{11}(\mathbf{x}) = 10^6 z_1^2 + \sum_{i=2}^{D} z_i^2 + f_{\text{opt}} \tag{11}$$

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

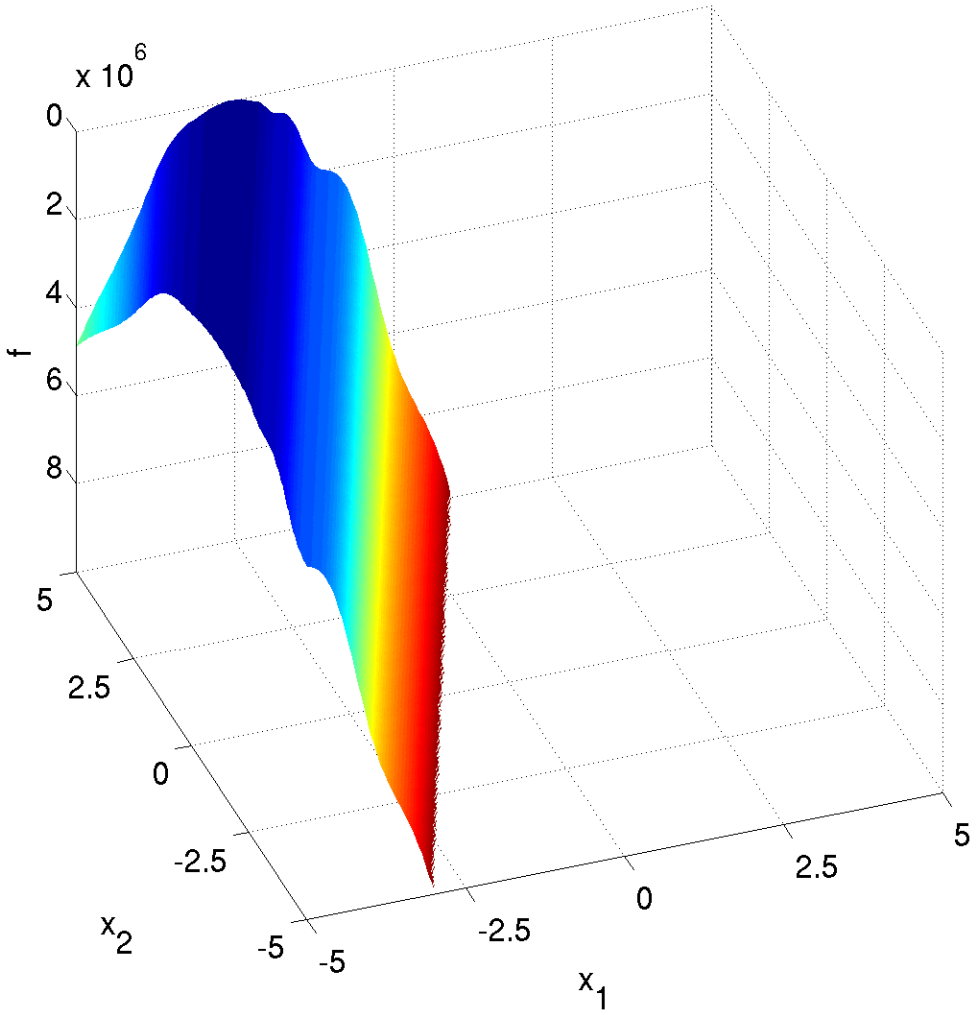
Properties Globally quadratic function with local irregularities. A single direction in search space is a thousand times more sensitive than all others.

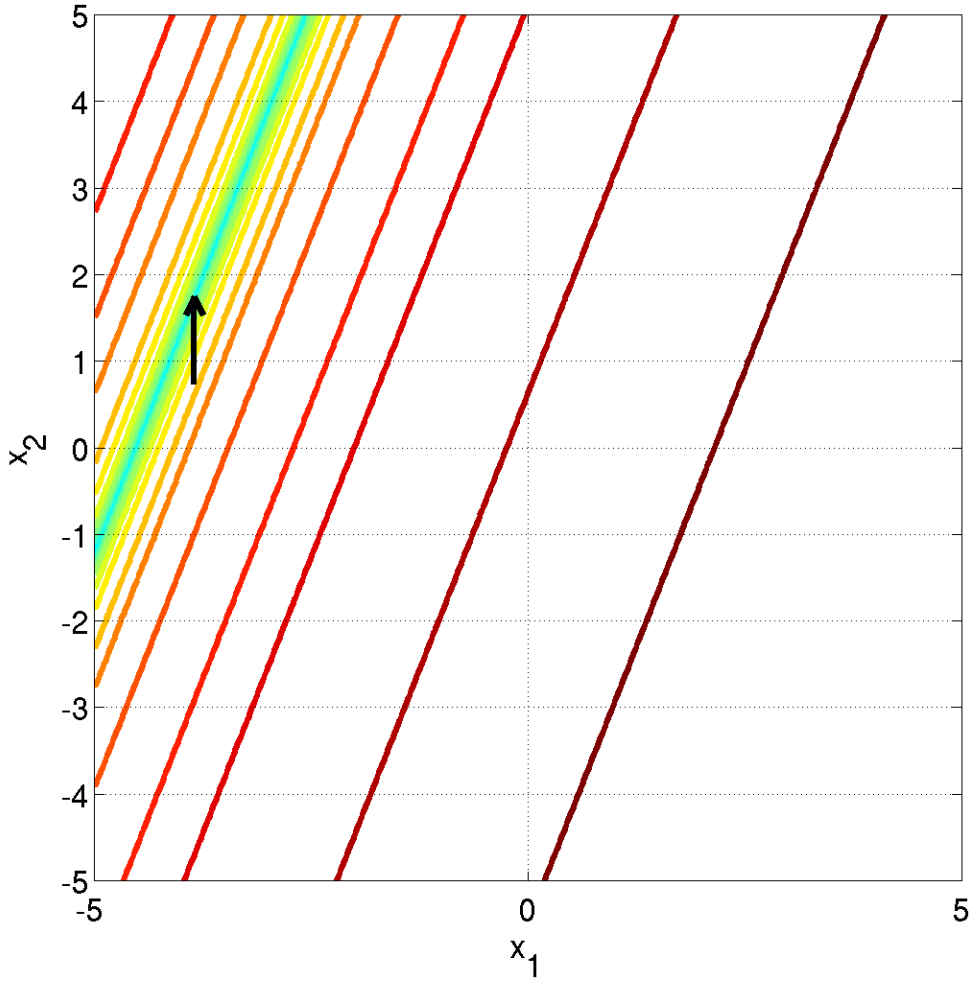
- conditioning is about 10^6

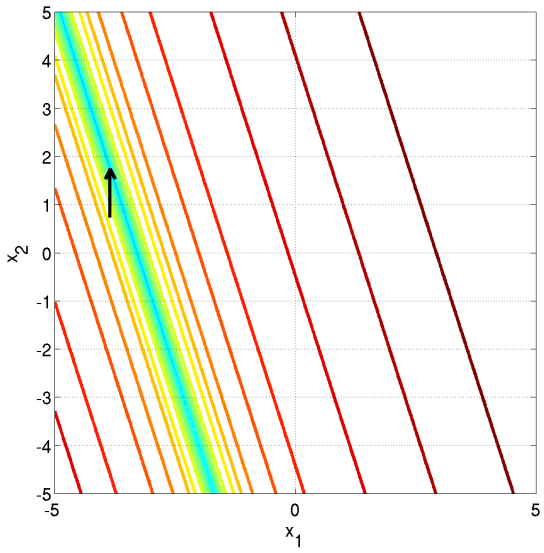
Information gained from this function:

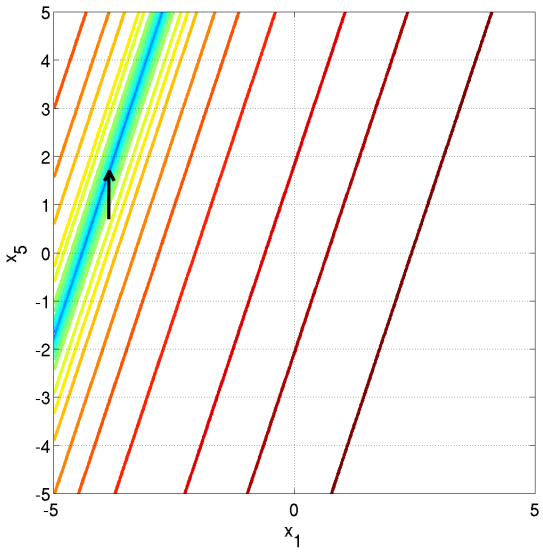
- In comparison to f1: What is the effect of constraint-like penalization?

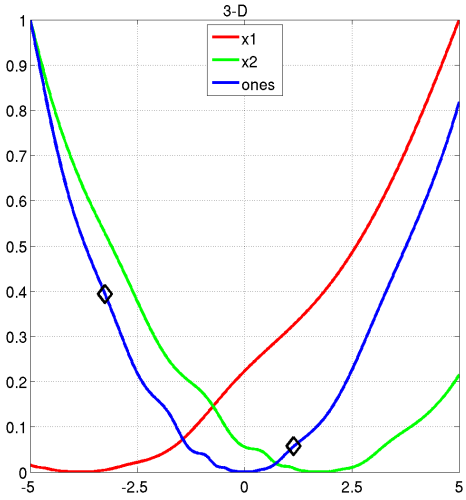
Note: The 3d plot shows only a part of the complete function in the vicinity of the optimum.

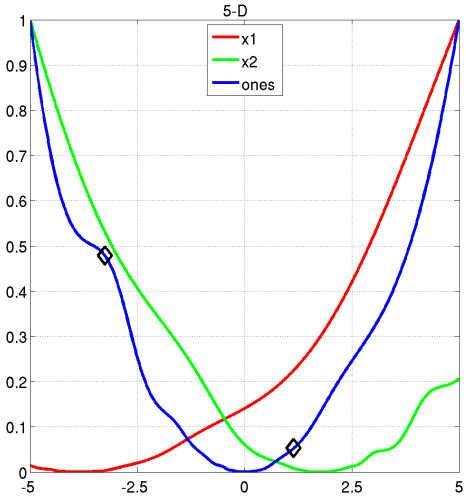


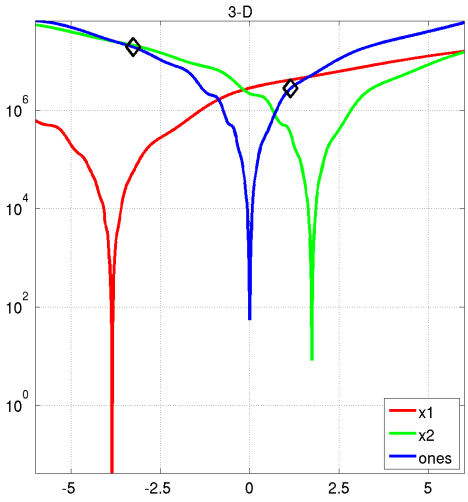

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

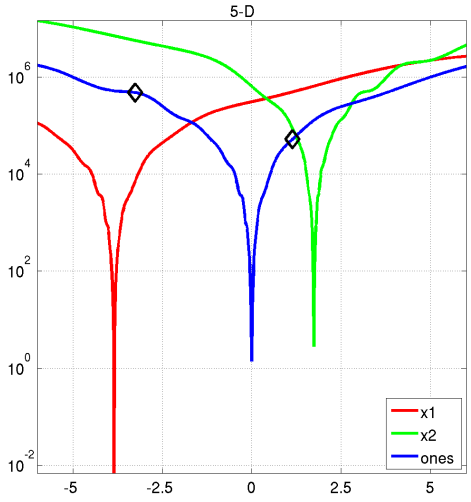

x2
x1

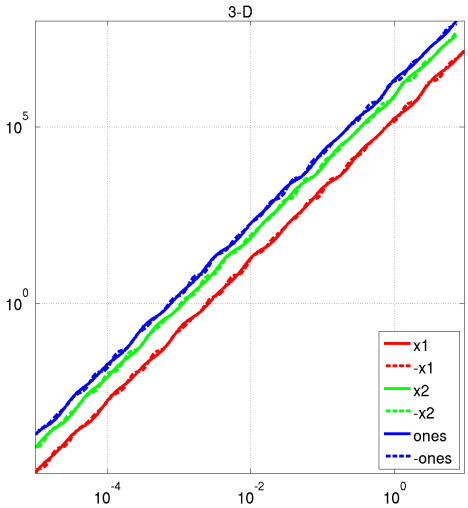

x5
x1

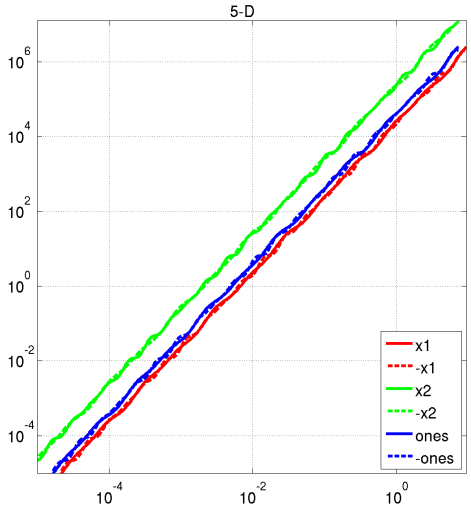

3-D
x1
x2
ones

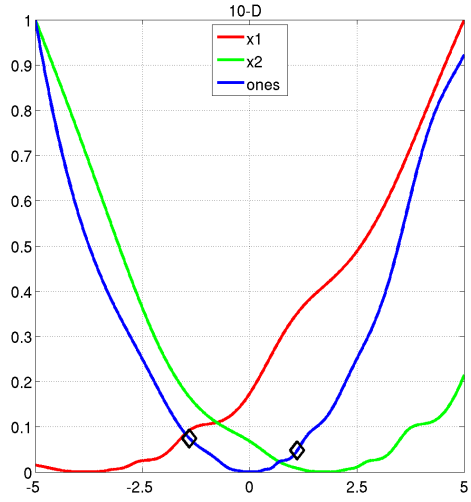

5-D
x1
x2
ones

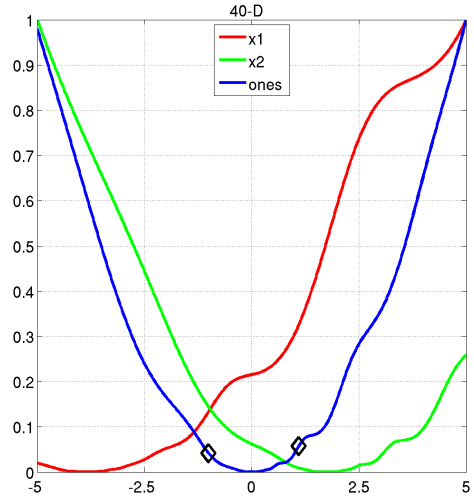

3-D
x1
x2
ones

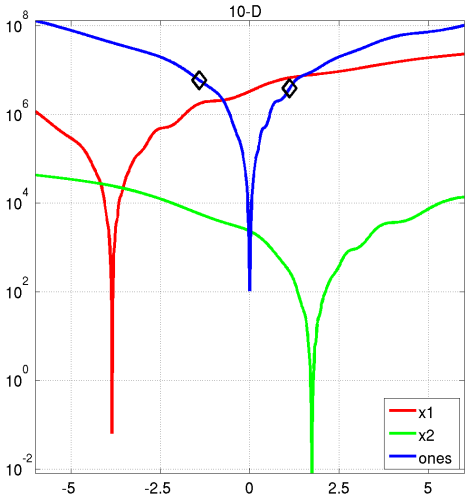

5-D
x1
x2
ones

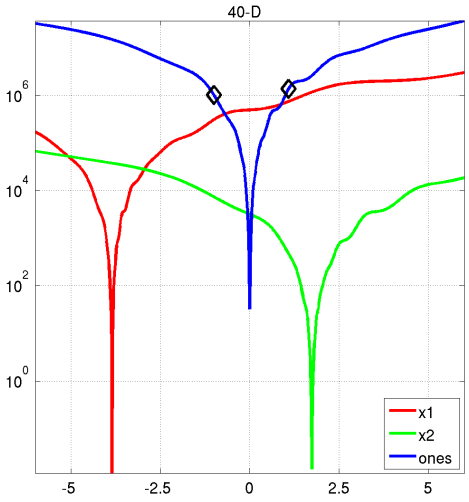

3-D
x1
-x1
x2
-x2
ones
-ones

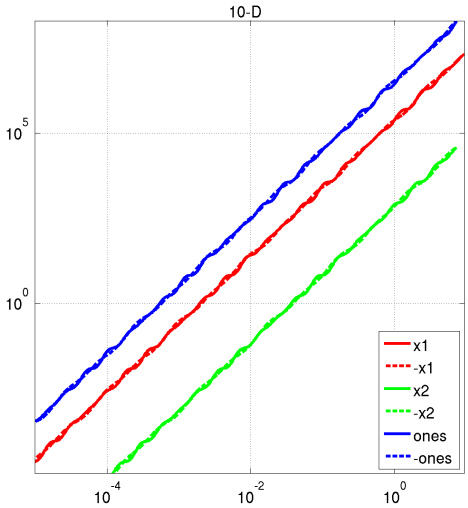

5-D
x1
-x1
x2
-x2
ones
-ones

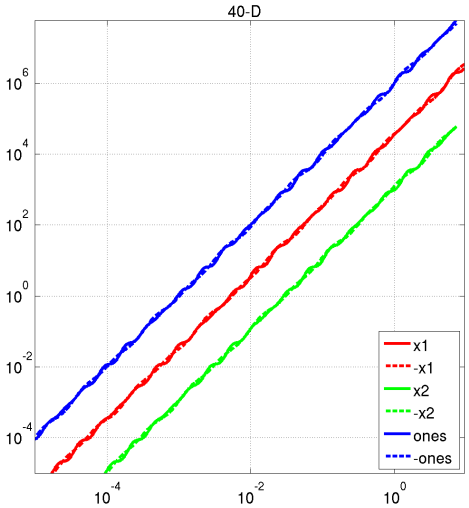

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

3.12 Bent Cigar Function

$$f_{12}(\mathbf{x}) = z_1^2 + 10^6 \sum_{i=2}^{D} z_i^2 + f_{\text{opt}} \tag{12}$$

- $\mathbf{z} = \mathbf{R}\, T_{\text{asy}}^{0.5}(\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}}))$

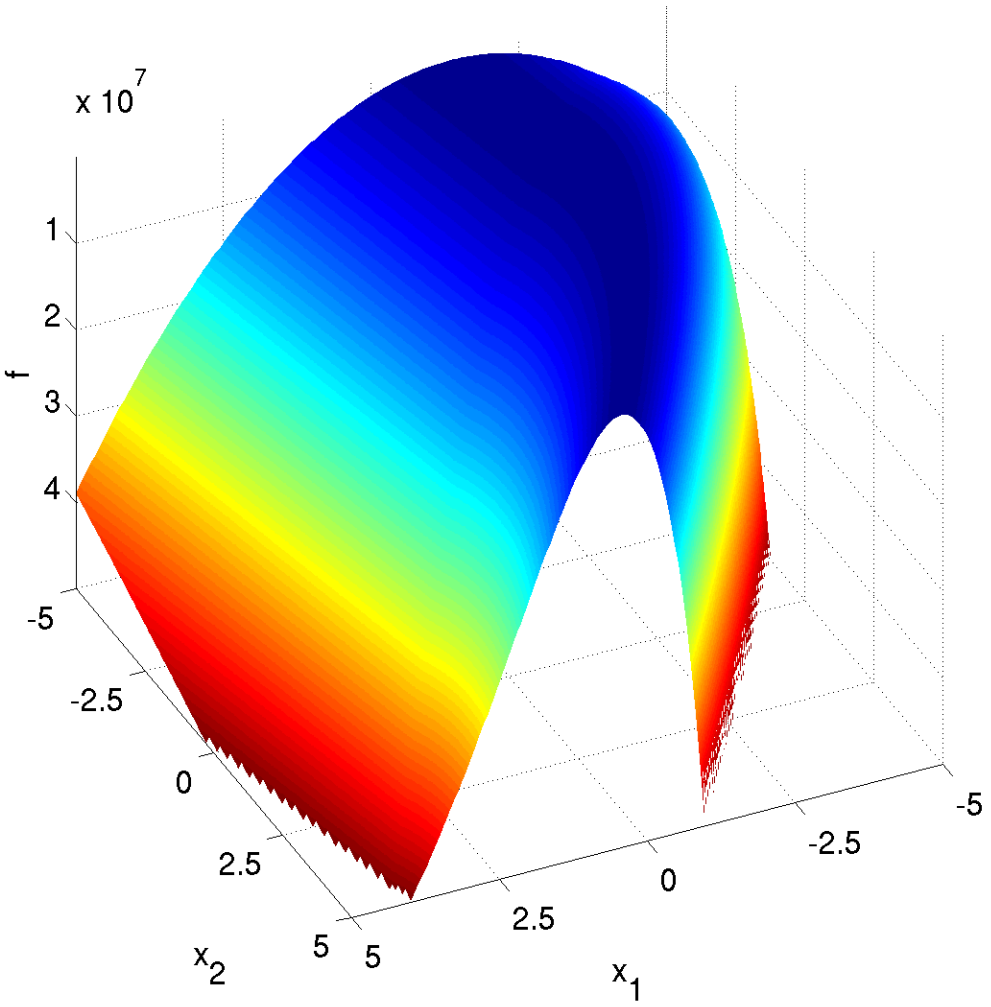
Properties A ridge defined as $\sum_{i=2}^{D} z_i^2 = 0$ needs to be followed. The ridge is smooth but very narrow. Due to $T_{\text{asy}}^{1/2}$ the overall shape deviates remarkably from being quadratic.

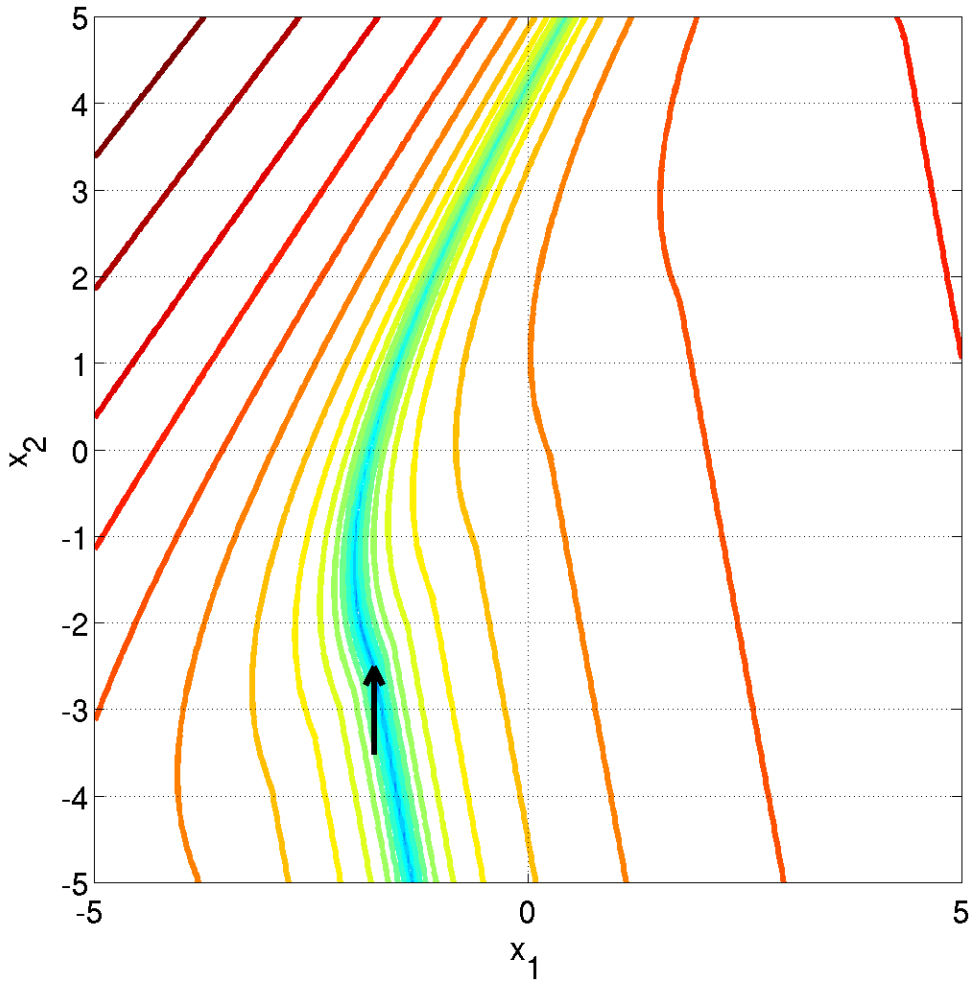
- conditioning is about 10^6, rotated, unimodal

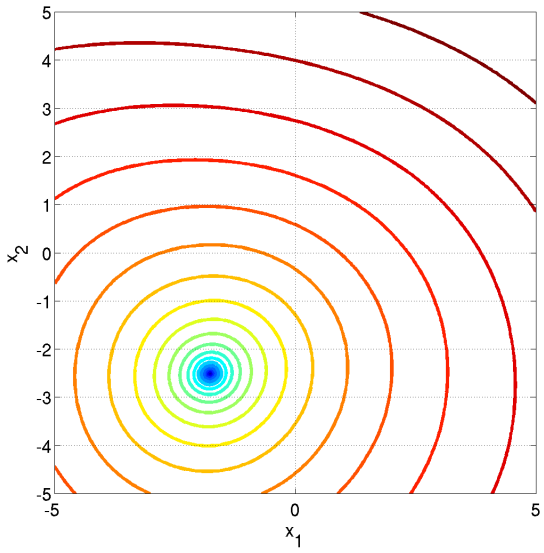
Information gained from this function:

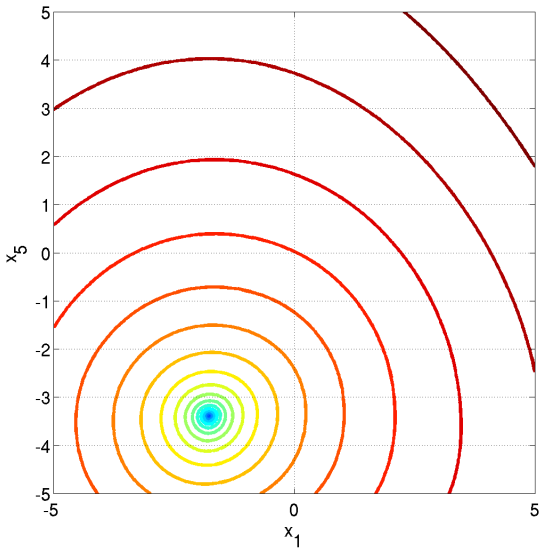
- Can the search continuously change its search direction?

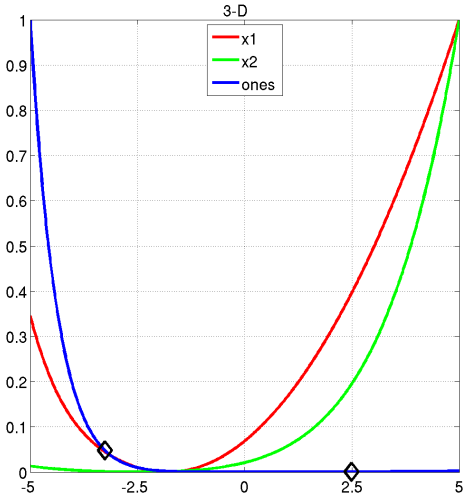
Note: The 3d plot shows only a part of the complete function in the vicinity of the optimum.

