Variable space transformation techniques and new algorithms for global optimization

Department of Computer, Control and Management Engineering "Antonio Ruberti"

Doctor of Philosopy in Automatic Control, Bioengineering and Operations Research – XXXIII Ciclo

Candidate

Francesco Romito
ID number 1198740

Thesis Advisor

Prof. Stefano Lucidi

January 2021

Thesis defended on May 19, 2021
in front of a Board of Examiners composed by:

Prof. Giuseppe Baselli, Politecnico di Milano
Prof. Stefano Panzieri, Università di Roma Tre
Prof. Fabio Tardella, Sapienza Università di Roma (chairman)

Variable space transformation techniques and new algorithms for global optimization

Ph.D. thesis. Sapienza – University of Rome

This thesis has been typeset by LATEX and the Sapthesis class.

Author's email: romito@diag.uniroma1.it

Abstract

Solving a global optimization problem is a hard task.

In the chapters of this thesis variable space transformation techniques and new algorithmic approaches are proposed to deal with such hard problems.

In the first research investigation some variable space transformation techniques are defined as a tool that can be helpfully integrated in (almost) all algorithm frameworks. In particular the focus will be on piecewise linear and non-linear transformations that allow to tackle the problem advantageously. After introducing the theory, preliminary numerical experiments are reported exploiting the transformations in a simple multi-start framework. The idea is to gather the information obtained during a multi-start approach and to apply a sequence of transformations in the variable space that makes the exploration easier. The aim is to expand the attraction basins of global minimizers shrinking those of the local minima already found. Preliminary considerations are made about the possibility to use these transformations as derivative-free preconditioner.

The second research investigation concerns the definition of an efficient algorithm on a wide spectrum of global optimization problems. In particular will be discussed how to do an accurate exploratory geometry and a space search reduction strategy, recently renamed in literature as *zoom-in* strategy, in a probabilistic algorithm that can speed up significantly the convergence towards better solutions. After introducing the algorithm framework named GABRLS, presented as the winner of the Generalization-based Contest in Global Optimization (GENOPT 2017, [61]), the approach is extended to handle also non-continuous variables. The resulting algorithm has been tested in a real case study of *design optimization* of electric motor. The case study provides evidences about the promising exploratory geometry of the approach in quickly finding feasible and optimal solutions of a mixed integer constrained problem.

In the last research investigation, a new black-box approach is proposed to tackle a real case study of the *spare part management problem* of a fleet of aircraft. In particular, for this specific type of inventory problem, a black-box model and a tailored global optimization algorithm is defined. The aim is to address the non-linearity of the problem as is, without any decomposition in sub-problem and without any approximation or necessity to check ex post the feasibility of the solution. The main contribution consists of advancing the existing literature for multi-item inventory systems through an enhanced time-effective optimization algorithm tested in a single-echelon system.

Acknowledgments

Contents

Introduction

Optimization is a basic tool in all areas of applied mathematics, engineering, medicine, economics and other sciences. Technological progress and the expansion of knowledge, constantly bring problems to light which need to be addressed in quantitative terms, solving *Global Optimization* problems [37, 38, 39, 41, 59]. Global optimization is a relatively young field of applied mathematics, studying theory, methods, and implementation of models and strategies for solving multiextremal optimization problems. Global optimization fundamentally dates back to 1975-78, when the two volumes *"Towards Global Optimisation"* [7, 8], by Szego and Dixon, appeared. These volumes are the first books containing a collection of papers that present different solution methods for problems with continuous variables. Through the years there was a wide dissemination of this field in books like *"Handbook of Global Optimization"* [37, 45] by Horst and Pardalos (vol.1), Pardalos and Romeijn (vol.2), and journal like *"Journal of Global Optimization"* [46] fully devoted to the developments in this area.

The scope of global optimization is to approach to problems that requires to find a point x^* and the corresponding value $f(x^*)$ being the global minimizer (or maximizer) of a function $f(x)$ defined over an $N-$dimensional set $\mathcal{X}$.

Since the function $f(x)$ can be hard non-differentiable, multimodal, and given as a "black box" from external routine or simulation software, traditional local optimization methods [34, 47] can not be exploited.

It is well known that the global optimum of a generic optimization problem can be computed only by an algorithm which has an everywhere asymptotic dense convergence to the feasible domain. The algorithms proposed in literature try to ensure such property by exploiting different approaches such as:

- to partition the feasible domain into a growing number of hyper-intervals [1, 2];

- to employ space-filling curves [3, 4, 40];

- to choose points at random according to a suitable distribution [5, 6, 8, 9, 10, 11].

Among these approaches, in the deterministic setting one can mention the so-called direct search methods [49], such as the DIRECT (DIvide RECTangles) [2, 21], the response surface [50], surrogate model methods [51] and pattern search methods [90, 63, 64].

While well-known techniques in the stochastic setting, that are mainly based on random sampling in the feasible set, are adaptive random search [6], simulated annealing [52, 53], evolutionary and genetic algorithms [54, 55, 56], tabu search [57, 58].

In both deterministic and stochastic global optimization algorithms it is possible to identify two separate phases [48]: the global phase and the local phase. The exhaustive exploration of the search space is delegated to the global phase. At each iteration in the global phase, a local optimization procedure is called to refine the current solution, this is known as the local phase, which is usually deterministic. A robust local optimization procedure is essential for a fast convergence. Most of the numerical calculus (solution refining) is performed in the local phase, which turns out to be the most computationally expensive step of the global optimization algorithm. Approaches in this framework are the so called multi-start strategies [8, 12] where in the first phase the solution is generated whereas in the second one the solution is typically (but not necessarily) improved. It aims to iterate local minimization starting from points provided by probabilistic [5] or deterministic methods [12]. It is shown under suitable assumption that local minimizations can be attracted by the global minimum of both gradient related algorithms [16] and derivative-free ones [17, 18].

Other approaches to global optimization problems are the filled function techniques [13, 14, 15]. In the filled function techniques the objective function is iteratively perturbed in order to avoid that local optimization methods getting stuck in local minima.

The aim of this thesis is precisely concerned with research in global optimization to which it intends to contribute. It is composed by three chapters. The first chapter focuses on exploiting variables space transformations as tool for global optimization algorithms. The second chapter presents an algorithm for the solution of general global optimization problems. The last one describes a tailored algorithm approach for a specific case study of spare part management problem.

Chapter 1

Exploiting variables space transformations in global optimization

1.1 Introduction

In this chapter a new strategy that can be useful to tackling hard global optimization problems is proposed. The rationale behind the approach is to gain an advantage by an iterative perturbation of the original problem. In particular it performs a suitable continuous transformation of the variable space without directly modifying the objective function.

Hard global optimization problems arise when the regions of attraction of local optimal solutions are stronger than the global ones. This situation, in which algorithms may fails, often occurs in one of the following circumstances:

   i) global minimum surrounded by a cluster of many local minima;

   ii) global minimum placed in steeper valley than those of local minimizer.

The idea for problems showing drawback i) is to perform a space dilation of the region containing the cluster to avoid that the global minimum is hidden by the attraction regions of the local minima meanwhile the rest of the space shrinks due to the continuity of the transformation.

For problems that shows drawback ii) the opposite action can be performed. Namely, for such problems a contraction of the attraction regions of local optimal solutions imply the expansion of the attraction region of the global minimum.

The integration of a dynamic strategy of expansion-contraction of the space within existent algorithmic schemes could lead to obtain a better exploration of the search space and hence hopefully to a speed up of the global search.

In the following sections two continuous bijective transformations are proposed. After some preliminary definitions that will be used throughout the chapter, a simple piecewise linear transformation (PLT) is described and its properties are outlined. Then a more complex non-linear transformation (NLT) is proposed. Finally, in order to have a feel of the possible interest of the proposed transformations, they are embedded in a simple multi-start algorithm.

1.2   Problem definition and Notation

Consider the following problem

$$\min_x f(x)$$
$$x \in \mathcal{X} \,, \tag{1.1}$$

where

- $\mathcal{X} = (x \in \mathbb{R}^n \ : \ l \leq \ x \ \leq u) \quad \text{with} \quad l, u \in \mathbb{R}^n;$

- f is a continuous function.

The aim is to find the global solution x^*, namely a point such that

$$f(x^*) \leq f(x), \ \forall x \in \mathcal{X}.$$

Consider n-dimensional Euclidean space $\mathbb{R}^n$ and denote the $l_2 - norm$ (Euclidean distance) as

$$\|x\|_2 = \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}},$$

and the $l_\infty - norm$ (Chebyshev distance) as

$$\|x\|_\infty = \max_i |x_i|, \ i = 1, \ldots, n.$$

Definition of hypercubes and hyperspheres

   We introduce a notation for the shape of the neighborhood of radius r of a generic solution $\hat{x}$. In particular, exploiting the $l_\infty - norm$, we define hyper-cubical neighborhoods

- $\mathcal{B}_\infty(\hat{x}, r) = (x \in \mathbb{R}^n, \ r > 0 \ : \|x - \hat{x}\|_\infty \leq r);$

and its boundary

- $\partial \mathcal{B}_\infty(\hat{x}, r) = (x \in \mathbb{R}^n, \ r > 0 \ : \|x - \hat{x}\|_\infty = r);$

and hyper-spherical neighborhoods by using Euclidean norm distance

- $\mathcal{B}_2(\hat{x}, r) = (x \in \mathbb{R}^n, \ r > 0 \ : \ \|x - \hat{x}\|_2 \leq r);$

and its boundary

- $\partial \mathcal{B}_2(\hat{x}, r) = (x \in \mathbb{R}^n, \ r > 0 \ : \|x - \hat{x}\|_2 = r).$

Definition of linear intervals

We also refer to linear interval as

- $I(\hat{x}, r) = \{ \ x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| \leq r, \ i = 1, \ldots, n \ \}.$

and its boundary

- $\partial I(\hat{x}, r) = \{ \ x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| = r, \ i = 1, \ldots, n \ \}.$

Definition of hyperrectagles and hyperellipses

Consider a square diagonal matrix with positive entries $D \in \mathbb{R}^{n \times n}$

$$D = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \end{pmatrix},$$

We define an hyper-rectangular neighborhood ($l_\infty - norm$ ellipsoidal neighborhood)

- $\mathcal{E}_\infty(\hat{x}, D, r) = (x \in \mathbb{R}^n, \ r > 0 \ : \ \|D(x - \hat{x})\|_\infty \leq r);$

and its boundary

- $\partial\mathcal{E}_\infty(\hat{x}, D, r) = (x \in \mathbb{R}^n, \ r > 0 \ : \ \|D(x - \hat{x})\|_\infty = r);$

and hyper-ellipsoidal neighborhood by using Euclidean norm distance

- $\mathcal{E}_2(\hat{x}, D, r) = (x \in \mathbb{R}^n, \ r > 0 \ : \ \|D(x - \hat{x})\|_2 \leq r);$

and its boundary

- $\partial\mathcal{E}_2(\hat{x}, D, r) = (x \in \mathbb{R}^n, \ r > 0 \ : \ \|D(x - \hat{x})\|_2 = r).$

Definition of direct and inverse transformation

We refer to

$$y = T_{\hat{x}}(x) \quad \text{or shortly} \quad y = \bar{y}(x) \;,$$

as the direct continuous bijective transformation from the original space $\mathcal{X}$ to a transformed space $\mathcal{Y} = (y \in \mathbb{R}^n \; : \; l \leq \; y \; \leq u)$ with $l, u \in \mathbb{R}^n$. The point $\hat{x}$ is the centroid of the transformation.

Conversely we refer to

$$x = T_{\hat{x}}^{-1}(y) \quad \text{or shortly} \quad x = \bar{x}(y)$$

as the inverse continuous bijective transformation from a space $\mathcal{Y}$ to the original space $\mathcal{X}$.

Definition of original and transformed problem

Given the inverse transformation $T_{\hat{x}}^{-1}(y)$, the problem (1.1)

$$\min_x f(x)$$
$$x \in \mathcal{X},$$

is transformed in

$$\min_y f(T_{\hat{x}}^{-1}(y)) \qquad\qquad (1.2)$$
$$y \in \mathcal{Y},$$

where $f(T_{\hat{x}}^{-1}(y))$ has suitable properties.

We point out that we never need to double-transform the variable space during the search of solution. In fact we only deal with problem (1.2), where we make the search for solutions in a transformed variable space $\mathcal{Y}$ and by using the inverse transformation, we always evaluate the objective function in the original variable space $\mathcal{X}$. By the way, as long as the direct and inverse equations we are going to define are both continuous and bijective (one-to-one and onto), one can switch between them.

1.3 Piecewise Linear Transformation (PLT)

In this section it is introduced a piecewise linear transformation $y = T_{\hat{x}}(x)$ between two variable spaces: the original space $\mathcal{X}$ and a transformed one $\mathcal{Y}$.

It is componentwise, and it is piecewise along a single component. In figure (1.1) the transformation is illustrated before going into the details of the definition of the transformation equations.

For the sake of simplicity consider a reference point $\hat{y} \in \mathcal{Y}$ as the origin of the Cartesian axes. This point is set as centroid of the hypercube $\mathcal{B}_\infty(\hat{y}, r_y)$ in which we concentrate the transformation impact range. The point $\hat{y}$ is also set as centroid of the smaller hypercube $\mathcal{B}_\infty(\hat{y}, \epsilon_y)$ in which the transformation impact has the main focus. The centroid $\hat{y}$ is a fixed point of the transformation, so that the equality $\hat{y} = \hat{x}$ holds, where $\hat{x}$ is the centroid of the corresponding hypercube $\mathcal{B}_\infty(\hat{x}, r_x)$ in the $\mathcal{X}$ space.

For each dimension $i = 1, 2, ..., n$ the piecewise linear transformation (blue line in fig. (1.1)) propagates outward from the centroid in the relative space.

The larger the angle between the bisector (green line in fig. (1.1)) and the piecewise linear transformation, the greater the effect of the transformation.

Parameters

- $\theta_1 = \tan(\alpha) > 0$;
- $\theta_2 = \tan(\beta) > 0$;
- $\theta_3 = \tan(\gamma) = 1$;
- $r_x \geq \epsilon_x \geq 0$;
- $r_y \geq \epsilon_y \geq 0$.

Vector Parameters

- $l_i \leq \hat{x}_i - r_x$;
- $\hat{x}_i + r_x \leq u_i$;
- $l_i \leq \hat{y}_i - r_y$;
- $\hat{y}_i + r_y \leq u_i$.
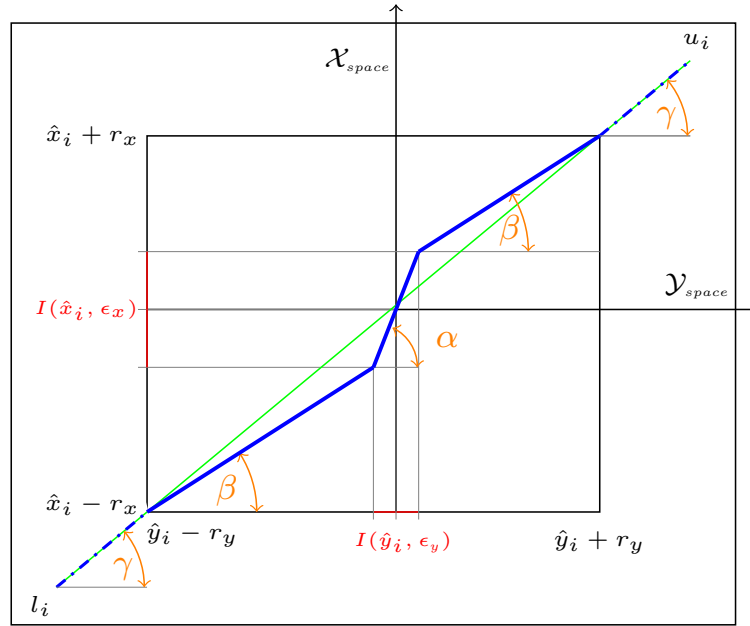
Figure 1.1. Graph of the piecewise linear transformation in the ith dimension.

In the next subsections the equations that described the piecewise linear transformation are formally defined.

1.3.1   Definition of the PLT over hypercubes

Let $\hat{x} \in \mathbb{R}^n$, ϵ_x, r_x, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_x \leq r_x$.

The following equations define the piecewise linear transformation $y = T_{\hat{x}}(x)$ over hyper-cubical space.

- For $x \in \mathbb{R}^n$: $|x_i - \hat{x}_i| \leq \epsilon_x$, $i = 1, \dots, n$,

$$\bar{y}\left(x\right)_i = \hat{y}_i + \theta_1 \left(x_i - \hat{x}_i\right). \tag{1.3}$$

- For $x \in \mathbb{R}^n$: $\epsilon_x \leq |x_i - \hat{x}_i| \leq r_x$, $i = 1, \dots, n$,

$$\bar{y}\left(x\right)_i = \hat{y}_i + \frac{\left(x_i - \hat{x}_i\right)}{|x_i - \hat{x}_i|} \left(\theta_1 \epsilon_x + \theta_2(|x_i - \hat{x}_i| - \epsilon_x)\right). \tag{1.4}$$

- For $x \in \mathbb{R}^n$: $|x_i - \hat{x}_i| \geq r_x$, $i = 1, \dots, n$,

$$\bar{y}\left(x\right)_i = \hat{y}_i + \frac{\left(x_i - \hat{x}_i\right)}{|x_i - \hat{x}_i|} \left(\theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x) + \theta_3 \left(|x_i - \hat{x}_i| - r_x\right)\right). \tag{1.5}$$

1.3.2 Properties of the PLT - hypercube

Here below are listed the properties and the necessary coupling condition to preserve the continuity of the transformation between the original space $\mathcal{X}$ and the transformed space $\mathcal{Y}$.

i) The reference point $\hat{x} \in \mathcal{X}$ is transformed in the corresponding one $\hat{y}$ in the transformed space $\mathcal{Y}$:

$$\hat{y} = T_{\hat{x}}(\hat{x}).$$

ii) The region $\{\, x \in \mathbb{R}^n \; : \; |x_i - \hat{x}_i| \le \epsilon_x, \; i = 1, \ldots, n \,\}$ is transformed in

$$\{\, y \in \mathbb{R}^n \; : \; |y_i - \hat{y}_i| \le \epsilon_y, \; i = 1, \ldots, n \,\},$$

where (1.3)

$$\bar{y}\,(x)_i - \hat{y}_i = \theta_1\,(x_i - \hat{x}_i)\,,$$

imposes at the boundary the following coupling condition

$$\epsilon_y = \theta_1 \epsilon_x. \tag{1.6}$$

iii) The region $\{\, x \in \mathbb{R}^n \; : \epsilon_x \le \; |x_i - \hat{x}_i| \le \; r_x, \; i = 1, \ldots, n \,\}$ is transformed in

$$\{\, y \in \mathbb{R}^n \; : \epsilon_y \le \; |y_i - \hat{y}_i| \le r_y, \; i = 1, \ldots, n \,\},$$

where (1.4)

$$\bar{y}\,(x)_i - \hat{y}_i = \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|}\,(\theta_1 \epsilon_x + \theta_2(|x_i - \hat{x}_i| - \epsilon_x))\,,$$

and (1.6) impose at the boundary the following coupling condition

$$r_y = \epsilon_y + \theta_2(r_x - \epsilon_x). \tag{1.7}$$

iv) The region $\{\, x \in \mathbb{R}^n \;:\; |x_i - \hat{x}_i| \geq\; r_x,\; i = 1, \ldots, n \,\}$ is transformed in

$$\{\, y \in \mathbb{R}^n \;:\; |y_i - \hat{y}_i| \geq r_y,\; i = 1, \ldots, n \,\},$$

where from (1.5)

$$\bar{y}\,(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|}\,(\theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x) + \theta_3\,(|x_i - \hat{x}_i| - r_x))\,,$$

substituting (1.6) and (1.7)

$$\bar{y}\,(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|}\,(r_y + \theta_3(|x_i - \hat{x}_i| - r_x))\,. \qquad (1.8)$$

As long as the (1.8) refers to the outermost region there is no need to further coupling condition. It means that θ_3 is free of choice.

1.3.3 Special case of PLT - hypercube

In this section a particular example of the transformation is considered. The focus is on a transformation that can be suitable in the context of global optimization. The idea is to get values of the parameters $\theta_1, \theta_2, \theta_3$ such that the transformation performs an expansion of the neighborhood of the global minimum. This task can be obtained either by shrinking or expanding the neighborhoods of the already found local minima. Therefore in this case it can be set:

- $\hat{y} = \hat{x}$.

The possibility of expanding or contracting the neighborhood of $\hat{x}$ follows from suitable choices of $\epsilon_x, \ \epsilon_y$. It implies from (1.6) that

- $\theta_1 = \frac{\epsilon_y}{\epsilon_x} \neq 1$.

Moreover for ensuring that the transformation impact has the main effect in a local area, inside the radius $r_x, \ r_y$ it has to be set

- $r_y = r_x = r$.

It implies from (1.7) that

- $\theta_2 = \frac{r - \epsilon_y}{r - \epsilon_x}$.

Finally in (1.8) by imposing that

- $\theta_3 = 1$;

for $x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| \geq r, \ i = 1, \ldots, n, \ $ it has

$$\bar{y}(x)_i = x_i.$$

From this choices, the equation (1.3), for $x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| \leq \epsilon_x, \ i = 1, \ldots, n,$

$$\bar{y}(x)_i = \hat{y}_i + \theta_1 (x_i - \hat{x}_i),$$

becomes

$$\bar{y}(x)_i = \hat{y}_i + \frac{\epsilon_y}{\epsilon_x} (x_i - \hat{x}_i).$$

The equation (1.4), for $x \in \mathbb{R}^n \ : \ \epsilon_x \leq |x_i - \hat{x}_i| \leq r, \ i = 1, \ldots, n,$

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|}\left(\theta_1 \epsilon_x + \theta_2(|x_i - \hat{x}_i| - \epsilon_x)\right),$$

becomes

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|}\left(\epsilon_y + \frac{r - \epsilon_y}{r - \epsilon_x}(|x_i - \hat{x}_i| - \epsilon_x)\right).$$

The equation (1.5), for $x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| \geq r, \ i = 1, \ldots, n,$

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|}\left(\theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x) + \theta_3(|x_i - \hat{x}_i| - r_x)\right),$$

becomes

$$\bar{y}(x)_i = x_i.$$

In summary the piecewise linear transformation in this special case:

- For $x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| \leq \epsilon_x, \ i = 1, \ldots, n,$

$$\bar{y}(x)_i = \hat{y}_i + \frac{\epsilon_y}{\epsilon_x}(x_i - \hat{x}_i). \tag{1.9}$$

- For $x \in \mathbb{R}^n \ : \ \epsilon_x \leq |x_i - \hat{x}_i| \leq r, \ i = 1, \ldots, n,$

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|}\left(\epsilon_y + \frac{r - \epsilon_y}{r - \epsilon_x}(|x_i - \hat{x}_i| - \epsilon_x)\right). \tag{1.10}$$

- For $x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| \geq r, \ i = 1, \ldots, n,$

$$\bar{y}(x)_i = x_i. \tag{1.11}$$

1.3.4   Definition of the inverse PLT over hypercubes

Let $\hat{y} \in \mathbb{R}^n$, ϵ_y, r_y, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_y \leq r_y$.