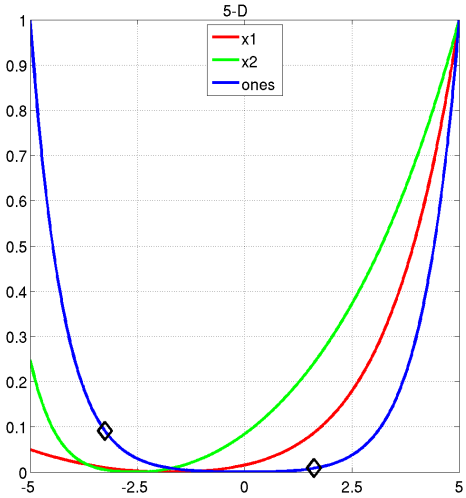


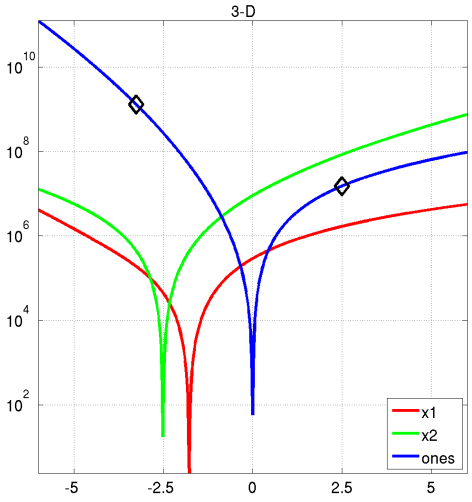

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

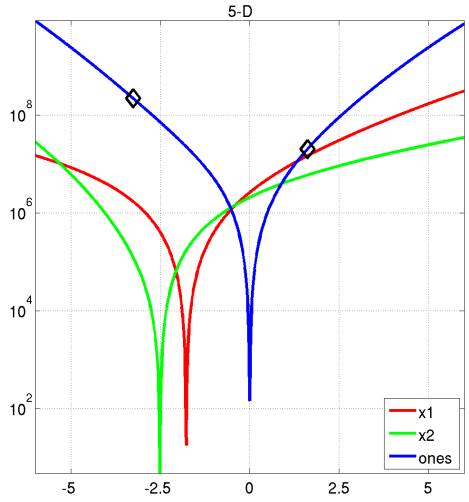

x2
x1

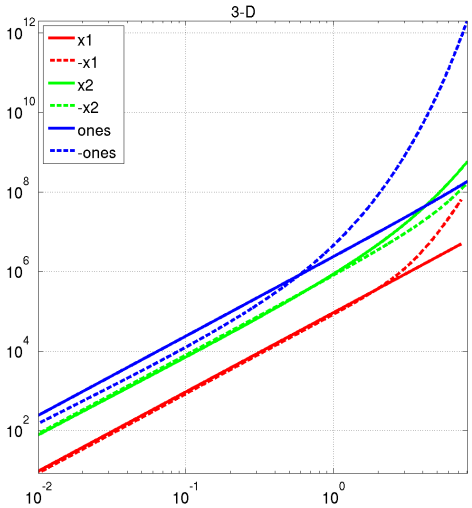

x5
x1

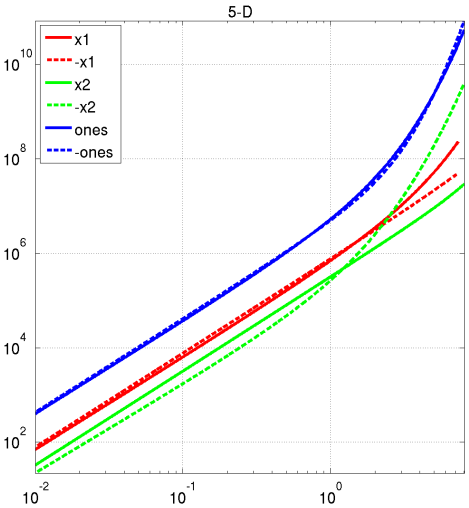

3-D
x1
x2
ones

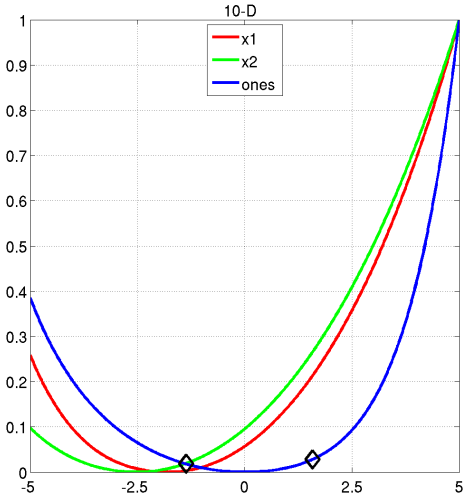

5-D
x1
x2
ones

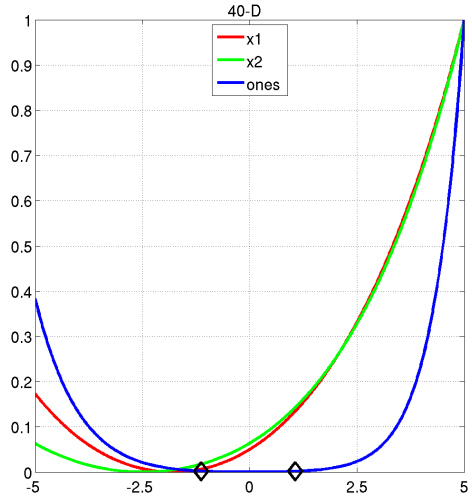

3-D
x1
x2
ones

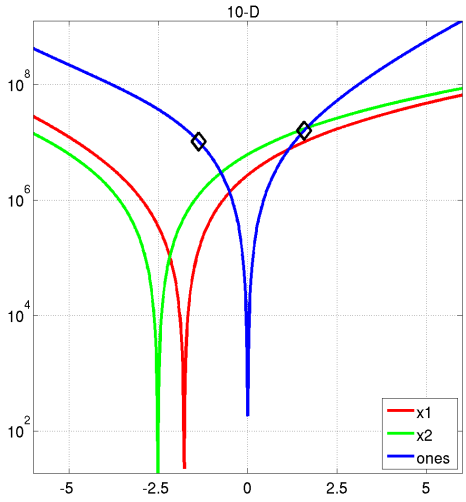

5-D
x1
x2
ones

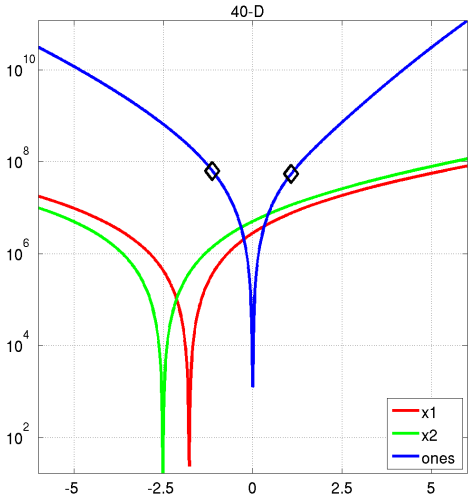

3-D
x1
-x1
x2
-x2
ones
-ones

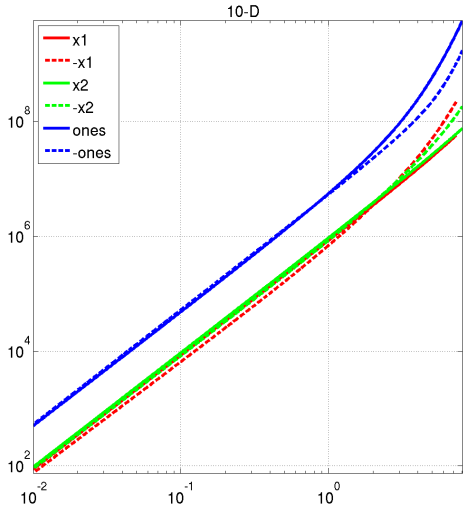

5-D
x1
-x1
x2
-x2
ones
-ones

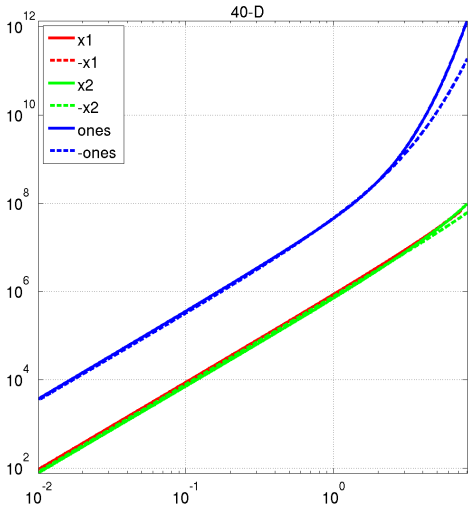

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

3.13 Sharp Ridge Function

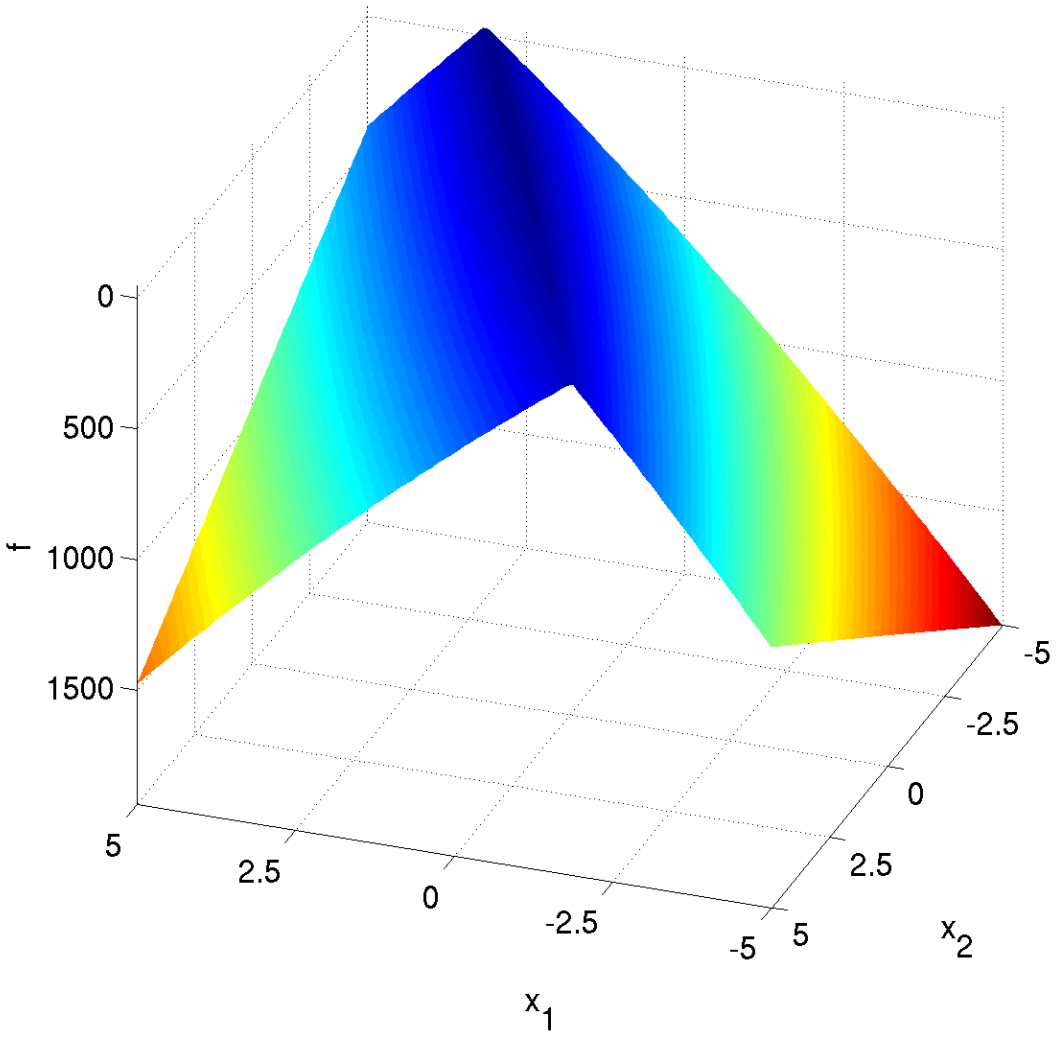
$$f_{13}(\mathbf{x}) = z_1^2 + 100\sqrt{\sum_{i=2}^{D} z_i^2} + f_{\text{opt}} \tag{13}$$

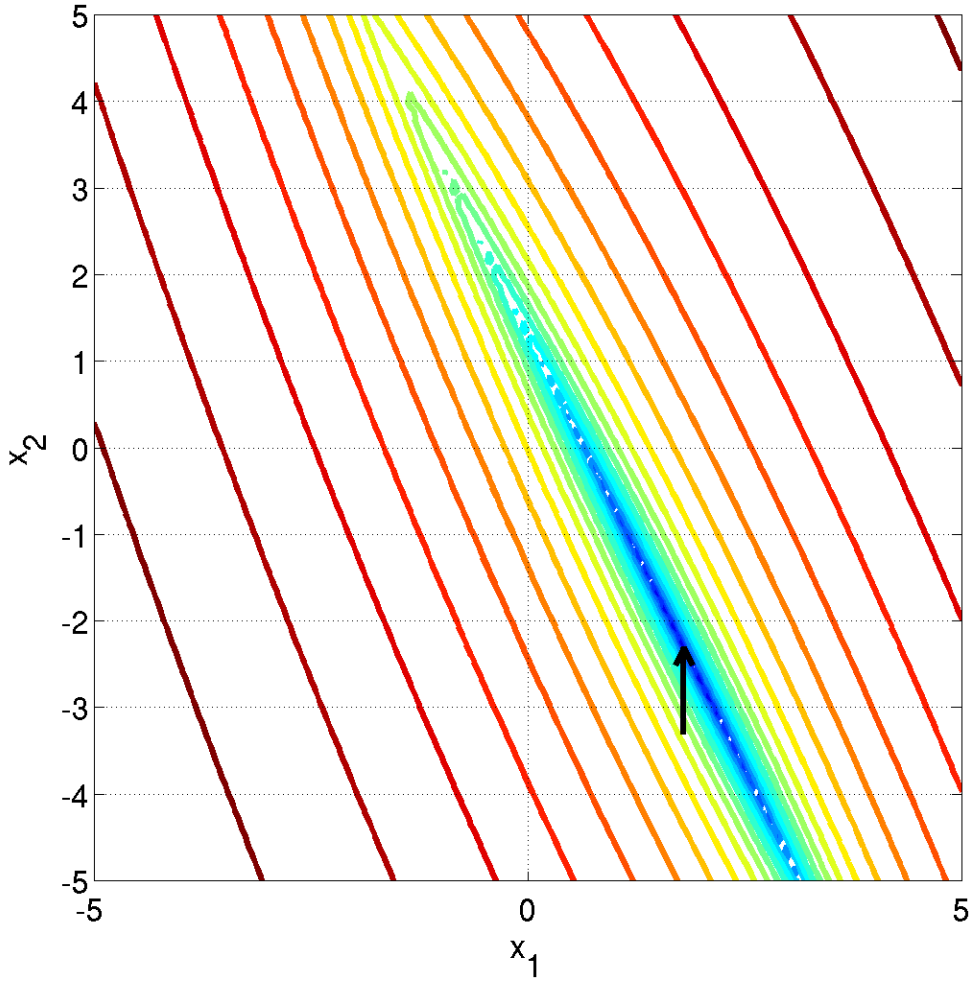
- $\mathbf{z} = \mathbf{Q}\Lambda^{10}\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}})$

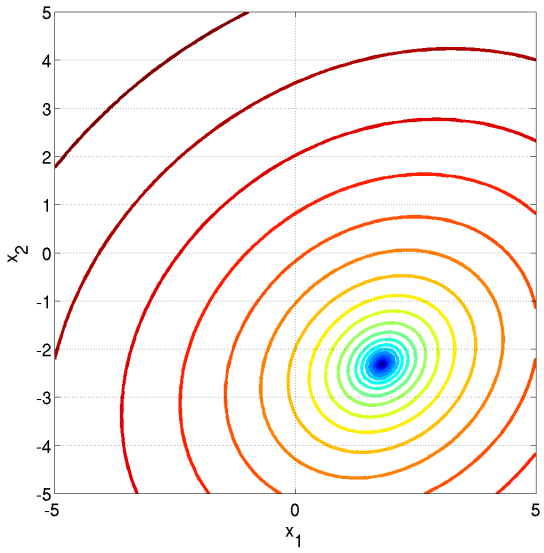
Properties As for the Bent Cigar function, a ridge defined as $\sum_{i=2}^{D} z_i^2 = 0$ must be followed. The ridge is non-differentiable and the gradient is constant when the ridge is approached from any given point. Following the gradient becomes ineffective close to the ridge where the ridge needs to be followed in z_1-direction to its optimum. The necessary change in "search behavior" close to the ridge is difficult to diagnose, because the gradient towards the ridge does not flatten out.

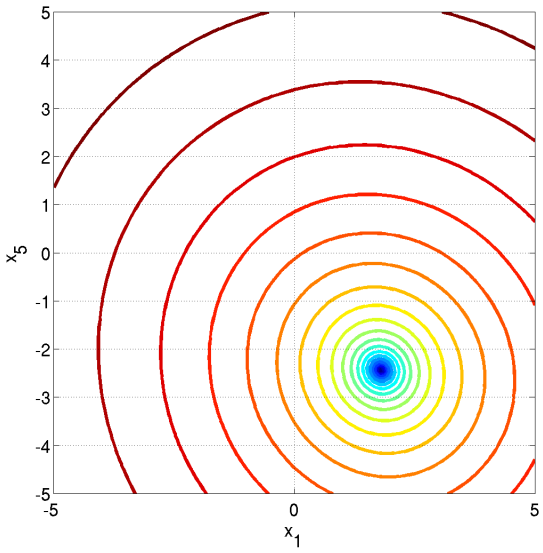
Information gained from this function:

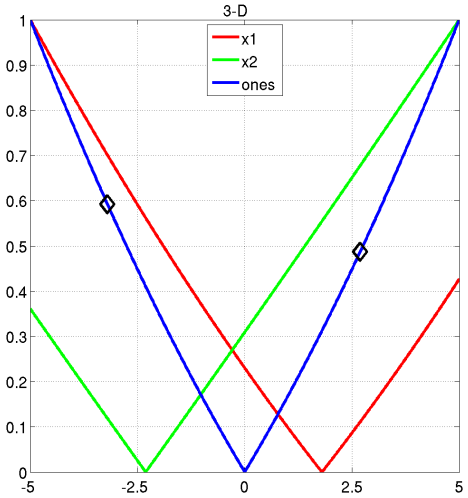
- In comparison to f12: What is the effect of non-smoothness, non-differentiabale ridge?

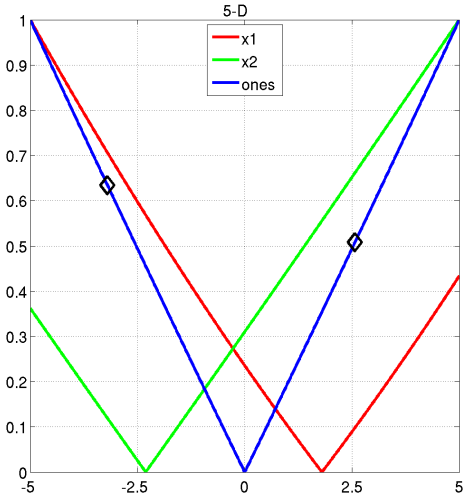


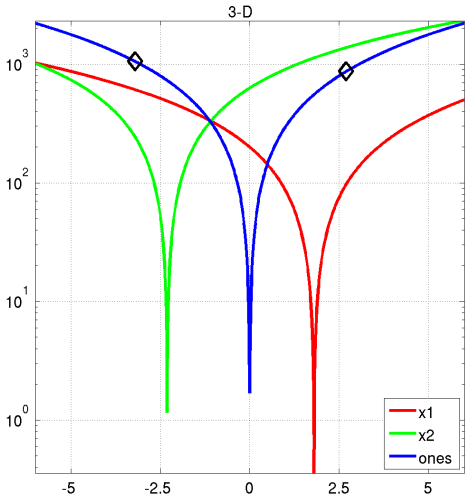

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

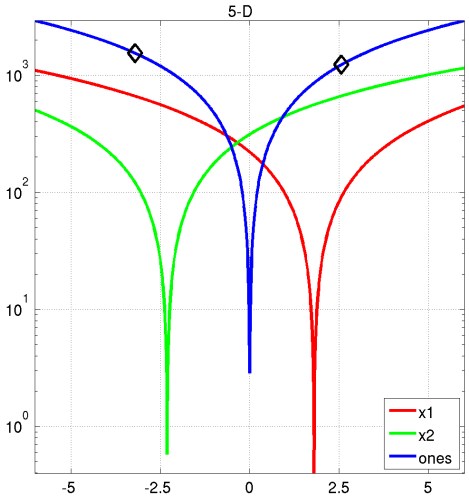

x2
x1

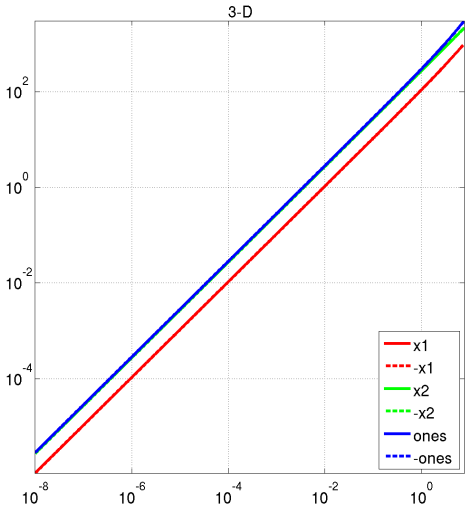

x5
x1

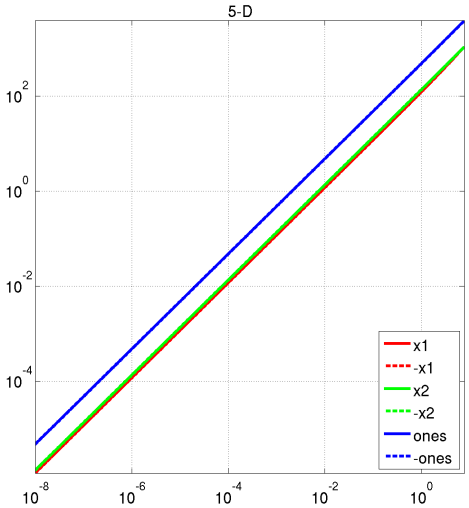

3-D
x1
x2
ones

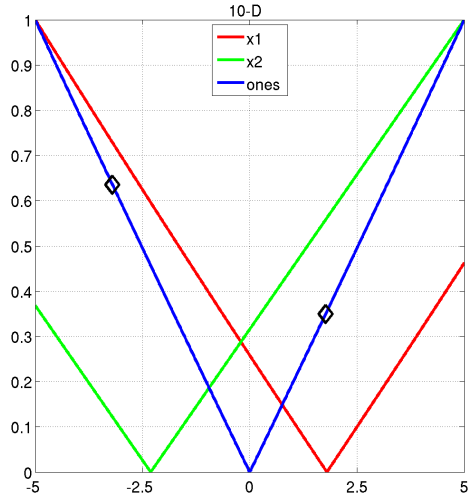

5-D
x1
x2
ones

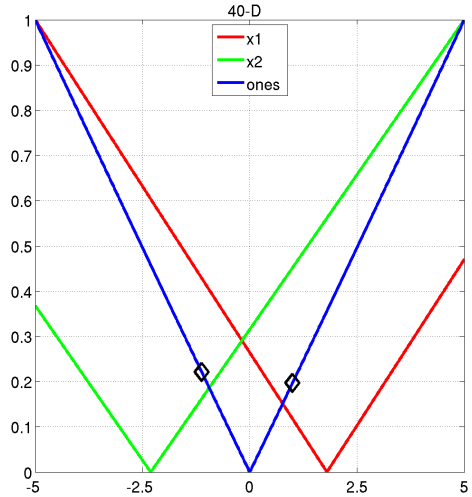

3-D
x1
x2
ones

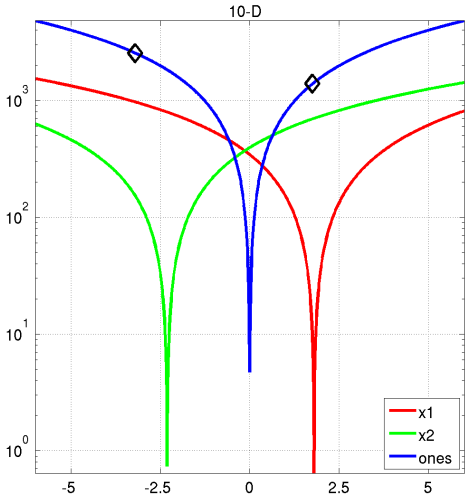

5-D
x1
x2
ones

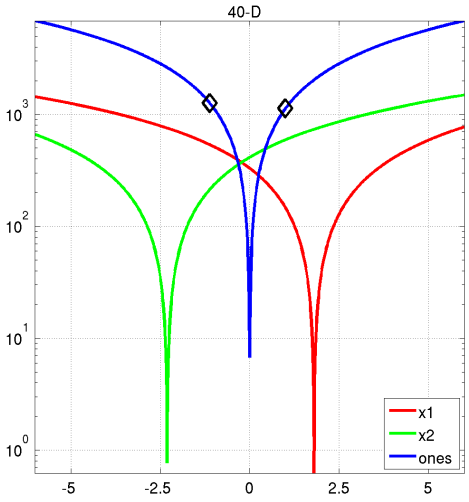

3-D
x1
-x1
x2
-x2
ones
-ones

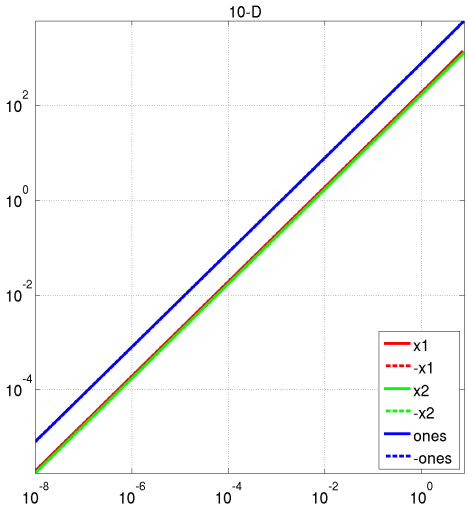

5-D
x1
-x1
x2
-x2
ones
-ones

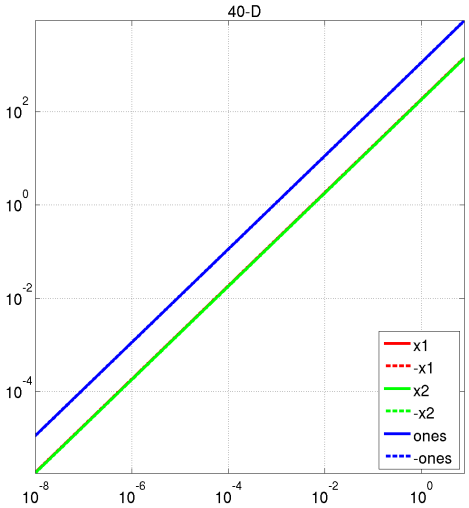

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones

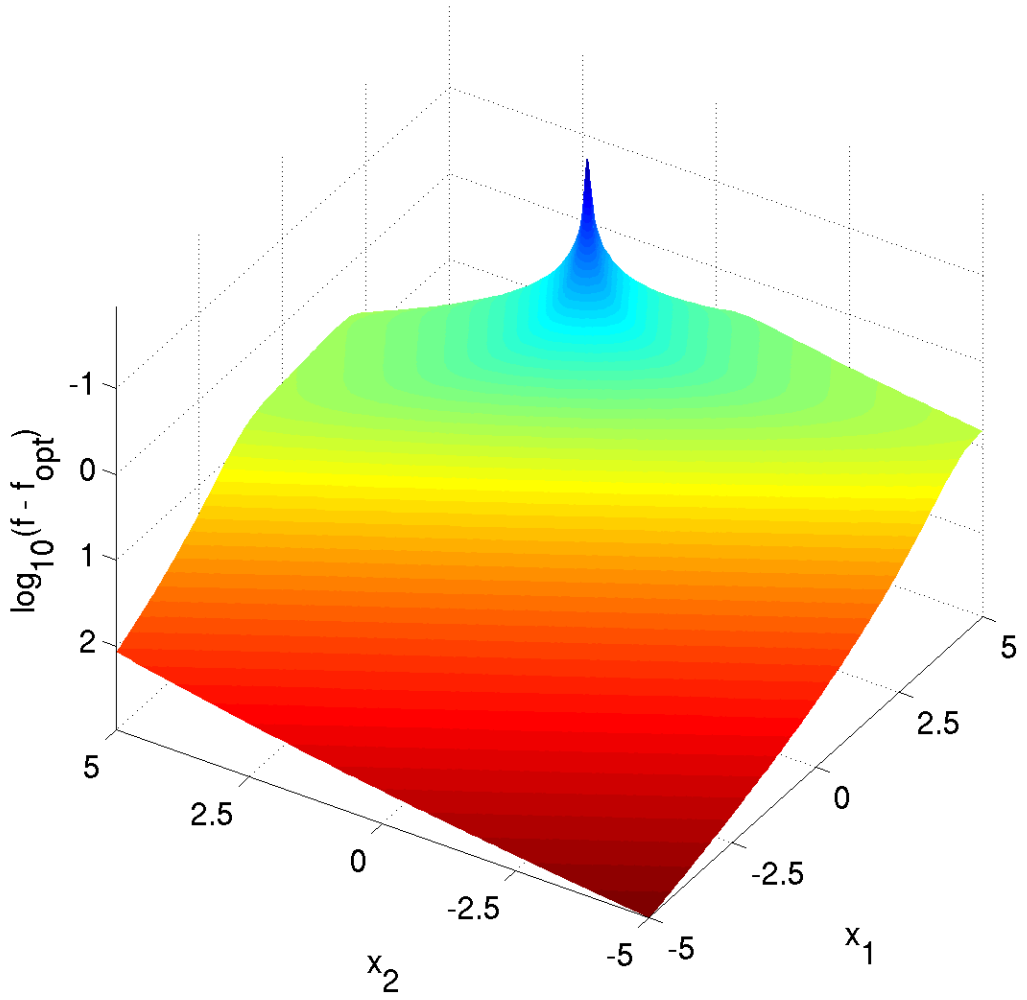

40-D
x1
-x1
x2
-x2
ones
-ones

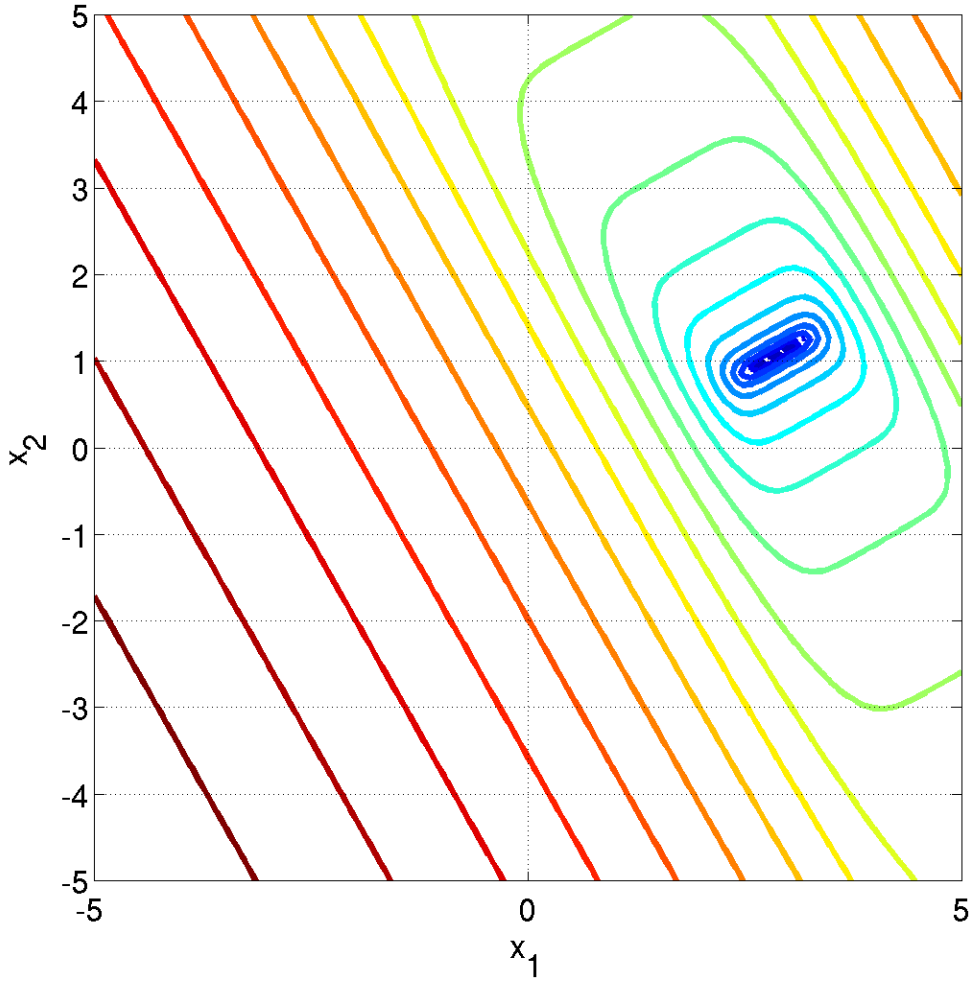
3.14 Different Powers Function

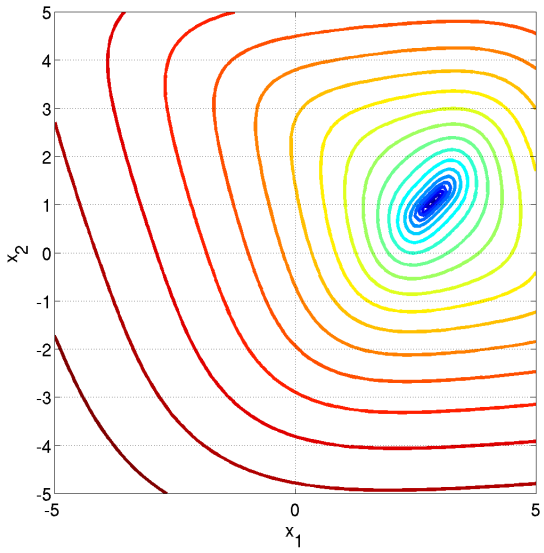
$$f_{14}(\mathbf{x}) = \sqrt{\sum_{i=1}^{D} |z_i|^{2+4\frac{i-1}{D-1}}} + f_{\text{opt}} \tag{14}$$

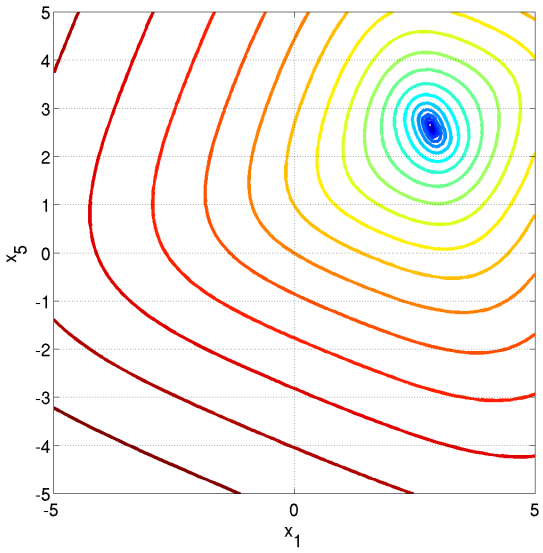
- $\mathbf{z} = \mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}})$

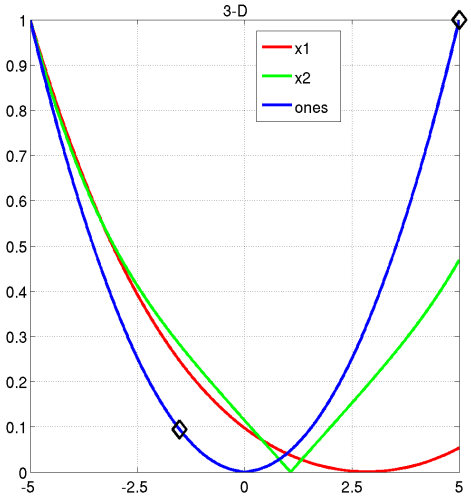
Properties Due to the different exponents the sensitivies of the z_i-variables become more and more different when approaching the optimum.

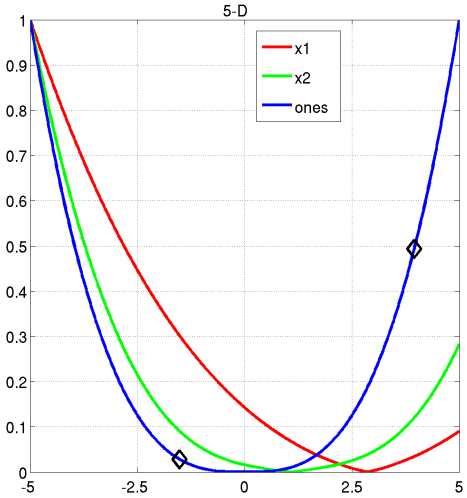


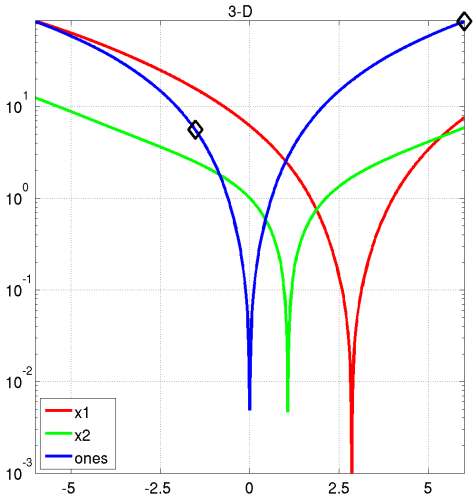

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

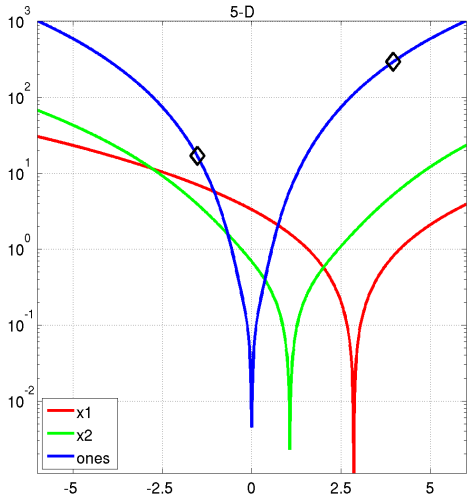

x2
x1

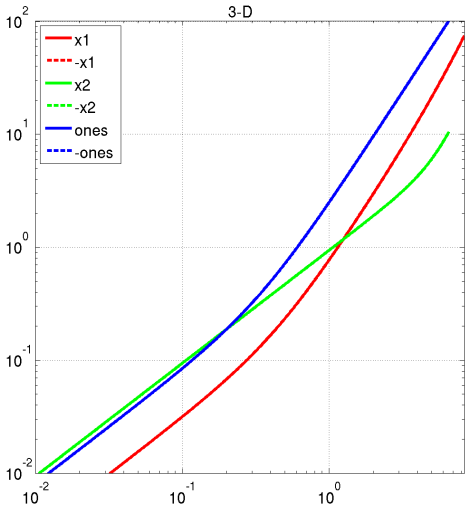

x5
x1

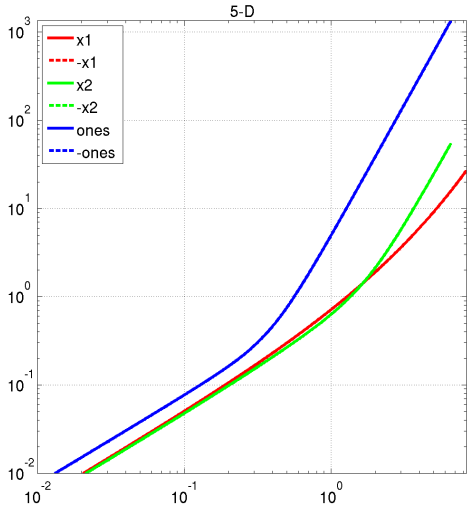

3-D
x1
x2
ones

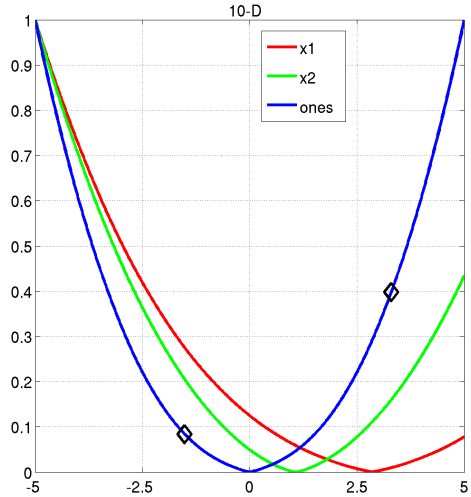

5-D
x1
x2
ones

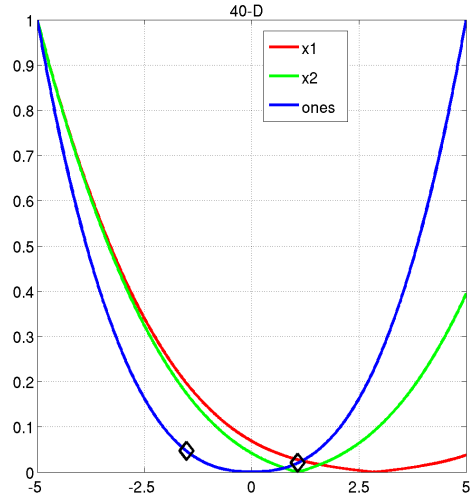

3-D
x1
x2
ones

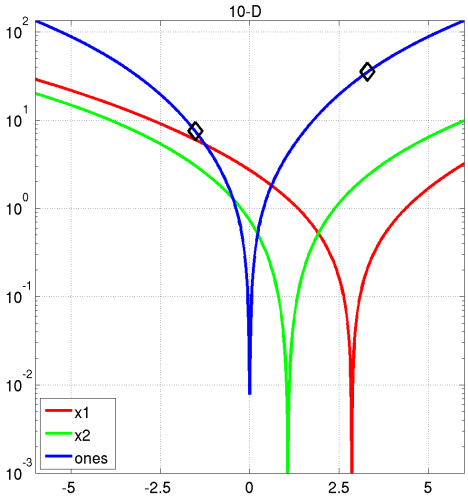

5-D
x1
x2
ones

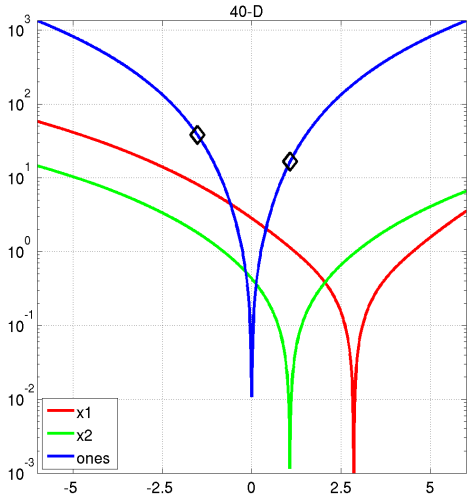

3-D
x1
-x1
x2
-x2
ones
-ones

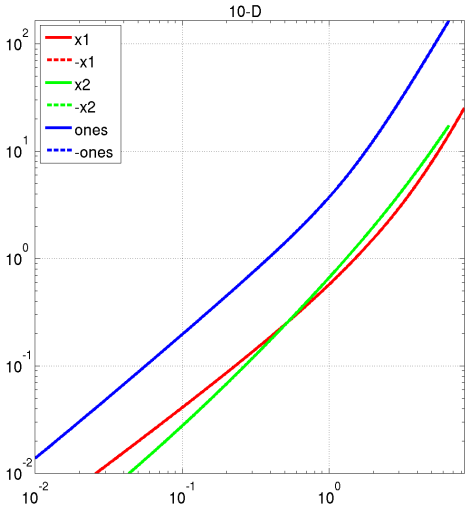

5-D
x1
-x1
x2
-x2
ones
-ones

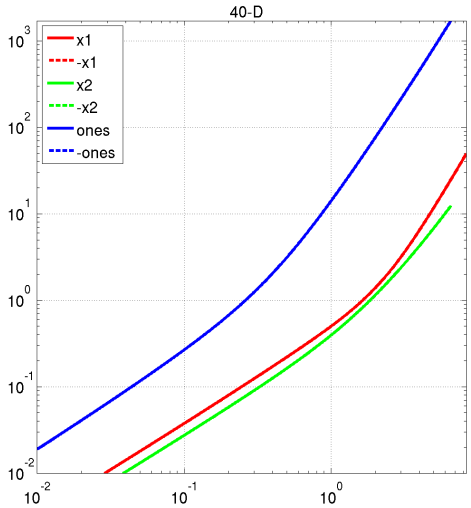

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

4 Multi-modal functions with adequate global structure

4.15 Rastrigin Function

$$f_{15}(\mathbf{x}) = 10\left(D - \sum_{i=1}^{D}\cos(2\pi z_i)\right) + \|\mathbf{z}\|^2 + f_{\text{opt}} \tag{15}$$

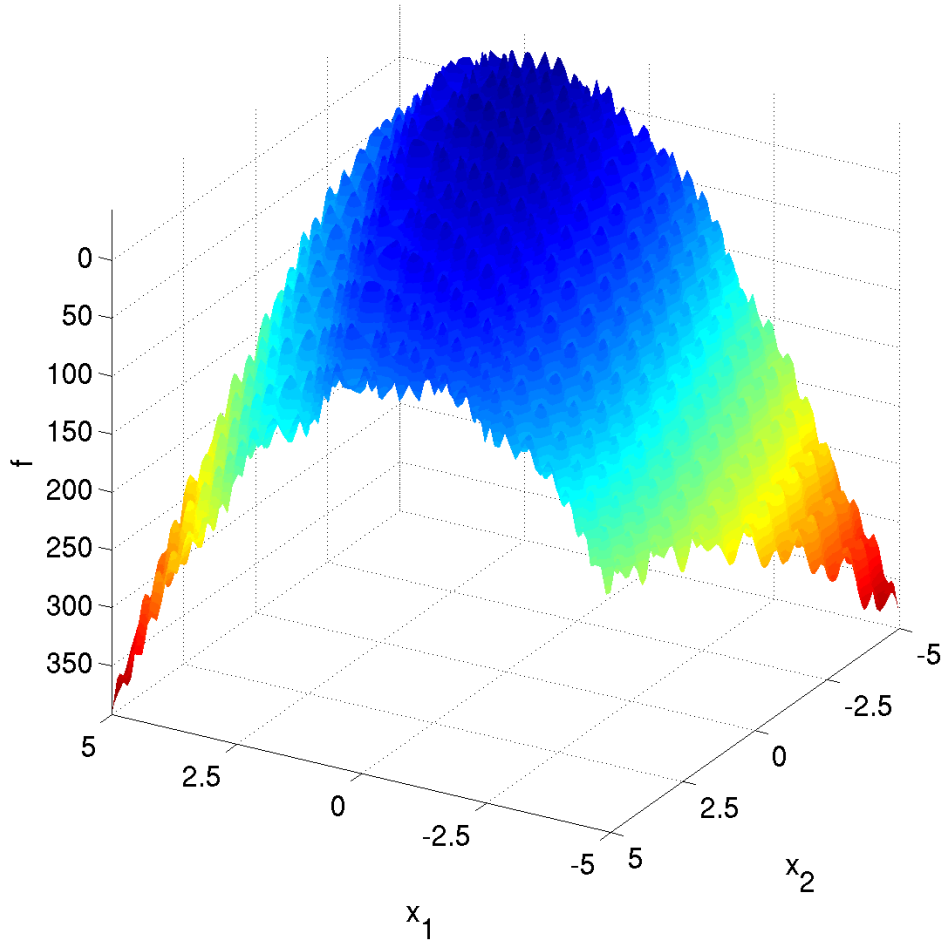
- $\mathbf{z} = \mathbf{R}\Lambda^{10}\mathbf{Q}\,T_{\text{asy}}^{0.2}(T_{\text{osz}}(\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}})))$

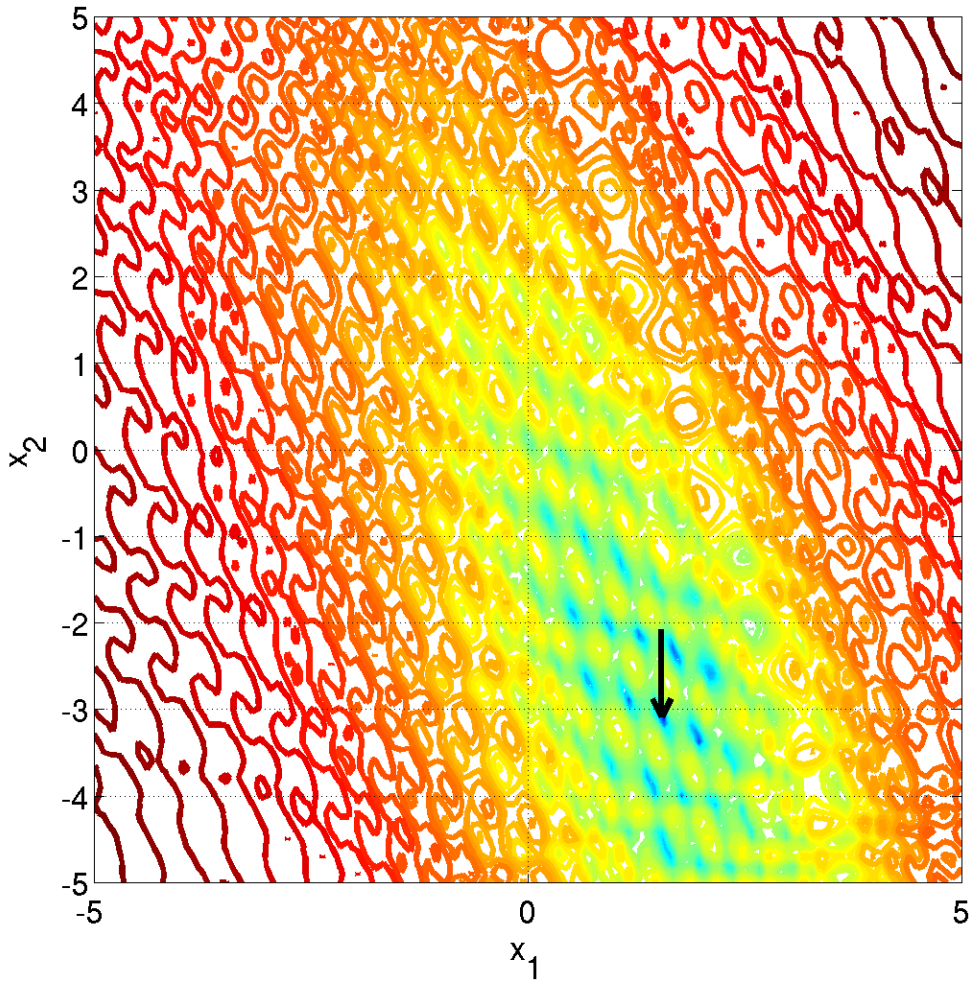
Properties Prototypical highly multimodal function which has originally a very regular and symmetric structure for the placement of the optima. The transformations T_{asy} and T_{osz} alleviate the symmetry and regularity of the original Rastrigin function.

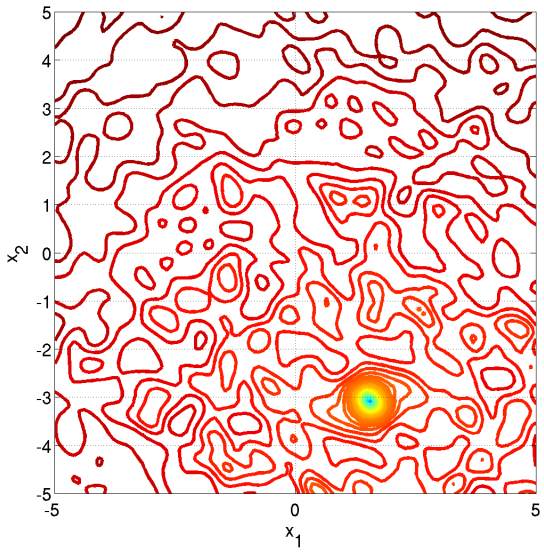
- non-separable less regular counterpart of f_3
- roughly 10^D local optima
- conditioning is about 10

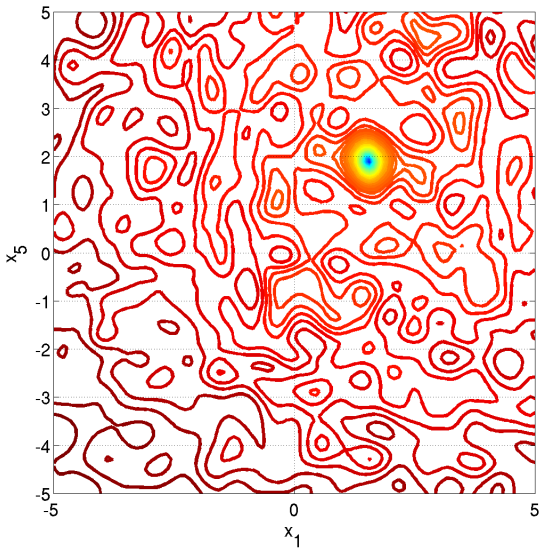
Information gained from this function:

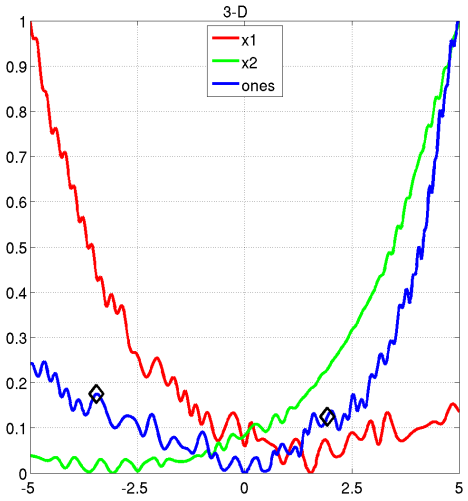
- in comparison to f3: What is the effect of non-separability for a highly multimodal function?

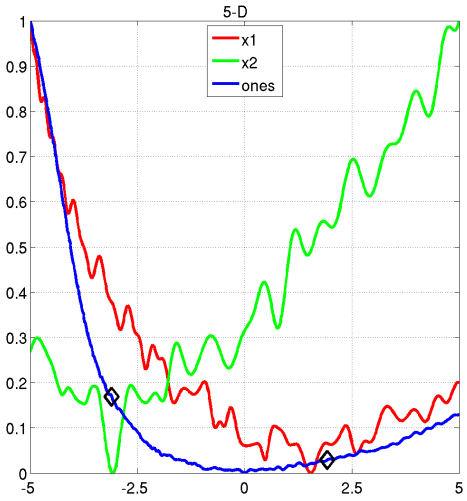


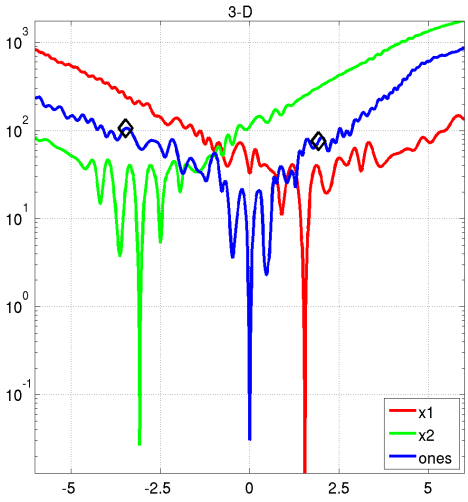

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

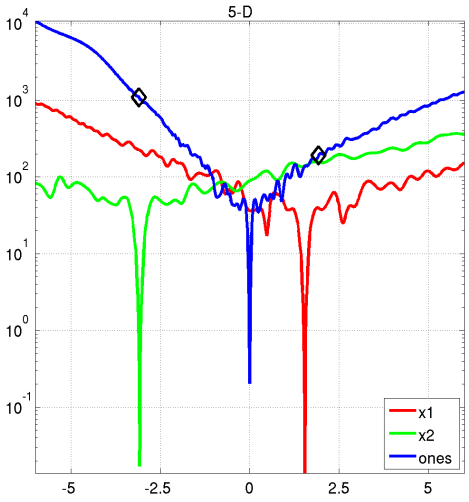

x2
x1

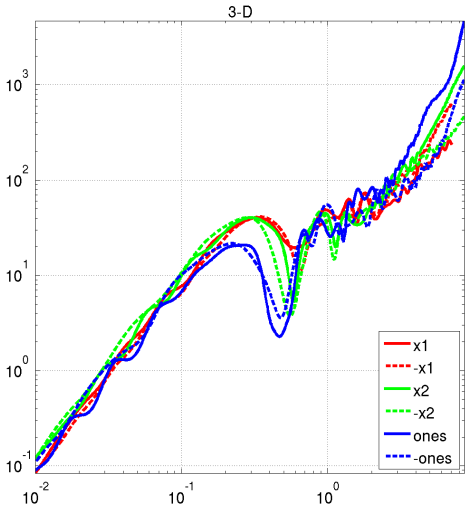

x5
x1

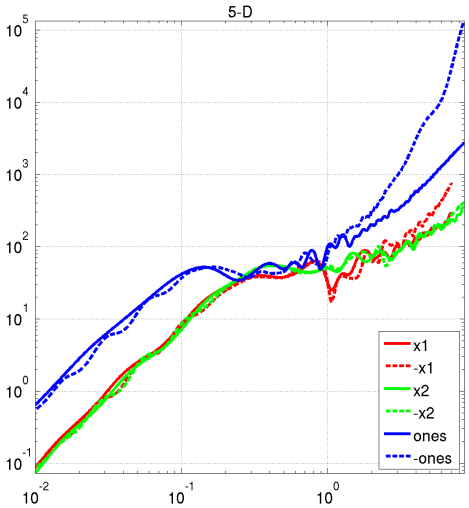

3-D
x1
x2
ones

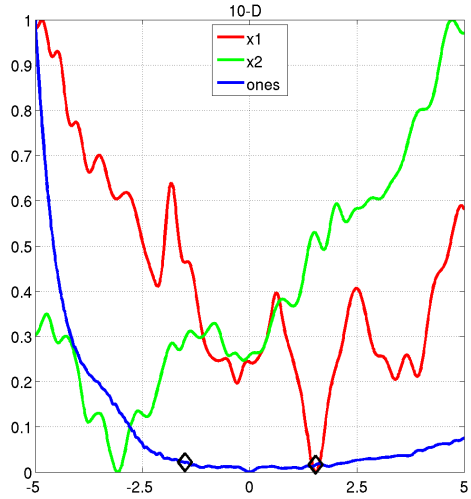

5-D
x1
x2
ones

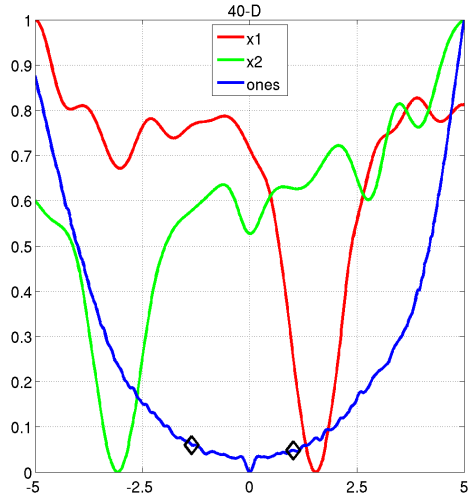

3-D
x1
x2
ones

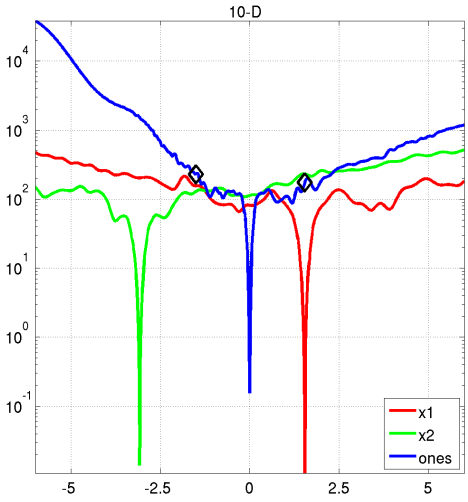

5-D
x1
x2
ones

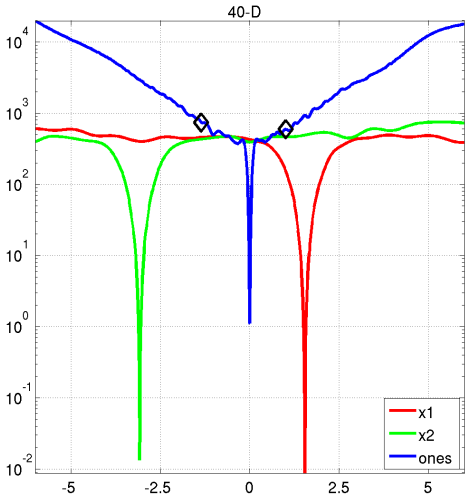

3-D
x1
-x1
x2
-x2
ones
-ones

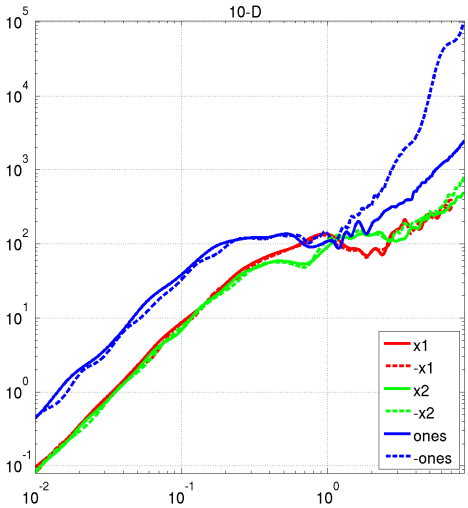

5-D
x1
-x1
x2
-x2
ones
-ones

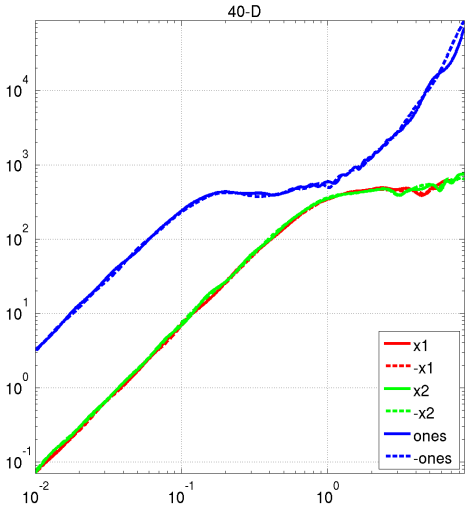

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

4.16 Weierstrass Function

$$f_{16}(\mathbf{x}) = 10\left(\frac{1}{D}\sum_{i=1}^{D}\sum_{k=0}^{11} 1/2^k \cos(2\pi 3^k(z_i + 1/2)) - f_0\right)^3 + \frac{10}{D} f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{16}$$

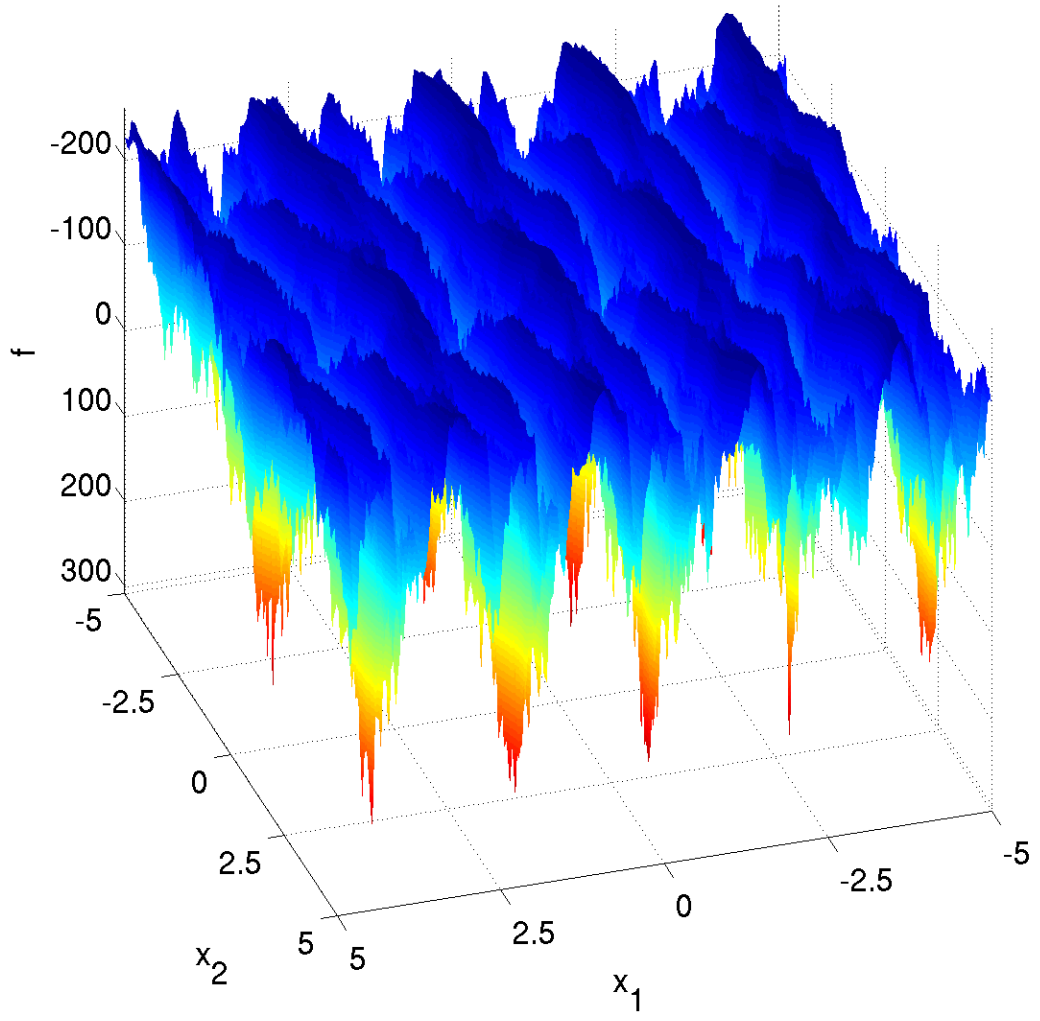
- $\mathbf{z} = \mathbf{R}\Lambda^{1/100}\mathbf{Q}\, T_{\text{osz}}(\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}}))$
- $f_0 = \sum_{k=0}^{11} 1/2^k \cos(2\pi 3^k 1/2)$

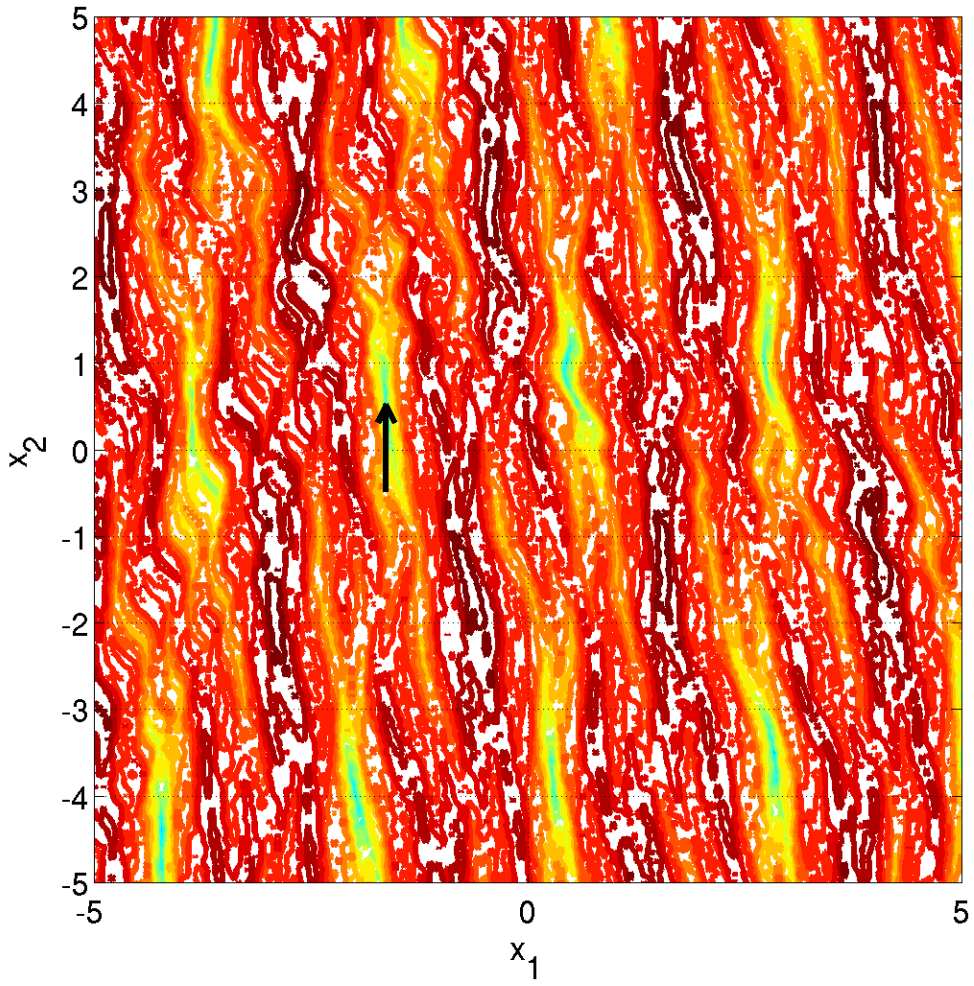
Properties Highly rugged and moderately repetitive landscape, where the global optimum is not unique.