The following equations define the inverse piecewise linear transformation.

- For $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \leq \epsilon_y$, $i = 1, \ldots, n$,

$$\bar{x}\,(y)_i = \hat{x}_i + \frac{1}{\theta_1}\,(y_i - \hat{y}_i)\,. \tag{1.12}$$

- For $y \in \mathbb{R}^n$: $\epsilon_y \leq |y_i - \hat{y}_i| \leq r_y$, $i = 1, \ldots, n$,

$$\bar{x}\,(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(|y_i - \hat{y}_i| - \epsilon_y)\right)\,. \tag{1.13}$$

- For $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \geq r_y$, $i = 1, \ldots, n$,

$$\bar{x}\,(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(r_y - \epsilon_y) + \frac{1}{\theta_3}(|y_i - \hat{y}_i| - r_y)\right)\,. \tag{1.14}$$

To have an easier reading and better clarity the properties and the special cases are in the appendix A in sections A.1 and A.1.2.

1.3.5   Representation of the PLT over hyper-cubical space

For a quick understanding a 2D representation of the transformation is drawn. Let us consider the multimodal continuous function Ackley :

$$f(x_1 \cdots x_n) = -20 exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - exp\left(\frac{1}{n}\sum_{i=1}^{n} cos(2\pi x_i)\right) + 20 + exp(1),$$

over

$$\mathcal{X} = (x \in \mathbb{R}^n \; : \; -3 \leq \; x_i \; \leq 3 \, , \; i = 1, \ldots, n \;),$$

with minimum at

$$f(0, \cdots, 0) = 0.$$

Let us consider a transformed variables space

$$\mathcal{Y} = (y \in \mathbb{R}^n \; : \; -3 \leq \; y_i \; \leq 3 \, , \; i = 1, \ldots, n \;).$$

Expansion   Consider to make an expansion of a hyper-cubical neighborhood $\mathcal{B}_\infty(\hat{x}, \epsilon_x)$ of the point $\hat{x} = \hat{y} = (0,0)$ with radius $\epsilon_x = 0.9$ bringing it to a radius of $\epsilon_y = 1.125$. The transformation boundary is $r_x = r_y = 1.8$. For the continuity of the transformation, the region in between the two hypercubes is shrunk. Summarizing, the transformation parameters and the other resulting quantities are the following:

- $\epsilon_x = 0.9$;
- $r_x = 1.8$;
- $\theta_1 = 0.8$;
- $\theta_3 = 1$;

- $\epsilon_y = 1.125$;
- $r_y = 1.8$;
- $\theta_2 = 1.33$;
- $\hat{x} = \hat{y} = (0,0)$.

In figure (1.2) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the transformation in the $\mathcal{Y}$ space.
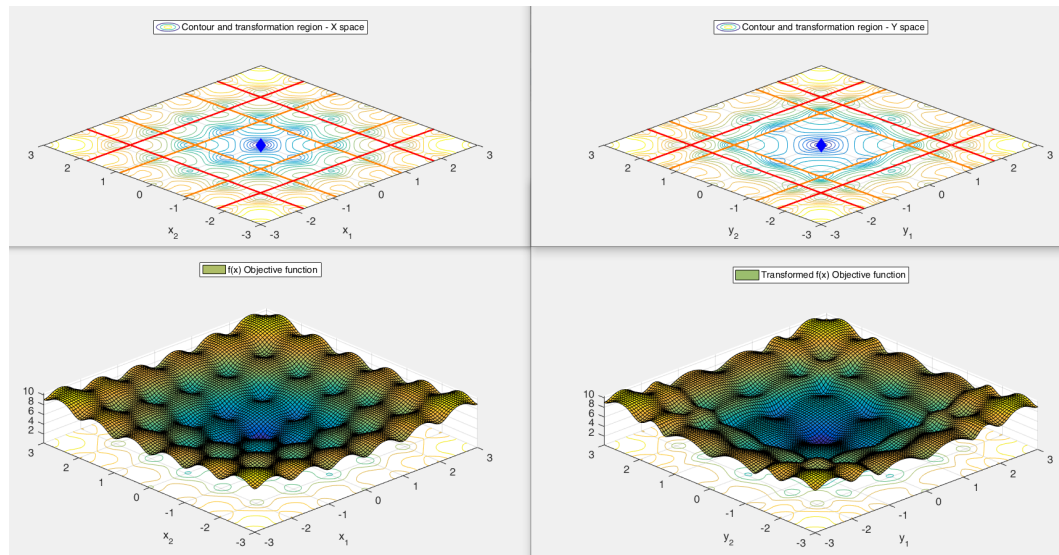
Figure 1.2. A hyper-cubical neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hypercubes. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

Contraction Now consider make a contraction of a hyper-cubical neighborhood of the point $\hat{x} = \hat{y} = (0,0)$ with radius $\epsilon_x = 1.44$ bringing it to a radius of $\epsilon_y = 1.125$. The transformation boundary is $r_x = r_y = 1.8$ For the continuity of the transformation, the space in between the two hypercube is expanded. Summarizing, the transformation parameters and the other resulting quantities are the following:

- $\epsilon_x = 1.44;$
- $\epsilon_y = 1.125;$

- $r_x = 1.8,$
- $r_y = 1.8;$

- $\theta_1 = 1.28;$
- $\theta_2 = 0.533;$

- $\theta_3 = 1;$
- $\hat{x} = \hat{y} = (0,0).$

In figure (1.3) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the transformation in the $\mathcal{Y}$ space.
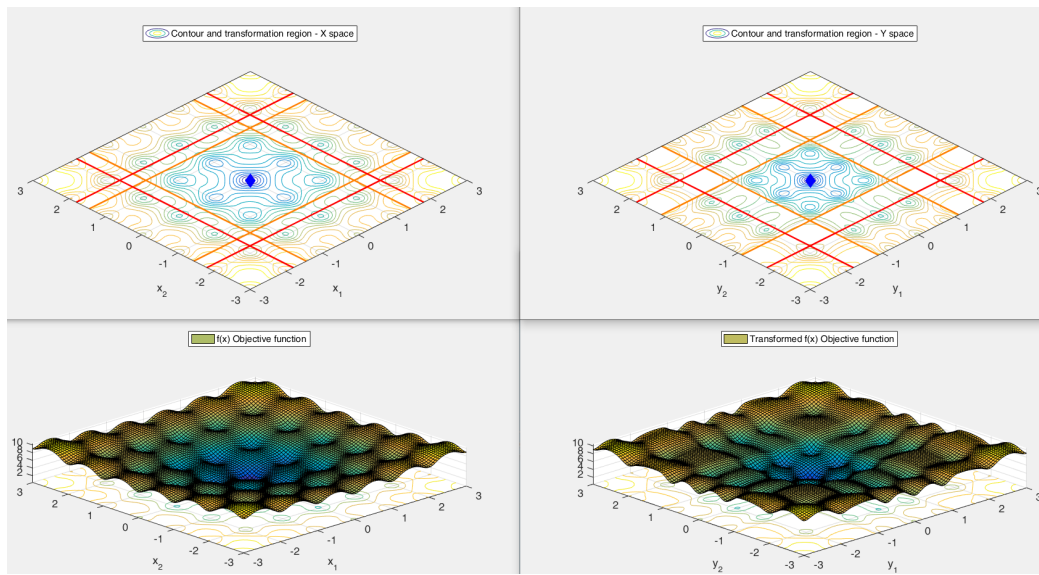
Figure 1.3. A hyper- neighborhood of $\hat{x}$ has been shrunk while expanding the one in between the two hypercubes. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

1.3.6   Definition of the PLT over hyperrectangles

The basic equations that describe the transformation over hypercubes can be suited for general hyper-rectangular space. Let us introduce two square diagonal matrices with positive entries:

$$
D_x = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \end{pmatrix}, \qquad D_y = \begin{pmatrix} \nu_1 & 0 & \dots & 0 \\ 0 & \nu_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \nu_n \end{pmatrix}.
$$

Let $\hat{x} \in \mathbb{R}^n$, ϵ_x, r_x, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_x \leq r_x$. The following equations define the inverse piecewise linear transformation over hyper-rectangles.

- For $x \in \mathbb{R}^n$: $|x_i - \hat{x}_i| \leq \frac{\epsilon_x}{\sigma_i}$, $i = 1, \dots, n$,

$$
\bar{y}(x)_i = \hat{y}_i + \theta_1 \left(\frac{\sigma_i}{\nu_i} \right) (x_i - \hat{x}_i). \tag{1.15}
$$

- For $x \in \mathbb{R}^n$: $\frac{\epsilon_x}{\sigma_i} \leq |x_i - \hat{x}_i| \leq \frac{r_x}{\sigma_i}$, $i = 1, \dots, n$,

$$
\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\frac{\sigma_i}{\nu_i} \right) \left(\theta_1 \left(\frac{\epsilon_x}{\sigma_i} \right) + \theta_2 \left(|x_i - \hat{x}_i| - \frac{\epsilon_x}{\sigma_i} \right) \right). \tag{1.16}
$$

- For $x \in \mathbb{R}^n$: $|x_i - \hat{x}_i| \geq \frac{r_x}{\sigma_i}$, $i = 1, \dots, n$,

$$
\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\frac{\sigma_i}{\nu_i} \right) \left(\theta_1 \left(\frac{\epsilon_x}{\sigma_i} \right) + \theta_2 \left(\frac{r_x - \epsilon_x}{\sigma_i} \right) + \theta_3 \left(|x_i - \hat{x}_i| - \frac{r_x}{\sigma_i} \right) \right). \tag{1.17}
$$

1.3.7 Properties of the PLT - hyperrectangle

Here below are listed the properties and the necessary coupling condition to preserve the continuity of the transformation between the original space $\mathcal{X}$ and the transformed space $\mathcal{Y}$.

i) The reference point $\hat{x} \in \mathcal{X}$ is transformed in the corresponding one $\hat{y}$ in the transformed space $\mathcal{Y}$:

$$\hat{y} = T_{\hat{x}}(\hat{x}).$$

ii) The region $\left\{ x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| \leq \dfrac{\epsilon_x}{\sigma_i}, \ i = 1, \ldots, n \right\}$ is transformed in

$$\left\{ y \in \mathbb{R}^n \ : \ |y_i - \hat{y}_i| \leq \frac{\epsilon_y}{\nu_i}, \ i = 1, \ldots, n \right\},$$

where (1.15)

$$\bar{y}\left(x\right)_i = \hat{y}_i + \theta_1 \left(\frac{\sigma_i}{\nu_i}\right)(x_i - \hat{x}_i),$$

imposes at the boundary the following coupling condition

$$\epsilon_y = \theta_1 \epsilon_x. \tag{1.18}$$

iii) The region $\left\{ x \in \mathbb{R}^n \ : \frac{\epsilon_x}{\sigma_i} \leq |x_i - \hat{x}_i| \leq \frac{r_x}{\sigma_i}, \ i = 1, \ldots, n \right\}$ is transformed in

$$\left\{ y \in \mathbb{R}^n \ : \frac{\epsilon_y}{\nu_i} \leq |y_i - \hat{y}_i| \leq \frac{r_y}{\nu_i}, \ i = 1, \ldots, n \right\},$$

where (1.16)

$$\bar{y}\left(x\right)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\frac{\sigma_i}{\nu_i}\right) \left(\theta_1 \left(\frac{\epsilon_x}{\sigma_i}\right) + \theta_2 \left(|x_i - \hat{x}_i| - \frac{\epsilon_x}{\sigma_i}\right)\right),$$

and (1.18) impose at the boundary the following coupling condition

$$r_y = \epsilon_y + \theta_2(r_x - \epsilon_x). \tag{1.19}$$

iv) The region $\left\{ x \in \mathbb{R}^n \; : \; |x_i - \hat{x}_i| \geq \dfrac{r_x}{\sigma_i} \; , \; i = 1, \ldots, n \right\}$ is transformed in

$$\left\{ y \in \mathbb{R}^n \; : \; |y_i - \hat{y}_i| \geq \dfrac{r_y}{\nu_i} \; , \; i = 1, \ldots, n \right\},$$

where from (1.17)

$$\bar{y}\left(x\right)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\frac{\sigma_i}{\nu_i}\right) \left(\theta_1 \left(\frac{\epsilon_x}{\sigma_i}\right) + \theta_2 \left(\frac{r_x - \epsilon_x}{\sigma_i}\right) + \theta_3 \left(|x_i - \hat{x}_i| - \frac{r_x}{\sigma_i}\right)\right),$$

substituting (1.18) and (1.19)

$$\bar{y}\left(x\right)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\frac{\sigma_i}{\nu_i}\right) \left(\frac{r_y}{\sigma_i} + \theta_3 \left(|x_i - \hat{x}_i| + \frac{r_x}{\sigma_i}\right)\right). \tag{1.20}$$

1.3.8   Special case of PLT - hyperrectangle

As well as in the transformation over hypercubes, this section consider a particular example of the transformation over hyperrectangles. Therefore in this case it can be set:

- $\hat{y} = \hat{x}.$

The hyperrectangle dimensions in the original space $\mathcal{X}$ and in the transformed space $\mathcal{Y}$ are described by

- $D_y = D_x$ and so $\nu_i = \sigma_i$ for $i = 1, \ldots, n;$

The possibility of expanding or contracting the neighborhood of $\hat{x}$ follows from suitable choices of ϵ_x, ϵ_y. It implies from (1.18) that

- $\theta_1 = \frac{\epsilon_y}{\epsilon_x} \neq 1.$

Moreover for ensuring that the transformation impact has the main effect in a local area, inside the radius r_x, r_y it has to be set

- $r_y = r_x = r.$

It implies from (1.19) that

- $\theta_2 = \frac{r - \epsilon_y}{r - \epsilon_x}.$

Finally in (1.20) by imposing that

- $\theta_3 = 1;$

for $x \in \mathbb{R}^n : |x_i - \hat{x}_i| \geq r, \ i = 1, \ldots, n$, it has

$$\bar{y}(x)_i = x_i.$$

From this choices, the equation (1.15), for $x \in \mathbb{R}^n : |x_i - \hat{x}_i| \leq \frac{\epsilon_x}{\sigma_i}, \ i = 1, \ldots, n,$

$$\bar{y}(x)_i = \hat{y}_i + \theta_1 \left(\frac{\sigma_i}{\nu_i}\right)(x_i - \hat{x}_i),$$

becomes

$$\bar{y}(x)_i = \hat{y}_i + \frac{\epsilon_y}{\epsilon_x}(x_i - \hat{x}_i).$$

The equation (1.16), for $x \in \mathbb{R}^n$: $\frac{\epsilon_x}{\sigma_i} \leq |x_i - \hat{x}_i| \leq \frac{r}{\sigma_i}$, $i = 1, \ldots, n$,

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\frac{\sigma_i}{\nu_i}\right) \left(\theta_1 \left(\frac{\epsilon_x}{\sigma_i}\right) + \theta_2 \left(|x_i - \hat{x}_i| - \frac{\epsilon_x}{\sigma_i}\right)\right),$$

becomes

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\left(\frac{\epsilon_y}{\sigma_i}\right) + \frac{r - \epsilon_y}{r - \epsilon_x}\left(|x_i - \hat{x}_i| - \frac{\epsilon_x}{\sigma_i}\right)\right).$$

The equation (1.17), for $x \in \mathbb{R}^n$: $|x_i - \hat{x}_i| \geq r$, $i = 1, \ldots, n$,

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\frac{\sigma_i}{\nu_i}\right) \left(\theta_1 \left(\frac{\epsilon_x}{\sigma_i}\right) + \theta_2 \left(\frac{r_x - \epsilon_x}{\sigma_i}\right) + \theta_3 \left(|x_i - \hat{x}_i| - \frac{r_x}{\sigma_i}\right)\right),$$

becomes

$$\bar{y}(x)_i = x_i.$$

In summary the piecewise linear transformation in this special case:

- For $x \in \mathbb{R}^n$: $|x_i - \hat{x}_i| \leq \frac{\epsilon_x}{\sigma_i}$, $i = 1, \ldots, n$,

$$\bar{y}(x)_i = \hat{y}_i + \frac{\epsilon_y}{\epsilon_x}(x_i - \hat{x}_i). \tag{1.21}$$

- For $x \in \mathbb{R}^n$: $\frac{\epsilon_x}{\sigma_i} \leq |x_i - \hat{x}_i| \leq \frac{r}{\sigma_i}$, $i = 1, \ldots, n$,

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} \left(\left(\frac{\epsilon_y}{\sigma_i}\right) + \frac{r - \epsilon_y}{r - \epsilon_x}\left(|x_i - \hat{x}_i| - \frac{\epsilon_x}{\sigma_i}\right)\right). \tag{1.22}$$

- For $x \in \mathbb{R}^n$: $|x_i - \hat{x}_i| \geq \frac{r}{\sigma_i}$, $i = 1, \ldots, n$,

$$\bar{y}(x)_i = x_i. \tag{1.23}$$

1.3.9 Definition of the inverse PLT over hyperrectangles

Let us introduce two square diagonal matrices with positive entries:

$$
D_y = \begin{pmatrix} \nu_1 & 0 & \dots & 0 \\ 0 & \nu_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \nu_n \end{pmatrix}, \qquad D_x = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \end{pmatrix}.
$$

Let $\hat{y} \in \mathbb{R}^n$, ϵ_y, r_y, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_y \leq r_y$. The following equations define the inverse piecewise linear transformation over hyperrectangles.

- For $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \leq \frac{\epsilon_y}{\nu_i}$, $i = 1, \dots, n$,

$$
\bar{x}(y)_i = \hat{x}_i + \frac{1}{\theta_1} \left(\frac{\nu_i}{\sigma_i} \right) (y_i - \hat{y}_i). \tag{1.24}
$$

- For $y \in \mathbb{R}^n$: $\frac{\epsilon_y}{\nu_i} \leq |y_i - \hat{y}_i| \leq \frac{r_y}{\nu_i}$, $i = 1, \dots, n$,

$$
\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|} \left(\frac{\nu_i}{\sigma_i} \right) \left(\frac{1}{\theta_1} \left(\frac{\epsilon_y}{\nu_i} \right) + \frac{1}{\theta_2} \left(|y_i - \hat{y}_i| - \frac{\epsilon_y}{\nu_i} \right) \right). \tag{1.25}
$$

- For $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \geq \frac{r_y}{\nu_i}$, $i = 1, \dots, n$,

$$
\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|} \left(\frac{\nu_i}{\sigma_i} \right) \left(\frac{1}{\theta_1} \left(\frac{\epsilon_y}{\nu_i} \right) + \frac{1}{\theta_2} \left(\frac{r_y - \epsilon_y}{\nu_i} \right) + \frac{1}{\theta_3} (|y_i - \hat{y}_i| - r_y) \right). \tag{1.26}
$$

The properties and the special cases are in the appendix A in sections A.1.3 and A.1.4.

1.3.10   Definition of diagonal entries of the matrix D_x

The aim is to find the matrix D_x such that

$$\{x \in \mathbb{R}^n : \|D_x(x - \hat{x})\|_\infty \le r_x\} \in \mathcal{X}. \tag{1.27}$$

From the definition of ellipsoidal norm it has:

$$\max_{i=1,\ldots,n} |(d_x)_i(x - \hat{x})_i| = \|D_x(x - \hat{x})\|_\infty \le r_x,$$

from which considering the index of max $\bar{i} \in \{1, \ldots, n\}$ and from (1.27), it follows that

$$\frac{r_x}{(d_x)_{\bar{i}}} \le \min\{x_{\bar{i}} - l_{\bar{i}}, u_{\bar{i}} - x_{\bar{i}}\}.$$

The same derivation can be followed for the diagonal element of matrix D_y.

1.3.11 Representation of the PLT over hyper-rectangular space

For a quick understanding a 2D representation of the transformation is drawn. Let us consider the multimodal continuous function of section (1.3.5).

Expansion   Consider to make an expansion of a hyper-rectangular neighborhood $\mathcal{E}_\infty(\hat{x}, D_x, \epsilon_x)$ of the point $\hat{x} = \hat{y} = (0,0)$ with radius $\epsilon_x = 0.9$ bringing it to a radius of $\epsilon_y = 1.125$. The transformation boundary is $r_x = r_y = 1.8$. For the continuity of the transformation, the space in between the two hyperrectangles is shrunk. Summarizing, the transformation parameters and the other resulting quantities are the following:

- $\sigma = [1.5\ 1.0]$;
- $\epsilon_x = 0.9$;
- $r_x = 1.8$;
- $\theta_1 = 0.7$;
- $\theta_3 = 1$;

- $\nu = [1.5\ 1.0]$;
- $\epsilon_y = 1.286$;
- $r_x = r_y = 1.8$;
- $\theta_2 = 1.75$;
- $\hat{x} = \hat{y} = (0,0)$.