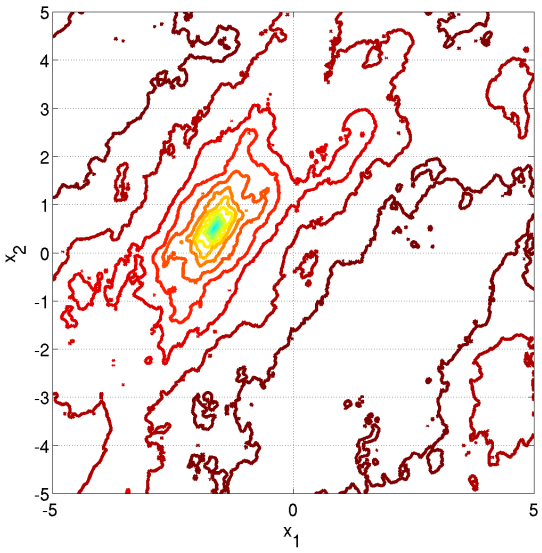
- the term $\sum_k 1/2^k \cos(2\pi 3^k \dots)$ introduces the ruggedness, where lower frequencies have a larger weight $1/2^k$.
- rotated, locally irregular, non-unique global optimum

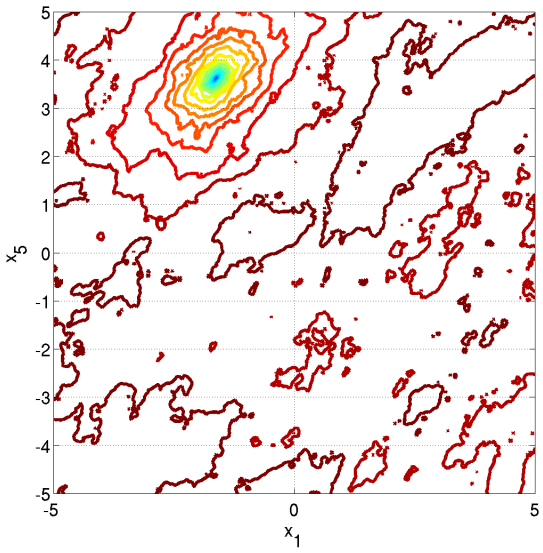
Information gained from this function:

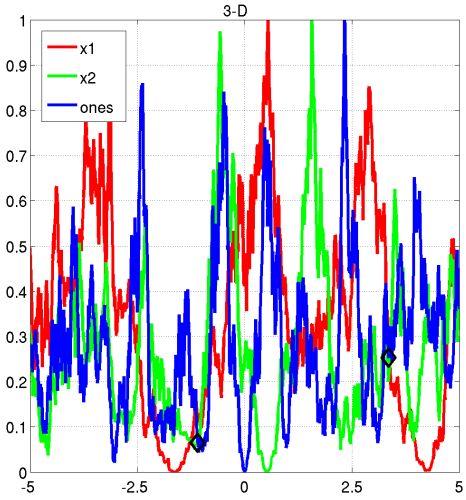
- in comparison to f17: Does ruggedness or a repetitive landscape deter the search behavior?

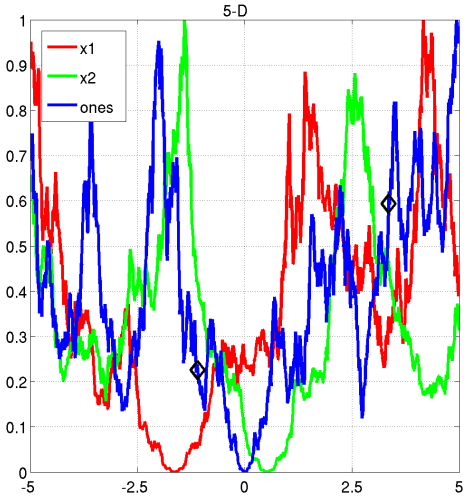


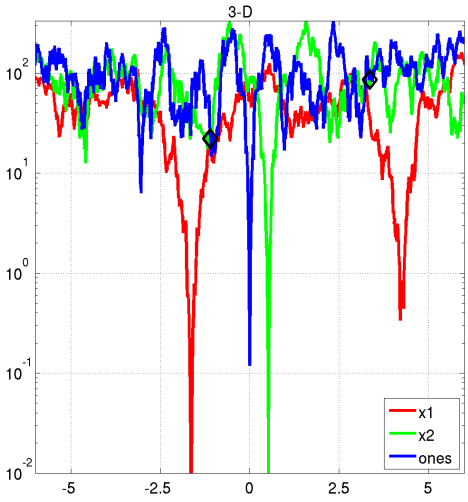

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

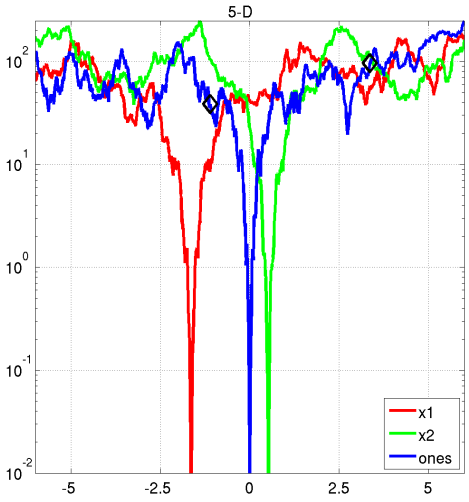

x2
x1

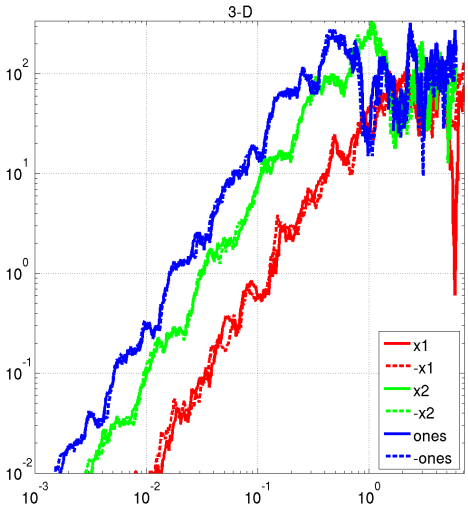

x5
x1

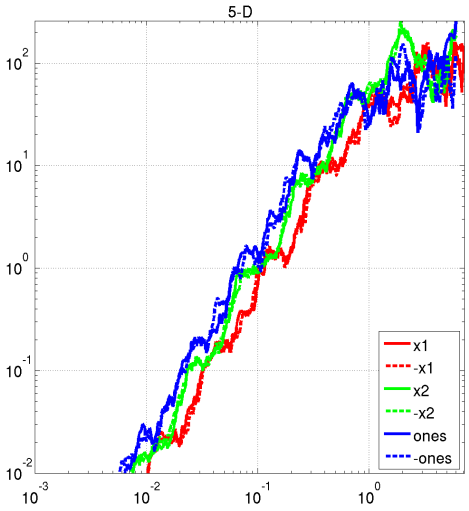

3-D
x1
x2
ones

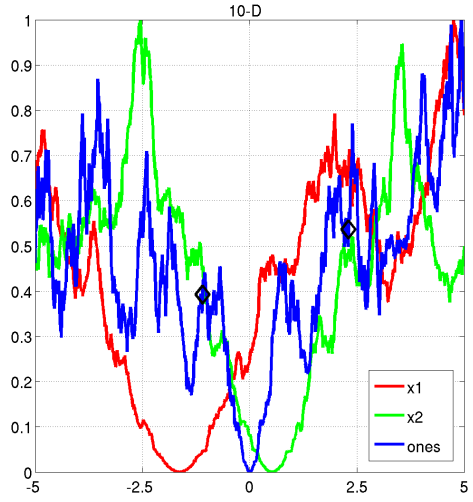

5-D
x1
x2
ones

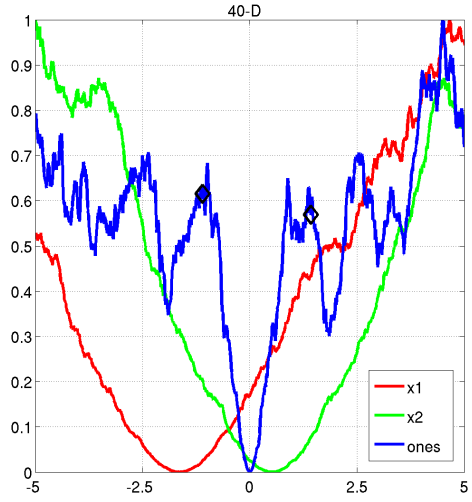

3-D
x1
x2
ones

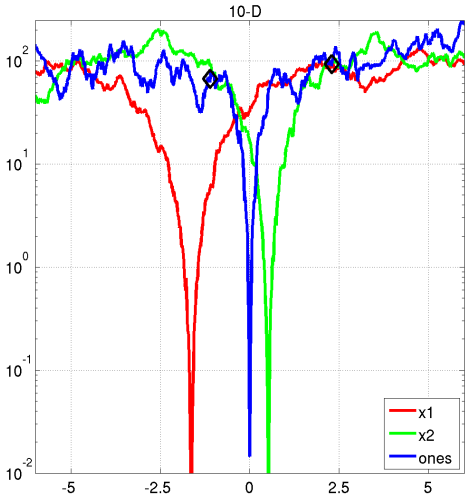

5-D
x1
x2
ones

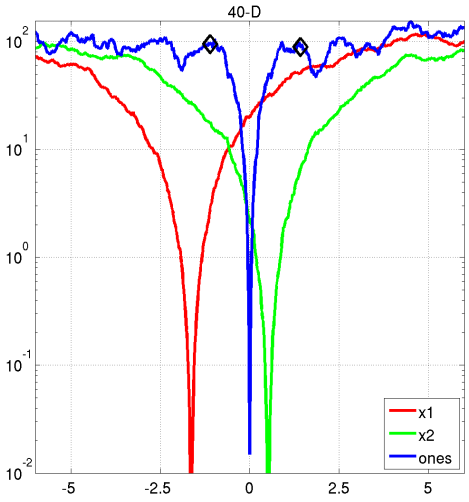

3-D
x1
-x1
x2
-x2
ones
-ones

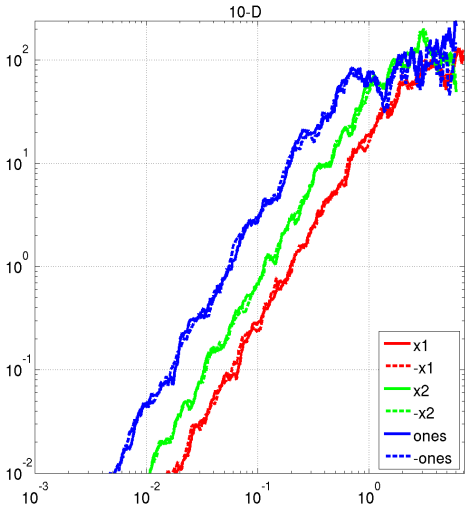

5-D
x1
-x1
x2
-x2
ones
-ones

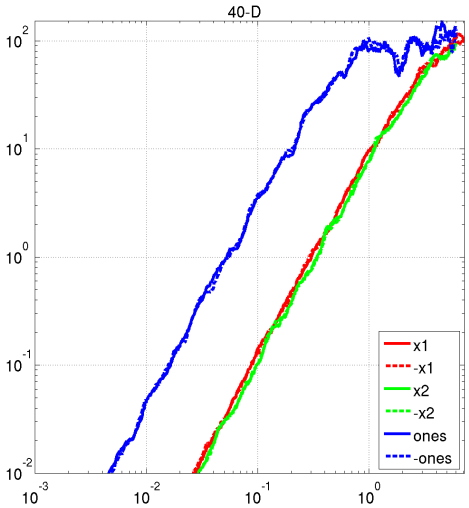

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

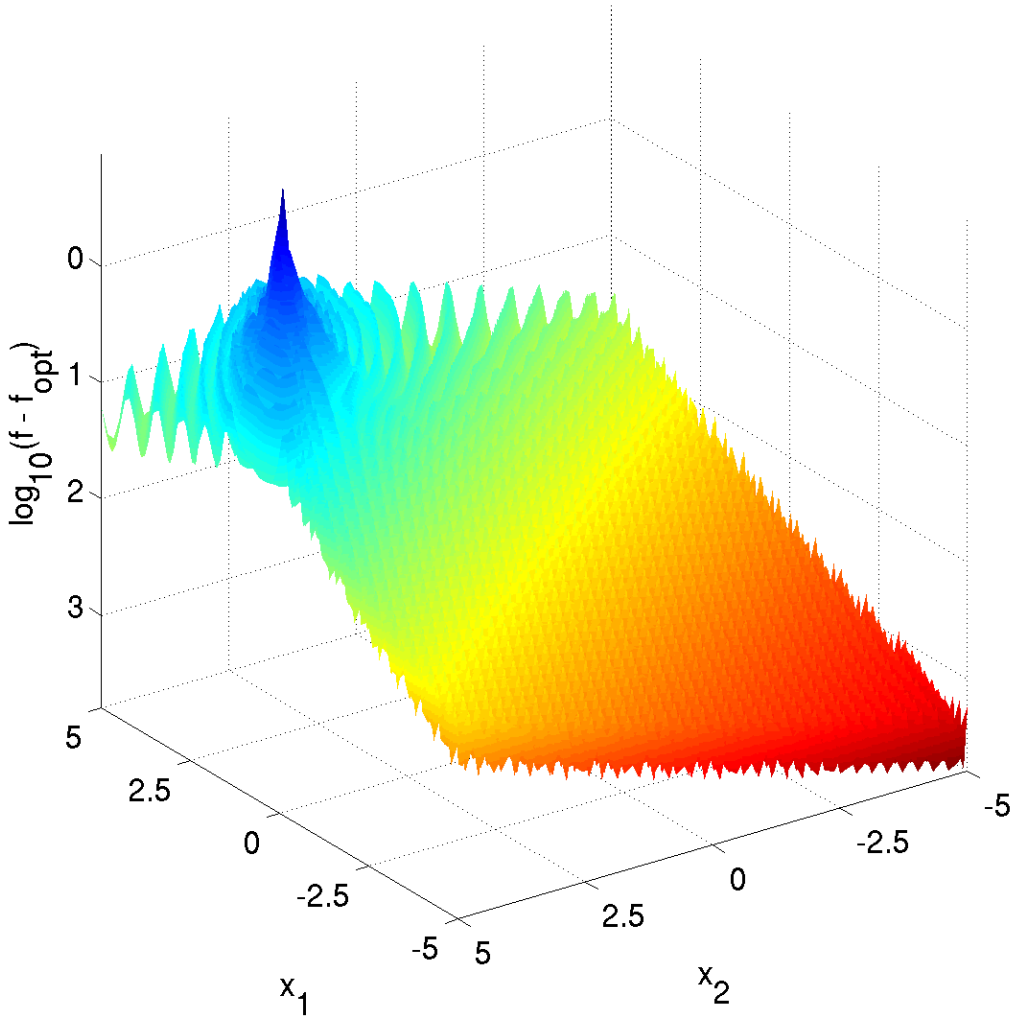
4.17 Schaffers F7 Function

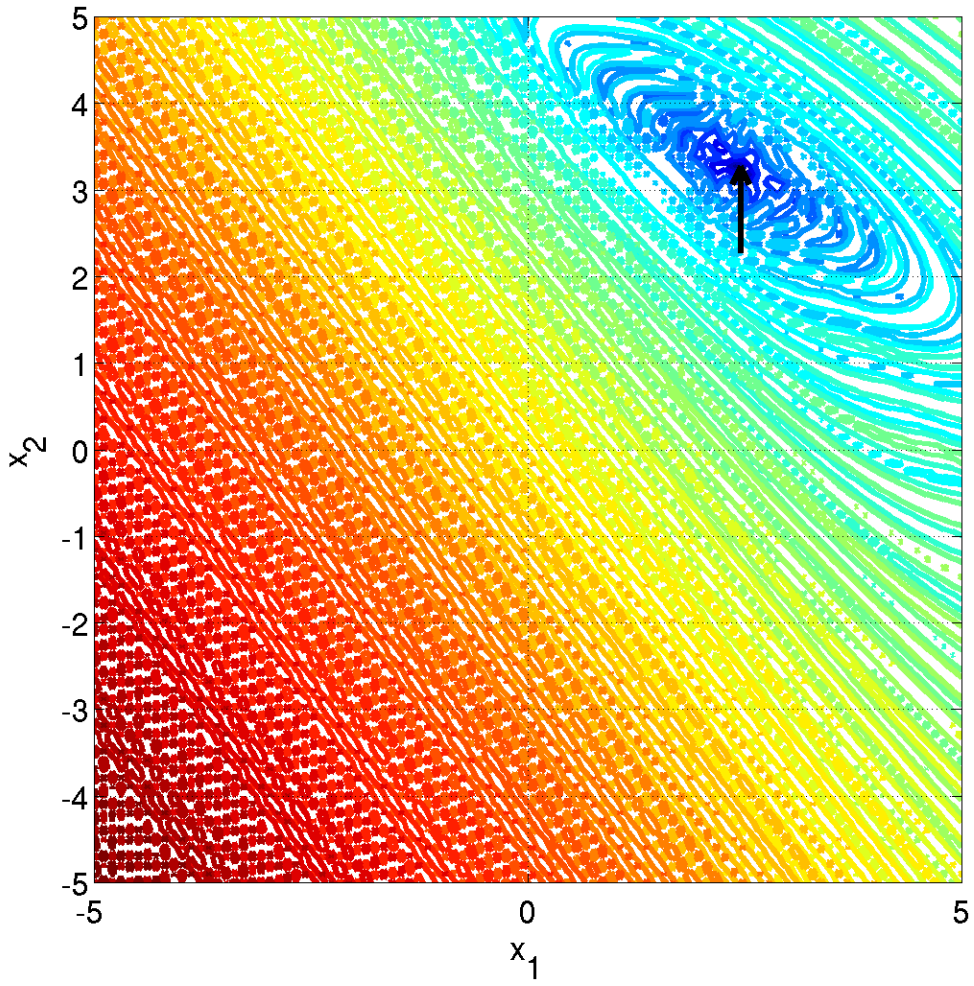
$$f_{17}(\mathbf{x}) = \left(\frac{1}{D-1} \sum_{i=1}^{D-1} \sqrt{s_i} + \sqrt{s_i} \sin^2\left(50\, s_i^{1/5}\right) \right)^2 + 10\, f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{17}$$

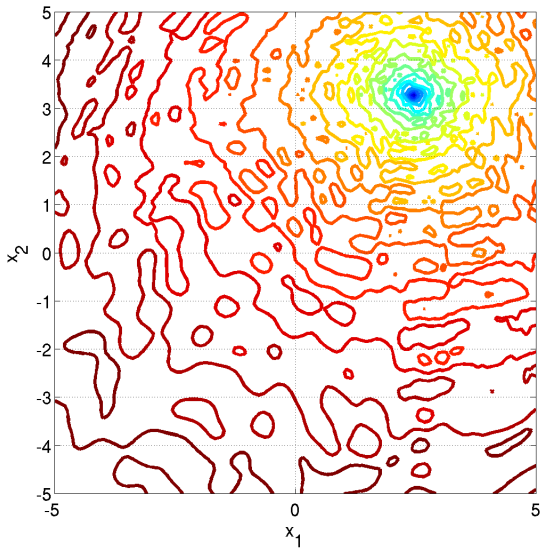
- $\mathbf{z} = \Lambda^{10} \mathbf{Q}\, T_{\text{asy}}^{0.5}(\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}}))$
- $s_i = \sqrt{z_i^2 + z_{i+1}^2}$ for $i = 1, \ldots, D$

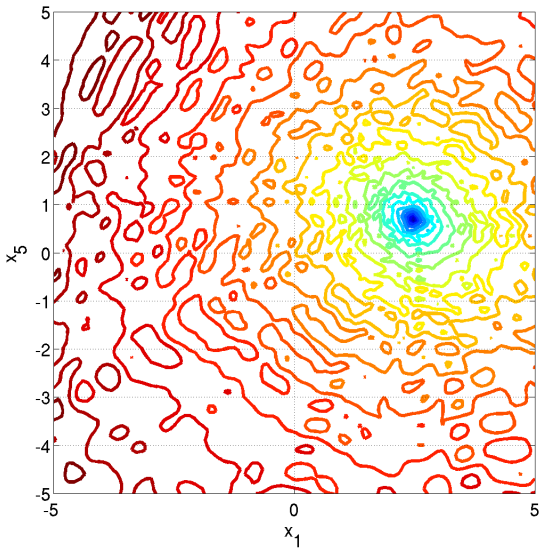
Properties A highly multimodal function where frequency and amplitude of the modulation vary.

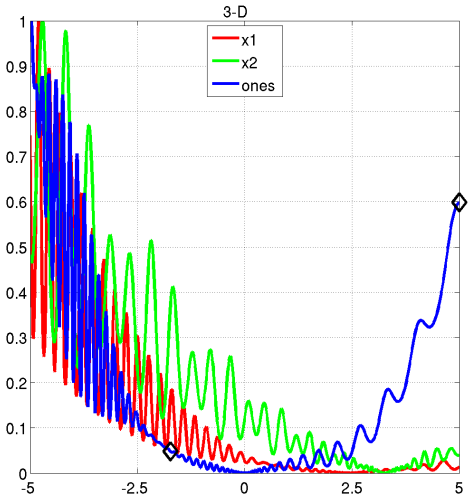
- asymmetric, rotated
- conditioning is low

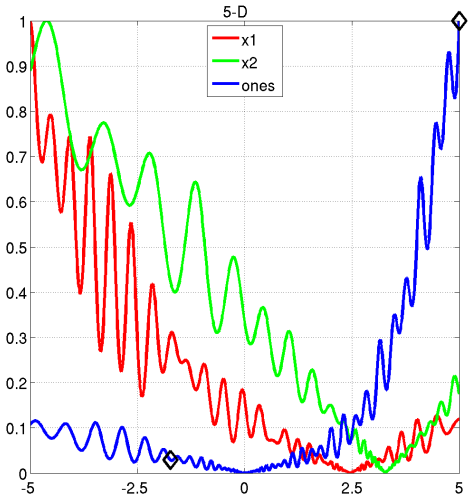


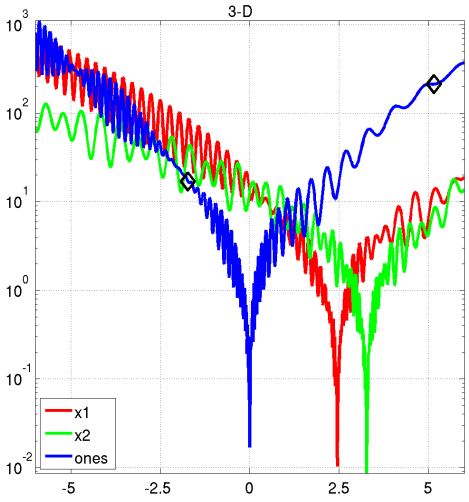

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

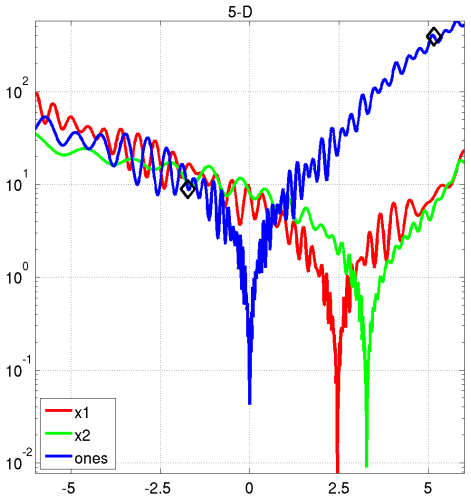

x2
x1

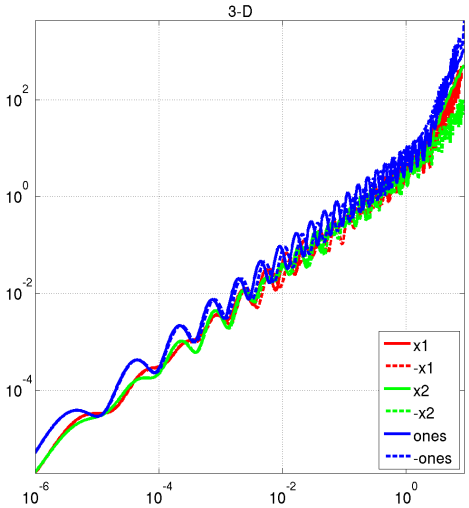

x5
x1

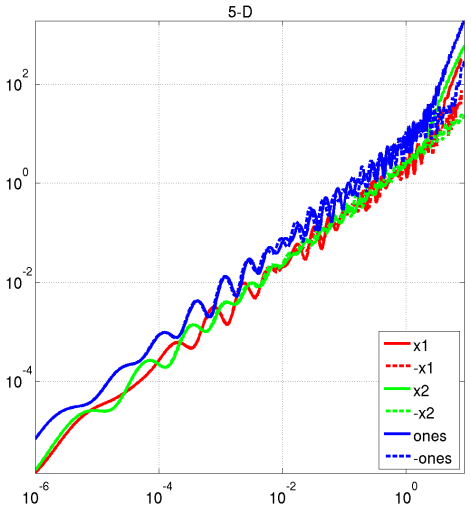

3-D
x1
x2
ones

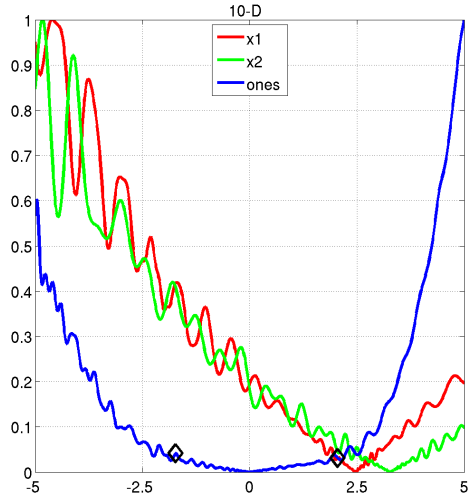

5-D
x1
x2
ones

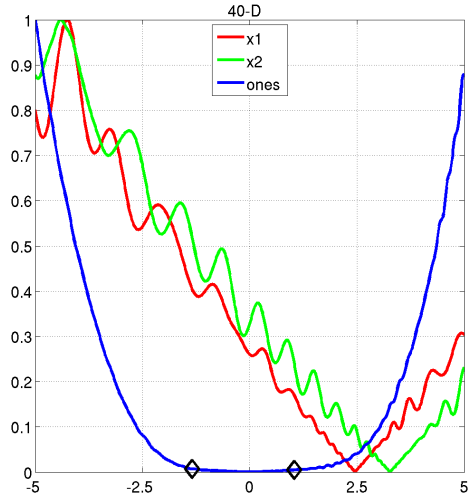

3-D
x1
x2
ones

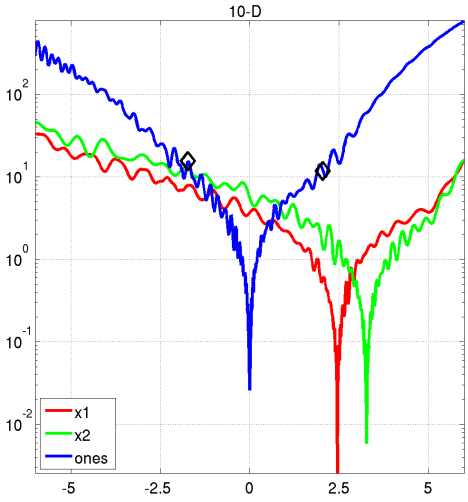

5-D
x1
x2
ones

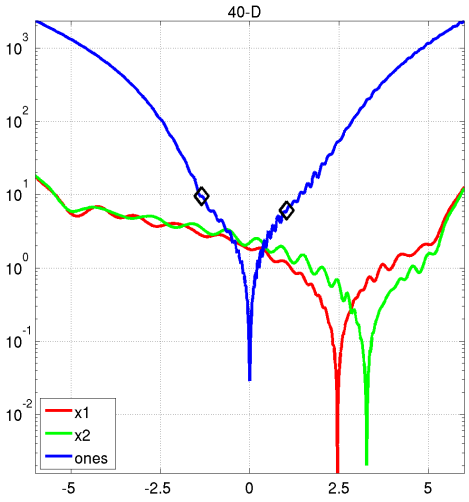

3-D
x1
-x1
x2
-x2
ones
-ones

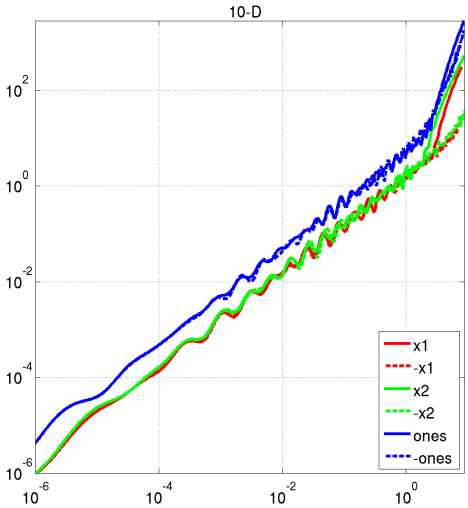

5-D
x1
-x1
x2
-x2
ones
-ones

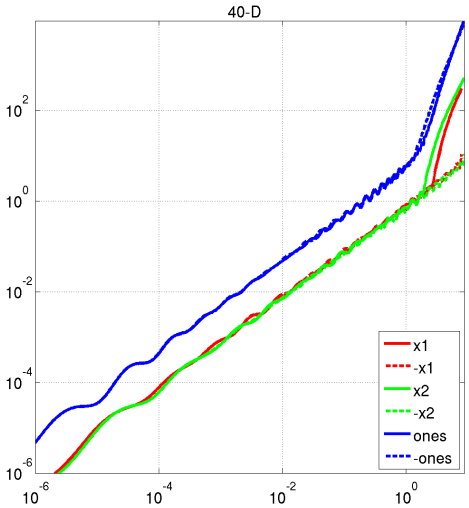

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

4.18 Schaffers F7 Function, moderately ill-conditioned

$$f_{18}(\mathbf{x}) = \left(\frac{1}{D-1} \sum_{i=1}^{D-1} \sqrt{s_i} + \sqrt{s_i} \sin^2\left(50\, s_i^{1/5}\right) \right)^2 + 10\, f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{18}$$

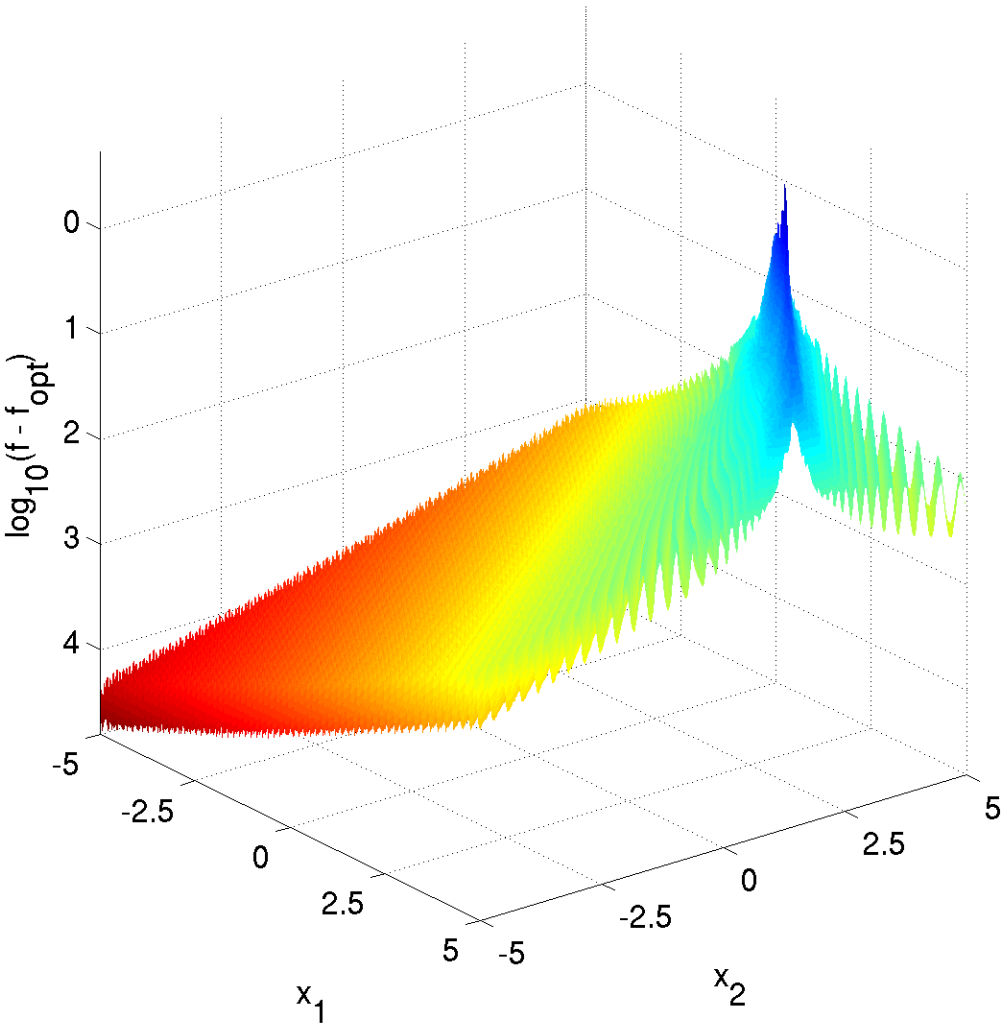
- $\mathbf{z} = \Lambda^{1000} \mathbf{Q}\, T_{\text{asy}}^{0.5}(\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}}))$
- $s_i = \sqrt{z_i^2 + z_{i+1}^2}$ for $i = 1, \ldots, D$