In figure (1.4) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the transformation in the $\mathcal{Y}$ space.
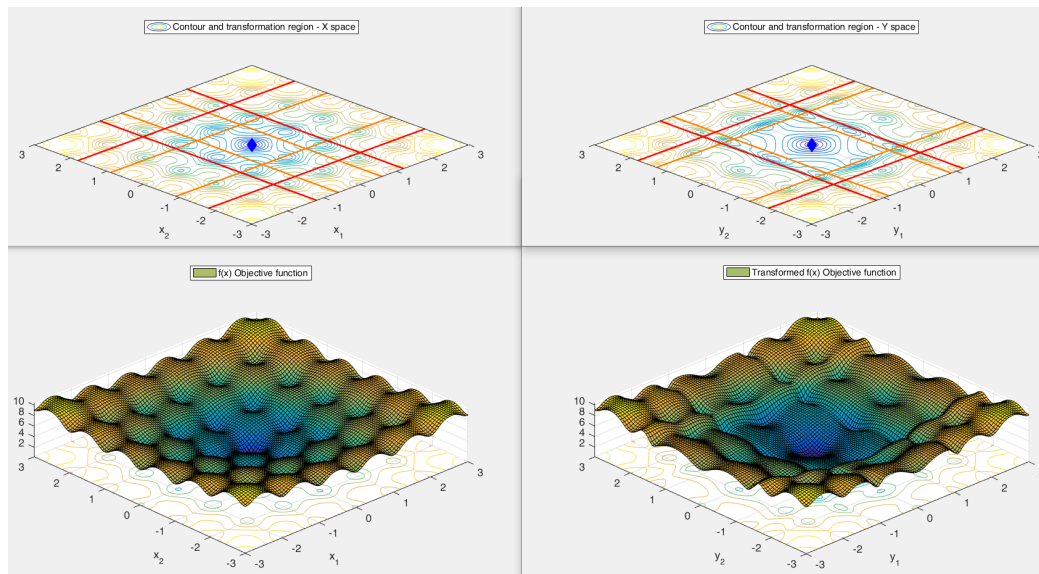
Figure 1.4. An hyper-rectangular neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperrectangles. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

Contraction   Now consider to make a contraction of a hyper-rectangular neighborhood of the point $\hat{x} = \hat{y} = (0,0)$ with radius $\epsilon_x = 1.44$ bringing it to a radius of $\epsilon_y = 1.125$. Summarizing, the transformation parameters and the other resulting quantities are the following:

- $\sigma = [1.5\ 1.0]$;
- $\epsilon_x = 1.44$;
- $r_x = 1.8$;
- $\theta_1 = 1.6$;
- $\theta_3 = 1.0$;

- $\nu = [1.5\ 1.0]$;
- $\epsilon_y = 0.9$;
- $r_y = 1.8$;
- $\theta_2 = 0.4$;
- $\hat{x} = \hat{y} = (0,0)$.

In figure (1.5) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the transformation in the $\mathcal{Y}$ space.
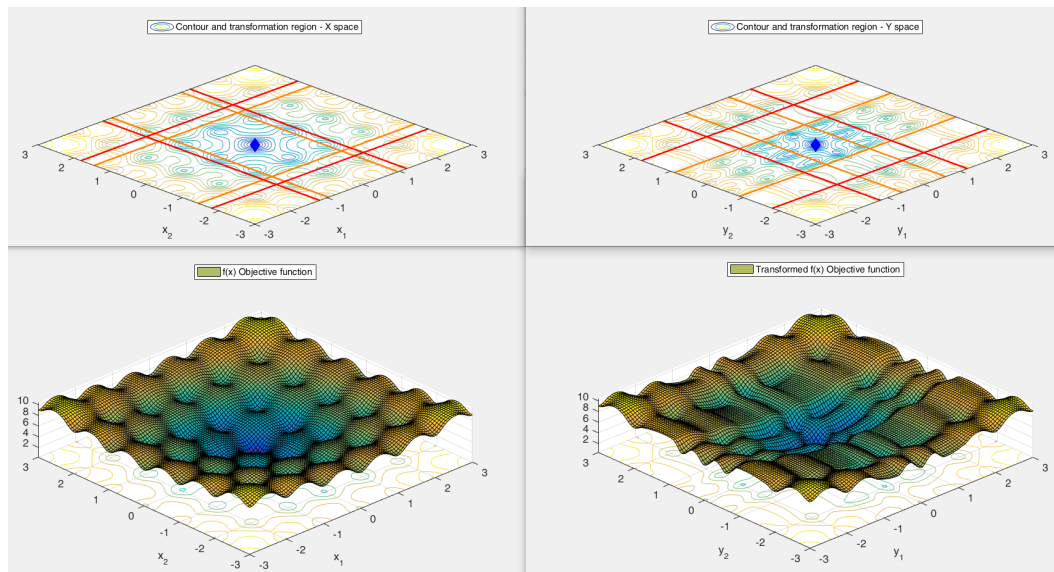
Figure 1.5. An hyper-rectangular neighborhood of $\hat{x}$ has been shrunk while expanding the one in between the two hyperrectangles. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

1.3.12   Pros and cons of the piecewise linear transformation

The piecewise linear transformation is well suited for all global optimization approach where the feasible region is partitioned in sets of hyper-rectangles.

Examples can be a deterministic algorithms such as DIRECT-type approach (DIviding RECTangle [2] and its variants [21]). In the modifications of the DIRECT, there are hybridization with local minimization algorithm [12] in order to speed up the search in promising partition. One can expand the current promising partition to let the local minimization algorithm be fast exploiting longer stepsize or to avoid coming across numerical issue if the partition region is already too small.
For example, the transformation can be applied to the subrectangles which are more promising, while the rest of subrectangles remain unchanged.

A possible drawback is that the transformation propagates outside the hyperrectangle. In particular the subsection (1.3.3) describes the special case of a transformation which has an effect as limited as possible within the region

$$\{x \in \mathbb{R}^n \ : \ \epsilon_x \le |x_i - \hat{x}_i| \le r, \ i = 1, \dots, n\}.$$

However also in this case the transformations perturbs all the space except for the region

$$\{x \in \mathbb{R}^n \ : \ |x_i - \hat{x}_i| \geq r, \ i = 1, \ldots, n\}.$$

This drawback can be noted in the figures (1.4) and (1.5) above where all the space is interested by the transformation except hyperrectangles at the corners.

1.4   Non Linear Transformation (NLT)

In this section we introduce a transformation that overcome the drawback of the piecewise linear one. It means that for such transformation there exist choices of parameters such that its effect can be limited in the region

$$\left\{ x \in \mathbb{R}^n \ : \ \|x - \hat{x}\|_2 \leq r_x, \ i = 1, \ldots, n \right\}.$$

In particular here below it is introduced a piecewise nonlinear transformation $y = T_{\hat{x}}(x)$ between two variable spaces: the original space $\mathcal{X}$ and a transformed one $\mathcal{Y}$. It is not component-wise (not separable in dimension higher that one), but it is still piecewise. Figure (1.6) illustrates the transformation before going into the details of the definition of the transformation equations.

     For the sake of simplicity consider a reference point $\hat{y} \in \mathcal{Y}$ as the origin of the Cartesian axes. This point is set as centroid of the hypersphere $\mathcal{B}_2(\hat{y}, r_y)$ in which we concentrate the transformation impact range. The point $\hat{y}$ is also set as centroid of the smaller hypersphere $\mathcal{B}_2(\hat{y}, \epsilon_y)$ in which the transformation impact has the main focus. The centroid $\hat{y}$ is a fixed point of the transformation, so the equality $\hat{y} = \hat{x}$ holds, where $\hat{x}$ is the centroid of the corresponding hypersphere $\mathcal{B}_2(\hat{x}, r_x)$ in the $\mathcal{X}$ space.

     In one dimension the nonlinear transformation (blue line in fig. (1.6)) coincides to the graph of the piecewise linear transformation. It propagates outward from the centroid in the relative space.
The larger the angle between the bisector (green line in fig. (1.6)) and the piecewise nonlinear transformation, the greater the effect of the transformation.

Parameters

- $\theta_1 = \tan(\alpha) > 0$;
- $\theta_2 = \tan(\beta) > 0$;
- $\theta_3 = \tan(\gamma) = 1$;
- $r_x \geq \epsilon_x \geq 0$;
- $r_y \geq \epsilon_y \geq 0$.

Vector Parameters

- $l_i \leq \hat{x}_i - r_x$;
- $\hat{x}_i + r_x \leq u_i$;
- $l_i \leq \hat{y}_i - r_y$;
- $\hat{y}_i + r_y \leq u_i$.
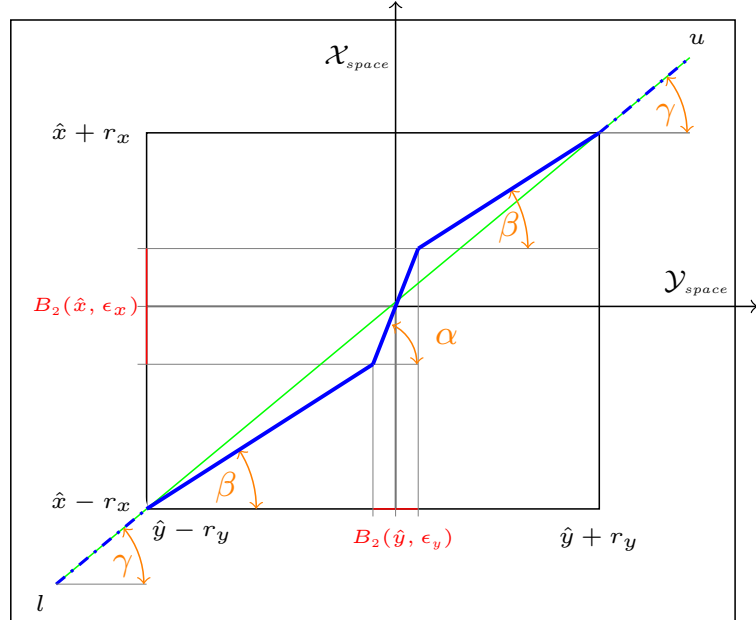
Figure 1.6. Graph of the piecewise nonlinear transformation in one dimension.

1.4.1 Definition of the NLT over hyperspheres

Let $\hat{x} \in \mathbb{R}^n$, ϵ_x, r_x, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_x \leq r_x$.

The following equations define the nonlinear transformation over hyper-spherical space.

- For $x \in \mathbb{R}^n$: $\|x - \hat{x}\|_2 \leq \epsilon_x$,

$$\bar{y}(x) = \hat{y} + \theta_1 (x - \hat{x}). \tag{1.28}$$

- For $x \in \mathbb{R}^n$: $\epsilon_x \leq \|x - \hat{x}\|_2 \leq r_x$,

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} (\theta_1 \epsilon_x + \theta_2 (\|x - \hat{x}\|_2 - \epsilon_x)). \tag{1.29}$$

- For $x \in \mathbb{R}^n$: $\|x - \hat{x}\|_2 \geq r_x$,

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} (\theta_1 \epsilon_x + \theta_2 (r_x - \epsilon_x) + \theta_3 (\|x - \hat{x}\|_2 - r_x)). \tag{1.30}$$

1.4.2   Properties of the NLT - hypersphere

Here below are listed the properties and the necessary coupling condition to preserve the continuity of the transformation between the original space $\mathcal{X}$ and the transformed space $\mathcal{Y}$.

i)   The reference point $\hat{x} \in \mathcal{X}$ is transformed in corresponding one $\hat{y}$ in the transformed space $\mathcal{Y}$:

$$\hat{y} = T_{\hat{x}}(\hat{x}).$$

ii)   The region $\{\ x \in \mathbb{R}^n \ : \|x - \hat{x}\|_2 \leq \epsilon_x \ \}$ is transformed in

$$\{\ y \in \mathbb{R}^n \ : \ \|y - \hat{y}\|_2 \leq \epsilon_y \ \}$$

where (1.28)

$$\bar{y}(x) = \hat{y} + \theta_1 (x - \hat{x}),$$

imposes at the boundary the following coupling condition

$$\epsilon_y = \theta_1 \epsilon_x. \tag{1.31}$$

iii)   The region $\{\ x \in \mathbb{R}^n \ : \epsilon_x \leq \ \|x - \hat{x}\|_2 \leq \ r_x \ \}$ is transformed in

$$\{\ y \in \mathbb{R}^n \ : \epsilon_y \leq \ \|y - \hat{y}\|_2 \leq r_y \ \}$$

where (1.29)

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right),$$

which can be rewritten as

$$\frac{(\bar{y}(x) - \hat{y})}{\|\bar{y}(x) - \hat{y}\|_2} \|\bar{y}(x) - \hat{y}\|_2 = \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right),$$

and applying the norm operator it has

$$\left\| \frac{(\bar{y}(x) - \hat{y})}{\|\bar{y}(x) - \hat{y}\|_2} \|\bar{y}(x) - \hat{y}\|_2 \right\|_2 = \left\| \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right) \right\|_2,$$

from which taking into account scalar terms

$$\frac{\|\bar{y}(x) - \hat{y}\|_2}{\|\bar{y}(x) - \hat{y}\|_2} \|\bar{y}(x) - \hat{y}\|_2 = \frac{\|x - \hat{x}\|_2}{\|x - \hat{x}\|_2} (\theta_1 \epsilon_x + \theta_2 (\|x - \hat{x}\|_2 - \epsilon_x)),$$

that is

$$\|\bar{y}(x) - \hat{y}\|_2 = \theta_1 \epsilon_x + \theta_2 (\|x - \hat{x}\|_2 - \epsilon_x),$$

and by substituting (1.31) it impose at the boundary the following coupling condition

$$r_y = \epsilon_y + \theta_2 (r_x - \epsilon_x). \tag{1.32}$$

iv) The region $\{ \|x - \hat{x}\|_2 \geq r_x \}$ is transformed in

$$\{ y \in \mathbb{R}^n \ : \ \|y - \hat{y}\|_2 \geq r_y \}$$

where from (1.30)

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} (\theta_1 \epsilon_x + \theta_2 (r_x - \epsilon_x) + \theta_3 (\|x - \hat{x}\|_2 - r_x)),$$

substituting (1.31) and (1.32)

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} (r_y + \theta_3 (\|x - \hat{x}\|_2 - r_x)). \tag{1.33}$$

As long as the (1.33) refers to the outermost region there is no need to further coupling condition. It means that θ_3 is free of choice.

1.4.3   Special case of NLT - hypersphere

This section consider a particular example of the transformation over hyperrectangles. Therefore in this case it can be set:

- $\hat{y} = \hat{x}$.

The possibility of expanding or contracting the neighborhood of $\hat{x}$ follows from suitable choices of ϵ_x, ϵ_y. It implies from (1.31) that

- $\theta_1 = \frac{\epsilon_y}{\epsilon_x} \neq 1$.

Moreover for ensuring that the transformation impact has the main effect in a local area, inside the radius r_x, r_y it has to be set

- $r_y = r_x = r$.

It implies from (1.32) that

- $\theta_2 = \frac{r - \epsilon_y}{r - \epsilon_x}$.

Finally in (1.33) by imposing that

- $\theta_3 = 1$;

for $x \in \mathbb{R}^n$: $\|x - \hat{x}\|_2 \geq r$, it has

$$\bar{y}(x) = x.$$

From this choices, the equation (1.28), for $x \in \mathbb{R}^n$: $\|x - \hat{x}\|_2 \leq \epsilon_x$,

$$\bar{y}(x) = \hat{y} + \theta_1(x - \hat{x}),$$

becomes

$$\bar{y}(x) = \hat{y} + \frac{\epsilon_y}{\epsilon_x}(x - \hat{x}).$$

The equation (1.29), for $x \in \mathbb{R}^n$: $\epsilon_x \leq \|x - \hat{x}\|_2 \leq r_x$,

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2}\left(\theta_1\epsilon_x + \theta_2(\|x - \hat{x}\|_2 - \epsilon_x)\right),$$

becomes

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2}\left(\epsilon_y + \left(\frac{r - \epsilon_y}{r - \epsilon_x}\right)(\|x - \hat{x}\|_2 - \epsilon_x)\right).$$

The equation (1.30), for $x \in \mathbb{R}^n \; : \; \|x - \hat{x}\|_2 \geq r_x,$

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\theta_1 \epsilon_x + \theta_2 (r_x - \epsilon_x) + \theta_3 \left(\|x - \hat{x}\|_2 - r_x\right)\right),$$

becomes

$$\bar{y}(x) = x.$$

In summary the nonlinear transformation in this special case:

- For $x \in \mathbb{R}^n \; : \; \|x - \hat{x}\|_2 \leq \epsilon_x,$

$$\bar{y}(x) = \hat{y} + \frac{\epsilon_y}{\epsilon_x} (x - \hat{x}). \tag{1.34}$$

- For $x \in \mathbb{R}^n \; : \; \epsilon_x \leq \|x - \hat{x}\|_2 \leq r,$

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\epsilon_y + \left(\frac{r - \epsilon_y}{r - \epsilon_x}\right) \left(\|x - \hat{x}\|_2 - \epsilon_x\right)\right). \tag{1.35}$$

- For $x \in \mathbb{R}^n \; : \; \|x - \hat{x}\|_2 \geq r,$

$$\bar{y}(x) = x. \tag{1.36}$$

1.4.4   Invertibility of the nonlinear transformation

We are going to prove that the direct equations of the nonlinear transformation introduced in section (1.4.1) are invertible. In particular the non-linearities appear in (1.29) and (1.30). Consider the equation (1.29):

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right),$$

substituting (1.31)

$$y - \hat{y} = \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\epsilon_y + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right), \tag{1.37}$$

applying the $2 - norm$ operator and taking into account that $\|x - \hat{x}\|_2 \geq \epsilon_x$

$$
\begin{aligned}
\|y - \hat{y}\|_2 &= \left\| \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\epsilon_y + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right) \right\|_2 \\
&= \left\| \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \right\|_2 \left(\epsilon_y + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right) \\
&= \left(\epsilon_y + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right),
\end{aligned}
$$

and so

$$\|y - \hat{y}\|_2 = \epsilon_y + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right), \tag{1.38}$$

by inverting the last expression we have

$$\|x - \hat{x}\|_2 = \epsilon_x + \frac{1}{\theta_2} \left(\|y - \hat{y}\|_2 - \epsilon_y \right). \tag{1.39}$$

Substituting (1.38) in (1.37)

$$y - \hat{y} = \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \|y - \hat{y}\|_2,$$

by inverting the last expression

$$x(y) - \hat{x} = \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \|x - \hat{x}\|_2,$$

and substituting (1.39) we have

$$x(y) - \hat{x} = \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \left(\epsilon_x + \frac{1}{\theta_2} \left(\|y - \hat{y}\|_2 - \epsilon_y \right) \right),$$

that is

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \left(\epsilon_x + \frac{1}{\theta_2} \left(\|y - \hat{y}\|_2 - \epsilon_y \right) \right). \tag{1.40}$$

Now consider the equation (1.30):

$$\bar{y}\left(x\right) = \hat{y} + \frac{\left(x - \hat{x}\right)}{\|x - \hat{x}\|_2} \left(\theta_1 \epsilon_x + \theta_2 (r_x - \epsilon_x) + \theta_3 \left(\|x - \hat{x}\|_2 - r_x\right)\right),$$

substituting (1.31) and (1.32)

$$y - \hat{y} = \frac{\left(x - \hat{x}\right)}{\|x - \hat{x}\|_2} \left(r_y + \theta_3 \left(\|x - \hat{x}\|_2 - r_x\right)\right), \tag{1.41}$$

applying the $2 - norm$ operator and taking into account that $\|x - \hat{x}\|_2 \geq r_x$

$$
\begin{aligned}
\|y - \hat{y}\|_2 \quad &= \quad \left\| \frac{\left(x - \hat{x}\right)}{\|x - \hat{x}\|_2} \left(r_y + \theta_3 \left(\|x - \hat{x}\|_2 - r_x\right)\right) \right\|_2 \\
&= \quad \left\| \frac{\left(x - \hat{x}\right)}{\|x - \hat{x}\|_2} \right\|_2 \left(r_y + \theta_3 \left(\|x - \hat{x}\|_2 - r_x\right)\right) \\
&= \quad \left(r_y + \theta_3 \left(\|x - \hat{x}\|_2 - r_x\right)\right),
\end{aligned}
$$

and so

$$\|y - \hat{y}\|_2 = r_y + \theta_3 \left(\|x - \hat{x}\|_2 - r_x\right), \tag{1.42}$$

by inverting the last expression we have

$$\|x - \hat{x}\|_2 = r_x + \frac{1}{\theta_3} (\|y - \hat{y}\|_2 - r_y). \tag{1.43}$$

Substituting (1.42) in (1.41)

$$y - \hat{y} = \frac{\left(x - \hat{x}\right)}{\|x - \hat{x}\|_2} \|y - \hat{y}\|_2,$$

by inverting the last expression

$$x\left(y\right) - \hat{x} = \frac{\left(y - \hat{y}\right)}{\|y - \hat{y}\|_2} \|x - \hat{x}\|_2,$$

and substituting (1.43) we have

$$x - \hat{x} = \frac{\left(y - \hat{y}\right)}{\|y - \hat{y}\|_2} \left(r_x + \frac{1}{\theta_3} (\|y - \hat{y}\|_2 - r_y)\right),$$

that is

$$\bar{x}\left(y\right) = \hat{x} + \frac{\left(y - \hat{y}\right)}{\|y - \hat{y}\|_2} \left(r_x + \frac{1}{\theta_3} (\|y - \hat{y}\|_2 - r_y)\right). \tag{1.44}$$

1.4.5   Definition of the inverse NLT over hypersphere

Let $\hat{y} \in \mathbb{R}^n$, ϵ_y, r_y, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_y \leq r_y$.

The following equations define the inverse nonlinear transformation over hyperspherical space.

- For $y \in \mathbb{R}^n \ : \ \|y - \hat{y}\|_2 \leq \epsilon_y$,

$$\bar{x}(y) = \hat{x} + \frac{1}{\theta_1}(y - \hat{y}). \tag{1.45}$$

- For $y \in \mathbb{R}^n \ : \ \epsilon_y \leq \|y - \hat{y}\|_2 \leq r_y$,

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2}\left(\frac{1}{\theta_1}\epsilon_y + \frac{1}{\theta_2}\left(\|y - \hat{y}\|_2 - \epsilon_y\right)\right). \tag{1.46}$$

- For $x \in \mathbb{R}^n \ : \ \|y - \hat{y}\|_2 \geq r_y$,

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2}\left(\frac{1}{\theta_1}\epsilon_y + \frac{1}{\theta_2}(r_y - \epsilon_y) + \frac{1}{\theta_3}\left(\|y - \hat{y}\|_2 - r_y\right)\right). \tag{1.47}$$

To have an easier reading and better clarity the properties and the special cases are in the appendix A in sections A.2 and A.2.2.

1.4.6   Representation of the NLT over hyper-spherical space

Here below a 2D representation of the nonlinear transformation is drawn. Let us consider the multimodal continuous function of section (1.3.5).

Expansion   Consider to make an expansion of a hyper-spherical neighborhood $\mathcal{B}_2(\hat{x}, \epsilon_x)$ of the point $\hat{x} = \hat{y} = (0,0)$ with radius $\epsilon_x = 1.5$ bringing it to a radius of $\epsilon_y = 2.14$. The transformation boundary is $r_x = r_y = 3.0$. For the continuity of the transformation, the space in between the two hyperrectangles shrunk. The transformation parameters has been set as follows:

- $\epsilon_x = 1.5;$
- $\epsilon_y = 2.14;$

- $r_x = 3.0$;
- $r_y = 3.0;$

- $\theta_1 = 0.7;$
- $\theta_2 = 1.75;$

- $\theta_3 = 1;$
- $\hat{x} = \hat{y} = (0,0).$
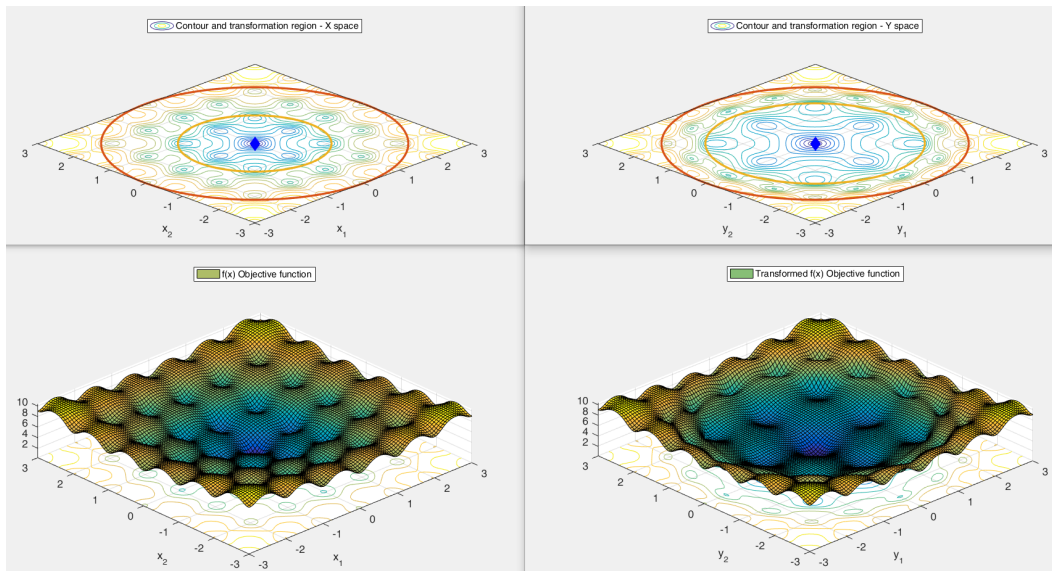
Figure 1.7. A spherical neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two spheres .

Contraction   Now consider to make a contraction of a hyper-spherical neighborhood $\mathcal{B}_2(\hat{x}, \epsilon_x)$ of the point $\hat{x} = \hat{y} = (0,0)$ with radius $\epsilon_x = 2.4$ bringing it to a radius of $\epsilon_y = 1.33$. The transformation boundary is $r_x = r_y = 3.0$. For the continuity of the transformation, the space in between the two hyperrectangles is expanded.

- $\epsilon_x = 2.4$;
- $\epsilon_y = 1.33$;
- $r_x = 3.0$;
- $r_y = 3.0$;
- $\theta_1 = 1.8$;
- $\theta_2 = 0.36$;
- $\theta_3 = 1$;
- $\hat{x} = \hat{y} = (0,0)$.
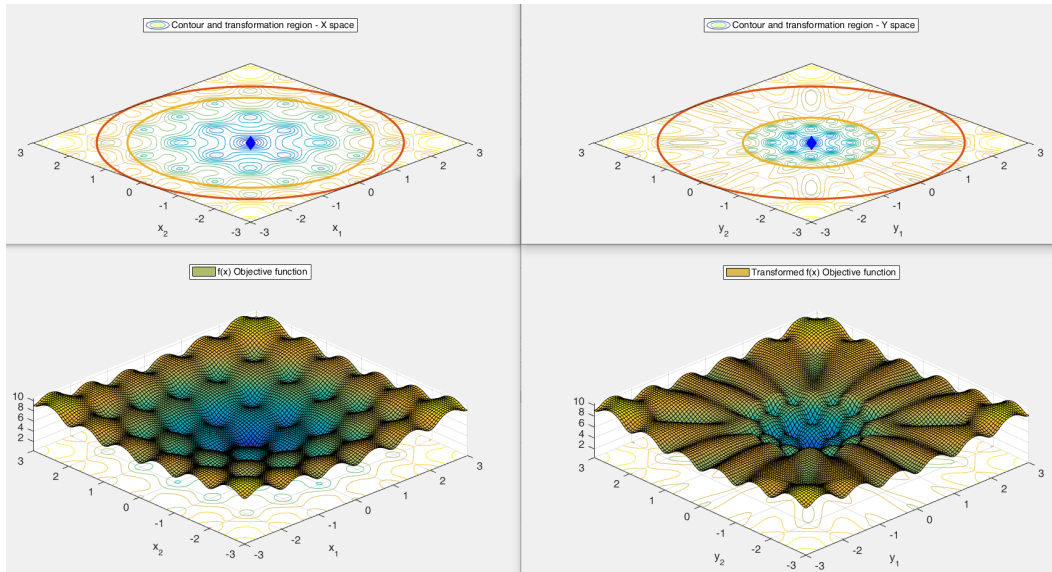
Figure 1.8. A spherical neighborhood of $\hat{x}$ has been shrunk while expanding the one in between the two spheres .

1.4.7 Definition of the NLT over hyperellipsoids

The equations that describe the nonlinear transformation over hyperspheres can be suited for general hyperellipsoidal space. Let us introduce two diagonal matrices with positive entries:

$$D_x = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \end{pmatrix}, \qquad D_y = \begin{pmatrix} \nu_1 & 0 & \ldots & 0 \\ 0 & \nu_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \nu_n \end{pmatrix}.$$

The matrices $D_y \in \mathbb{R}^{n \times n}$ e $D_x \in \mathbb{R}^{n \times n}$ are positive definite and invertible and can be used for specifying the ellipsoidal $l_2 - norm$ distance over the $\mathcal{Y}$ space in this way

$$\|D_x(x - \hat{x})\|_2 = \sqrt{(x - \hat{x})^T D_x D_x (x - \hat{x})},$$

and over the $\mathcal{X}$ space

$$\|D_y(y - \hat{y})\|_2 = \sqrt{(y - \hat{y})^T D_y D_y (y - \hat{y})}.$$

Let $\hat{x} \in \mathbb{R}^n$, ϵ_x, r_x, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_x \leq r_x$. The following equations define the nonlinear transformation over hyper-ellipsoidal space.

- For $x \in \mathbb{R}^n$: $\|D_x(x - \hat{x})\|_2 \leq \epsilon_x$,

$$\bar{y}(x) = \hat{y} + \theta_1 D_y^{-1} D_x (x - \hat{x}). \tag{1.48}$$

- For $x \in \mathbb{R}^n$: $\epsilon_x \leq \|D_x(x - \hat{x})\|_2 \leq r_x$,

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x\right)\right). \tag{1.49}$$

- For $x \in \mathbb{R}^n$: $\|D_x(x - \hat{x})\|_2 \geq r_x$,

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2 (r_x - \epsilon_x) + \theta_3 \left(\|D_x(x - \hat{x})\|_2 - r_x\right)\right). \tag{1.50}$$

1.4.8   Properties of the NLT - hyperellipse

Here below are listed the properties and the necessary coupling condition to preserve the continuity of the transformation between the original space $\mathcal{X}$ and the transformed space $\mathcal{Y}$.

i) The reference point $\hat{x} \in \mathcal{X}$ is transformed in the corresponding one $\hat{y}$ in the transformed space $\mathcal{Y}$:

$$\hat{y} = T_{\hat{x}}(\hat{x}).$$

ii) The region $\{ x \in \mathbb{R}^n : \|D_x(x - \hat{x})\|_2 \leq \epsilon_x \}$ is transformed in

$$\{ y \in \mathbb{R}^n : \|D_y(y - \hat{y})\|_2 \leq \epsilon_y \},$$

where (1.48)

$$\bar{y}(x) = \hat{y} + \theta_1 D_y^{-1} D_x(x - \hat{x}),$$

imposes at the boundary the following coupling condition

$$\epsilon_y = \theta_1 \epsilon_x. \tag{1.51}$$

iii) The region $\{ x \in \mathbb{R}^n : \epsilon_x \leq \|D_x(x - \hat{x})\|_2 \leq r_x \}$ is transformed in

$$\{ y \in \mathbb{R}^n : \epsilon_y \leq \|D_y(y - \hat{y})\|_2 \leq r_y \},$$

where (1.49)

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x \right) \right),$$

which can be rewritten as

$$\frac{D_y(\bar{y}(x) - \hat{y})}{\|D_y(\bar{y}(x) - \hat{y})\|_2} \|D_y(\bar{y}(x) - \hat{y})\|_2 = \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x \right) \right),$$

and applying the norm operator it has

$$\left\| \frac{D_y(\bar{y}(x) - \hat{y})}{\|D_y(\bar{y}(x) - \hat{y})\|_2} \|D_y(\bar{y}(x) - \hat{y})\|_2 \right\|_2 = \left\| \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x \right) \right) \right\|_2,$$

from which taking into account scalar terms

$$\frac{\|D_y(\bar{y}(x) - \hat{y})\|_2}{\|D_y(\bar{y}(x) - \hat{y})\|_2} \|D_y(\bar{y}(x) - \hat{y})\|_2 = \frac{\|D_x(x - \hat{x})\|_2}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x \right) \right),$$

that is

$$\|D_y(\bar{y}(x) - \hat{y})\|_2 = \theta_1 \epsilon_x + \theta_2 \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x\right),$$

and by substituting (1.51) it imposes at the boundary the following coupling condition

$$r_y = \epsilon_y + \theta_2(r_x - \epsilon_x). \tag{1.52}$$

iv) The region $\{\ \|D_x(x - \hat{x})\|_2 \geq\ r_x\ \}$ is transformed in

$$\{\ y \in \mathbb{R}^n\ :\ \|D_y(y - \hat{y})\|_2 \geq r_y\ \},$$

where from (1.50)

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x) + \theta_3 \left(\|D_x(x - \hat{x})\|_2 - r_x\right)\right),$$

substituting (1.51) and (1.52)

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(r_y + \theta_3(\|D_x(x - \hat{x})\|_2 - r_x)\right). \tag{1.53}$$

As long as the (1.53) refers to the outermost region there is no need to further coupling condition. It means that θ_3 is free of choice.

1.4.9 Special case of NLT - hyperellipse

As well as in the transformation over hyperspheres, this section consider a particular example of the transformation over hyperellipses. Therefore in this case it can be set:

- $\hat{y} = \hat{x}$.

The hyperellipse dimensions in the original space $\mathcal{X}$ and in the transformed space $\mathcal{Y}$ are described by

- $D_y = D_x$ and so $\nu_i = \sigma_i$ for $i = 1, \ldots, n$;

The possibility of expanding or contracting the neighborhood of $\hat{x}$ follows from suitable choices of ϵ_x, ϵ_y. It implies from (1.51) that

- $\theta_1 = \frac{\epsilon_y}{\epsilon_x} \neq 1$.

Moreover for ensuring that the transformation impact has the main effect in a local area, inside the radius r_x, r_y it has to be set

- $r_y = r_x = r$.

It implies from (1.52) that

- $\theta_2 = \frac{r - \epsilon_y}{r - \epsilon_x}$.

Finally in (1.53) by imposing that

- $\theta_3 = 1$;

for $x \in \mathbb{R}^n \; : \; \|D_x(x - \hat{x})\|_2 \geq r$, it has

$$\bar{y}(x) = x.$$

From this choices, the equation (1.48), for $x \in \mathbb{R}^n \; : \; \|D_x(x - \hat{x})\|_2 \leq \epsilon_x$,

$$\bar{y}(x) = \hat{y} + \theta_1 D_y^{-1} D_x(x - \hat{x}),$$

becomes

$$\bar{y}(x) = \hat{y} + \frac{\epsilon_y}{\epsilon_x} D_y^{-1} D_x(x - \hat{x}).$$

The equation (1.49), for $x \in \mathbb{R}^n \; : \; \epsilon_x \leq \|D_x(x - \hat{x})\|_2 \leq r_x,$

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x \right) \right),$$

becomes

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\epsilon_y + \left(\frac{r - \epsilon_y}{r - \epsilon_x} \right) \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x \right) \right).$$

The equation (1.50), for $x \in \mathbb{R}^n \; : \; \|D_x(x - \hat{x})\|_2 \geq r_x,$

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\theta_1 \epsilon_x + \theta_2 (r_x - \epsilon_x) + \theta_3 \left(\|D_x(x - \hat{x})\|_2 - r_x \right) \right),$$

becomes

$$\bar{y}(x) = x.$$

In summary the nonlinear transformation in this special case:

- For $x \in \mathbb{R}^n \; : \; \|D_x(x - \hat{x})\|_2 \leq \epsilon_x,$

$$\bar{y}(x) = \hat{y} + \frac{\epsilon_y}{\epsilon_x} D_y^{-1} D_x(x - \hat{x}). \tag{1.54}$$

- For $x \in \mathbb{R}^n \; : \; \epsilon_x \leq \|D_x(x - \hat{x})\|_2 \leq r,$

$$\bar{y}(x) = \hat{y} + D_y^{-1} \frac{D_x(x - \hat{x})}{\|D_x(x - \hat{x})\|_2} \left(\epsilon_y + \left(\frac{r - \epsilon_y}{r - \epsilon_x} \right) \left(\|D_x(x - \hat{x})\|_2 - \epsilon_x \right) \right). \tag{1.55}$$

- For $x \in \mathbb{R}^n \; : \; \|D_x(x - \hat{x})\|_2 \geq r,$

$$\bar{y}(x) = x. \tag{1.56}$$

1.4.10 Definition of the inverse NLT over hyperellipses

Let us introduce two diagonal matrices with positive entries:

$$
D_y = \begin{pmatrix}
\nu_1 & 0 & \dots & 0 \\
0 & \nu_2 & \dots & 0 \\
\vdots & \vdots & \ddots & 0 \\
0 & 0 & 0 & \nu_n
\end{pmatrix}, \qquad
D_x = \begin{pmatrix}
\sigma_1 & 0 & \dots & 0 \\
0 & \sigma_2 & \dots & 0 \\
\vdots & \vdots & \ddots & 0 \\
0 & 0 & 0 & \sigma_n
\end{pmatrix}.
$$

The matrices $D_y \in \mathbb{R}^{n \times n}$ e $D_x \in \mathbb{R}^{n \times n}$ are positive definite and invertible and can be used for specifying the ellipsoidal $l_2 - norm$ distance over the $\mathcal{Y}$ space in this way

$$
\|D_y(y - \hat{y})\|_2 = \sqrt{(y - \hat{y})^T D_y D_y (y - \hat{y})},
$$

and over the $\mathcal{X}$ space

$$
\|D_x(x - \hat{x})\|_2 = \sqrt{(x - \hat{x})^T D_x D_x (x - \hat{x})}.
$$

Let $\hat{y} \in \mathbb{R}^n$, ϵ_y, r_y, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_y \leq r_y$. The following equations define the inverse nonlinear transformation over hyperellipsoidal space.

- For $y \in \mathbb{R}^n$: $\|D_y(y - \hat{y})\|_2 \leq \epsilon_y$,

$$
\bar{x}(y) = \hat{x} + \frac{1}{\theta_1} D_x^{-1} D_y (y - \hat{y}). \tag{1.57}
$$

- For $y \in \mathbb{R}^n$: $\epsilon_y \leq \|D_y(y - \hat{y})\|_2 \leq r_y$,

$$
\bar{x}(y) = \hat{x} + D_x^{-1} \frac{D_y(y - \hat{y})}{\|D_y(y - \hat{y})\|_2} \left(\frac{1}{\theta_1} \epsilon_y + \frac{1}{\theta_2} \left(\|D_y(y - \hat{y})\|_2 - \epsilon_y \right) \right). \tag{1.58}
$$

- For $x \in \mathbb{R}^n$: $\|D_y(y - \hat{y})\|_2 \geq r_y$,

$$
\bar{x}(y) = \hat{x} + D_x^{-1} \frac{D_y(y - \hat{y})}{\|D_y(y - \hat{y})\|_2} \left(\frac{1}{\theta_1} \epsilon_y + \frac{1}{\theta_2}(r_y - \epsilon_y) + \frac{1}{\theta_3} \left(\|D_y(y - \hat{y})\|_2 - r_y \right) \right). \tag{1.59}
$$

The properties and the special cases are in the appendix A in sections A.2.3 and A.2.4.

1.4.11 Definition of diagonal entries of the matrix D_x

The aim is to find the matrix D_x such that

$$\{x \in \mathbb{R}^n : \|D_x(x - \hat{x})\|_2 \leq r_x\} \in \mathcal{X}. \tag{1.60}$$

From the definition of ellipsoidal norm it has:

$$\sum_{i=1}^{2} (d_x)_i (x - \hat{x})_i^2 = \|D_x(x - \hat{x})\|_2^2 \leq r_x^2,$$

from which $\forall \bar{i} \in \{1, \dots, n\}$ it has

$$(d_x)_{\bar{i}}(x - \hat{x})_{\bar{i}} \leq \|D_x(x - \hat{x})\|_2 \leq r_x.$$

Finally (1.60) it follows that

$$\frac{r_x}{(d_x)_{\bar{i}}} \leq \min\{x_{\bar{i}} - l_{\bar{i}}, u_{\bar{i}} - x_{\bar{i}}\}.$$

The same derivation can be followed for the diagonal element of matrix D_y.

1.4.12   Representation of the piecewise nonlinear transformation over hyperellipses

Expansion   Consider to make an expansion of a hyperellipsoidal neighborhood $\mathcal{E}(\hat{x}, D_x, \epsilon_x)$ of the point $\hat{x} = \hat{y} = (0,0)$ with semi-axes $\epsilon_x/\sigma = [1.5\ 0.9]$ bringing it to semi-axes $\epsilon_y/\nu = [2.14\ 1.28]$. The transformation boundary is $r_x = r_y = 3.0$. For the continuity of the transformation, the space in between the two hyperrectangles is expanded. The transformation parameters has been set as follows:

- $\sigma = [1.0\ 1.67]$;
- $\epsilon_x = 1.5$;
- $r_x = 3.0$;
- $\theta_1 = 0.7$;
- $\theta_3 = 1.0$;

- $\nu = [1.0\ 1.67]$;
- $\epsilon_y = 2.14$;
- $r_y = 3.0$;
- $\theta_2 = 1.75$;
- $\hat{x} = \hat{y} = (0,0)$.
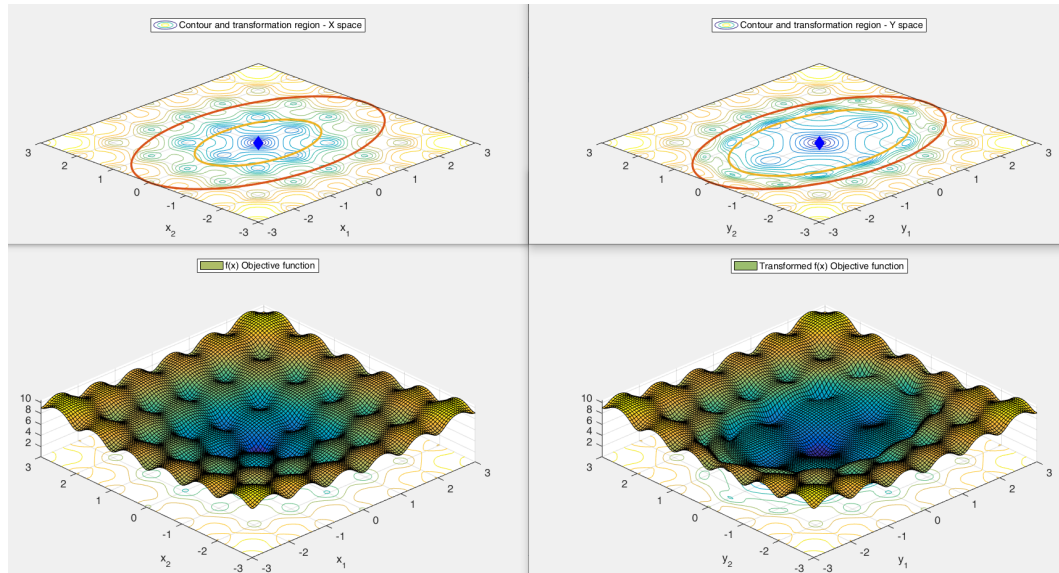
Figure 1.9. An ellipsoidal neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two ellipsoid.

Contraction   Consider to make a contraction of a hyperellipsoidal neighborhood $\mathcal{E}(\hat{x}, D_x, \epsilon_x)$ of the point $\hat{x} = \hat{y} = (0,0)$ with semi-axes $\epsilon_x/\sigma = [2.4\ 1.44]$ bringing it to semi-axes $\epsilon_y/\nu = [1.71\ 1.03]$. The transformation boundary is $r_x = r_y = 3.0$. For the continuity of the transformation, the space in between the two hyperrectangles is expanded.

- $\sigma = [1.0\ 1.67]$;
- $\epsilon_x = 2.4$;
- $r_x = 3.0$;
- $\theta_1 = 1.4$;
- $\theta_3 = 1.0$;

- $\nu = [1.0\ 1.67]$;
- $\epsilon_y = 1.71$;
- $r_y = 3.0$;
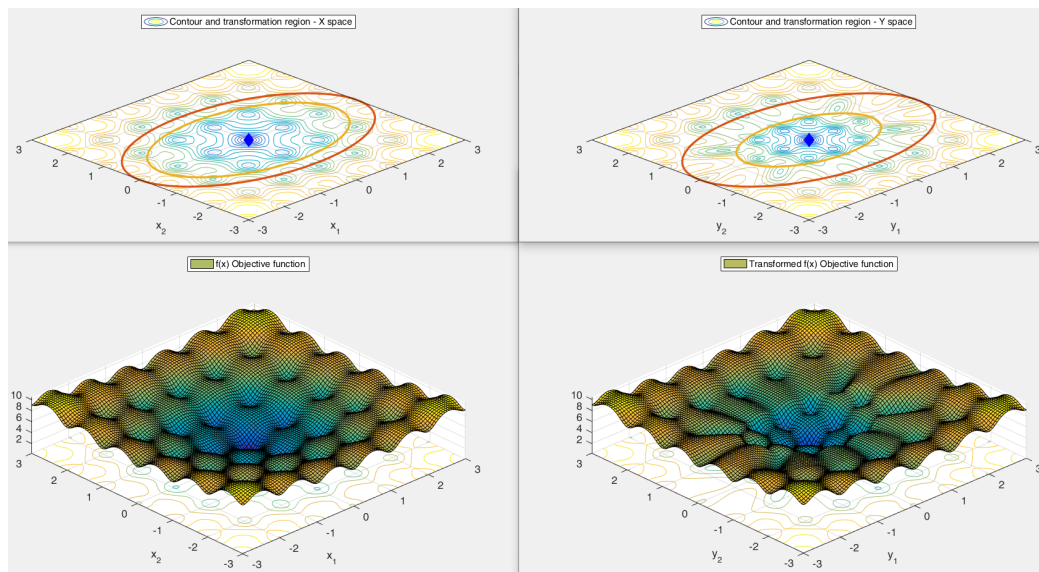- $\theta_2 = 0.47$;
- $\hat{x} = \hat{y} = (0,0)$.