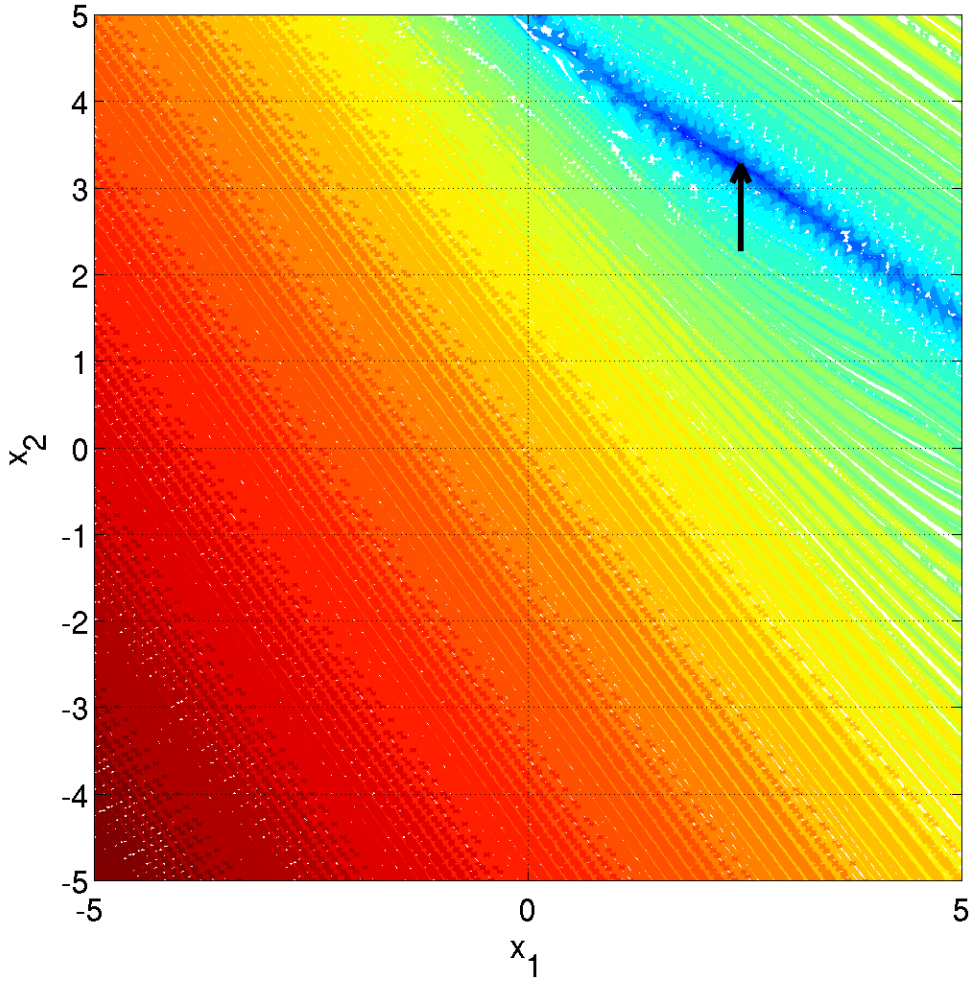
Properties Moderately ill-conditioned counterpart to f_{17}

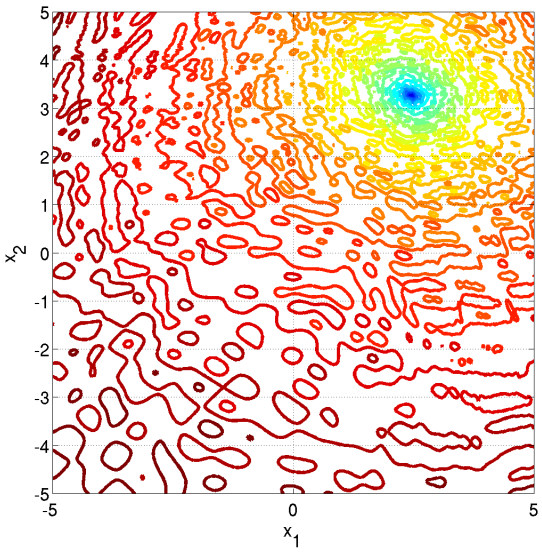
- conditioning of about 1000

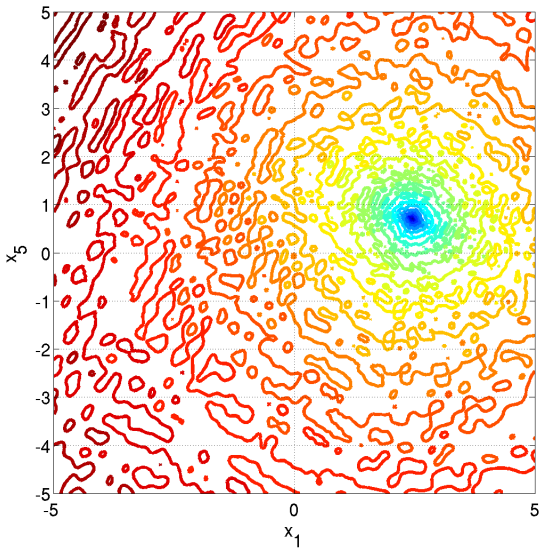
Information gained from this function:

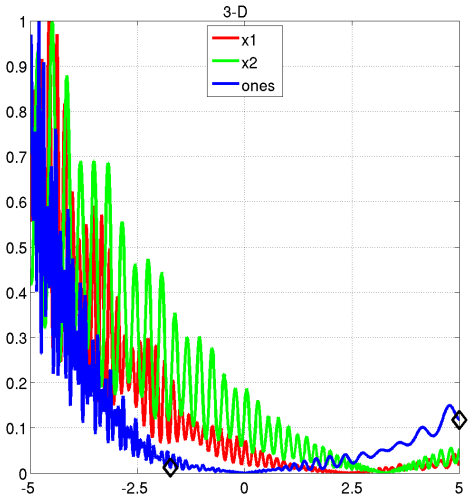
- In comparison to f17: What is the effect of ill-conditioning?

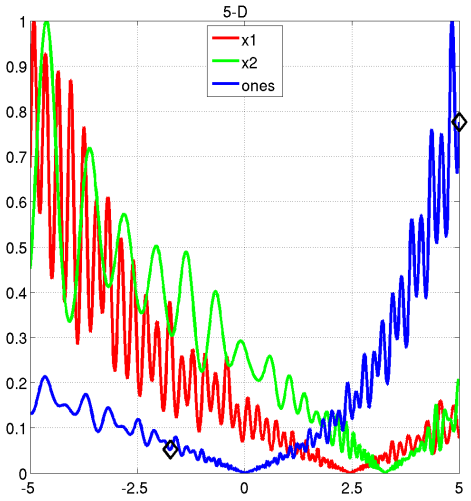


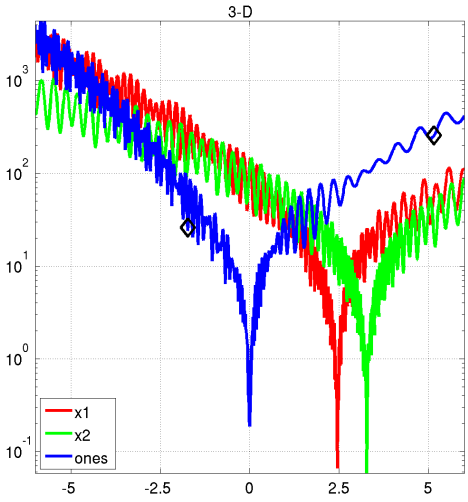

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

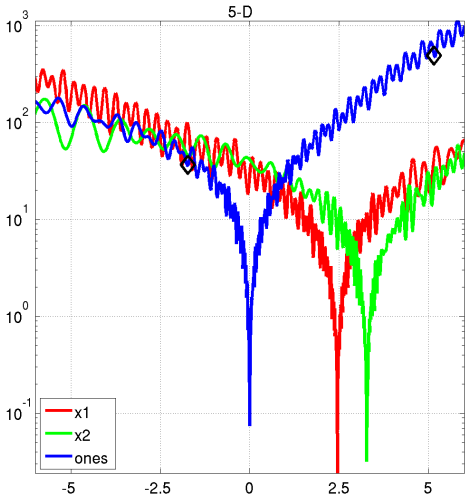

x2
x1

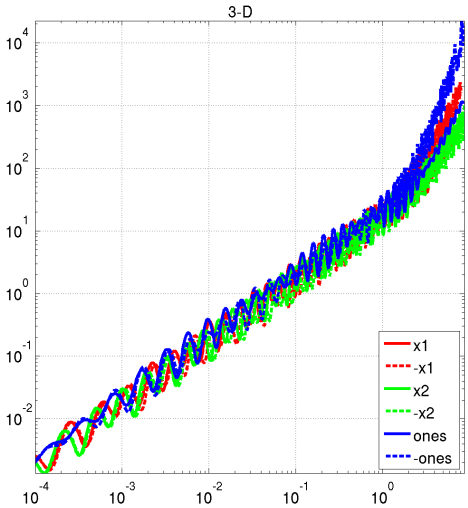

x5
x1

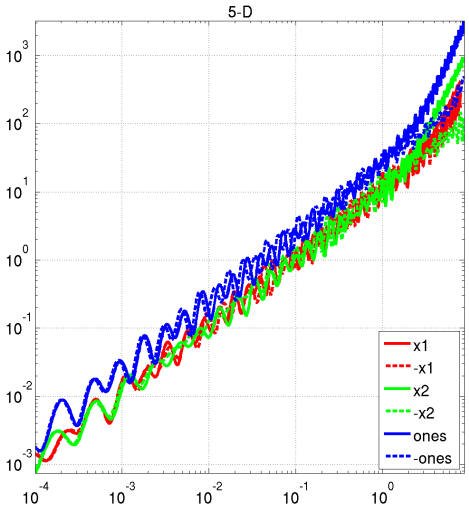

3-D
x1
x2
ones

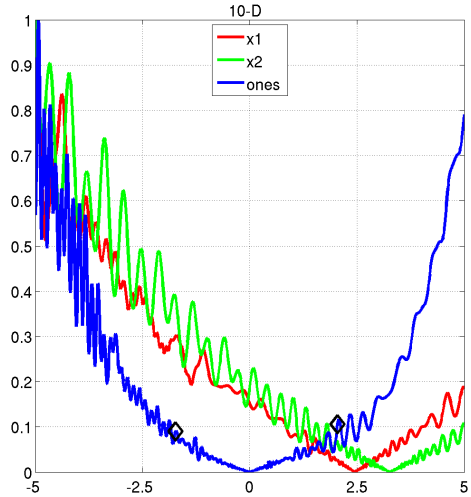

5-D
x1
x2
ones

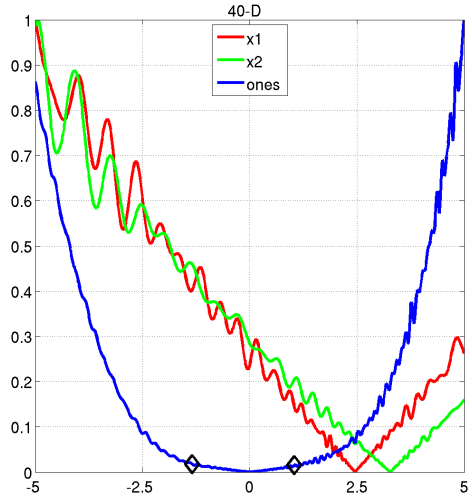

3-D
x1
x2
ones

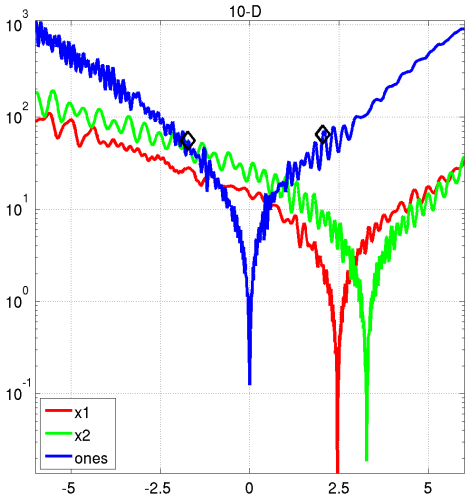

5-D
x1
x2
ones

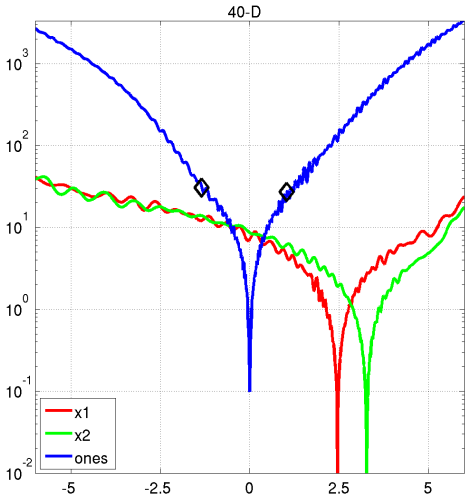

3-D
x1
-x1
x2
-x2
ones
-ones

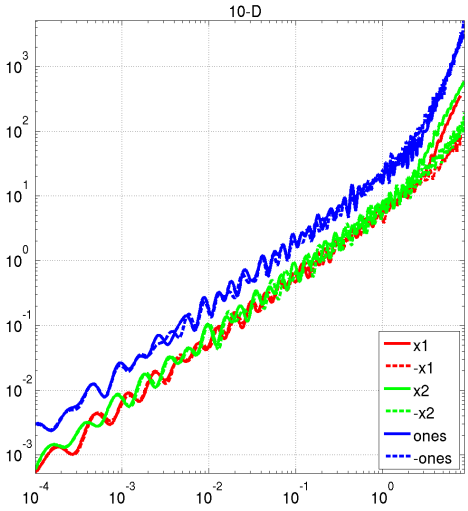

5-D
x1
-x1
x2
-x2
ones
-ones

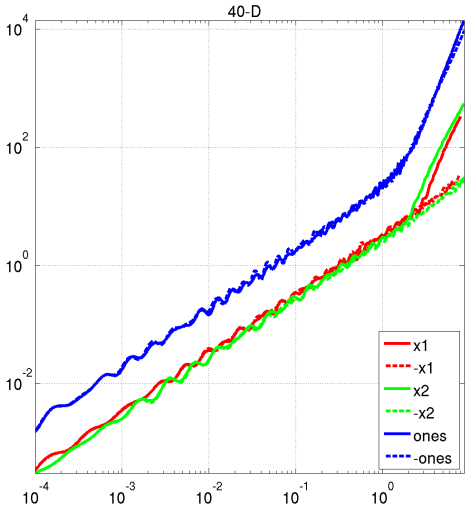

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones

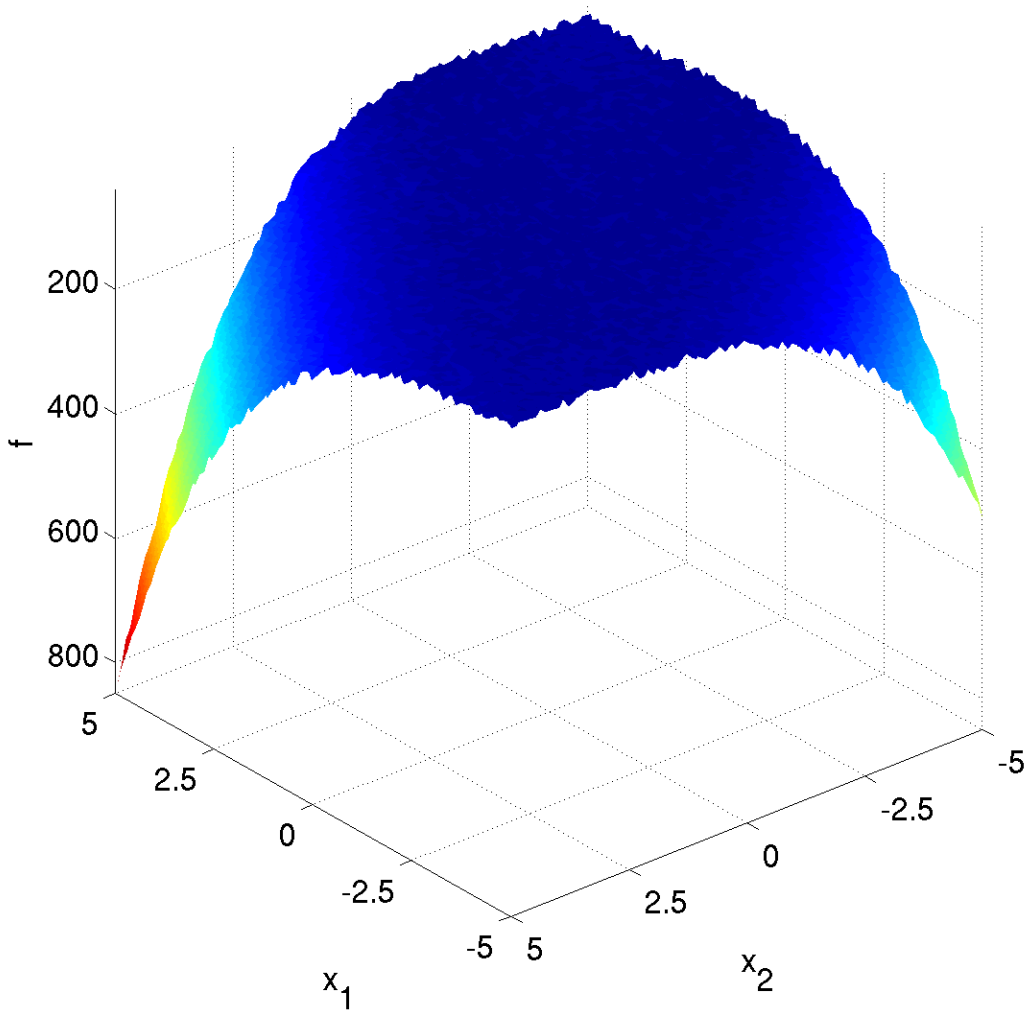

40-D
x1
-x1
x2
-x2
ones
-ones

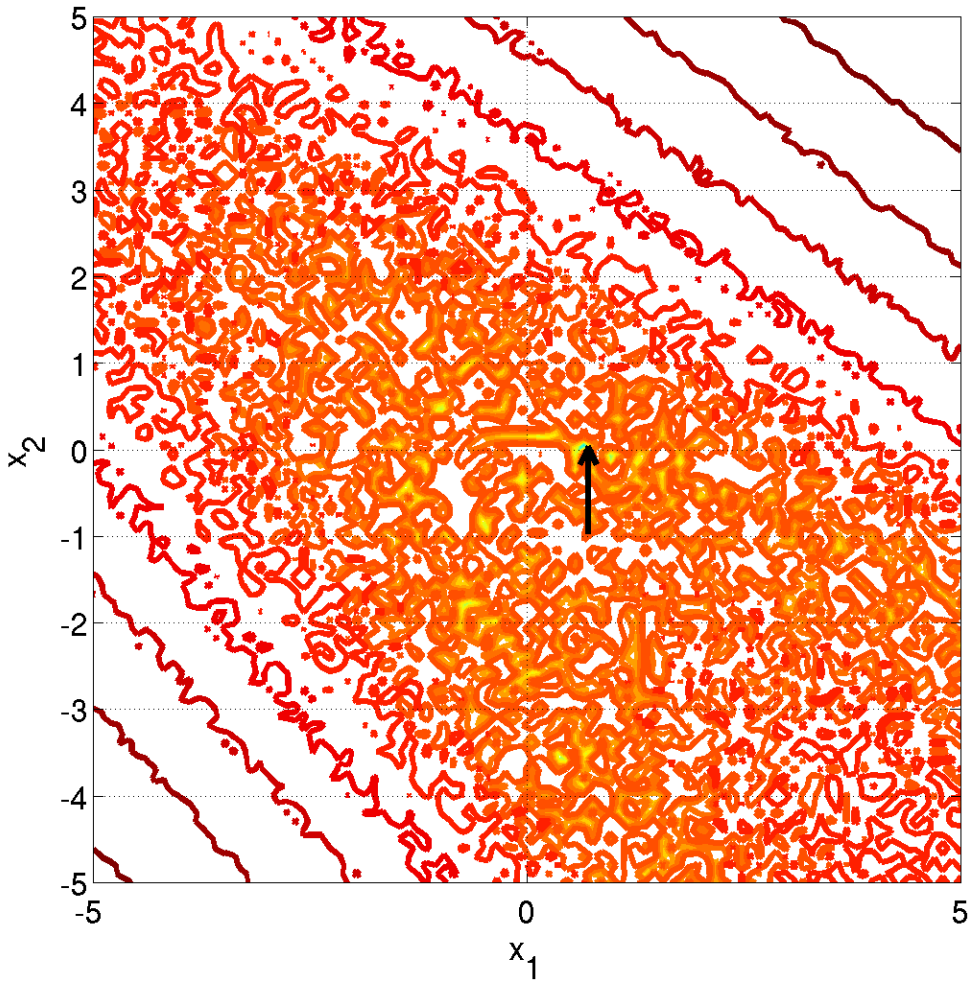
4.19 Composite Griewank-Rosenbrock Function F8F2

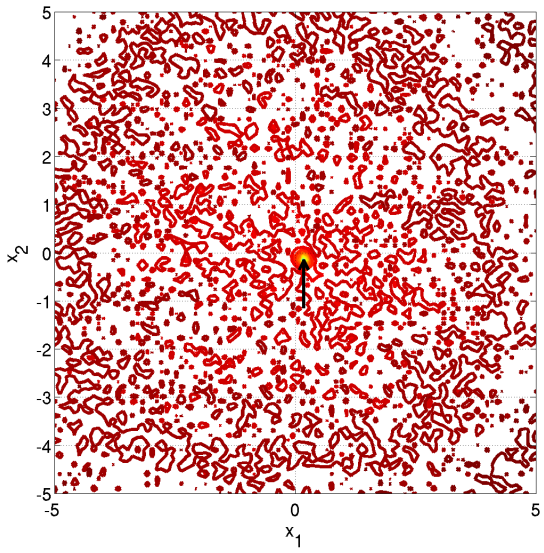
$$f_{19}(\mathbf{x}) = \frac{10}{D-1} \sum_{i=1}^{D-1} \left(\frac{s_i}{4000} - \cos(s_i) \right) + 10 + f_{\text{opt}} \tag{19}$$

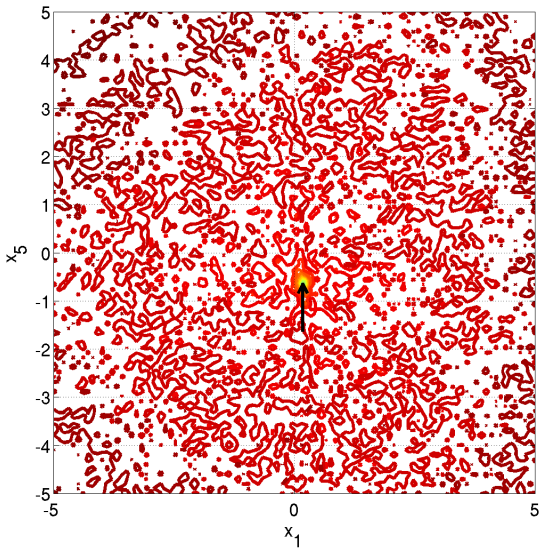
- $\mathbf{z} = \max\left(1, \frac{\sqrt{D}}{8}\right) \mathbf{R}\mathbf{x} + 0.5$
- $s_i = 100\,(z_i^2 - z_{i+1})^2 + (z_i - 1)^2$ for $i = 1, \ldots, D$
- $\mathbf{z}^{\text{opt}} = \mathbf{1}$

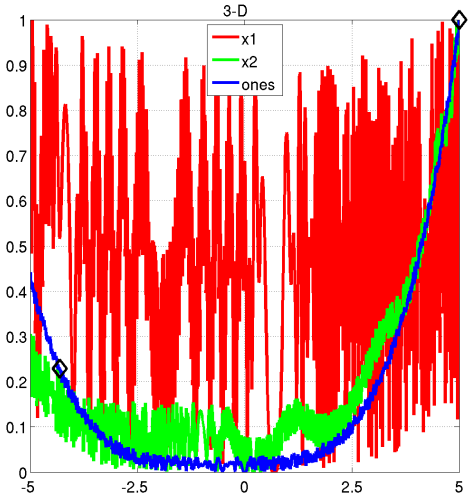
Properties Resembling the Rosenbrock function in a highly multimodal way.