Figure 1.10. An ellipsoidal neighborhood of $\hat{x}$ has been shrunk while expanding the one in between the two ellipsoid .

1.5   Recursive formula

Since the transformations proposed in chapter (1.3) and chapter (1.4) are continuous and bijective we can apply it recursively.

As long as we want to define a recursion we need to transit in intermediate variable spaces a number of times. Let us define the $k-th$ intermediate variable space $\mathcal{Z}^k$

$$\mathcal{Z}^k = \left(z^k \in \mathbb{R}^n \ : \ l \leq \ z^k \ \leq u \right)$$

with $l, u \in \mathbb{R}^n$. Given a sequence of $K > 1$ transformations

$$z^k = \bar{x} \left(z^{k-1} \right)^k, \quad k = 1, \dots, K$$

that $\forall k$ have its own parameters $D_x^k, D_y^k, \epsilon_x^k, \ \epsilon_y^k, \ r_x^k, \ r_y^k, \ \theta_1^k, \ \theta_2^k$, and reference point $\hat{z}^k$. We start from the transformed variable space $\mathcal{Y}$ to arrive to the original variable space $\mathcal{X}$. It means that for the first intermediate variable space the equality $\mathcal{Z}^k = \mathcal{Y}$ holds, whereas for the last one $\mathcal{Z}^K = \mathcal{X}$. Algorithm 1 summarize the recursive formula.

Algorithm 1 Recursive transformation formulas

1: $x \in \mathcal{X}, \ y \in \mathcal{Y}, \ z^k \in \mathcal{Z}^k, k = 1, \dots, K, \ z^0 = y, \ \hat{z}^0 = \hat{y}$
2: **for** $k = 1, \dots, K$ **do**
3:      set $D_x^k, D_y^k, \epsilon_x^k, \ \epsilon_y^k, \ r_x^k, \ r_y^k, \ \theta_1^k, \ \theta_2^k, \ \hat{z}^k = \hat{z}^{k-1}$
4:      $z^k = \bar{x} \left(z^{k-1} \right)^k$
5: **end for**
6: $x = z^K$
7: **return** x

The following figures show examples of application of two iteration of the recursive transformation in both cases of expansion and contraction. Respectively hyperrectangle shaped space for piecewise linear transformation and hyperellipsoidal shaped space for piecewise non linear transformation.

PWL Expansion   Let us start considering to make two recursive expansion of hyper-rectangular neighborhoods $\mathcal{B}_\infty(\hat{y}^k, \epsilon_y^k)$. The first neighborhood is characterized by the point $\hat{x}^1 = \hat{y}^1 = (2,3)$ with radius $\epsilon_x^1 = 0.8$ bringing it to a radius of $\epsilon_y^1 = 1.14$. The transformation boundary is $r_x^1 = r_y^1 = 1.6$. The same action is done for the second neighborhood at the point $\hat{x}^2 = \hat{y}^2 = (0,0)$ with radius $\epsilon_x^2 = 0.7$ bringing it to a radius of $\epsilon_y^2 = 1.0$. The transformation boundary is $r_x^2 = r_y^2 = 1.4$. For the continuity of the transformation, the space in between the two hyperrectangles is shrunk.

The transformation parameters for the recursive hyper-rectangular expansion at iteration $k = 1$ has been set as follows:

- $\sigma^1 = [0.9 \ 1.0];$
- $\epsilon_x^1 = 0.8;$
- $r_x^1 = 1.6;$
- $\theta_1^1 = 0.7;$
- $\theta_3^1 = 1.0;$

- $\nu^1 = [0.9 \ 1.0];$
- $\epsilon_y^1 = 1.14;$
- $r_y^1 = 1.6;$
- $\theta_2^1 = 1.75;$
- $\hat{x} = \hat{y} = (2,3).$

In figure (1.11) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the recursion at iteration $k = 1$, in the $\mathcal{Y}$ space.

Figure 1.11. Recursion at iteration $k = 1$. An hyper-rectangular neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperrectangles. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

The transformation parameters for the recursive hyper-rectangular expansion at iteration $k = 2$ has been set as follows:

- $\sigma^2 = [1.0 \ 0.7]$;
- $\epsilon_x^2 = 0.7$;
- $r_x^2 = 1.4$;
- $\theta_1^2 = 0.7$;
- $\theta_3^2 = 1.0$;
- $\nu^2 = [1.0 \ 0.7]$
- $\epsilon_y^2 = 1.0$;
- $r_y^2 = 1.4$;
- $\theta_2^2 = 1.75$;
- $\hat{x} = \hat{y} = (0, 0)$.

In figure (1.12) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the recursion at iteration $k = 2$, of the transformation in the $\mathcal{Y}$ space.

Figure 1.12. Recursion at iteration $k = 2$. Another hyper-rectangular neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperrectangles. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

PWL Contraction   Considering now to make two recursive contraction of hyper-rectangular neighborhoods $\mathcal{B}_\infty(\hat{y}^k, \epsilon_y^k)$. The first neighborhood is characterized by the point $\hat{x}^1 = \hat{y}^1 = (2,3)$ with radius $\epsilon_x^1 = 1.28$ bringing it to a radius of $\epsilon_y^1 = 0.71$. The transformation boundary is $r_x^1 = r_y^1 = 1.6$. The same action is done for the second neighborhood at the point $\hat{x}^2 = \hat{y}^2 = (0,0)$ with radius $\epsilon_x^2 = 1.12$ bringing it to a radius of $\epsilon_y^2 = 0.62$. The transformation boundary is $r_x^2 = r_y^2 = 1.4$. For the continuity of the transformation, the space in between the two hyperrectangles is shrunk.

The transformation parameters for the recursive hyper-rectangular contraction at iteration $k = 1$ has been set as follows:

- $\sigma^1 = [0.9\ 1.0]$
- $\epsilon_x^1 = 1.28$
- $r_x^1 = 1.4$
- $\theta_1^1 = 1.8;$
- $\theta_3^1 = 1.0;$

- $\nu^1 = [0.9\ 1.0];$
- $\epsilon_y^1 = 0.71;$
- $r_y^1 = 1.6;$
- $\theta_2^1 = 0.36;$
- $\hat{x} = \hat{y} = (2,3).$

In figure (1.13) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the recursion at iteration $k = 2$, in the $\mathcal{Y}$ space.
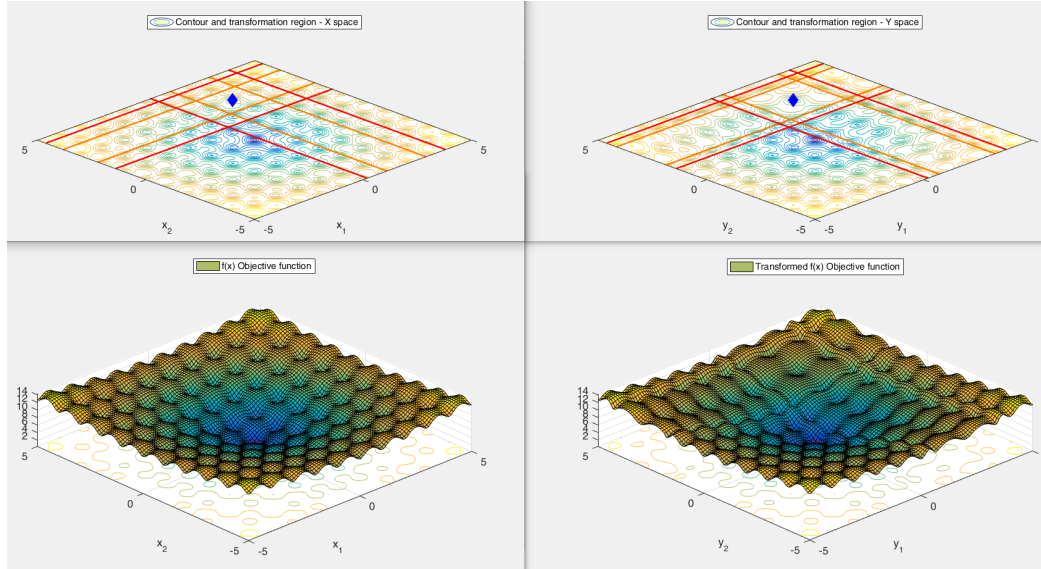
Figure 1.13. Recursion at iteration $k = 1$. An hyper-rectangular neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperrectangles. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

The transformation parameters for the recursive at iteration $k = 2$ has been set as follows:

- $\sigma^2 = [1.0\ 0.9]$; • $\epsilon_x^2 = 1.28$;   • $r_x^2 = 1.4$;   • $\theta_1^2 = 1.8$;   • $\theta_3^2 = 1.0$;

- $\nu^2 = [0.9\ 1.0]$; • $\epsilon_y^2 = 0.71$;   • $r_y^2 = 1.4$;   • $\theta_2^2 = 0.36$;   • $\hat{x} = \hat{y} = (0, 0)$.

In figure (1.14) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the recursion at iteration $k = 2$, of the transformation in the $\mathcal{Y}$ space.

Figure 1.14. Recursion at iteration $k = 2$. Another hyper-rectangular neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperrectangles. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

NLT Expansion With regard to piecewise non-linear transformation, consider to make two recursive expansion of hyperellipsoidal neighborhoods $\mathcal{B}_2(\hat{y}^k, \epsilon_y^k)$. The first neighborhood is of the point $\hat{x}^1 = \hat{y}^1 = (2,3)$ with radius $\epsilon_x^1 = 0.5$ bringing it to a radius of $\epsilon_y^1 = 1.1$. The transformation boundary is $r_x^1 = r_y^1 = 2.0$. The same action is done for the second neighborhood at the point $\hat{x}^2 = \hat{y}^2 = (0,0)$ with radius $\epsilon_x^2 = 0.6$ bringing it to a radius of $\epsilon_y^2 = 1.27$. The transformation boundary is $r_x^2 = r_y^2 = 2.2$. For the continuity of the transformation, the space in between the two hyperrectangles is shrunk.

The transformation parameters for the recursive hyperellipsoidal expansion at iteration k=1 has been set as follows:

- $\sigma^1 = [0.83 \ 1.0]$;
- $\epsilon_x^1 = 0.9$;
- $r_x^1 = 2.0$;
- $\theta_1^1 = 0.45$;
- $\theta_3^1 = 1.0$;

- $\nu^1 = [0.83 \ 1.0]$;
- $\epsilon_y^1 = 1.286$;
- $r_y^1 = 2.0$;
- $\theta_2^1 = 1.69$;
- $\hat{x} = \hat{y} = (2,3)$.

In figure (1.15) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the recursion at iteration $k = 1$, in the $\mathcal{Y}$ space.

Figure 1.15. Recursion at iteration $k = 1$. An hyperellipsoidal neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperellipsoidal. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

The transformation parameters for the recursive hyperellipsoidal expansion at iteration $k = 2$ has been set as follows:

- $\sigma^2 = [1.0 \ 0.88]$;
- $\epsilon_x^2 = 0.6$;
- $r_x^2 = 2.2$;
- $\theta_1^2 = 0.47$;
- $\theta_3^2 = 1.0$;

- $\nu^2 = [1.0 \ 0.88]$;
- $\epsilon_y^2 = 1.27$;
- $r_y^2 = 2.2$;
- $\theta_2^2 = 1.72$;
- $\hat{x} = \hat{y} = (0,0)$.

In figure (1.16) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the recursion at iteration $k = 2$, of the transformation in the $\mathcal{Y}$ space.
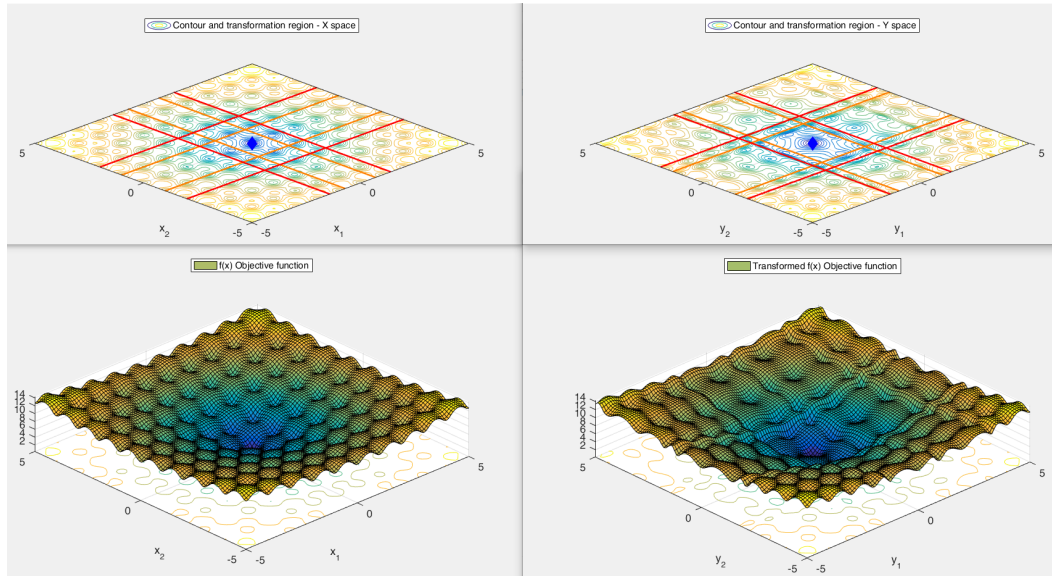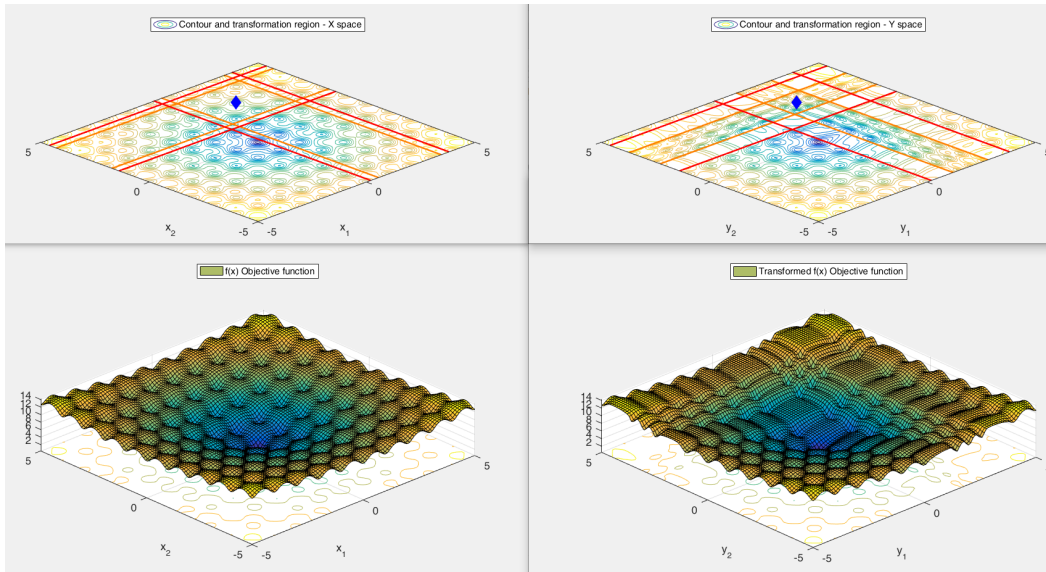
Figure 1.16. Recursion at iter $k = 2$. Another hyperellipsoidal neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperellipsoidal. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

NLT Contraction Considering now to make a contraction of two hyperellipsoidal neighborhoods. The first neighborhood is $\mathcal{B}_\infty(\hat{y}, \epsilon_y)$ of the point $\hat{x} = \hat{y} = (2, 3)$ with radius $\epsilon_x = 1.28$ bringing it to a radius of $\epsilon_y = 0.71$. The transformation boundary is $r_x = r_y = 1.6$. The same action is done for $\mathcal{B}_\infty(\hat{y}, \epsilon_y)$ of the point $\hat{x} = \hat{y} = (0, 0)$ with radius $\epsilon_x = 1.12$ bringing it to a radius of $\epsilon_y = 0.62$. The transformation boundary is $r_x = r_y = 1.4$. For the continuity of the transformation, the space in between the two hyperellipsoidal is shrunk.

The transformation parameters for the recursive hyperellipsoidal contraction iteration k=1 has been set as follows:

- $\sigma^1 = [0.9\ 1.0]$;
- $\epsilon_x^1 = 1.28$;
- $r_x^1 = 1.4$;
- $\theta_1^1 = 1.8$;
- $\theta_3^1 = 1.0$;

- $\nu^1 = [0.9\ 1.0]$;
- $\epsilon_y^1 = 0.71$;
- $r_y^1 = 1.6$;
- $\theta_2^1 = 0.36$;
- $\hat{x} = \hat{y} = (2, 3)$.

In figure 1.17 on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the recursion at iteration $k = 1$, in the $\mathcal{Y}$ space.
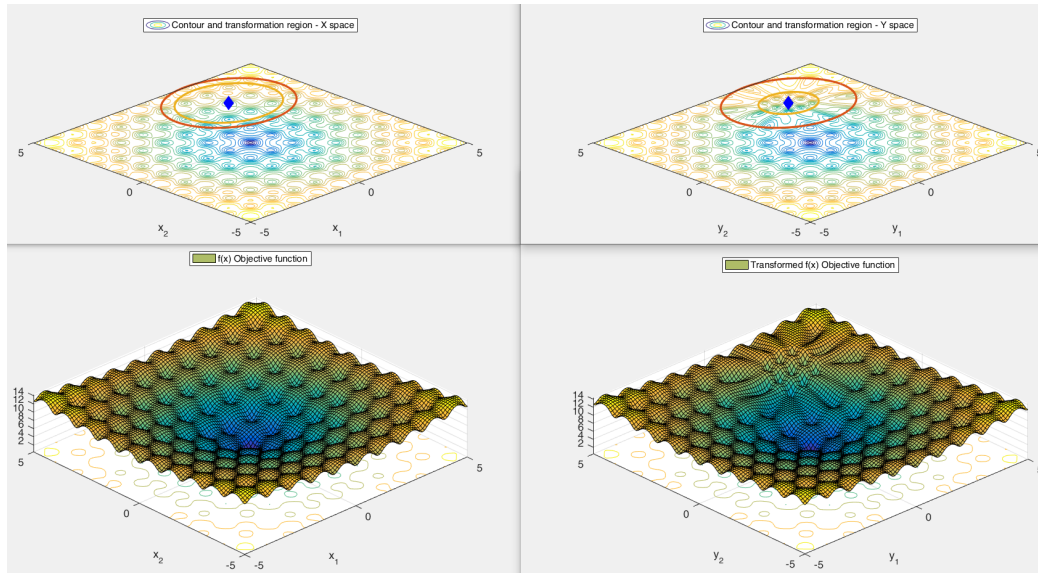
Figure 1.17. Recursion at iteration $k = 1$. An hyperellipsoidal neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperrectangles. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

The transformation parameters for the recursive hyperellipsoidal contraction iteration k=2 has been set as follows:

- $\sigma^2 = [1.0\ 0.9]$;
- $\epsilon_x^2 = 1.28$;
- $r_x^2 = 1.4$;
- $\theta_1^2 = 1.8$;
- $\theta_3^2 = 1.0$;

- $\nu^2 = [0.9\ 1.0]$;
- $\epsilon_y^2 = 0.71$;
- $r_y^2 = 1.4$;
- $\theta_2^2 = 0.36$;
- $\hat{x} = \hat{y} = (0, 0)$.

In figure (1.18) on the left there is the starting function and contour plot in the $\mathcal{X}$ space. On the right the result of the recursion at iteration $k = 2$, of the transformation in the $\mathcal{Y}$ space.
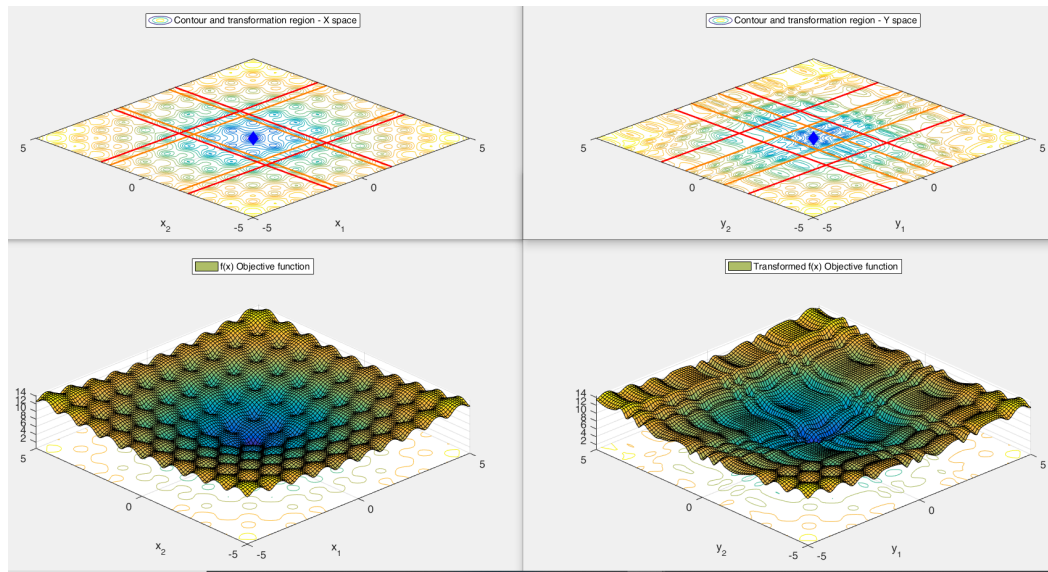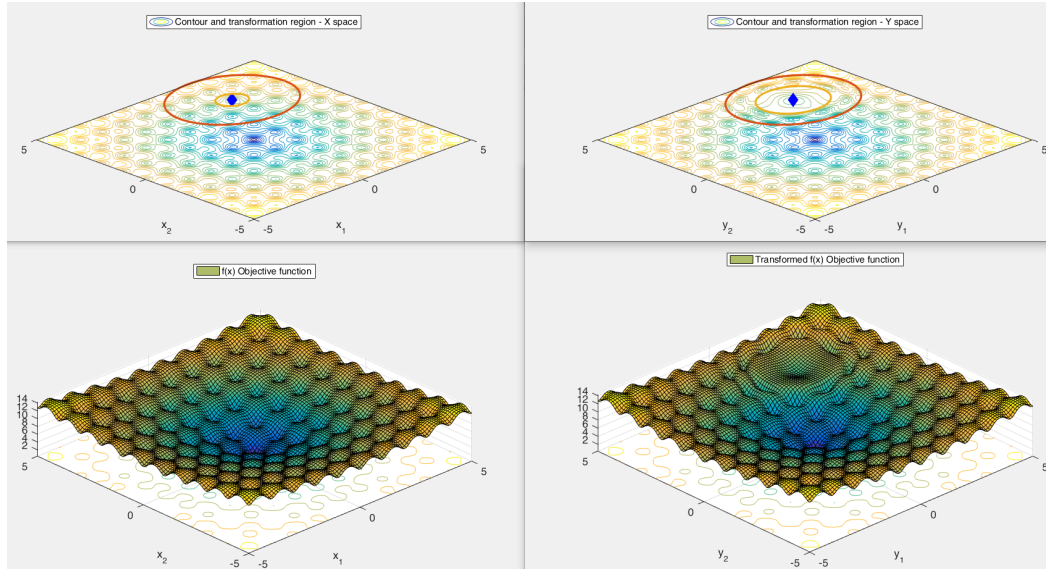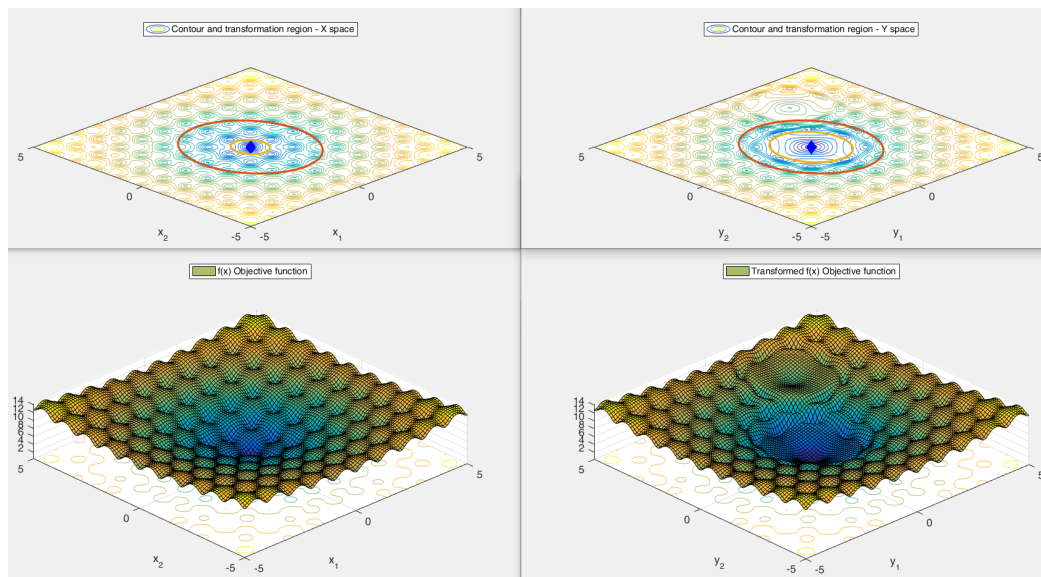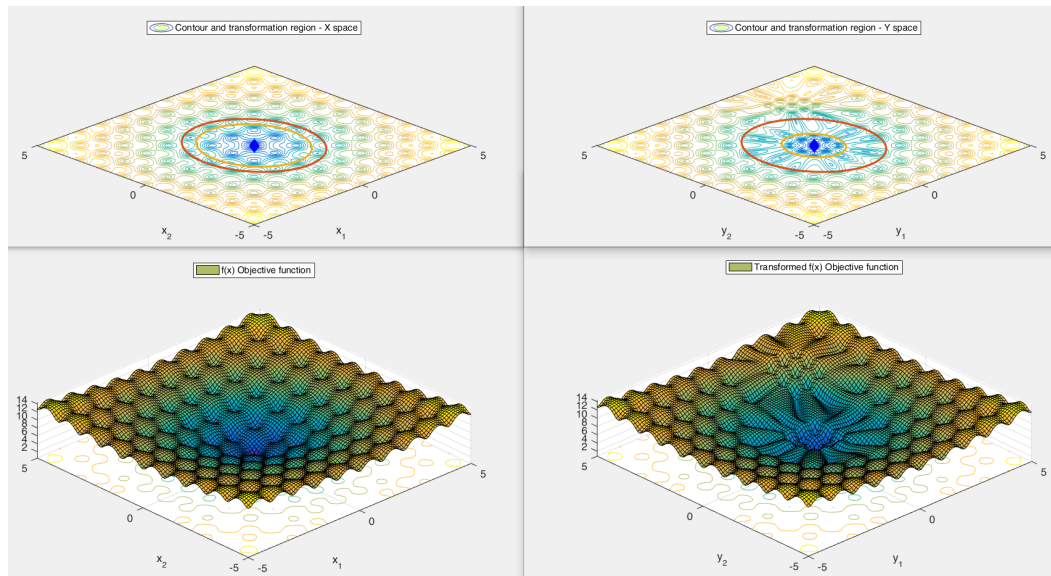
Figure 1.18. Recursion at iteration $k = 2$. Another hyperellipsoidal neighborhood of $\hat{x}$ has been expanded while shrinking the one in between the two hyperellipsoidal. Original $\mathcal{X}$ space on the left, transformed $\mathcal{Y}$ space on the right.

1.6   Scaling and Preconditioning

1.6.1   Scaling

In addition to the contraction and the expansion we have seen so far, the equations allow us to relate hyperspheres $\mathcal{B}(\hat{x}, r_x)$ in the original space, to scaled hyperspheres $\mathcal{B}(\hat{y}, r_y)$ in the transformed space. The scaling factor can be tuned by setting θ_3 to the ratio of the outer radius r_x/r_y of the considered hyperspheres. In figure (1.19) there is an example of an hypersphere in the $\mathcal{X}$ space centered in $\hat{x} = \hat{y} = (0, 0)$ with outer radius $r_x = 2.5$ that is related to an hypersphere in the $\mathcal{Y}$ space with outer radius $r_x = 3$.

The transformation parameters has been set as follows:

- $\epsilon_x = 2.0$;
- $r_x = 2.5$;
- $\theta_1 = 1.0$;
- $\theta_3 = 1.2$;

- $\epsilon_y = 2.0$;
- $r_y = 3.0$;
- $\theta_2 = 0.5$;
- $\hat{x} = \hat{y} = (0, 0)$.
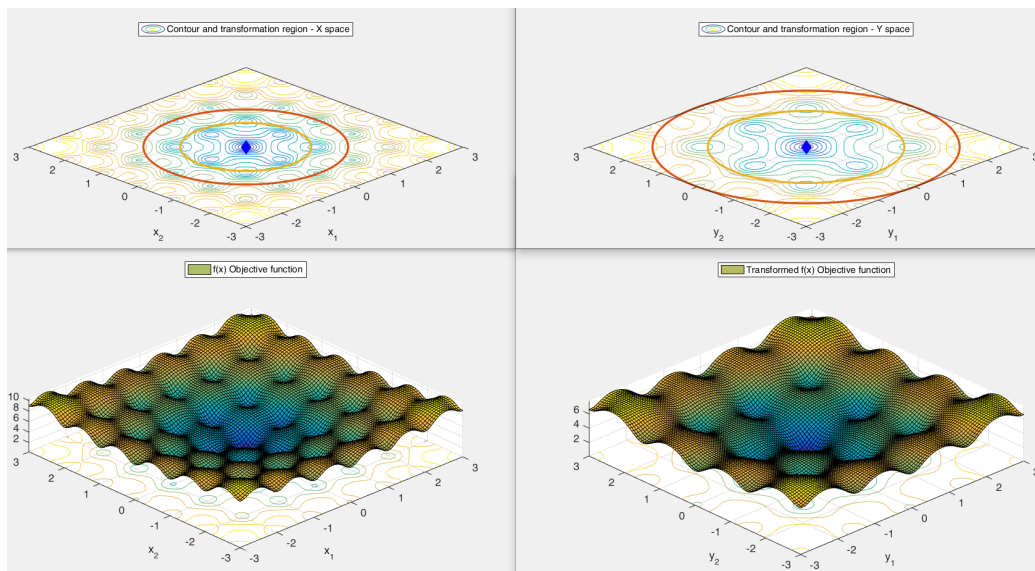
Figure 1.19. A spherical neighborhood of $\hat{x}$ has been scaled to an expanded one in the transformed space.

The same can be made for a contraction scaling by inverting the previous ratio and so by setting $\theta_3 = r_y/r_x$. The example is in figure (1.20) there is an example of an hypersphere in the $\mathcal{X}$ space centered in $\hat{x} = \hat{y} = (0, 0)$ with outer radius $r_x = 2.5$ that is related to an hypersphere in the $\mathcal{Y}$ space with outer radius $r_x = 3$.

The transformation parameters has been set as follows:

- $\epsilon_x = 2.0$;
- $r_x = 3.0$;
- $\theta_1 = 1.0$;
- $\theta_3 = 0.83$;

- $\epsilon_y = 2.0$;
- $r_y = 2.5$;
- $\theta_2 = 2.0$;
- $\hat{x} = \hat{y} = (0, 0)$.

Figure 1.20. An entire spherical neighborhood of $\hat{x}$ has been scaled to an expanded one in the transformed space. The innermost region of radius $\epsilon_x = \epsilon_x$ remain unchanged.

Scaling with unchanged innermost region.
The transformation equation can also be tuned to perform a generalized scaling but isolating a specific region to be unchanged. Let us consider the innermost transformed region

$$\mathcal{B}(\hat{y}, \epsilon_y) = (y \in \mathbb{R}^n, \ \epsilon_y > 0 \ : \ \|y - \hat{y}\| \leq \epsilon_y),$$

where the transformation equation involved is

$$\bar{x}(y) = \hat{x} + \theta_1 (y - \hat{y}),$$

with the coupling condition

$$\theta_1 \epsilon_y = \epsilon_x.$$

In order to keep the region unchanged we must set $\theta_1 = 1$, while setting $\theta_3 = r_x/r_y$. With this transformation to the hypersphere $\mathcal{B}(\hat{y}, \epsilon_y) \subseteq \mathcal{Y}$, corresponds an identical the hypersphere in the original space

$$\mathcal{B}(\hat{x}, \epsilon_x) = (x \in \mathbb{R}^n, \ \epsilon_x > 0 \ : \ \|x - \hat{x}\| \leq \epsilon_x),$$

The transformation parameters has been set as follows:

- $\epsilon_x = 2.0;$
- $r_x = 2.5;$
- $\theta_1 = 1.0;$
- $\theta_3 = 1.2;$

- $\epsilon_y = 2.0;$
- $r_y = 3.0;$
- $\theta_2 = 0.5;$
- $\hat{x} = \hat{y} = (0,0).$

- $\hat{x} = \hat{y} = (0,0).$
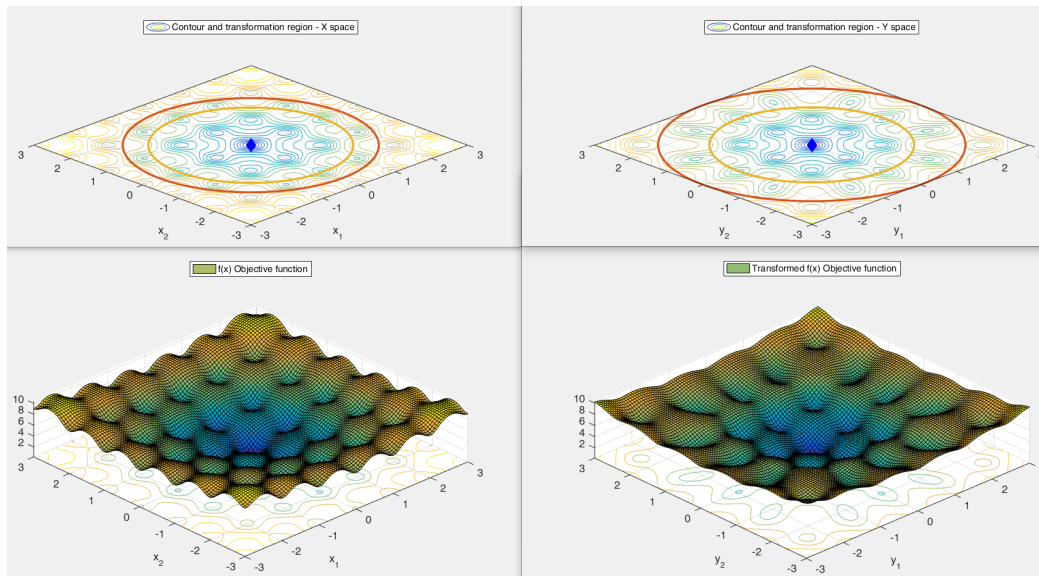
Figure 1.21. A spherical neighborhood of $\hat{x}$ has been scaled to an expanded one in the transformed space. The innermost region of radius $\epsilon_x = \epsilon_y$ remains unchanged.

The same can be made for a contraction scaling by inverting the previous the ratio and so by setting $\theta_3 = r_y/r_x$. The example is in figure (1.22) there is an example of an hypersphere in the $\mathcal{X}$ space centered in $\hat{x} = \hat{y} = (0,0)$ with outer radius $r_x = 2.5$ that is related to an hypersphere in the $\mathcal{Y}$ space with outer radius $r_x = 3$. The transformation parameters has been set as follows:

- $\epsilon_x = 2.0$;
- $r_x = 3.0$;
- $\theta_1 = 1.0$;
- $\theta_3 = 0.83$;

- $\epsilon_y = 2.0$;
- $r_y = 2.5$;
- $\theta_2 = 2.0$;
- $\hat{x} = \hat{y} = (0,0)$.

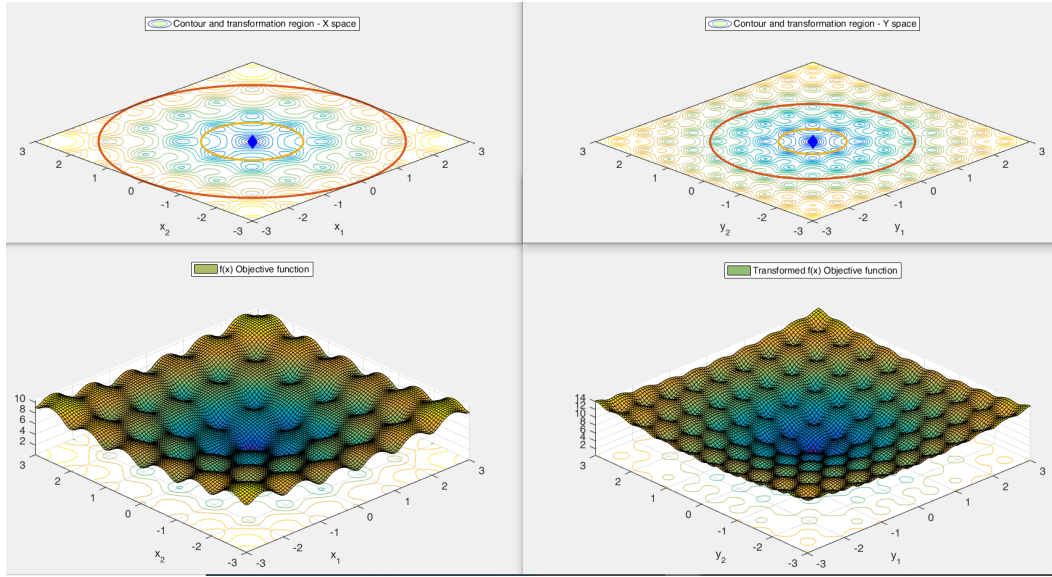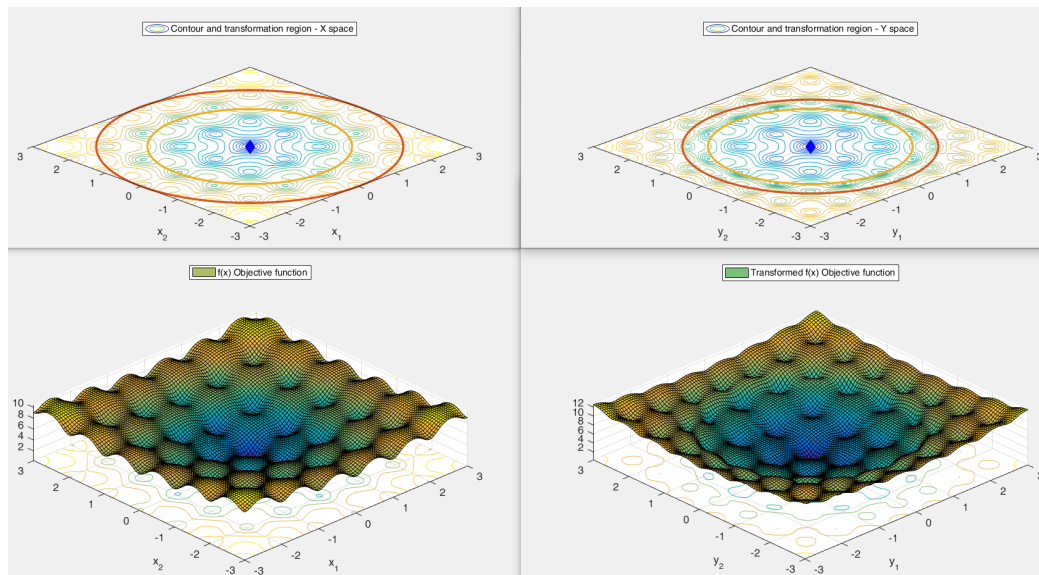**Figure 1.22.** A spherical neighborhood of $\hat{x}$ has been scaled to a contracted one in the transformed space. The innermost region of radius $\epsilon_x = \epsilon_y$ remains unchanged.

1.6.2   Preconditioning

The conditioning of a problem can be defined as the range (over a level set) of the maximum improvement of objective function value in a ball of small radius centered on a given level set. In the case of convex quadratic functions ($f(x) = \frac{1}{2}x^T H x$ where H is a symmetric definite matrix), the conditioning can be exactly defined as the condition number of the Hessian matrix H, i.e., the ratio between the largest and smallest eigenvalue. Since level sets associated to a convex quadratic function are ellipsoids, the condition number corresponds to the squared ratio between the largest and shortest axis lengths of the ellipsoid.

In optimization, the preconditioning is a technique exploited by algorithms which seeks to let an ill-conditioned problem be more straightforward to be tackled. In literature there are many methods for preconditioning but all are based on derivatives [27, 28, 29] or an approximation of them, such as finite differences [30, 31].

At the time of writing this thesis, it is under exploration how to apply the proposed transformations as preconditioning technique that neither use derivatives nor an approximation of them. In this subsection it is reported a possible way to proceed and a numerical example.

It considers the well known Rosenbrock function, also referred to as the Valley or Banana function, defined as follows

$$f(x_1 \cdots x_n) = \sum_{i=1}^{n-1} (\lambda(x_i^2 - x_{i+1})^2 + (1 - x_i)^2)$$

$$-3.0 \leq x_i \leq 3.0, \qquad \lambda \in \{1, \dots, 10^{10}\}$$

$$\text{minimum at } f(1, 1, \cdots, 1) = 0$$

The function is unimodal, non-separable, and the global minimum lies in a narrow, parabolic valley. However, even though this valley is easy to find, convergence to the minimum is difficult [32, 33]. Moreover for large enough λ and n, has one local minimum close to $x = [-1, 1, ..., 1]$, see also [35]. In figure (1.23) the contour lines of the 2D Rosenbrock function show a bent ridge that guides to the global optimum and the parameter λ controls the width o the ridge. In the classical Rosenbrock function λ is equals to 100. For smaller λ the ridge becomes wider and the function becomes less difficult to solve.

It considers a derivative-free algorithm DFL such as [34] (S. Lucidi, M. Sciandrone, 2002). The algorithm investigates the local behaviour of the objective function on the feasible set by sampling it along the coordinate directions and by performing a linesearch along suitable descend direction. When the stepsize α_i along a coordinate direction $i = 1, \dots, n$, differs too much, say, over a threshold τ, from the stepsize of the other coordinate directions, it might suggest that an ill-conditioning occurs in a neighborhood of the current best point $\hat{x}$.