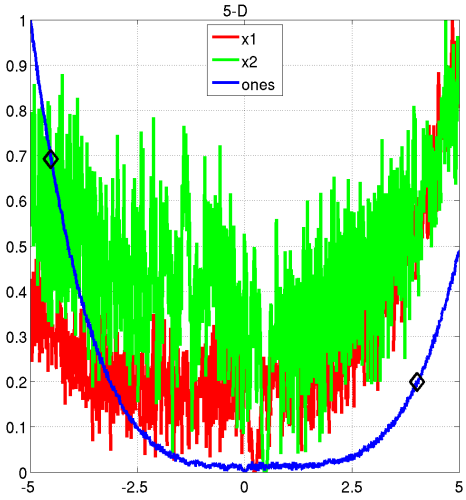


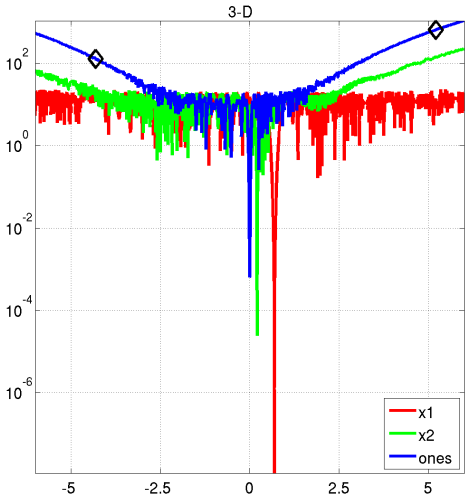

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

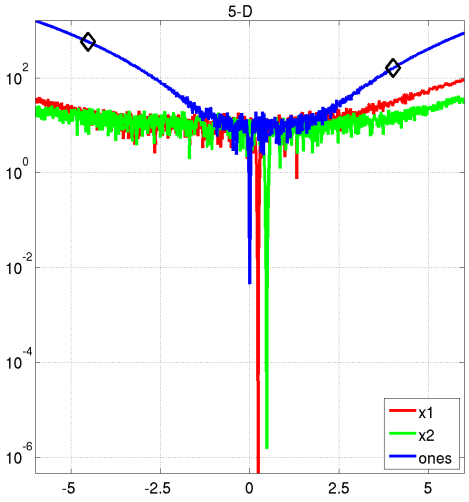

x2
x1

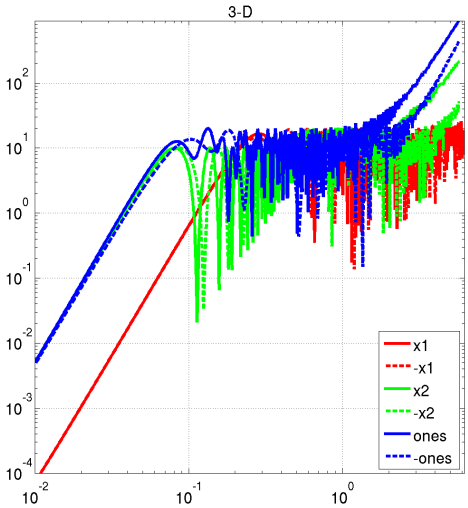

x5
x1

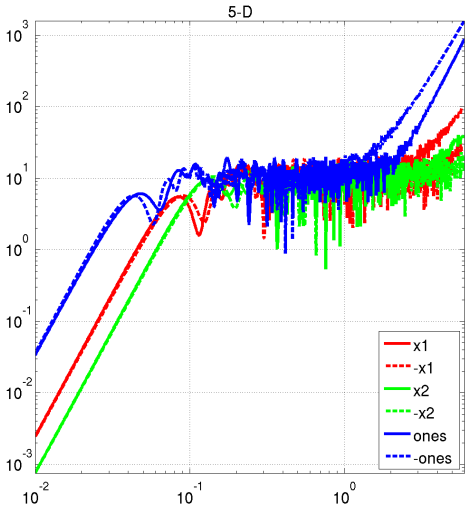

3-D
x1
x2
ones

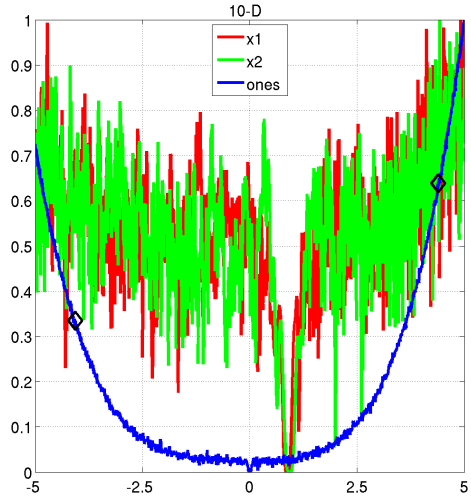

5-D
x1
x2
ones

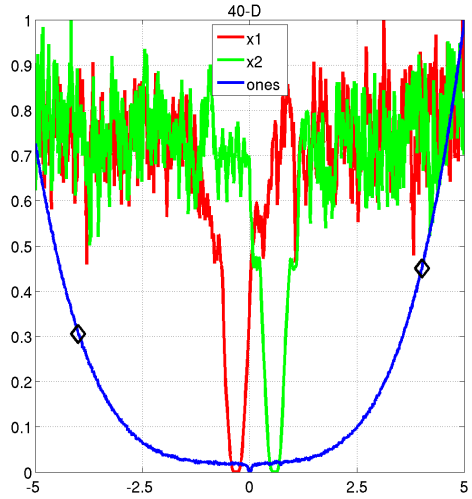

3-D
x1
x2
ones

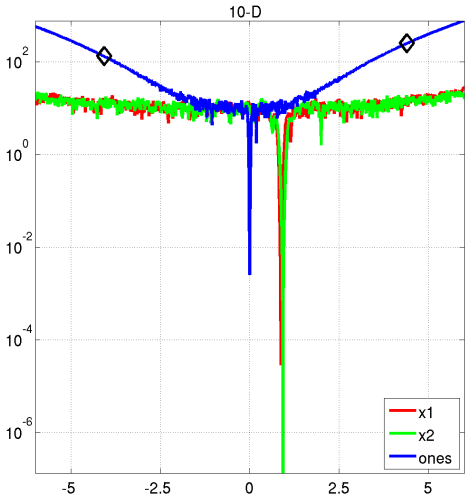

5-D
x1
x2
ones

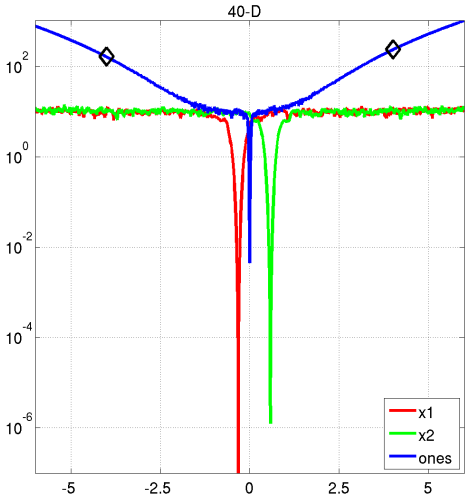

3-D
x1
-x1
x2
-x2
ones
-ones

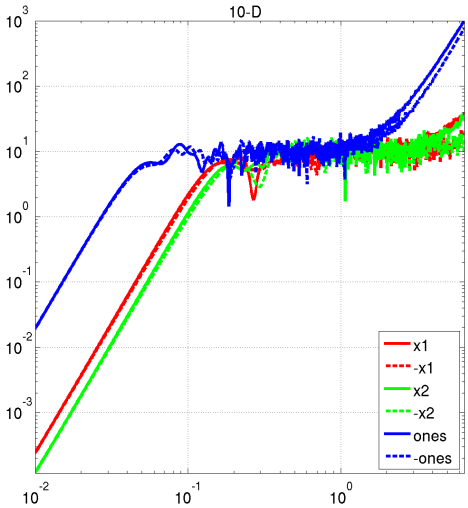

5-D
x1
-x1
x2
-x2
ones
-ones

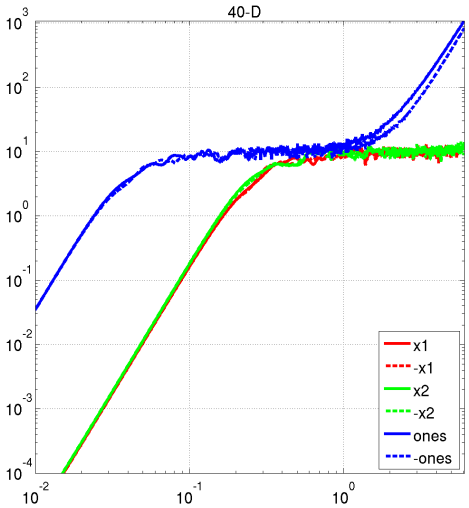

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

5 Multi-modal functions with weak global structure

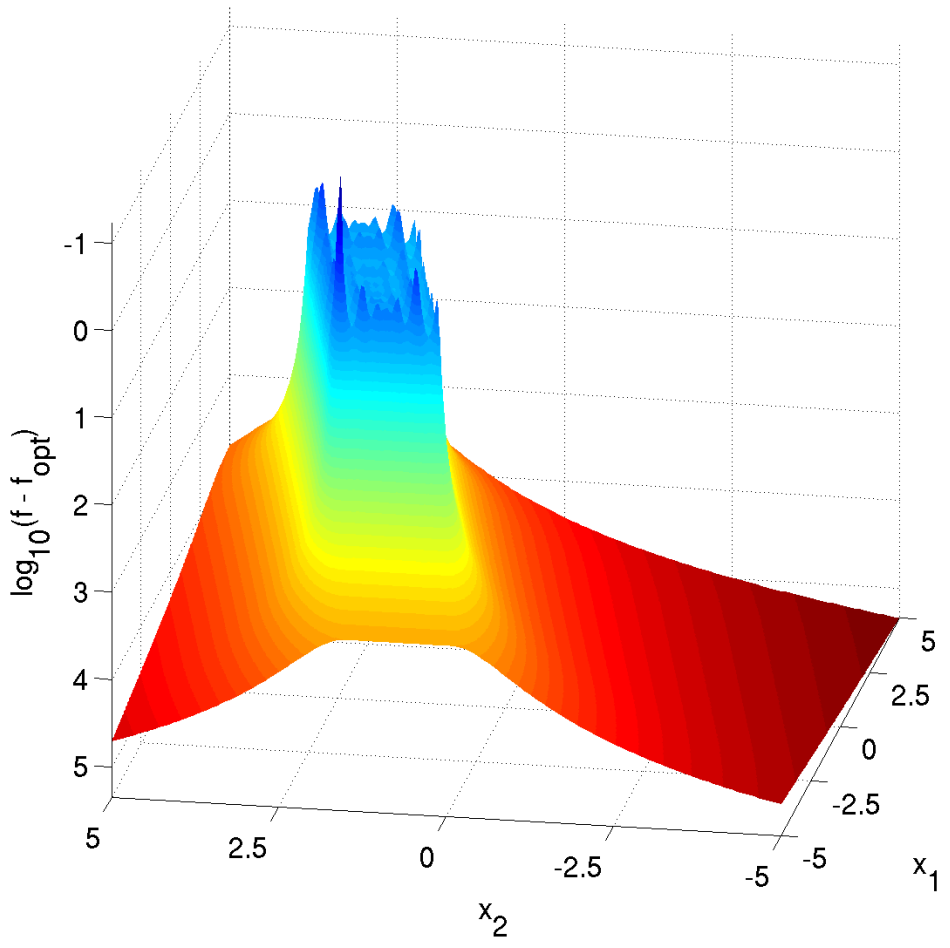
5.20 Schwefel Function

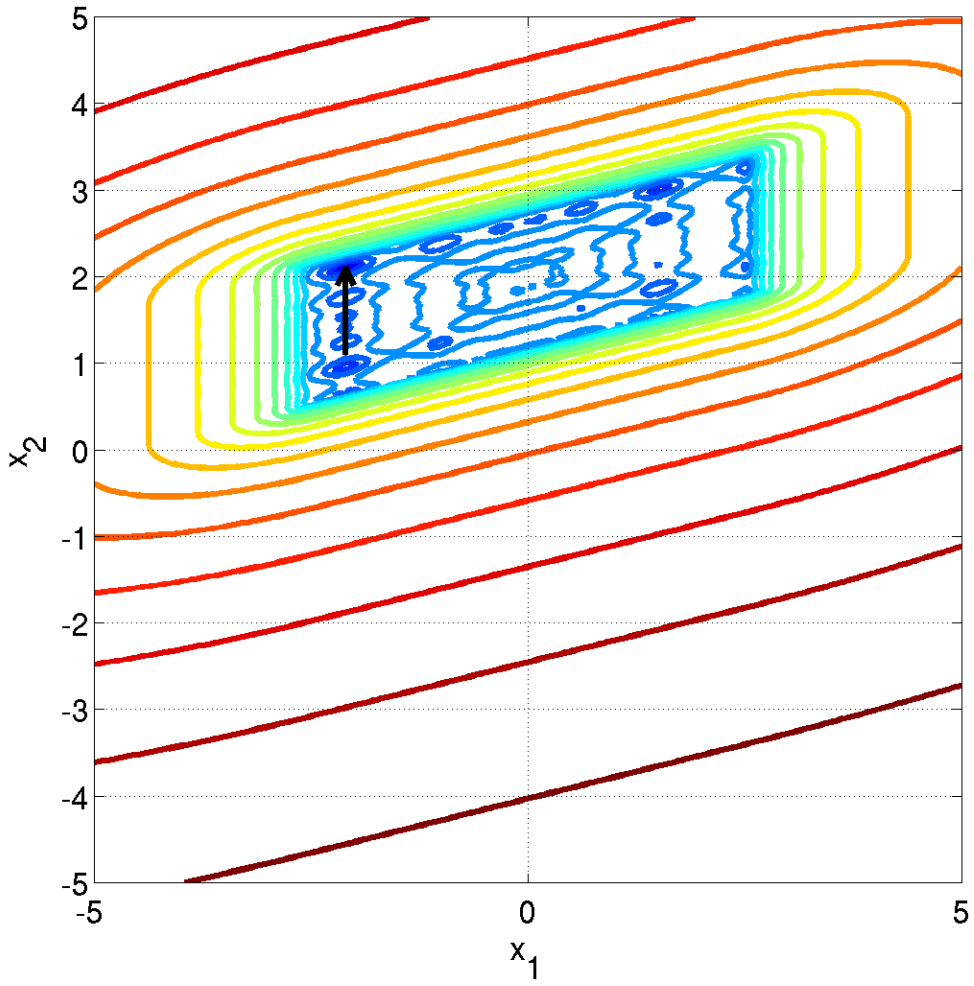
$$f_{20}(\mathbf{x}) = -\frac{1}{100D}\sum_{i=1}^{D} z_i \sin(\sqrt{|z_i|}) + 4.189828872724339 + 100 f_{\text{pen}}(\mathbf{z}/100) + f_{\text{opt}} \qquad (20)$$

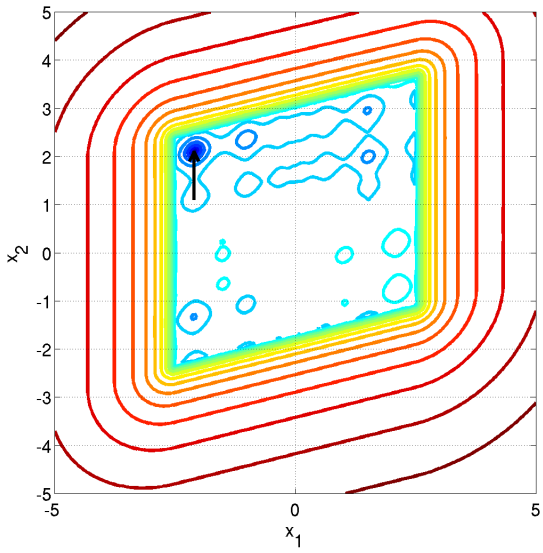
- $\hat{\mathbf{x}} = 2 \times \mathbf{1}_-^+ \otimes \mathbf{x}$
- $\hat{z}_1 = \hat{x}_1$, $\hat{z}_{i+1} = \hat{x}_{i+1} + 0.25\left(\hat{x}_i - [x_i^{\text{opt}} \longrightarrow] \; 2\,|x_i^{\text{opt}}|\right)$ for $i = 1, \ldots, D-1$
- $\mathbf{z} = 100\left(\Lambda^{10}(\hat{\mathbf{z}} - [\mathbf{x}^{\text{opt}} \longrightarrow] \; 2\,|\mathbf{x}^{\text{opt}}|) + [\mathbf{x}^{\text{opt}} \longrightarrow] \; 2\,|\mathbf{x}^{\text{opt}}|\right)$
- $\mathbf{x}^{\text{opt}} = 4.2096874633/2 \; \mathbf{1}_-^+$, where $\mathbf{1}_-^+$ is the same realization as above

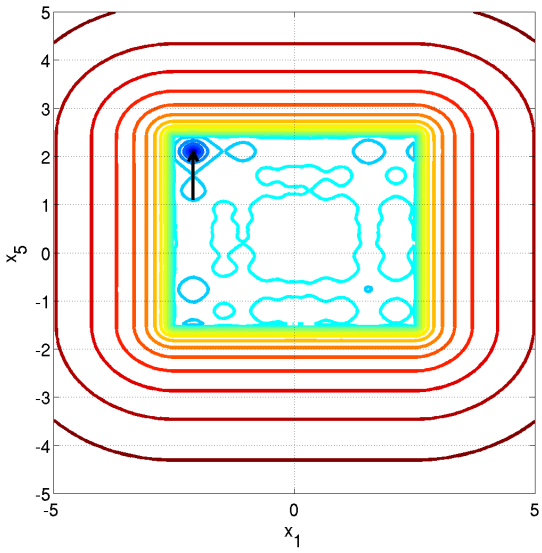
Properties The most prominent 2^D minima are located comparatively close to the corners of the unpenalized search area, based on [9].

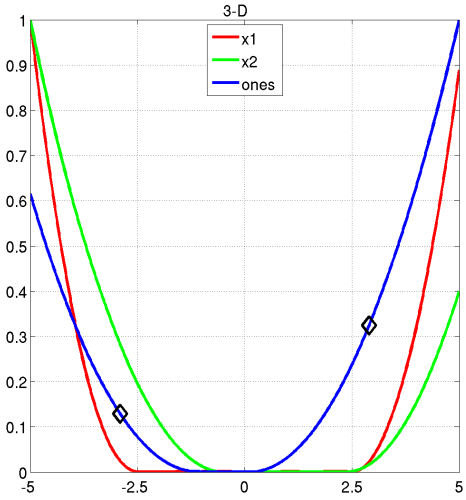
- the penalization is essential, as otherwise more and better minima occur further away from the search space origin

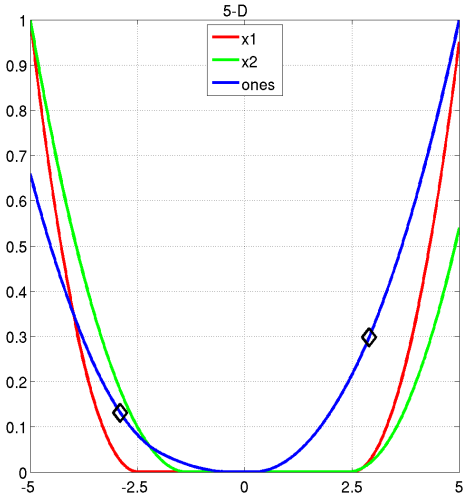


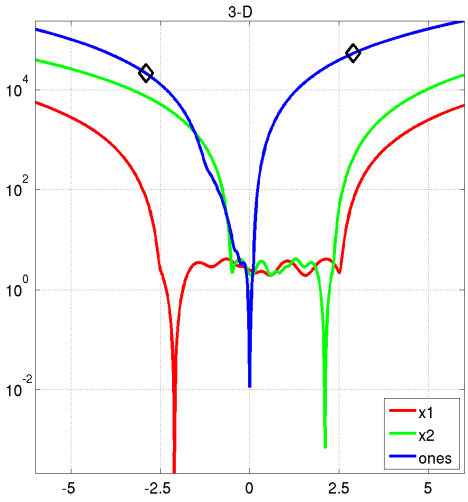

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

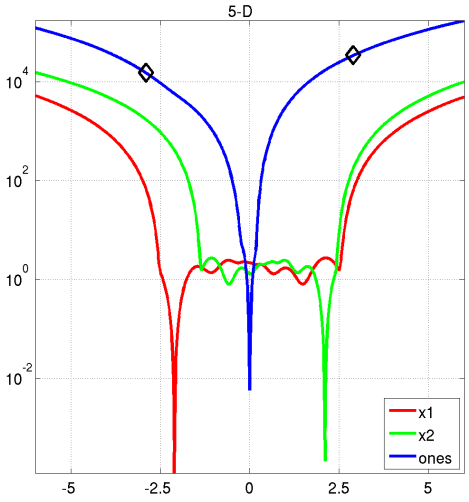

x2
x1

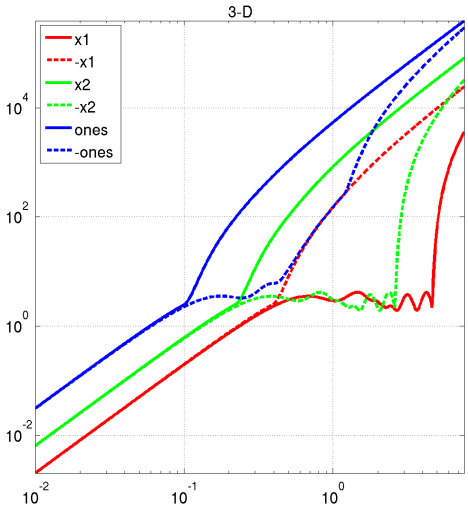

x5
x1

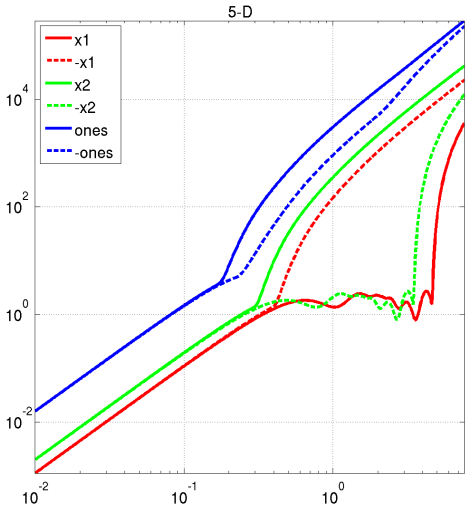

3-D
x1
x2
ones

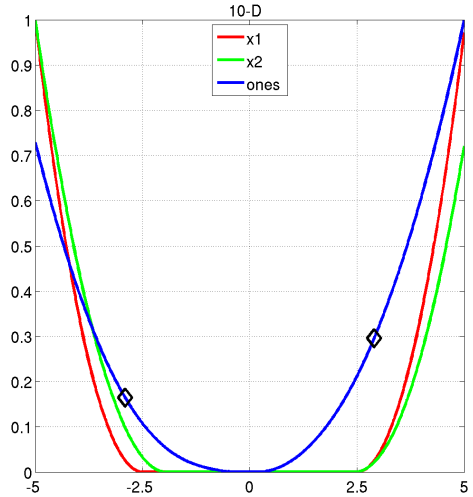

5-D
x1
x2
ones

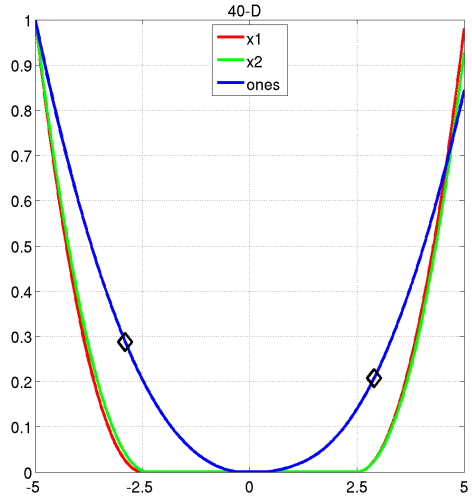

3-D
x1
x2
ones

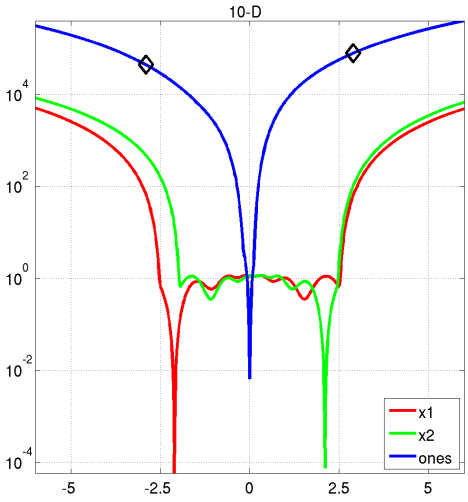

5-D
x1
x2
ones

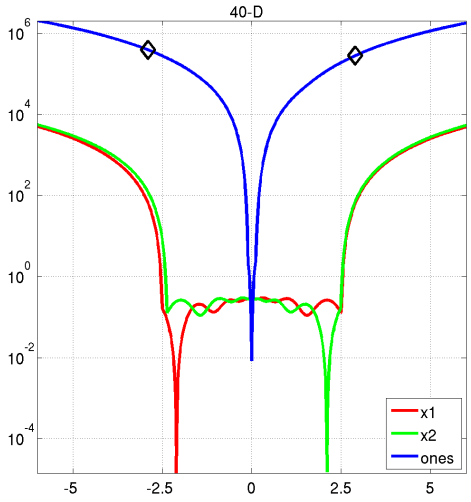

3-D
x1
-x1
x2
-x2
ones
-ones

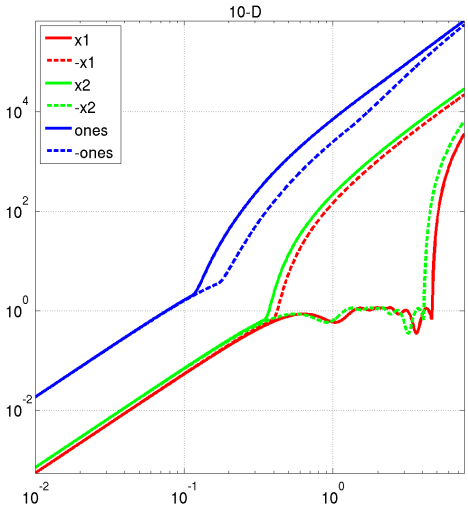

5-D
x1
-x1
x2
-x2
ones
-ones

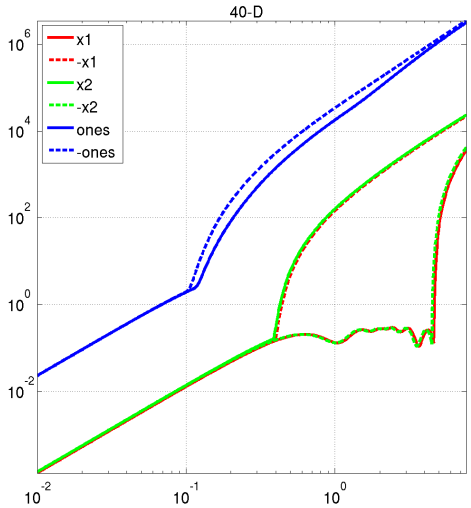

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

5.21 Gallagher's Gaussian 101-me Peaks Function

$$f_{21}(\mathbf{x}) = T_{\text{osz}}\left(10 - \max_{i=1}^{101} w_i \exp\left(-\frac{1}{2D}(\mathbf{x}-\mathbf{y}_i)^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{C}_i\mathbf{R}(\mathbf{x}-\mathbf{y}_i)\right)\right)^2 + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \quad (21)$$

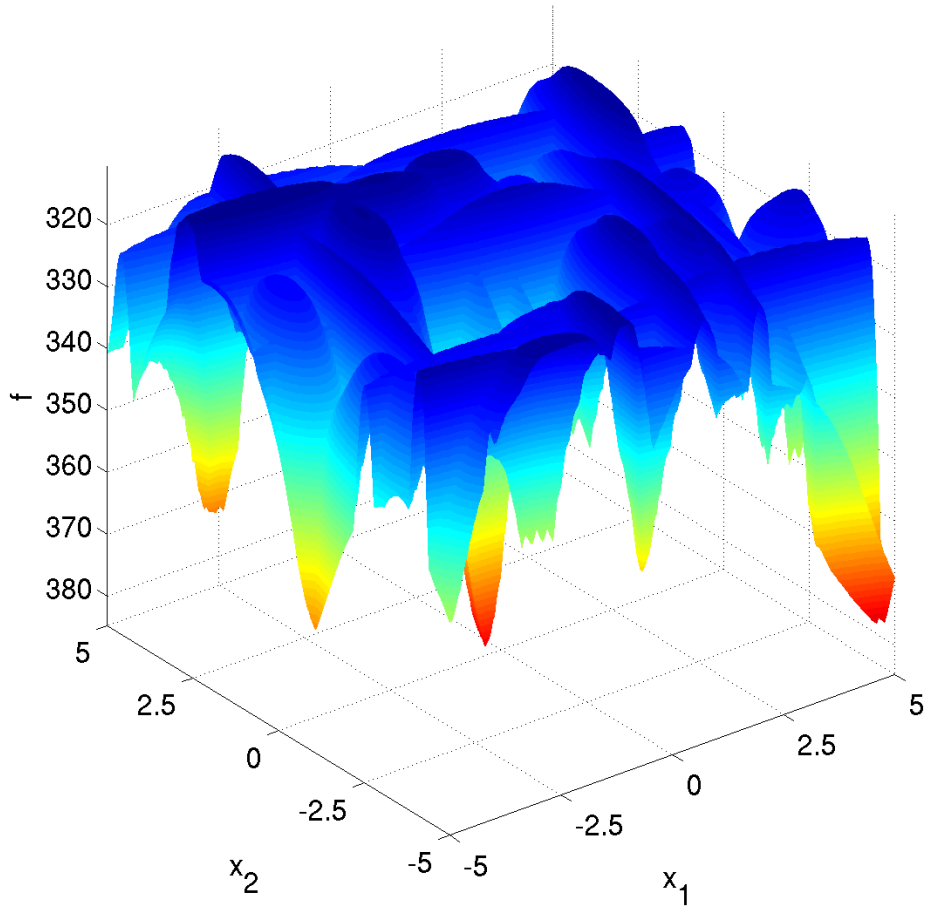
- $w_i = \begin{cases} 1.1 + 8 \times \frac{i-2}{99} & \text{for } i = 2, \ldots, 101 \\ 10 & \text{for } i = 1 \end{cases}$, three optima have a value larger than 9
- $\mathbf{C}_i = \Lambda^{\alpha_i}/\alpha_i^{1/4}$ where Λ^{α_i} is defined as usual (see Section 0.2), but with randomly permuted diagonal 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 $[[-4.9, 4.9]^D \longrightarrow] [-5, 5]^D$ for $i = 2, \ldots, 101$ and $\mathbf{y}_1 \in [-4, 4]^D$. The global optimum is at $\mathbf{x}^{\text{opt}} = \mathbf{y}_1$.

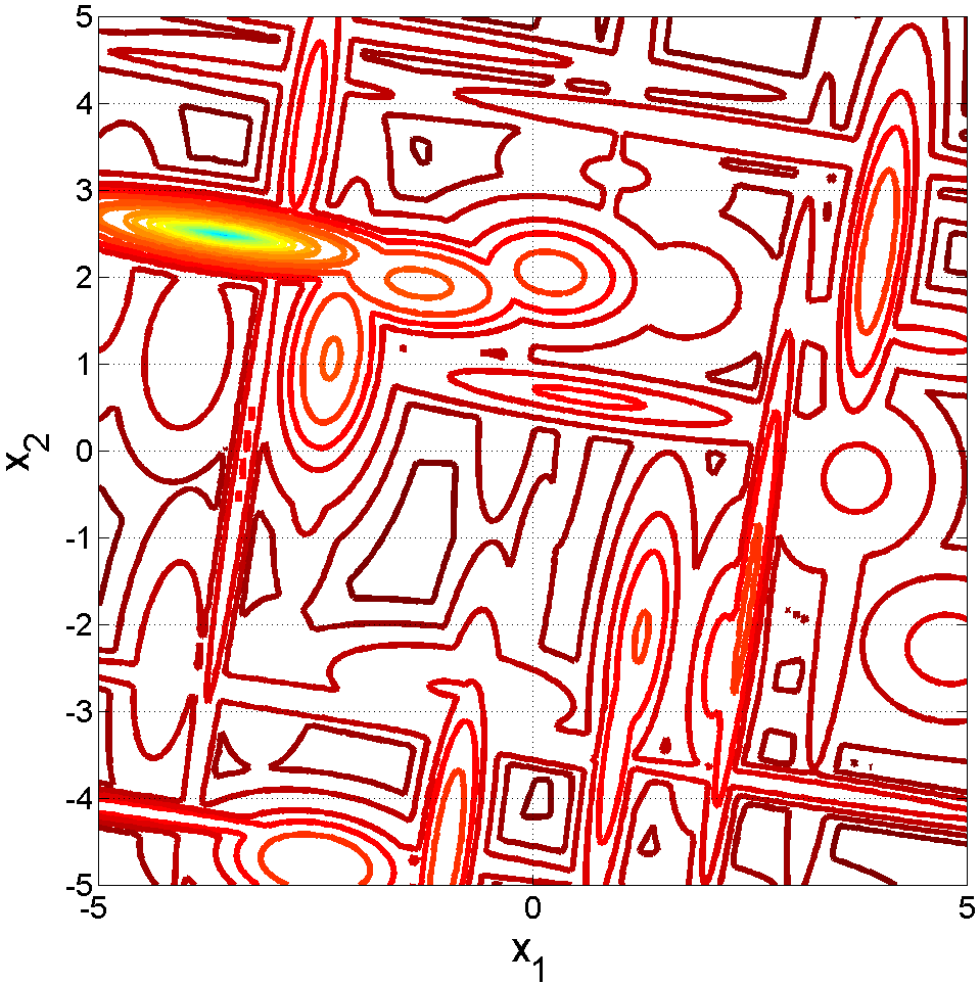
Properties The function consists of 101 optima with position and height being unrelated and randomly chosen (different for each instantiation of the function), based on [2].

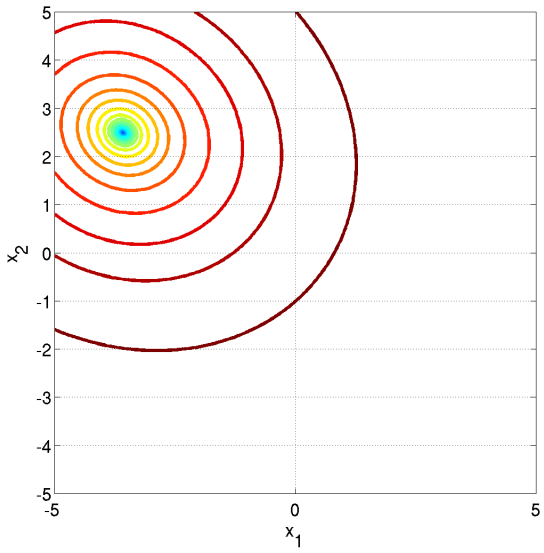
- the conditioning around the global optimum is about 30

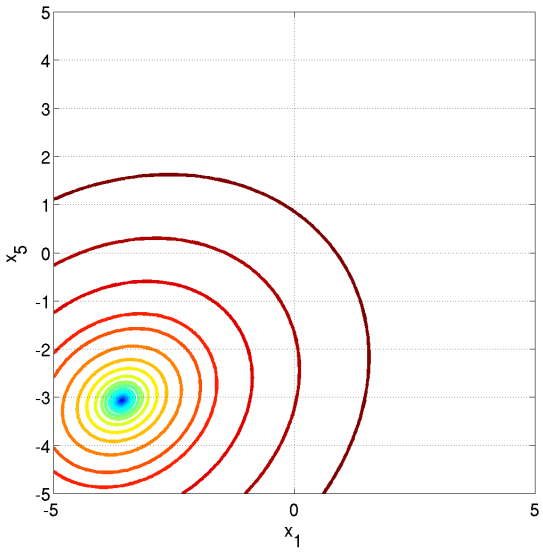
Information gained from this function:

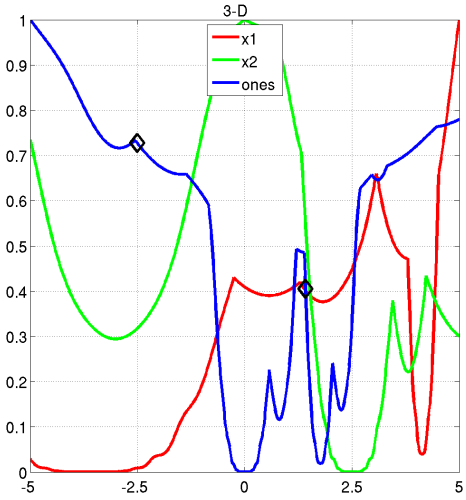
- Is the search effective without any global structure?

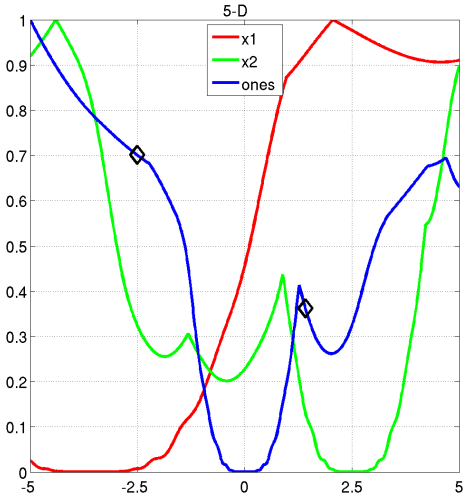


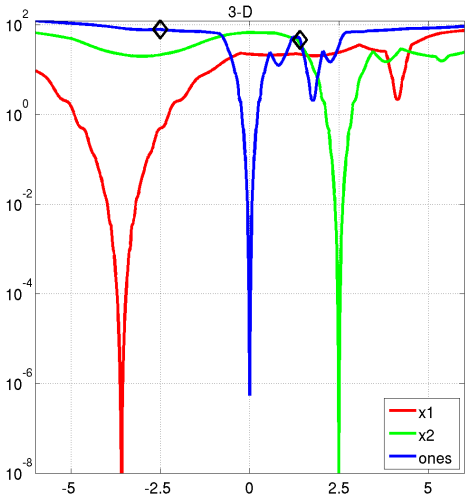

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

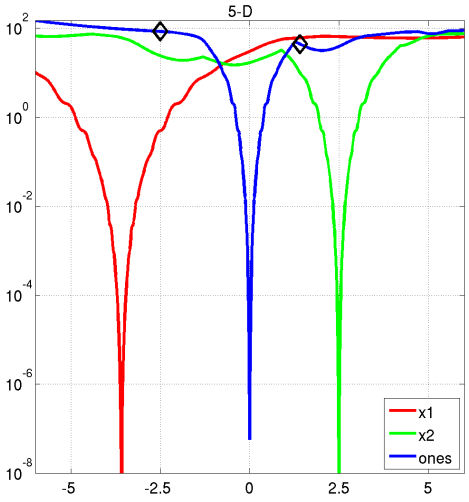

x2
x1

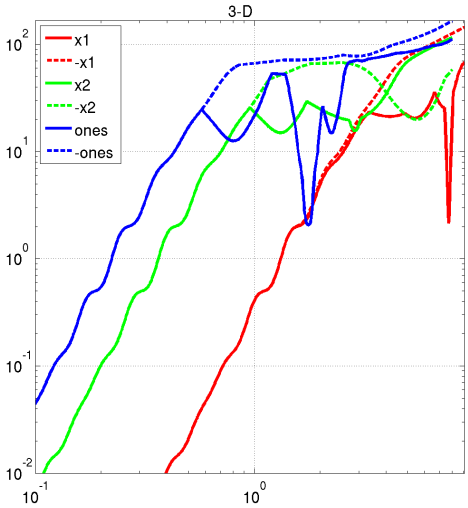

x5
x1

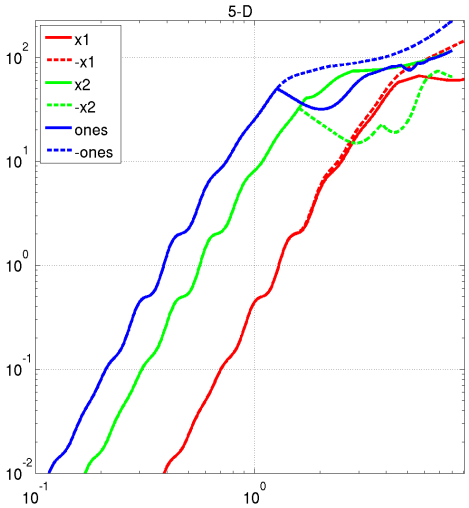

3-D
x1
x2
ones

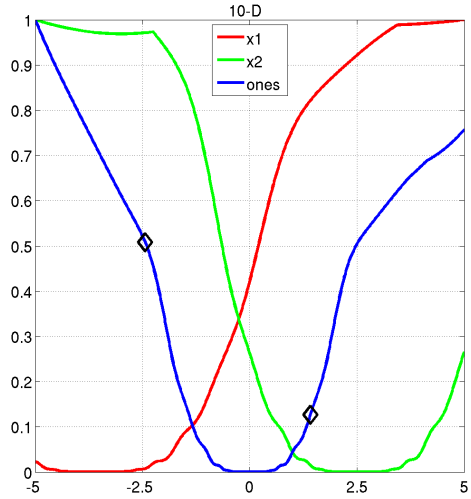

5-D
x1
x2
ones

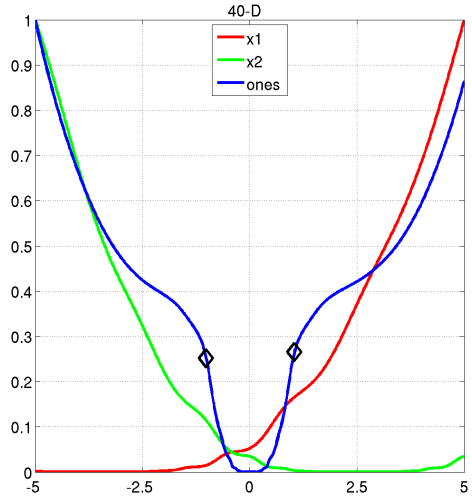

3-D
x1
x2
ones

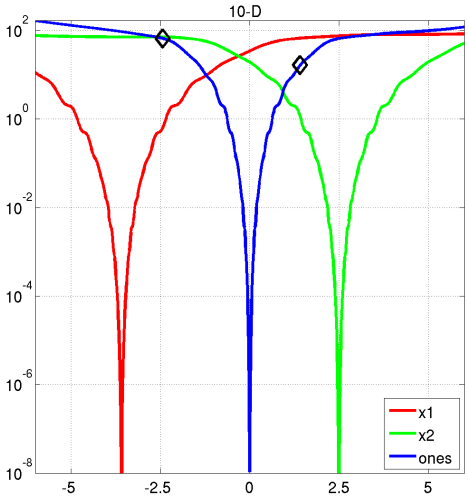

5-D
x1
x2
ones

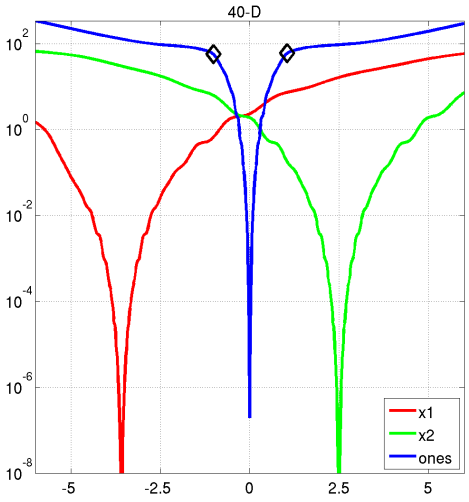

3-D
x1
-x1
x2
-x2
ones
-ones

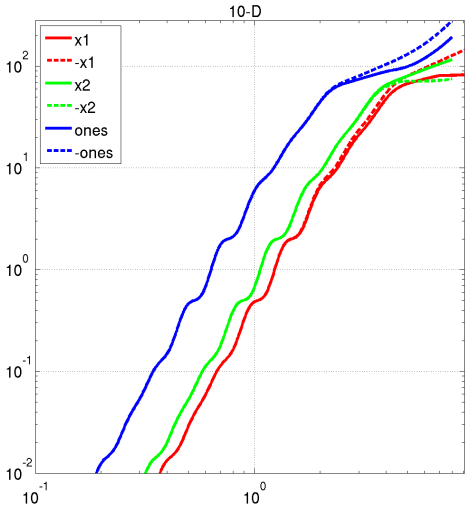

5-D
x1
-x1
x2
-x2
ones
-ones

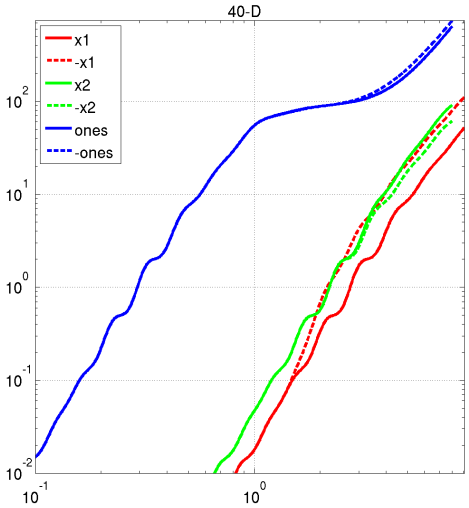

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

5.22 Gallagher's Gaussian 21-hi Peaks Function

$$f_{22}(\mathbf{x}) = T_{\text{osz}}\left(10 - \max_{i=1}^{21} w_i \exp\left(-\frac{1}{2D}(\mathbf{x}-\mathbf{y}_i)^{\mathrm{T}}\mathbf{R}^{\mathrm{T}}\mathbf{C}_i\mathbf{R}(\mathbf{x}-\mathbf{y}_i)\right)\right)^2 + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \quad (22)$$

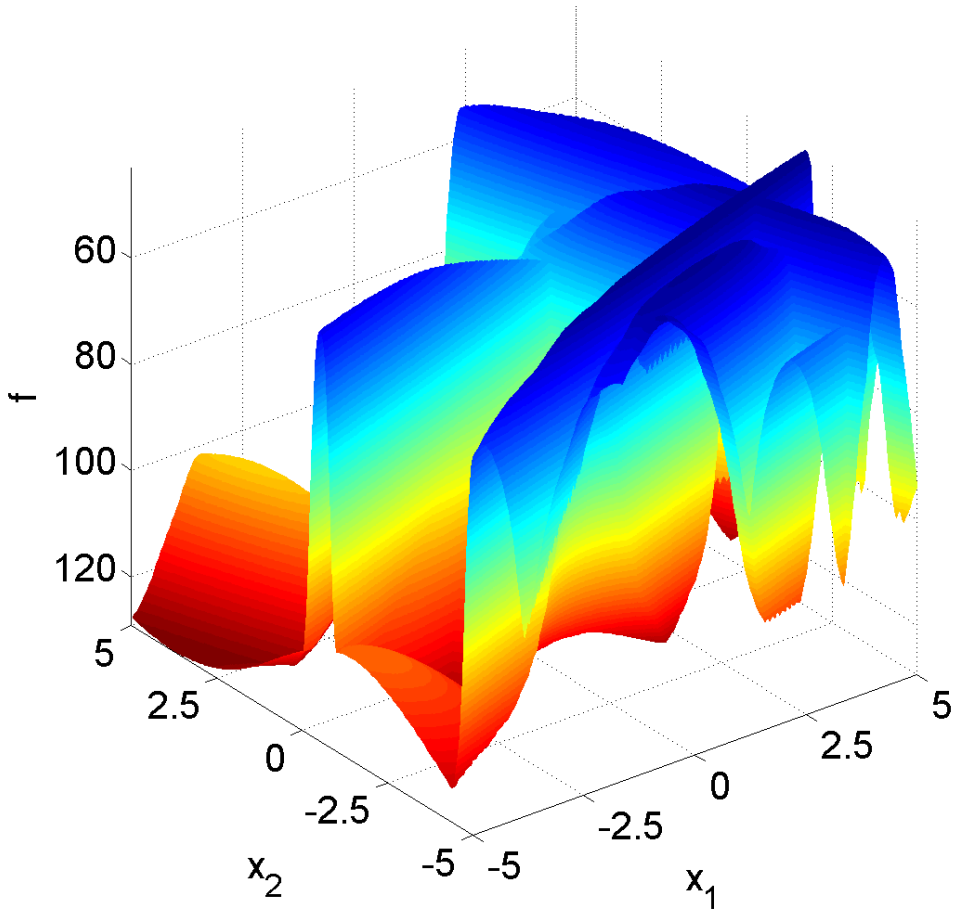
- $w_i = \begin{cases} 1.1 + 8 \times \frac{i-2}{19} & \text{for } i = 2, \ldots, 21 \\ 10 & \text{for } i = 1 \end{cases}$, two optima have a value larger than 9
- $\mathbf{C}_i = \Lambda^{\alpha_i}/\alpha_i^{1/4}$ where Λ^{α_i} is defined as usual (see Section 0.2), but with randomly permuted diagonal elements. For $i = 2, \ldots, 21$, α_i is drawn uniformly randomly from the set $\left\{1000^{2\frac{j}{19}} \mid j = 0, \ldots, 19\right\}$ without replacement, and $\alpha_i = 1000^2$ for $i = 1$.
- the local optima $\mathbf{y}_i$ are uniformly drawn from the domain $[-4.9, 4.9]^D$ for $i = 2, \ldots, 21$ and $\mathbf{y}_1 \in [\sout{[-4,4]^D \longrightarrow}] [-3.92, 3.92]^D]$. The global optimum is at $\mathbf{x}^{\text{opt}} = \mathbf{y}_1$.

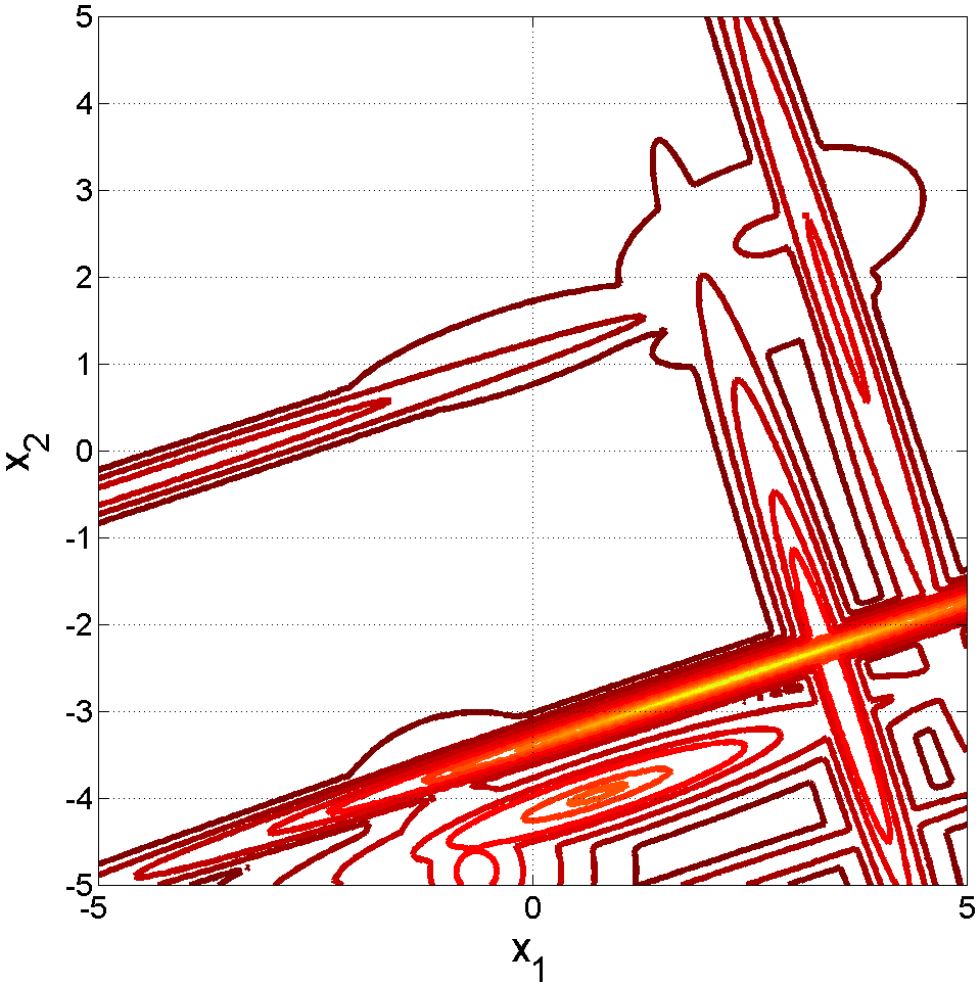
Properties The function consists of 21 optima with position and height being unrelated and randomly chosen (different for each instantiation of the function), based on [2].

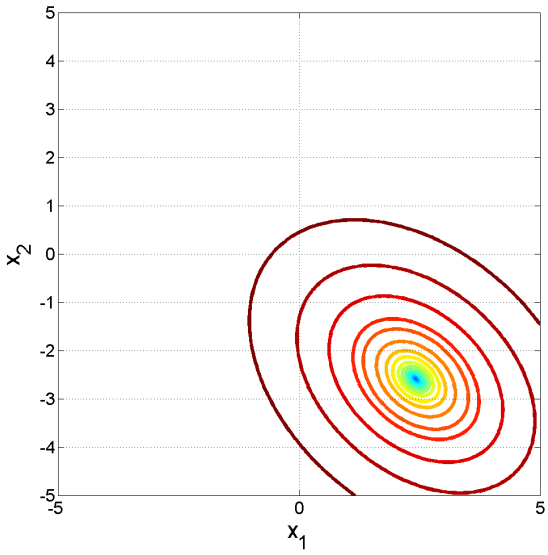
- the conditioning around the global optimum is about 1000

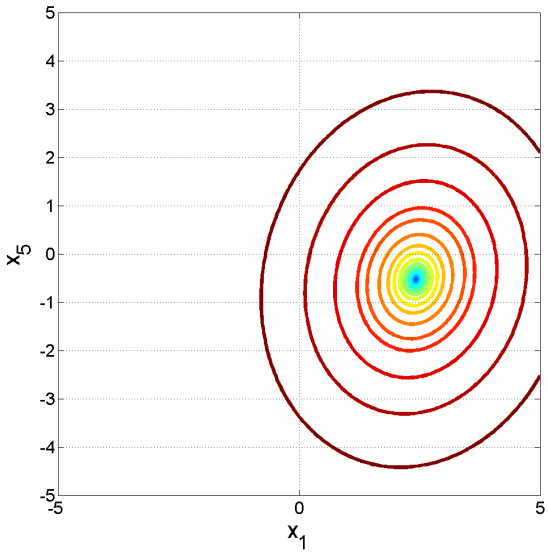
Information gained from this function:

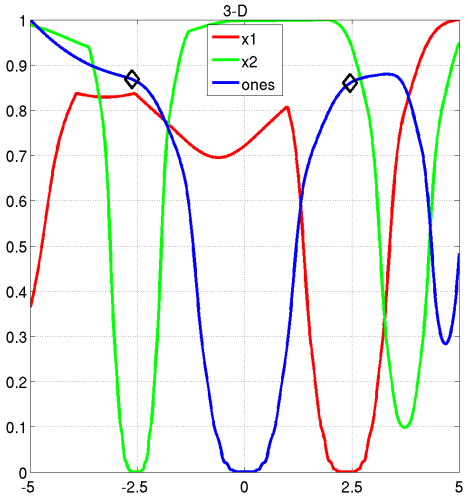
- In comparison to f21: What is the effect of higher condition?

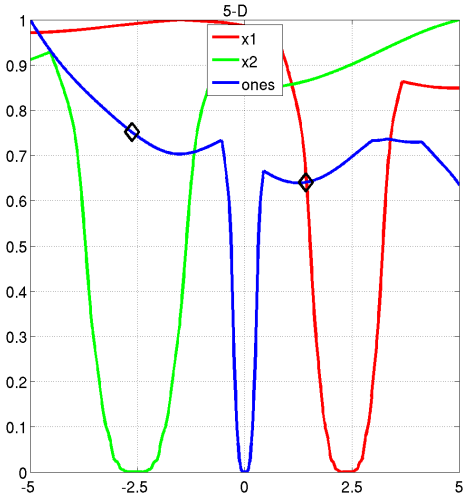


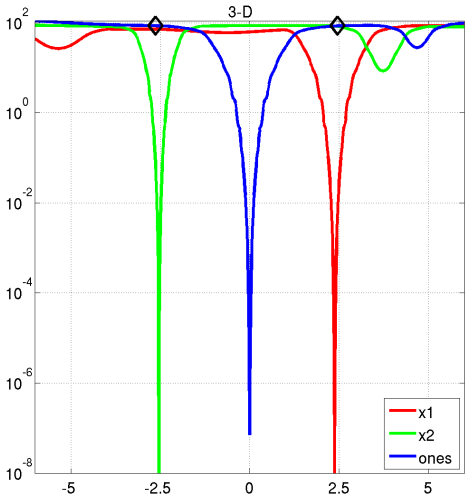

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

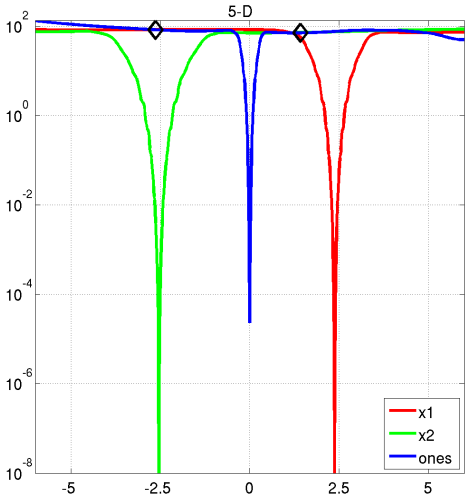

x2
x1

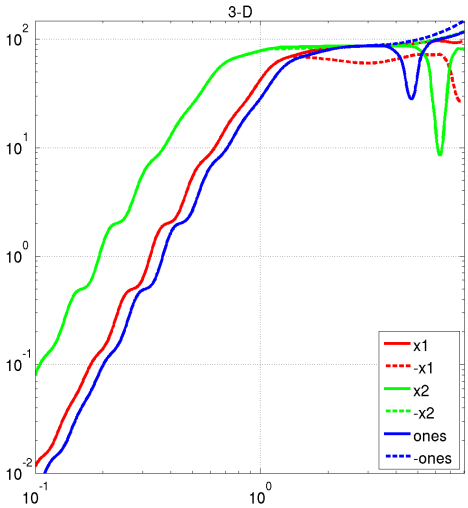

x5
x1

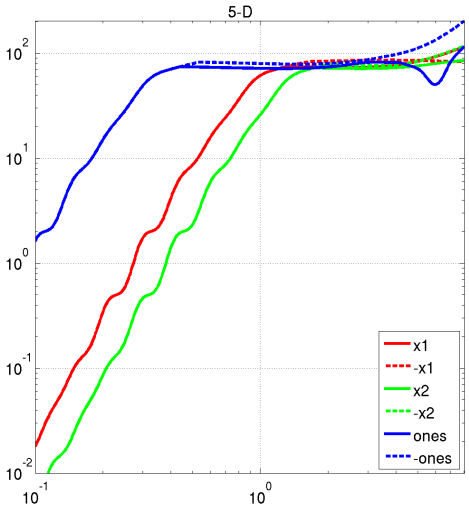

3-D
x1
x2
ones

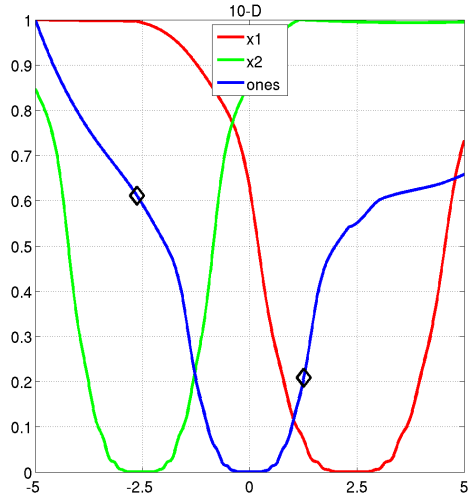

5-D
x1
x2
ones

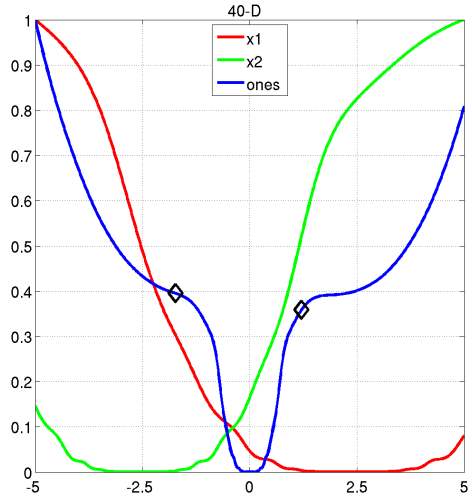

3-D
x1
x2
ones

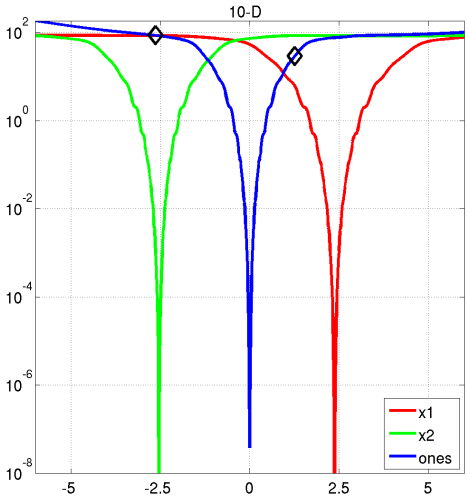

5-D
x1
x2
ones

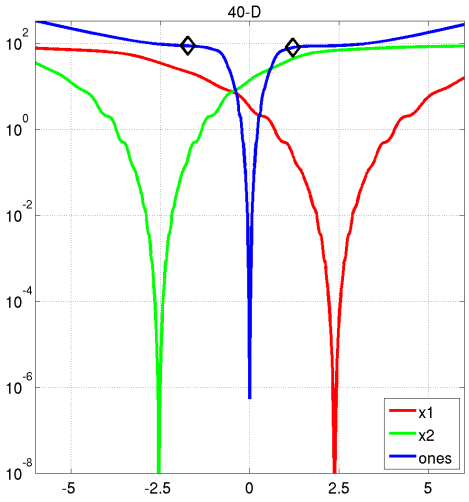

3-D
x1
-x1
x2
-x2
ones
-ones

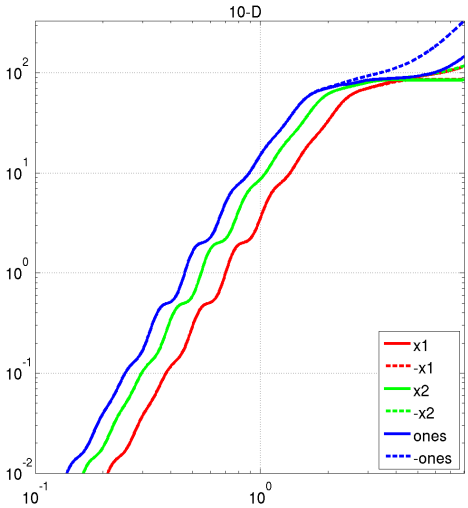

5-D
x1
-x1
x2
-x2
ones
-ones

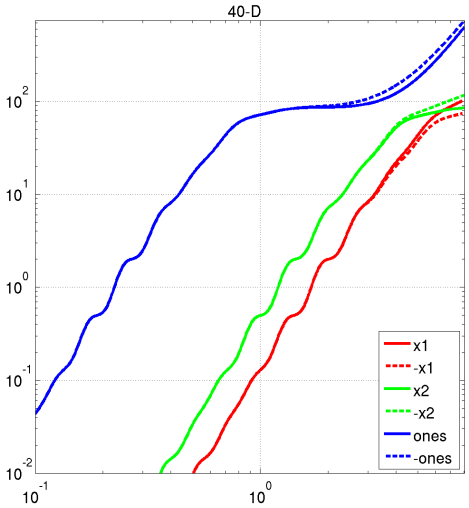

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones

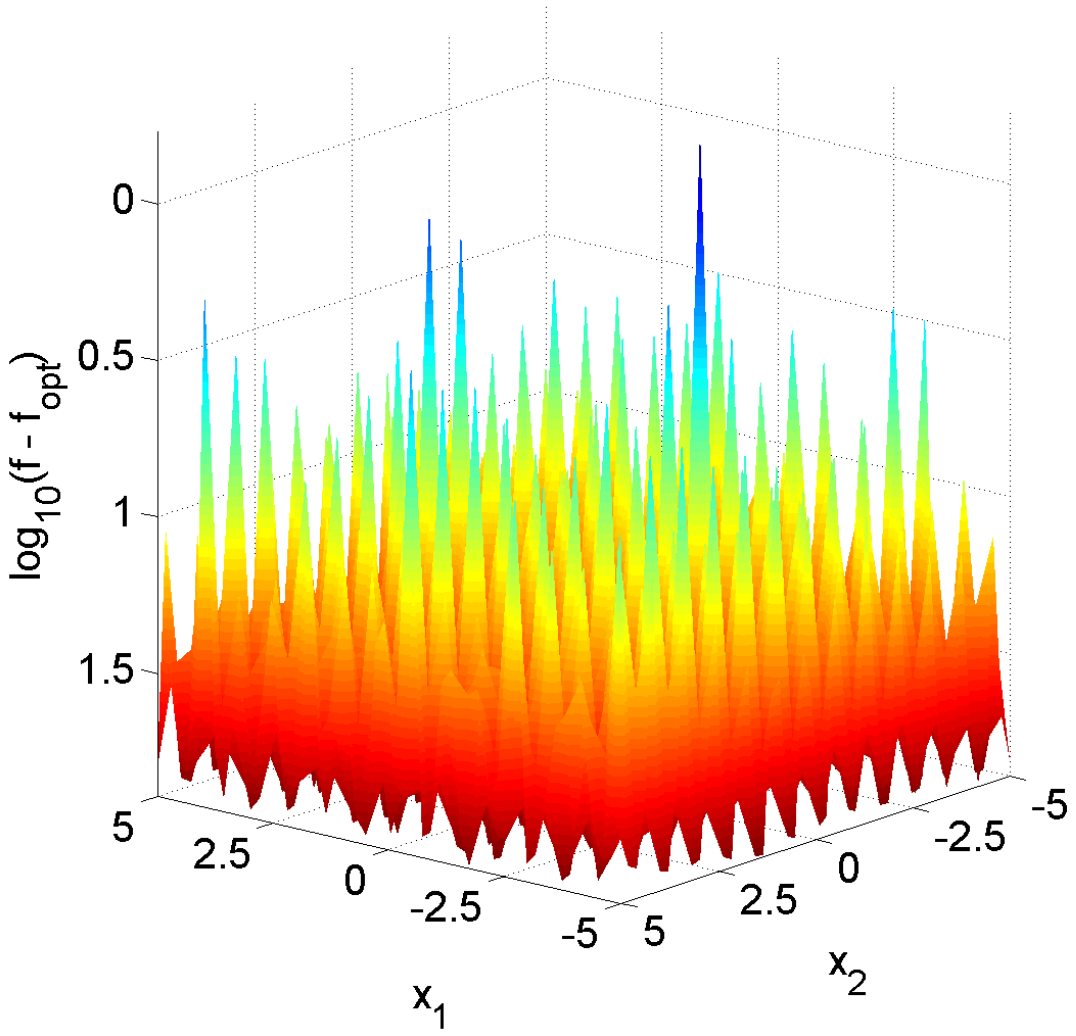

40-D
x1
-x1
x2
-x2
ones
-ones

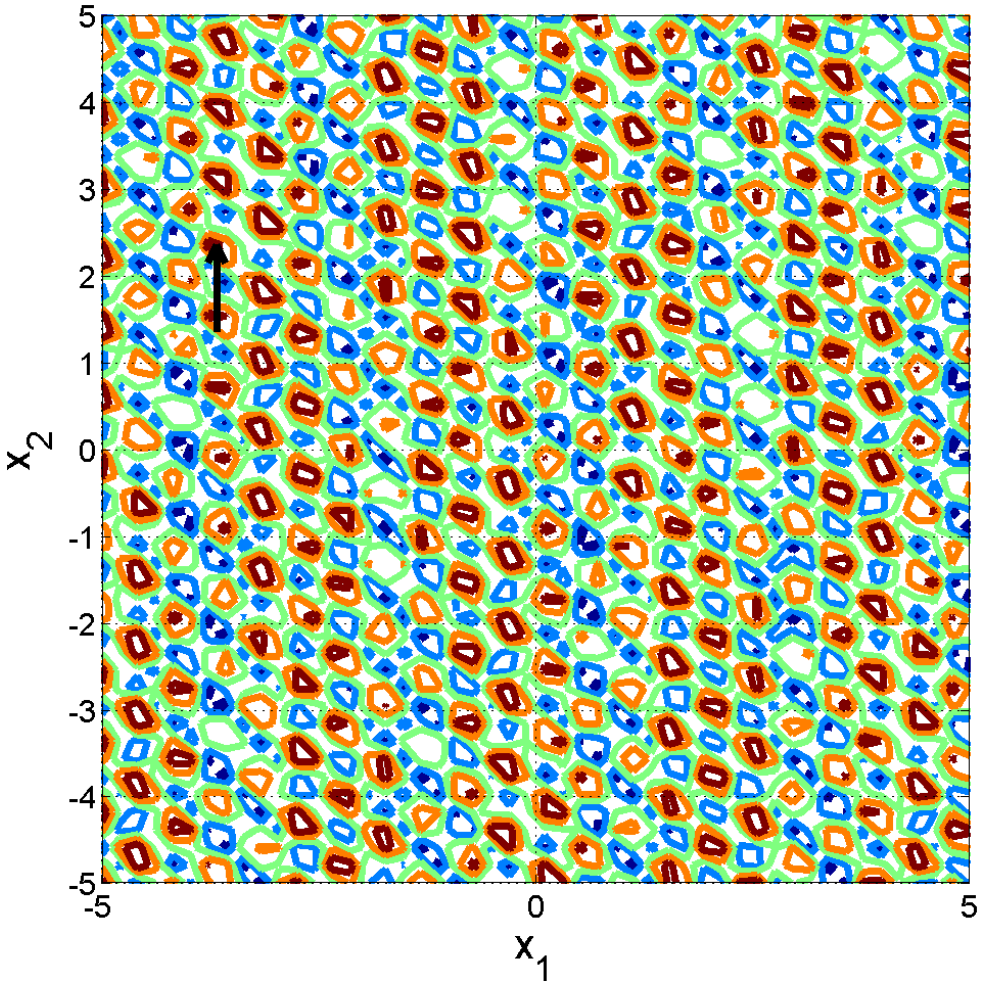
5.23 Katsuura Function

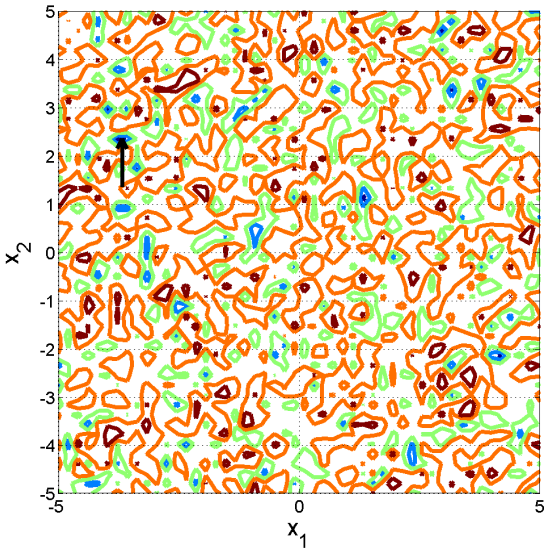
$$f_{23}(\mathbf{x}) = \frac{10}{D^2} \prod_{i=1}^{D} \left(1 + i \sum_{j=1}^{32} \frac{\left| 2^j z_i - [2^j z_i] \right|}{2^j} \right)^{10/D^{1.2}} - \frac{10}{D^2} + f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{23}$$

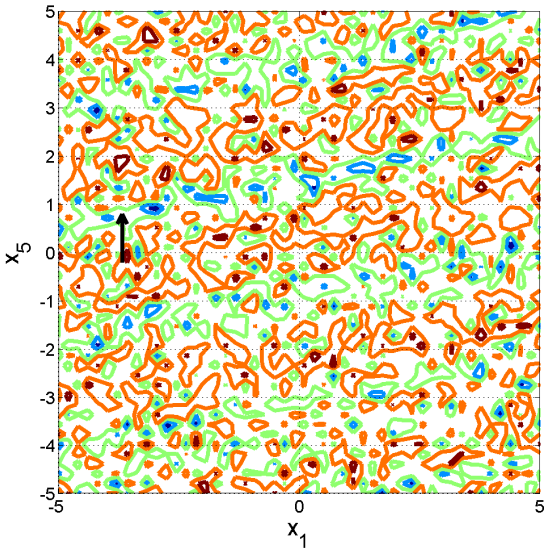
- $\mathbf{z} = \mathbf{Q}\,\Lambda^{100}\mathbf{R}(\mathbf{x} - \mathbf{x}^{\text{opt}})$

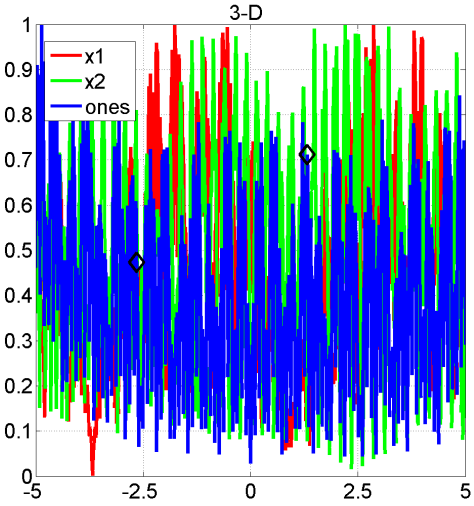
Properties Highly rugged and highly repetitive function with more than 10^D global optima, based on the idea in [6].