Now consider to apply the non-linear transformation over hyperellipsoids introduced in subsection 1.4.7 in such a way as to prevent the ill-conditioning. It can
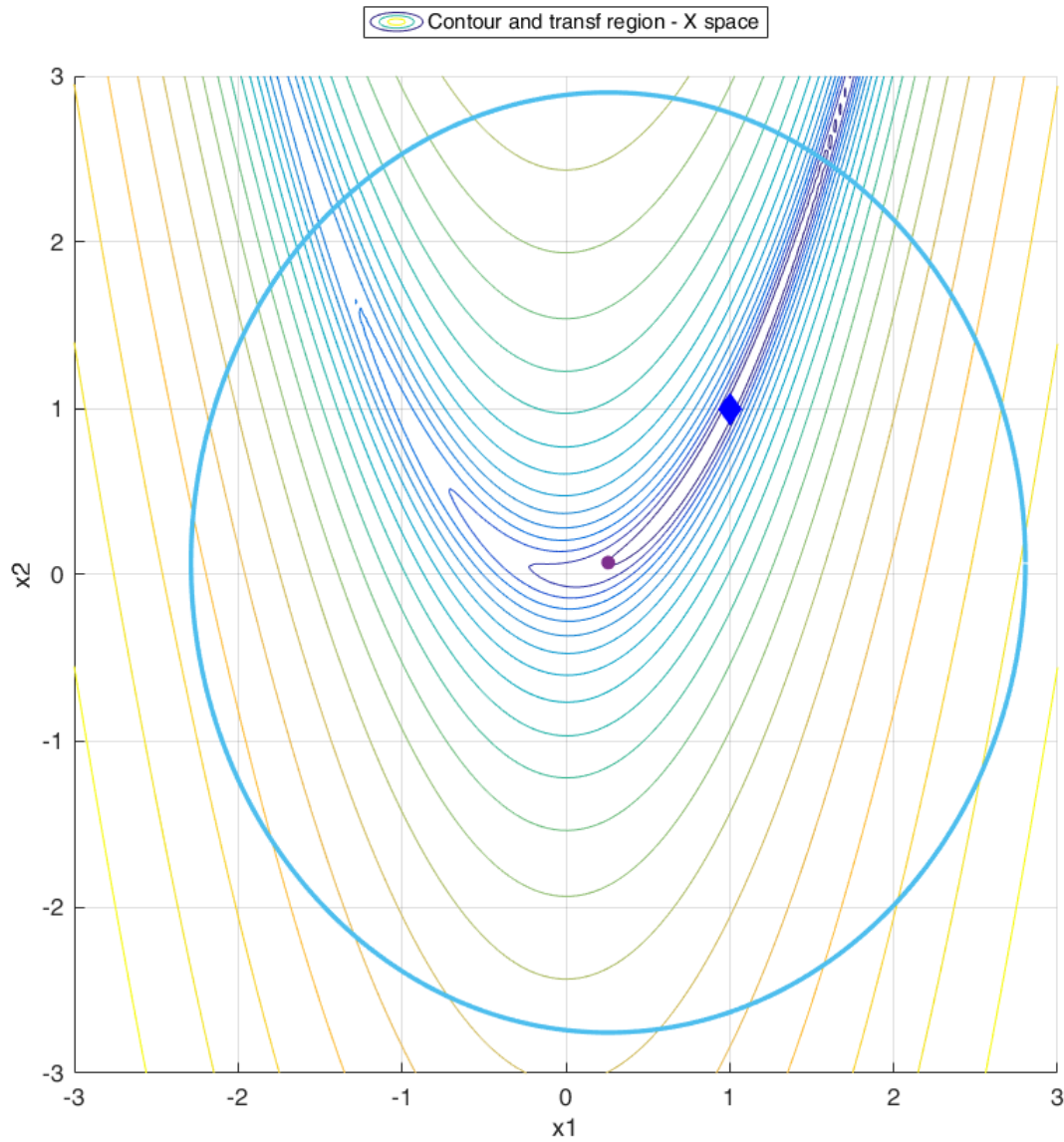
Figure 1.23. Contour plot of Rosenbrock function in 2D

.

be done by a suitable choice of the diagonal entries of the matrices D_x and D_y. In particular a possible choice is

$$
D_x = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n \end{pmatrix}, \quad D_y = \begin{pmatrix} \nu_1 & 0 & \dots & 0 \\ 0 & \nu_2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \nu_n \end{pmatrix},
$$

where $\sigma_i = \dfrac{\max\limits_i (\alpha_i)}{\alpha_i}, \quad \nu_i = 1, \quad i = 1, \dots, n.$

In figure (1.24), with the function parameter $\lambda = 10^2$, it is drawn the trace of 2102 improving points explored by the algorithm mentioned above in standard setting, until the stopping criterion of $\max_{i}(\alpha_i) < 10^{-6}$ is satisfied, without the use of any transformation.
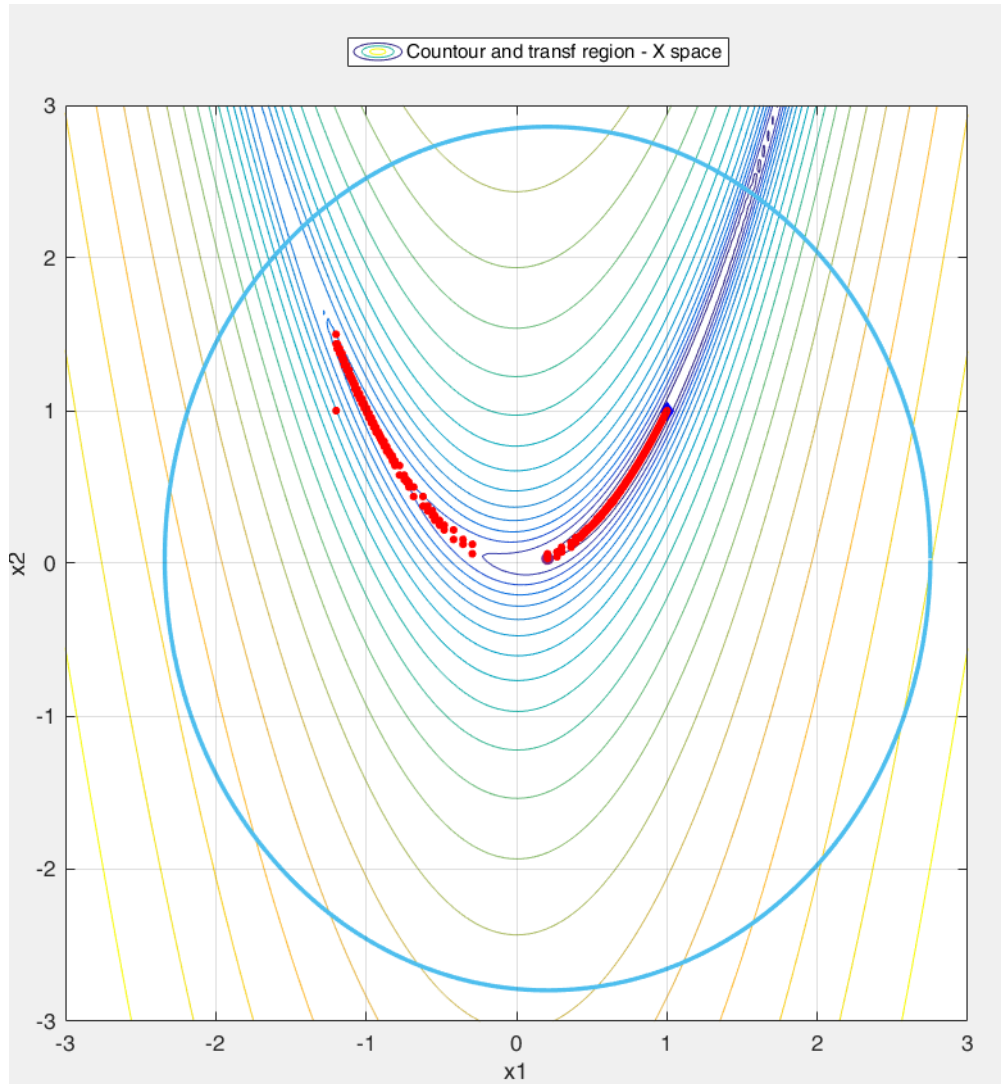
Figure 1.24. Trace of the improving points explored by the algorithm DFL
.

In figure (1.25), it is drawn the trace of the 1246 improving points explored by DFL algorithm with the use of the non-linear transformation.
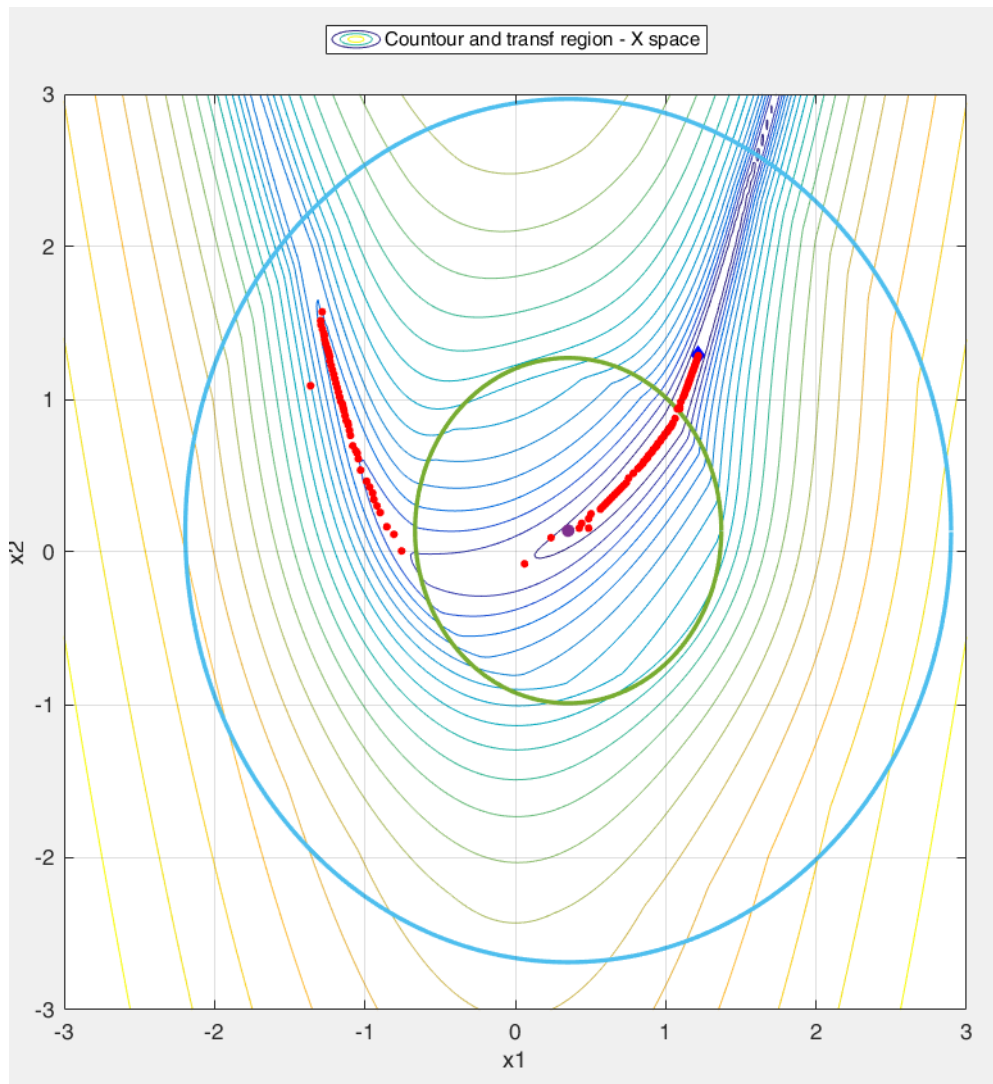
Figure 1.25. Trace of the improving points explored by the DFL with transformation.

The transformation is applied iteratively in a neighborhood of the current best solution $\hat{x}$ such that the path followed by the algorithm turns out to be smoother and it could exploit larger stepsize, meaning less function evaluations, until the stopping criterion is met. Tables (1.1, 1.2) report the results of DFL algorithm on the Rosenbrock, with and without the non-linear transformation, with different values of the function parameter λ and different thresholds $\tau = \frac{\alpha_i}{\alpha_{i+1}}$, $\frac{1}{\tau} = \frac{\alpha_{i+1}}{\alpha_i}$, $i = 1, \ldots, n-1$ that trigger the use of the transformation.

Table 1.1. DFL not exploiting non-linear transformation.

Function	Cond. Param.	Iter.	Funct. Eval.	Min. Value	Dim.	Threshold τ
Rosenbrock	100	2102	6270	0.000000	2	-
Rosenbrock	500	8992	26943	0.000002	2	-

Table 1.2. DFL exploiting non-linear transformation.

Function	Cond. Param.	Iter.	Funct. Eval.	Min. Value	Dim.	Threshold τ
Rosenbrock	100	1274	3784	0.000000	2	2
Rosenbrock	100	**1246**	**3700**	0.000000	2	**5**
Rosenbrock	100	1246	3702	0.000000	2	10
Rosenbrock	500	**5100**	**15264**	0.000001	2	**2**
Rosenbrock	500	5176	15492	0.000001	2	5
Rosenbrock	500	8992	26943	0.000002	2	10

The results seems promising. In particular the identification of a tailored dynamic strategy to select the τ threshold, and an accurate tuning of the transformation could be the keys to pursue the research.

1.7   Performance results

This section aims to provide preliminary results about the promising use of the transformation in global optimization algorithms. It is considered the DFL algorithm mentioned in the previous subsection in a multi-start framework that leverages variables transformations to gain an advantage for the search of the global minimum.

The goal is to find a good solution of multi-modal black-box global optimization problems. It is compared a simple multi-start approach against a multi-start that exploit the transformations. Both the algorithms were stopped either when the number of local searches exceeds the budget or when a point $\bar{x}$ is found such that:

$$\frac{f(\bar{x}) - f^*}{max\left(1, |f^*|\right)} \leq 10^{-5} \tag{1.61}$$

where $f(\bar{x})$ is the approximations of the global minimum value of the objective function f^* found by the multi-start algorithm.

For the testbed it is chosen the library *CEC' 2013 Benchmark Set for Real parameter Optimization*. The CEC'13 test problems library [24] was released on the occasion of the Special Session on Real-Parameter Optimization held in Cancun, Mexico. 20 - 23 June 2013, during IEEE Congress on Evolutionary Computation (CEC 2013). In figure (1.26) the details of the 28 problem classes available in dimension 2, 5, 10, 20, 50, 100.

As long as in the multi-start approach exploits randomly generated points as staring point for the local searches, the tests on these libraries were set in a stochastic fashion and it is summarized as follows

- Test problem: 28;

- Dimensions investigated 2, 5, 10;

- Budget local searches: 100 * dimensions;

- Stochastic runs per problem (random starting point): 10;

- Fixed seed for reproducibility of the pseudorandom number sequence (Mersenne Twister generator [36]).

It means that the total number of problems was 28 * 10 = 280. The following tables report the comparison of a simple multi-start approach against a multi-start that exploit the transformation. The dimensions investigated are 2, 5 and 10 for both the piecewise linear transformation and non-linear transformation and both the strategy of space expansion and space contraction. For each algorithm three key performance indicator are analysed:

- The mean value of the objective function over the 10 stochastic runs;

- The best value of the objective function over the 10 stochastic runs;

- The ability to find the global minimum value of the objective function.

	No.	Functions	$f_i^*=f_i(x^*)$
Unimodal Functions	1	Sphere Function	-1400
	2	Rotated High Conditioned Elliptic Function	-1300
	3	Rotated Bent Cigar Function	-1200
	4	Rotated Discus Function	-1100
	5	Different Powers Function	-1000
Basic Multimodal Functions	6	Rotated Rosenbrock's Function	-900
	7	Rotated Schaffers F7 Function	-800
	8	Rotated Ackley's Function	-700
	9	Rotated Weierstrass Function	-600
	10	Rotated Griewank's Function	-500
	11	Rastrigin's Function	-400
	12	Rotated Rastrigin's Function	-300
	13	Non-Continuous Rotated Rastrigin's Function	-200
	14	Schwefel's Function	-100
	15	Rotated Schwefel's Function	100
	16	Rotated Katsuura Function	200
	17	Lunacek Bi_Rastrigin Function	300
	18	Rotated Lunacek Bi_Rastrigin Function	400
	19	Expanded Griewank's plus Rosenbrock's Function	500
	20	Expanded Scaffer's F6 Function	600
Composition Functions	21	Composition Function 1 (n=5,Rotated)	700
	22	Composition Function 2 (n=3,Unrotated)	800
	23	Composition Function 3 (n=3,Rotated)	900
	24	Composition Function 4 (n=3,Rotated)	1000
	25	Composition Function 5 (n=3,Rotated)	1100
	26	Composition Function 6 (n=5,Rotated)	1200
	27	Composition Function 7 (n=5,Rotated)	1300
	28	Composition Function 8 (n=5,Rotated)	1400
Search Range: $[-100,100]^D$			

Figure 1.26. CEC13 problems. https://www.ntu.edu.sg/CEC2013/CEC2013.htm

The results are presented in terms of number of wins over the 28 function classes, it means that each value in the tables reveals the number of times an algorithm wins against the other. If this situation doesn't occur it means that both algorithms has the same performance.

Table 1.3. Hyperrectangle expansion in 2D. Number of of wins.

	Multi-Start Transformation	Multi-Start Simple	Both algorithms
Mean values Objective Function	9	1	18
Best values Objective Function	1	0	27
Global minimum values found	1	0	24

Table 1.4. Hyperrectangle expansion in 5D. Number of of wins.

	Multi-Start Transformation	Multi-Start Simple	Both algorithms
Mean values Objective Function	24	1	3
Best values Objective Function	19	2	7
Global minimum values found	0	0	4

Table 1.5. Hyperrectangle expansion in 10D. Number of of wins.

	Multi-Start Transformation	Multi-Start Simple	Both algorithms
Mean values Objective Function	23	0	5
Best values Objective Function	14	0	14
Global minimum values found	0	0	4

Table 1.6. Hyperrectangle shrinking in 2D. Number of of wins.

	Multi-Start Transformation	**Multi-Start Simple**	**Both algorithms**
Mean values Objective Function	8	1	19
Best values Objective Function	1	0	27
Global minimum values found	1	0	24

Table 1.7. Hyperrectangle shrinking in 5D. Number of of wins.

	Multi-Start Transformation	**Multi-Start Simple**	**Both algorithms**
Mean values Objective Function	24	0	4
Best values Objective Function	10	1	17
Global minimum values found	1	0	7

Table 1.8. Hyperrectangle shrinking in 10D. Number of of wins.

	Multi-Start Transformation	**Multi-Start Simple**	**Both algorithms**
Mean values Objective Function	23	0	5
Best values Objective Function	16	0	12
Global minimum values found	0	0	4

Table 1.9. Ellipsoidal expansion in 2D. Number of of wins.

	Multi-Start Transformation	Multi-Start Simple	Both algorithms
Mean values Objective Function	9	0	19
Best values Objective Function	4	0	24
Global minimum values found	1	0	24

Table 1.10. Ellipsoidal expansion in 5D. Number of of wins.

	Multi-Start Transformation	Multi-Start Simple	Both algorithms
Mean values Objective Function	24	0	4
Best values Objective Function	12	1	15
Global minimum values found	2	0	7

Table 1.11. Ellipsoidal expansion in 10D. Number of of wins.

	Multi-Start Transformation	Multi-Start Simple	Both algorithms
Mean values Objective Function	23	0	5
Best values Objective Function	15	0	13
Global minimum values found	0	0	4

Table 1.12. Ellipsoidal shrinking in 2D. Number of of wins.

	Multi-Start Transformation	**Multi-Start Simple**	**Both algorithms**
Mean values Objective Function	4	1	23
Best values Objective Function	1	0	27
Global minimum values found	1	0	24

Table 1.13. Ellipsoidal shrinking in 5D. Number of of wins.

	Multi-Start Transformation	**Multi-Start Simple**	**Both algorithms**
Mean values Objective Function	24	0	4
Best values Objective Function	10	1	17
Global minimum values found	1	0	7

Table 1.14. Ellipsoidal shrinking in 10D. Number of of wins.

	Multi-Start Transformation	**Multi-Start Simple**	**Both algorithms**
Mean values Objective Function	23	0	5
Best values Objective Function	14	0	14
Global minimum values found	0	0	4

1.8   Comments

The obtained numerical results show that the performance of the multi-start approach could be improved by exploiting the transformations.

It is clear that the proposed transformations can be integrated with more complex algorithm schemes, with a stronger ability in finding global solution than the one of the multi-start. The study of more powerful global optimization algorithms and the integration with a space expansion-contraction strategy will be the objective of future works.

Chapter 2

Exploratory geometries and space search reduction in GABRLS algorithm

The adoption of an accurate exploratory geometry and a space search reduction strategy can speed up significantly the convergence towards better solutions.

In this chapter it is presented a significant advancement of the global optimization algorithm GABRLS introduced in [113] (Romito, 2017). The GABRLS algorithm was the winner of the Generalization-based Contest in Global Optimization (GENOPT 2017) [61]. It is originally designed for continuous problems in the black-box environment, meaning problems for which the analytic formulation it is not known, or it is not convenient to directly solve it in terms of time to spend and amount of resource needed. The advanced version is able to handle continuous and discrete variables over a bound constrained set.

The problem of minimizing a continuously differentiable function of several variables, where some of them are restricted to take discrete or integer values, arises frequently in many industrial and scientific applications and this motivates the increasing interest in the study of new derivative-free methods for their solution.

The approach has been tested in a real case study of *design optimization* of electric motors.

2.1 Preliminary Concepts

Let $f(x)$ be continuously differentiable function within a feasible hyper-interval $\mathcal{X}$. The bound constrained minimization problem can be stated as follows:

$$\min_{x \in \mathcal{X}} f(x),$$
$$\mathcal{X} = \begin{cases} x_i \in \mathbb{R} : lb_i \leq x_i \leq ub_i \ i \in I_c \\ x_i \in \mathbb{D} : lb_i \leq x_i \leq ub_i \ i \in I_d \end{cases}, \tag{2.1}$$

where $lb < ub$ are the lower and upper bound in the variables, while I_c, I_d are the index sets of real and discrete variables respectively. The feasible set is

assumed dense (and not sparse as in the integer programming). It is common in many real-life applications where, despite the continuous nature of the underlying models, a number of practical issues prescribe rounding of values. It is pointed out that the work environment is totally black-box, meaning the exact analytical expression of $f(x)$ and its first order derivatives cannot be explicitly calculated or approximated. Usually, in this context, direct search methods are widely used. They are based on appropriate sampling strategies of the feasible set. Some of the techniques to approach to the problem (2.1) can be included in the methods based on one-dimensional searches that do not use derivatives [75, 76] or well structured set of points [69].

2.1.1 Searches based on sets of directions

In derivative-free optimization the selection of an appropriate set of directions is fundamental to fill the lack of derivatives. A simple way to identify a finite number of search direction is the use of the so called Positively Spanning Set that are widely used in direct search algorithms. In [62] can be found an extensive analysis. A well-known result is that if the gradient of a continuously differentiable objective function on $\mathbb{R}^n$ is nonzero at a point, then one of the vectors in any positive spanning set of $\mathbb{R}^n$ is a descent direction for the objective function from that point. A common used definition is the following.

Definition 1 Given the set of vectors $\mathcal{D}_+ = \{d_1, ..., d_m\} \in \mathbb{R}^n$, we say that $\mathcal{D}_+$ is a Positively Spanning Set (PSS) if for $n + 1 \leq m \leq 2n$ and any vector $v \in \mathbb{R}^n$ it has:

$$v = \sum_{j=1}^{m} \alpha_j d_j, \quad \alpha_j \geq 0, \tag{2.2}$$

i.e any vector $v \in \mathbb{R}^n$ can be expressed as the weighted sum of the vectors in D_+, using nonnegative weights. In particular if the size of a positive basis is $n + 1$ (the minimal positive bases) the maximal value of the cosine measure is $1/n$. A straightforward corollary is that the maximal cosine measure for any positive spanning set of size $n + 1$ is $1/n$. If the size of a positive basis is $2n$ (the maximal positive bases) the maximal cosine measure is $\sqrt{1/n}$. Upper bounds of the cosine measure are analysed in [81]. Examples of popular algorithms that make use of PSS include Pattern Search [63], Mesh Adaptive Direct Search [64] and Implicit Filtering [65].