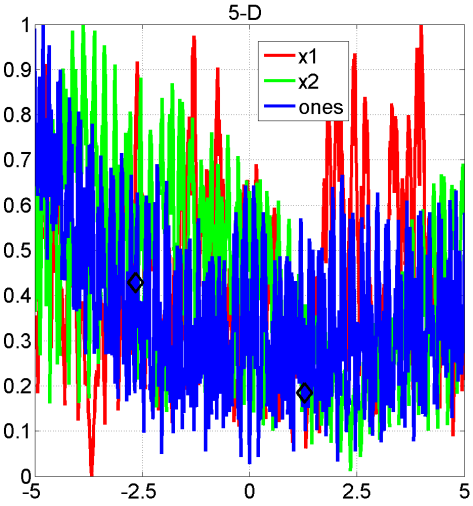


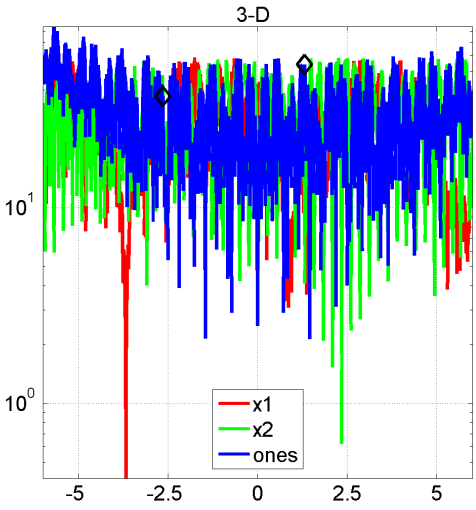

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

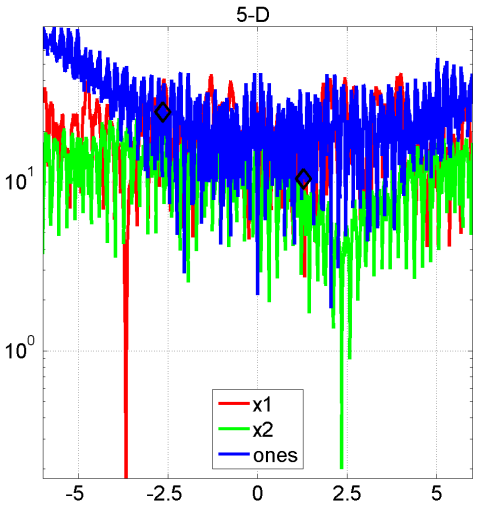

x2
x1

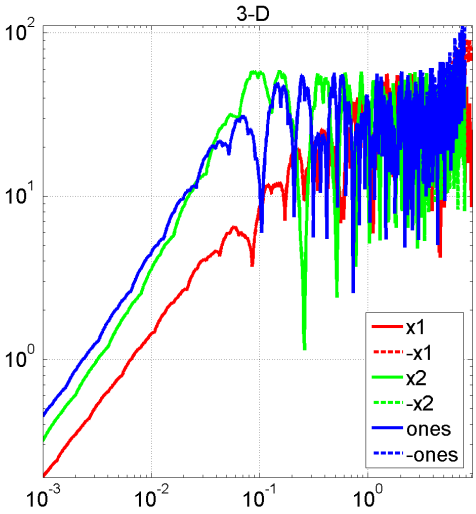

x5
x1

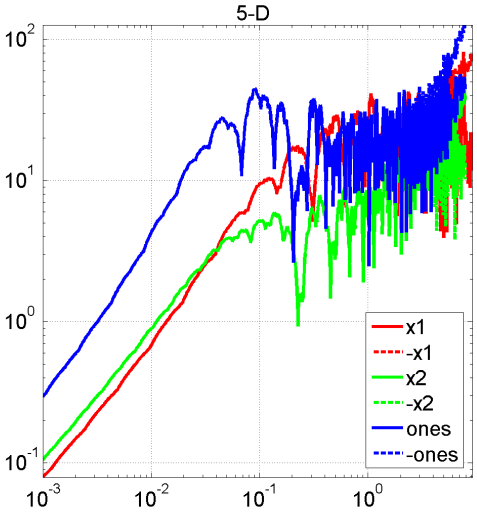

3-D
x1
x2
ones

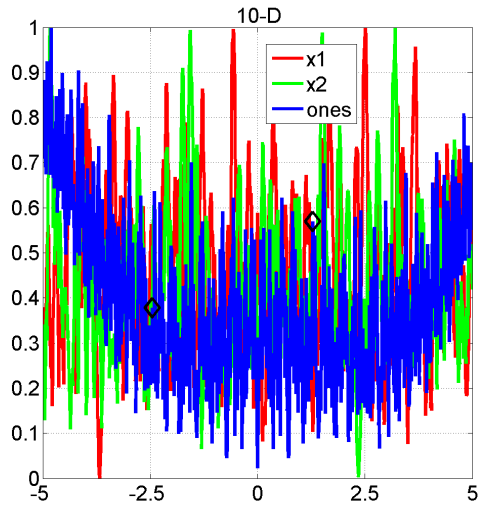

5-D
x1
x2
ones

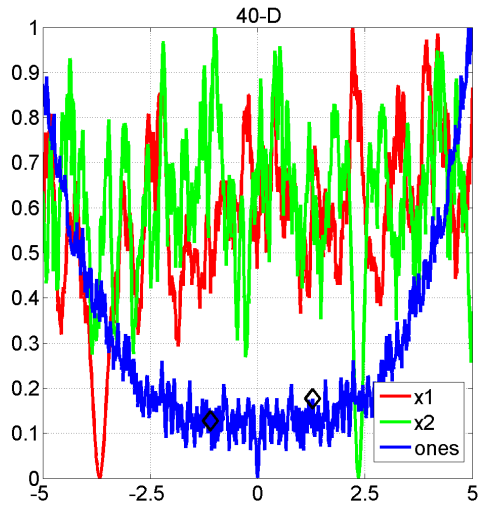

3-D
x1
x2
ones

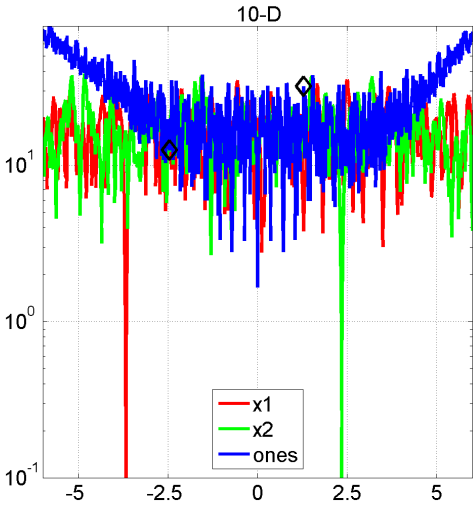

5-D
x1
x2
ones

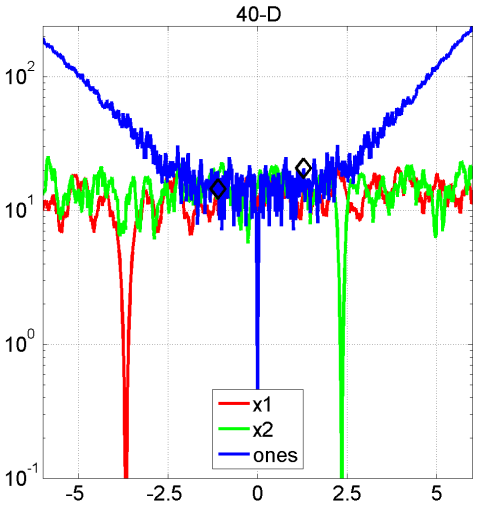

3-D
x1
-x1
x2
-x2
ones
-ones

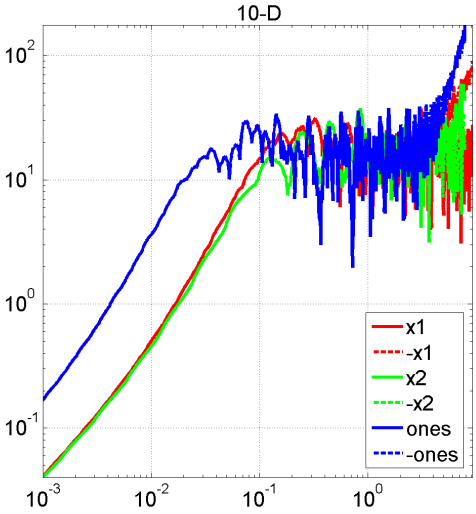

5-D
x1
-x1
x2
-x2
ones
-ones

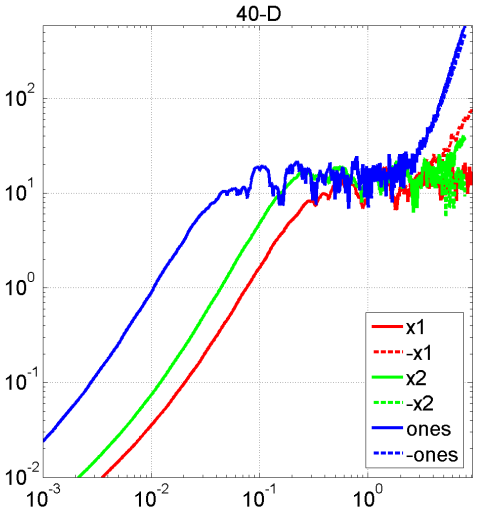

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

5.24 Lunacek bi-Rastrigin Function

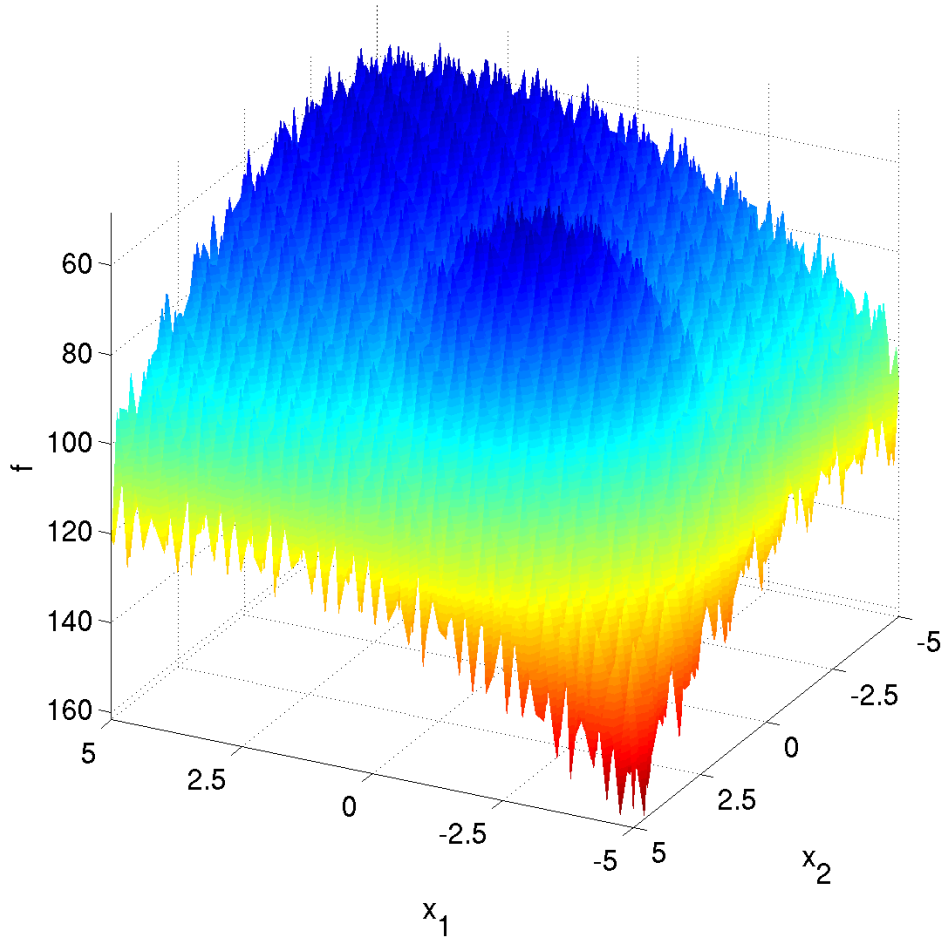
$$f_{24}(\mathbf{x}) = \min\left(\sum_{i=1}^{D}(\hat{x}_i - \mu_0)^2, d\,D + s\sum_{i=1}^{D}(\hat{x}_i - \mu_1)^2\right) + 10\left(D - \sum_{i=1}^{D}\cos(2\pi z_i)\right) + 10^4 f_{\text{pen}}(\mathbf{x}) + f_{\text{opt}} \tag{24}$$

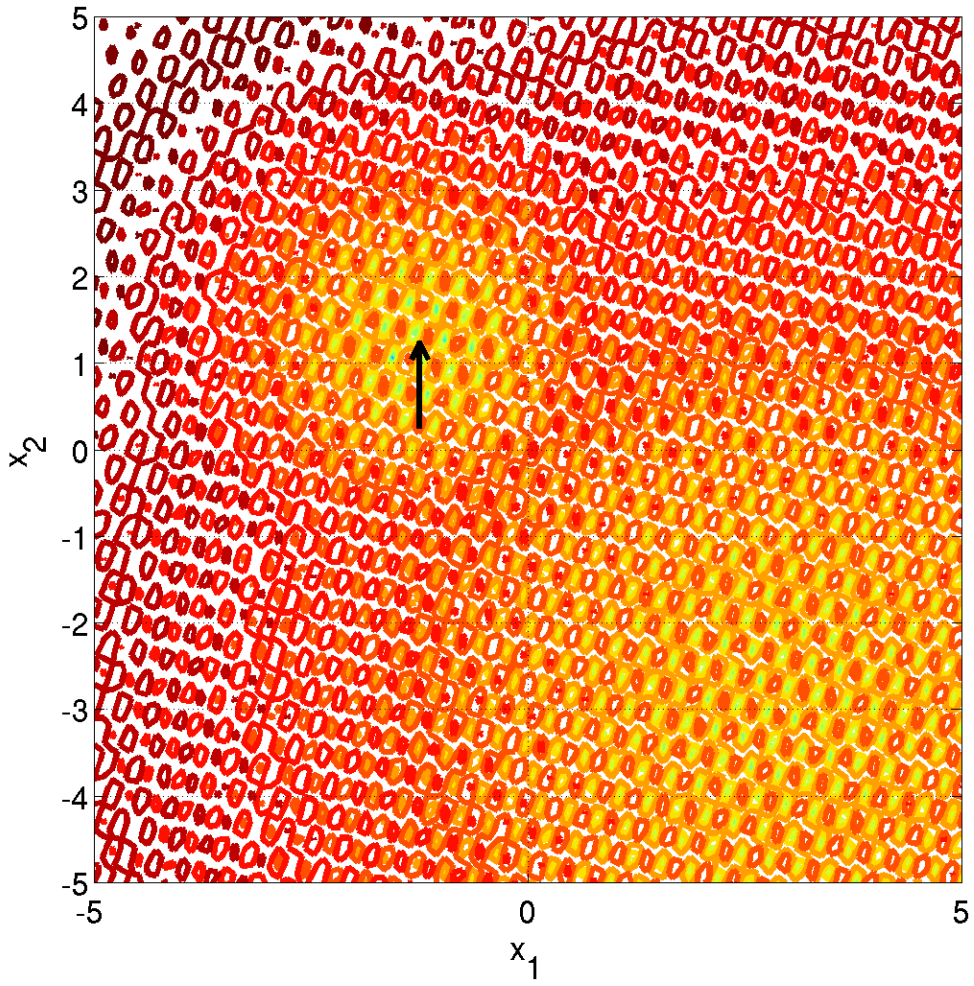
- $\hat{\mathbf{x}} = 2\,\text{sign}(\mathbf{x}^{\text{opt}}) \otimes \mathbf{x}$, $\mathbf{x}^{\text{opt}} = [\mu_0 \longrightarrow] \frac{\mu_0}{2}\mathbf{1}^+_-$
- $\mathbf{z} = \mathbf{Q}\Lambda^{100}\mathbf{R}(\hat{\mathbf{x}} - \mu_0\,\mathbf{1})$
- $\mu_0 = 2.5$, $\mu_1 = -\sqrt{\frac{\mu_0^2 - d}{s}}$, $s = 1 - \frac{1}{2\sqrt{D+20} - 8.2}$, $d = 1$

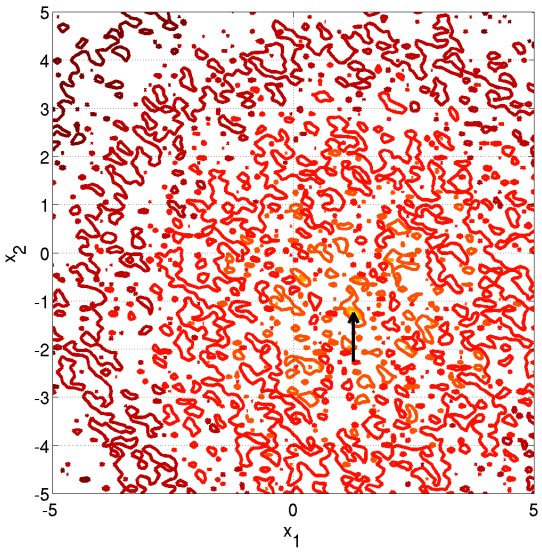
Properties Highly multimodal function based on [7] with two funnels around $[\mu_0\mathbf{1}^+_- \longrightarrow] \frac{\mu_0}{2}\mathbf{1}^+_-$ and $[-\mu_1\mathbf{1}^+_- \longrightarrow] \frac{\mu_1}{2}\mathbf{1}^+_-$ being superimposed by the cosine. Presumably different approaches need to be used for "selecting the funnel" and for search the highly multimodal function "within" the funnel. The function was constructed to be deceptive for evolutionary algorithms with large population size.

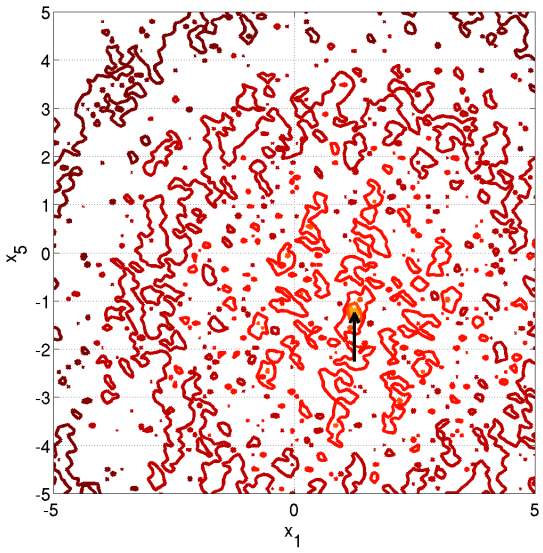
- the funnel of the local optimum at $[-\mu_1\mathbf{1}^+_-] \frac{\mu_1}{2}\mathbf{1}^+_-$ has roughly 70% of the search space volume within $[-5, 5]^D$.

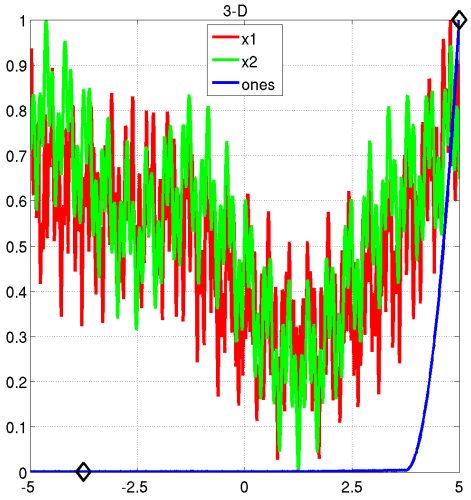
Information gained from this function: Can the search behavior be local on the global scale but global on a local scale?

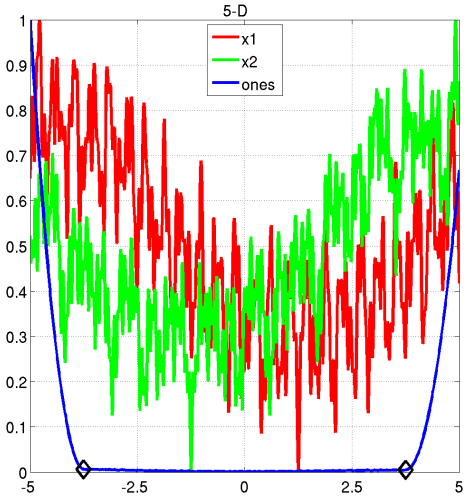


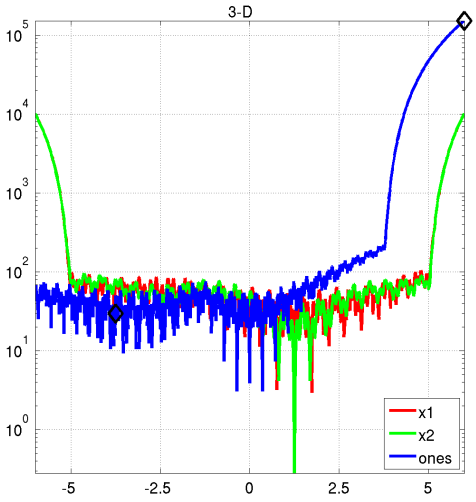

5
4
3
2
1
0
-1
-2
-3
-4
-5
x2
-5
0
5
x1

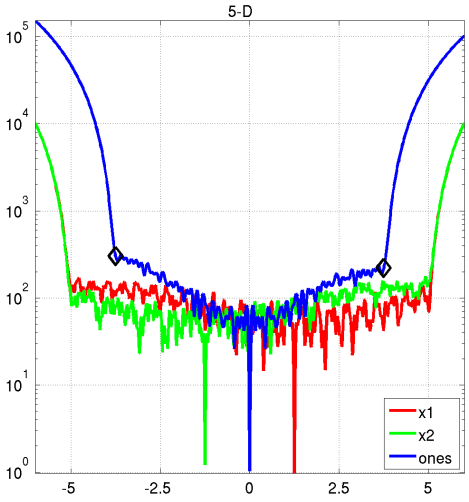

x2
x1

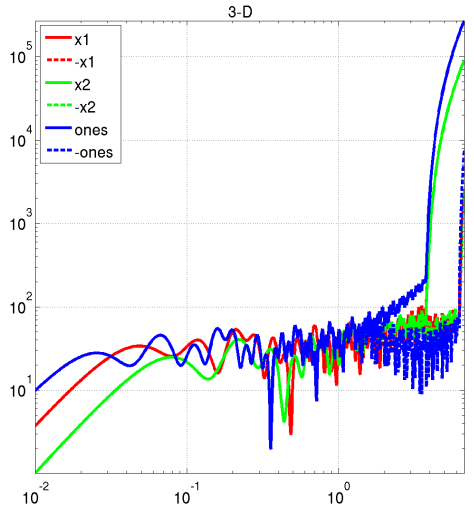

x5
x1

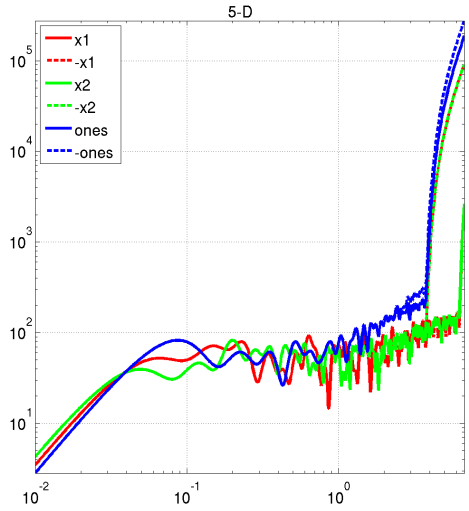

3-D
x1
x2
ones

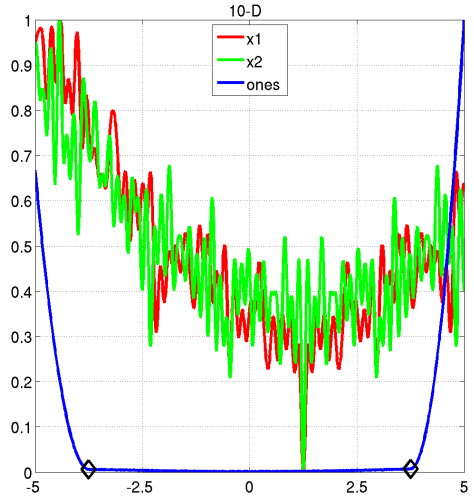

5-D
x1
x2
ones

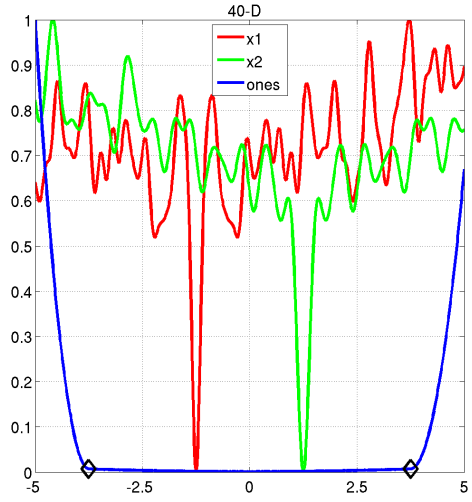

3-D
x1
x2
ones

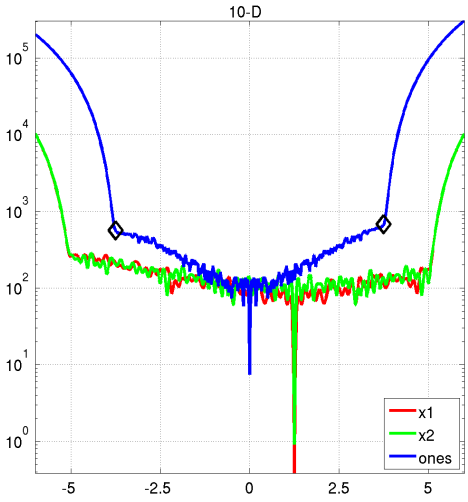

5-D
x1
x2
ones

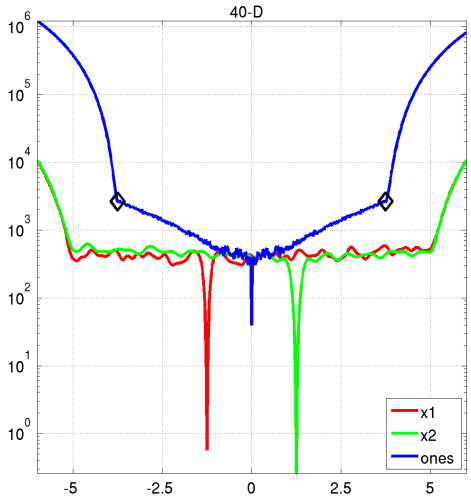

3-D
x1
-x1
x2
-x2
ones
-ones

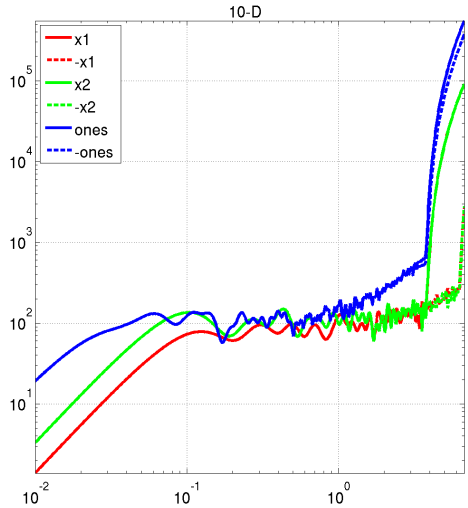

5-D
x1
-x1
x2
-x2
ones
-ones

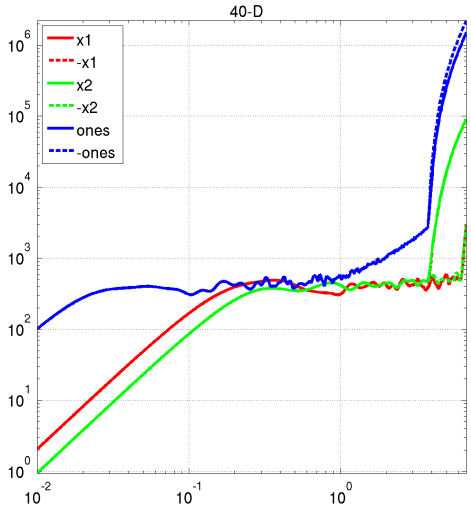

10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
x2
ones


40-D
x1
x2
ones


10-D
x1
-x1
x2
-x2
ones
-ones


40-D
x1
-x1
x2
-x2
ones
-ones

Acknowledgments

The authors would like to thank Arnold Neumaier for his instructive comments, Verena Heidrich-Meisner for her contribution on information gain, and Dimo Brockhoff and Duc Manh Nguyen for relentlessly spotting mistakes. Steffen Finck was supported by the Austrian Science Fund (FWF) under grant P19069-N18.

References

[1] Steffen Finck, Nikolaus Hansen, Raymond Ros, and Auger Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009.

[2] Marcus Gallagher and Bo Yuan. A general-purpose tunable landscape generator. *IEEE transactions on evolutionary computation*, 10(5):590–603, 2006.

[3] Nikolaus Hansen, Anne Auger, Dimo Brockhoff, Dejan Tušar, and Tea Tušar. Coco: performance assessment. *ArXiv e-prints*, arXiv:1605.03560 [cs.NE], 2016.

[4] Nikolaus Hansen, Anne Auger, Steffen Finck, and Raymond Ros. Real-parameter black-box optimization benchmarking 2009: Experimental setup. Technical Report RR-6828, INRIA, 2009.

[5] Nikolaus Hansen, Tea Tusar, Olaf Mersmann, Anne Auger, and Dimo Brockhoff. COCO: the experimental procedure. *ArXiv e-prints*, arXiv:1603.08776 [cs.AI], 2016.

[6] Hidefumi Katsuura. Continuous nowhere-differentiable functions–an application of contraction mappings. *The American Mathematical Monthly*, 98(5):411–416, 1991.

[7] Monte Lunacek, Darrell Whitley, and Andrew Sutton. The impact of global structure on search. In *International Conference on Parallel Problem Solving from Nature*, pages 498–507. Springer, 2008.

[8] H.H. Rosenbrock. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3):175–184, 1960.

[9] Hans-Paul Schwefel. *Numerical optimization of computer models*. John Wiley & Sons, Inc., 1981.

APPENDIX

A Function Properties

A.1 Deceptive Functions

All "deceptive" functions provide, beyond their deceptivity, a "structure" that can be exploited to solve them in a reasonable procedure.

A.2 Ill-Conditioning

Ill-conditioning is a typical challenge in real-parameter optimization and, besides multimodality, probably the most common one. Conditioning of a function can be rigorously formalized in the case of convex quadratic functions, $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{H}\mathbf{x}$ where $\mathbf{H}$ is a symmetric definite positive matrix, as the condition number of the Hessian matrix $\mathbf{H}$. Since contour lines associated to a convex quadratic function are ellipsoids, the condition number corresponds to the square root of the ratio between the largest axis of the ellipsoid and the shortest axis. For more general functions, conditioning loosely refers to the square of the ratio between the largest direction and smallest of a contour line. The testbed contains ill-conditioned functions with a typical conditioning of 10^6. We believe this a realistic requirement, while we have seen practical problems with conditioning as large as 10^{10}.

A.3 Regularity

Functions from simple formulas are often highly regular. We have used a non-linear transformation, T_{osz}, in order to introduce small, smooth but clearly visible irregularities. Furthermore, the testbed contains a few highly irregular functions.

A.4 Separability

In general, separable functions pose an essentially different search problem to solve, because the search process can be reduced to D one-dimensional search procedures. Consequently, non-separable problems must be considered much more difficult and most benchmark functions are designed being non-separable. The typical well-established technique to generate non-separable functions from separable ones is the application of a rotation matrix $\mathcal{R}$.

A.5 Symmetry

Stochastic search procedures often rely on Gaussian distributions to generate new solutions and it has been argued that symmetric benchmark functions could be in favor of these operators. To avoid a bias in favor of highly symmetric operators we have used a symmetry breaking transformation, T_{asy}. We have also included some highly asymmetric functions.

A.6 Target function value to reach

The typical target function value for all functions is $f_{\text{opt}} + 10^{-8}$. On many functions a value of $f_{\text{opt}} + 1$ is not very difficult to reach, but the difficulty versus function value is not uniform for all functions. These properties are not intrinsic, that is $f_{\text{opt}} + 10^{-8}$ is not intrinsically "very good". The value mainly reflects a scalar multiplier in the function definition.