2.1.2 Searches based on grid points

As well as line-search methods [34, 66] or trust region methods [67], grid-based methods are strongly investigated for both constrained and unconstrained optimization. In the studies of I. D. Coope and C. J. Price [68, 70, 71] can be found a full overview. They provided different general schemes of algorithms and proved the global convergence. It is redrafted two definitions of the points generated by grid-based methods in presence of bound constraints that are useful for the algorithmic scheme that will be introduced in the next section.

Definition 2 Given a maximal PPS of directions $\mathcal{D}_+ = \{d_1, ..., d_{2n}\} \in \mathcal{X} \subset \mathbb{R}^n$, the feasible points of a grid $\mathcal{G}^{(k)}$ are

$$\mathcal{G}^{(k)} = \left\{ x \in \mathcal{X} : x = \hat{x}^{(k)} + h^{(k)} \left(\sum_{j=1}^{n} \eta_j d_j + \sum_{j=1}^{n} \eta_{n+j}(-d_{n+j}) \right) \right\}, \qquad (2.3)$$

where

- $\hat{x}^{(k)} \in \mathcal{X}$ is the offset between different grids generated for $k = 0, 1...$;

- η is any integers such that $\forall j$

  - $|\eta_{n+j} d_{n+j}| \leq \left\lfloor \dfrac{\hat{x}_j^{(k)} - lb_j}{h^{(k)}} \right\rfloor, \quad$ if d_{n+j} is pointing to lb_j,

  - $|\eta_j d_j| \leq \left\lfloor \dfrac{ub_j - \hat{x}_j^{(k)}}{h^{(k)}} \right\rfloor, \quad$ if d_j is pointing to ub_j;

- $h^{(k)}$ is a positive scalar parameter

  - $0 < h^{(k)} \leq \min_{j=1,...,n} \left(\hat{x}_j^{(k)} - lb_j , ub_j - \hat{x}_j^{(k)} \right),$

    that adjust the mesh size as k is increased and it is fundamental to establish convergence by ensuring that consecutive mesh are suitably finer.

Now, a particular set of grid points called the local grid minimizers can be characterized.

Definition 3 Consider a maximal PPS $\mathcal{D}_+$. A point $x \in \mathcal{G}^{(k)}$ is a grid local minimum of $f(x)$ with respect to $\mathcal{D}_+$ if and only if

$$f(x + h^{(k)} d_j) \geq f(x) \quad \forall d_j \in \mathcal{D}_+ . \qquad (2.4)$$

Grid local minimizers are nothing other than finite difference approximations of the local optima in continuous optimization.

Now the attention is addressed to the minimization of the bound constrained optimization problem (2.1). The black-box context in real life problems hardly permit to design and solve more detailed models. In fact eventually one can make use of penalty [74] or barrier [60] approach to manage general constraints and so make the assumption that the non-linear constrained problem is approximated by a sequence of box-constrained problems. The presence of general nonlinear constraints is handled by using an exact penalty approach. Since only the violation of

such constraints is included in the penalty function, the algorithms developed for bound constrained problems can be used to minimize the penalized problem, which is proved to be equivalent to the original one under reasonable conditions. It is important to point out that the strategy of penalizing only the general nonlinear constraints have been successfully adopted in many researches from the literature related to derivative-free optimization [77, 78, 79].

2.2 The GABRLS algorithm

The algorithm GABRLS is a modified genetic algorithm (GA) in which have been successfully introduced an effective exploratory geometry (BR) and hybridization with local search (LS) to speed up the classical GA scheme. Hereinafter, we refer to GABR as to identify the global search strategy, disregarding the hybridization with any local search (LS). The salient points that let GABR be distinct from the other GAs are essentially two:

   (i) Effective geometry of the set of points iteratively generated (population).

   (ii) Space search reduction strategy for steering the search towards the most promising area.

The original version of (i) can be found in [15] in sec. 3.1, whereas (ii) in sec. 3.2, where the novel Bounding Restart (BR) technique has been introduced. The following two subsections provide a recap of the original version and the details of the updated version of the algorithm.

2.2.1 Adjustments on the GA phase

As in [15] starting from a classical formal scheme of GA (see Algorithm (2.1) below) it is detailed the change at each step.

Algorithm 3.1. Classical GA scheme

---

   1. Initial Population   (points generated in $\mathcal{X}$)

   2. **for** $k = 1 \longrightarrow generations$ (max iterations)

   3.   Evaluate Population

   4.   Selection Criterion   (Tournament, Elitism, etc.)

   5.   Genetic Operators   (Crossover – Mutation)

6.   Evolved Population   (new points generated in $\mathcal{X}$)

7. **end for**

Initial population. It is introduced a discretization technique to place the initial points (initial population) in the feasible domain, in such a way as to lie on nodes of a multi-dimensional grid. The original AxialPI routine has been modified in this regard. It allows the discrete positioning of points along axes that are parallel to the coordinate ones, unless an offset. The offset $\hat{x}$ is the center of the current feasible hyper-rectangle (Fig. 1 provides an example in a 3D box). Denoting with $Space_j$ the amplitude of the interval along the jth dimension, with $Increase_j$ the step along the jth dimension between two consecutive points and with Pop the number of initial points where the objective function f is evaluated, the result is the Algorithm (2.2).
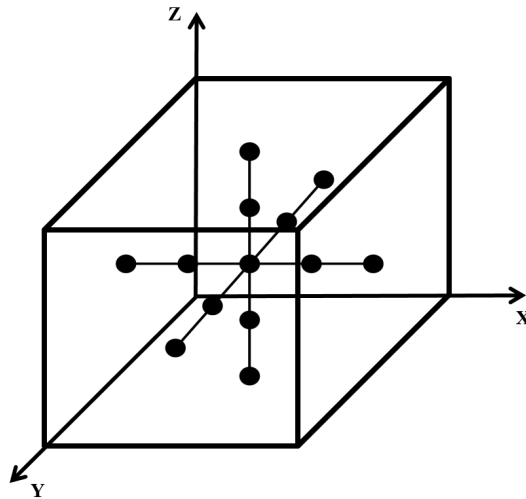
Figure 2.1. Graphic view of the points generated with AxialPI in a 3D box.

Algorithm 3.2. AxialPI

---

1. $n = |\mathcal{X}|,\ AxialPop = (Pop - 1)/n,\ Pop = (p \cdot 2n) + 1,\ p \geq 1$ integer

2. **for** $j = 1 \longrightarrow n$

3. $Space_j = (ub_j - lb_j)$

4. $Increase_j = Space_j/(AxialPop)$

5. $\hat{x}_j = (ub_j + lb_j)/2$

6. $Point_{1j} = \hat{x}_j$

7. **end for**

8. $z = 0$

9. $t = 1$

10. **for** $i = 2 \longrightarrow k = (Pop - 1)/2$

11.   **if** $(z = AxialPop/2)$ **then**

12. $z = 0$

13. $t = t + 1$

14.   **end if**

15. $z = z + 1$

16.   **for** $j = 1 \longrightarrow n$

17.   **if** $(j = t)$ **then**

18. $Point_{i,j} = \hat{x}_j + (z) \cdot Increase_j$

19. $Point_{k+i,j} = \hat{x}_j - (z) \cdot Increase_j$

20.   **else**

21. $Point_{i,j} = \hat{x}_j$

21. $Point_{k+i,j} = \hat{x}_j$

22. **end if**

23. **end for**

24. **end for**

Evaluation and Selection criterion. At the step 2 of Algorithm (2.1) the GA main loop starts. The step 3 and 4 are free, meaning one may use any suitable strategy as evaluation and selection criterion. A comparative analysis on the selection rules for GA can be found in [72].

Genetic operators. With respect to classical GA operators, it is employed a single-point crossover operator without recombination of blocks (see more in [15], section 3.1) in order to move the points lieing on a grid, while it is let mutation vary in any point of $\mathcal{X}$. In other words, the mutation operator grants to reach (in probability) any feasible point (therefore also global optima) in $\mathcal{X}$, while the crossover operator in conjunction with AxialPI routine allow to move a point to any other point in $\mathcal{X} \cap \mathcal{G}^{(k)}$. Let us consider points $x_i^{(k)}(j)$, $i = 1 \rightarrow Pop$, $j = 1 \rightarrow n$ and $k \geq 0$ is the index of the current population (iteration). Equations (2.5) show how the crossover works with two points.

$$x_1^{(k+1)} = \{x_1^{(k)}(1), ..., x_1^{(k)}(j-1), x_2^{(k)}(j), ..., x_2^{(k)}(n)\}, \qquad (2.5a)$$

$$x_2^{(k+1)} = \{x_2^{(k)}(1), ..., x_2^{(k)}(j-1), x_1^{(k)}(j), ..., x_1^{(k)}(n)\}. \qquad (2.5b)$$

All possible positions in which the crossover will be able to move the points are known a priori and are all nodes of a multi-dimensional mesh (Fig. 2.2).
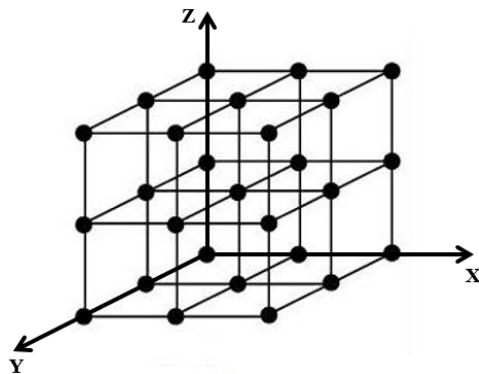
Figure 2.2. All possible points that crossover allows to reach through the recombination of the initial ones generated by AxialPI in a 3D box.

New population. In the main loop of Algorithm (2.1), for each $k > 0$ this step identify an evolved population of points, among all possible nodes of the lattice, increasingly in a neighborhood of the most promising area. It means that the evolution of the initial population of points let the lattice be more dense where the objective function values become lower.

Termination criterion. The criterion of a max number of generations to allow the evolution of the initial population is met and Algorithm (2.1) ends. Actually one can develop any suitable criterion to stop the GA search.

Since GAs are global optimization algorithms, other routines can be integrated in the main loop in order to speed up the search and to enhance the efficiency. After a number of iterations one can check the homogeneity of points and enforce a diversification to avoid premature convergence or, for very expensive black box problems, one can save time not doing evaluations of identical points. Both these strategy are very useful but it would be a good idea to make use of them after a sufficient number of iterations. As it comes up from experimental results, a reasonable choice is at least one half of the budget iterations avoiding to deceive the algorithm. The following formal scheme sums up all the analysis carried out until now.

Algorithm 3.3. Modified GA scheme

---

1.   Initial Population   (Discrete points positioning - AxialPI)

2.   **for** $k = 1 \longrightarrow generations$ (max iterations)

3.   Evaluate Population   (at least $|\mathcal{D}_+|$ points if $k = 1$)

4.   Selection Criterion   (any suitable routines)

5.   Genetic Operators   (CrossOver – Mutation)

6.   **if** (k>generations/2) **then**

7.   Check premature convergence   (any suitable routines)

8.   **end if**

9.   New Population

10. **end for**

At this point, it is recalled the Bounding Restart technique ([15] sec 3.2) and then we show how it is able to manage finer and finer grids iteratively generated satisfying the condition that all points, excepted the mutated ones, lie on these grids.

2.2.2 Properties of Bounding Restart (BR)

We all known that GAs perform a weaker search when the dimension of problems increase or the search space is large. An iterative space reduction technique can be useful to avoid too many function evaluations to locate a good solution.

BR is a two-step technique that fulfils the requirement of efficiency, in which GA can be successfully integrated. In literature this technique is recently called "zoom-in" strategy by Jones D.R. in his work *"The DIRECT algorithm: 25 years Later"* [21] and exploited to speed-up the efficiency of the well known deterministic partitioning algorithm DIRECT. We call the first step of BR as bounding step. Assuming that the genetic algorithm has the ability to quickly identify a promising area, however large, it's reasonable to focus the search in the given area temporarily. On that basis, by means of hyper-intervals that are dynamically resized, according to the need to lead the search towards the most promising area, the bounding step at the generic iteration k is carried out through the following equations:

$$LB^{(k)} = \frac{ub + lb}{2} - \frac{ub - lb}{2 \cdot CF^\lambda}, \quad CF \in \mathbb{R}: \ CF = constant > 1, \ \lambda \in \mathbb{N}, \quad (2.6a)$$

$$UB^{(k)} = \frac{ub + lb}{2} + \frac{ub - lb}{2 \cdot CF^\lambda}, \quad CF \in \mathbb{R}: \ CF = constant > 1, \ \lambda \in \mathbb{N}. \quad (2.6b)$$

CF is a convergence factor that has a high impact on reducing the bounds. The reduction is managed by increasing λ according to the speed one want to proceed to the search of a solution. After updating lower and upper bounds, the reduced set is centred in the best point $\bar{x}^{(k)}$ currently known. Let $\hat{x}^{(k)}$ be the offset between the current best solution $\bar{x}^{(k)}$ and the centre of the kth reduced hyper-interval then, taking into account the bound constraints, the set can be put centrally as follows:

$$\hat{x}^{(k)} = \bar{x}^{(k)} - \frac{UB^{(k)} + LB^{(k)}}{2}, \quad (2.7)$$

$$LB^{(k+1)} = \max\left\{lb, LB^{(k)} + \hat{x}^{(k)}\right\}, \quad (2.8a)$$

$$UB^{(k+1)} = \min\left\{ub, UB^{(k)} + \hat{x}^{(k)}\right\}. \quad (2.8b)$$

The second step of BR is the restart of the modified GA (Algorithm (2.3)) in the new reduced hyper-interval. A suitable stopping criterion for the reduction of the

hyper-interval can be the precision with which approximate the solution. Considering $\delta > 0$ as acceptable threshold for an approximated solution $\bar{x}$. Let assume that the threshold has been attained at the $\bar{k}^{th}$ BR reduction cycle. Considering $\lambda = \bar{\lambda}$, where $\bar{\lambda}$ is the number of bounding step performed until now, an upper bound for the reduced feasible space is

$$\max_{j \in \{1,...,n\}} Space_j^{(\bar{k})} = |UB_{j_M}^{(\bar{k})} - LB_{j_M}^{(\bar{k})}| = \frac{|ub_{j_M} - lb_{j_M}|}{CF^{\bar{\lambda}}} \leq 2\delta, \qquad (2.9)$$

where j_M is the index of the maximum length among the dimensions of the current hyper-interval. It is the same as identifying an hyper-interval neighborhood $\mathcal{H}_\infty(\bar{x}, \delta)$ about $\bar{x}$ defined by the maximum norm $\|x_j - \bar{x}_j\|_\infty \leq \delta, x \in \mathcal{X}$. The following Algorithm (2.4) shows the overall integration scheme of GABR. It is a generalized scheme in which only a simple decrease condition is required as to declare an iteration successful.

Algorithm 3.4. GABR scheme - Simple decrease acceptance criterion

---

1. $k = 0, \; \delta > 0, \; D^0 \subseteq \mathcal{X}, \; x^0 = (UB^0 + LB^0)/2, \; LB^0 = lb, \; UB^0 = ub,$

    $Space^0 = (ub - lb)/CF^\lambda, \; CF = constant > 1, \; \lambda = 0$

2. **while** $(Space^{(k)} > 2\delta)$ **or** (other stopping criterion)

3.   run **Algorithm 3** in $D^{(k)}$ and found the current minimizer $\bar{x}^{(k)}$

4.   **if** $(\, f(\bar{x}^{(k)}) < f(x^{(k)}) \,)$ **then**

5. $x^{(k)} = \bar{x}^{(k)}$

6. $k = k + 1$, **continue**

7.   **else**

8. $k = k + 1$

9. $\lambda = \lambda + 1$, perform a **Bounding step**, eqs. (2.6),(2.7),(2.8)

10. **end if**

11. **end**

___

In the Algorithm (2.4), until the stopping criterion is not verified, the GA phase is restarted. At each iteration, if a simple decrease in the objective function does not occur, the feasible hyper-interval is reduced, it remain the same otherwise. Thus, for each k, the step 1 of Algorithm (2.3) is repeated to generate a new set of initial points through AxialPI routine in the reduced set. It means that for each unsuccessful iteration we have the same number of initial points placed closer and closer one to another along axes that are parallel to the coordinate ones and the current best solution is their origin. It is an extended search over each dimension that is performed along the axes at the current iteration k in which the stepsize $\alpha_j^{(k)}$ and trial points $^p x_j^{(k)}$, $j = 1 \rightarrow n$, $p = 1 \rightarrow AxialPop$, are defined by

$$^p x_j^{(k)} = \ ^p \bar{x}_j^{(k)} + \alpha_j^{(k)} d_j = \ ^p \bar{x}_j^{(k)} + p\, \beta_j^{(k)} d_j, \quad d_j \in \{e_j, -e_j\}, \tag{2.10}$$

$$\beta_j^{(k)} = \frac{ub_j - lb_j}{CF^\lambda(AxialPop - 1)} = \frac{UB_j^{(k)} - LB_j^{(k)}}{(AxialPop - 1)}, \quad AxialPop > 1, \tag{2.11}$$

where, recalling Algorithm (2.2), $AxialPop$ is the number of search points along an axis. For each k, the equation (2.11) is the minimal distance between points along the jth axis, whereas the equation (2.10) identify all points along the jth axis of a PSS and for $p = 0$ we have the starting point that is the current best solution. Every point except ones on the bounds have two adjacent points that are a simultaneous trial expansion and reduction of the stepsize. Figure (2.3) shows the points produced by (2.10), (2.11) in two consecutive iteration of the one-dimensional searches identified by iterative restarting AxialPI in a reduced hyper-interval. In particular, in the example figure the iteration k is a failure, hence the feasible set is reduced and the GA phase restarted (and so also AxialPI) for the iteration $k + 1$.

BR is not only a simple technique to speed up the search of a genetic algorithm. Another property of BR is that it is able to proficiently manage grids. It is not immediate if one look at BR separately from the modified GA scheme and from the definition of grid. For each BR step an instance of (Algorithm 2.3) is run and so for each inner iteration (GA main loop), the operator of crossover is carried out through equations (2.5) such that the initial points can be moved lieing on a grid. Now following the steps of the Algorithm (2.4) it can be identified a reformulation of the grid (Definition 2, (2.3)) by using the equations (2.6) that iteratively redefine its bounds and meshsize.

Definition 5 Given Algorithm (2.4) and a maximal PPS $\mathcal{D}_+ = \{d_1, ..., d_{2n}\}$ that positively span the hyper-intervals $D^{(k)} \subseteq \mathcal{X}$ iteratively generated. The feasible
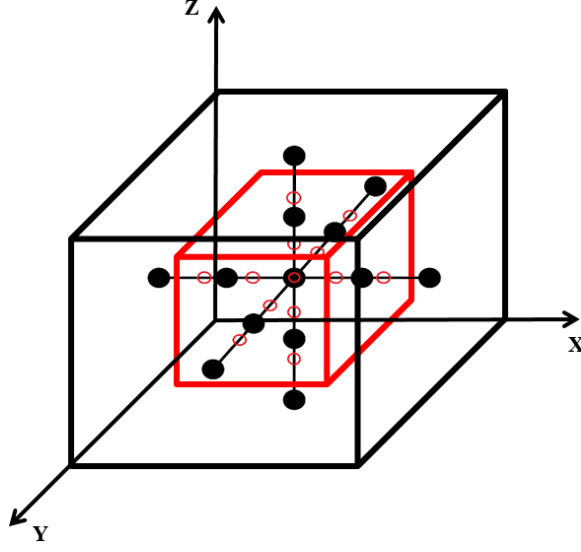
Figure 2.3. The one-dimensional searches performed by AxialPI in $D^{(k)}$.
Black points for $k = 0$, red circles for $k = 1$.

points of a grid $\mathcal{G}^{(k)}$ are

$$\mathcal{G}^{(k)} = \left\{ x \in D^{(k)} : x = \hat{x}^{(k)} + \sum_{j=1}^{n} h_j^{(k)} \eta_j d_j + \sum_{j=1}^{n} h_j^{(k)} \eta_{n+j}(-d_{n+j}) \right\}, \qquad (2.12)$$

where

- $\hat{x}^{(k)} \in D^{(k)}$ is the offset between different grids, given by (2.7);

- η_j is any integers such that $\forall j$

  - $|\eta_{n+j} d_{n+j}| \leq \left\lfloor \dfrac{\hat{x}_j^{(k)} - LB_j^{(k)}}{h_j^{(k)}} \right\rfloor$, $LB_j^{(k)}$ given by (2.6a), if d_j is pointing to lb_j,

  - $|\eta_j d_j| \leq \left\lfloor \dfrac{UB_j^{(k)} - \hat{x}_j^{(k)}}{h_j^{(k)}} \right\rfloor$, $UB_j^{(k)}$ given by (2.6b), if d_j is pointing to ub_j;

- $h_j^{(k)}$ is a positive scalar parameter such that

  - $h_j^{(k)} = \beta_j^{(k)}$ *given by* (2.11), if the iteration k is a failure,

  - $h_j^{(k)} = h_j^{(k-1)}$ otherwise,

The mesh size parameter is an $n - dimensional$ vector that considers different scale along different dimensions. Since $\mathcal{X}$ is a bounded set, the scale between the dimensions is a well defined constant.

The GABR algorithms has the ability of a quick identification of the most promising area in the feasible domain. A local optimality guarantee it is obtained by hybridizing the GABR algorithm with a local search strategy with proved convergence.

2.2.3 Hybridizing GABR with Local Searches

The global search represented by Algorithm (2.4) can be expensive in terms of function evaluations if the goal is to identify an optimal solution with a high precision. The reason is that the higher is the precision required, the more BR iterations must be performed. The refining of the solution can be delegated to a local search algorithm. With the aim to make the whole algorithm more efficient, a derivative free local searches (DFL) has been introduced. In literature there are many ideas on the hybridization of the global search. Examples of automatic balancing techniques can be found in [80] where the estimates of local Lipschitz constants allow to accelerate significantly the global search. A suitable strategy has been implemented in the GABR algorithm, similar to that one described in [22] where local minimizations are started in the most promising area of the feasible space. Since the number of LSs performed affects the efficiency, it is necessary to locate when a LS should be started. A reasonable choice is to perform LS at every BR iteration, only after the end of GA main loop and only if an improved solution is found with respect to previous BR iteration.

The comprehensive GABRLS algorithmic scheme is the following:

Algorithm 5. GABRLS scheme

---

1. Run **Algorithm 3** in $D^{(k)}$, $D^k \subseteq \mathcal{X}$, $k = 0$

2. Start LS from the current best point

3. **while** (stopping criterion)

4.   Bounding step (2.6), $k = k + 1$

5.   Restart **Algorithm 3** in $D^{(k)}$

6.   **if** (an improved solution has been found) **then**
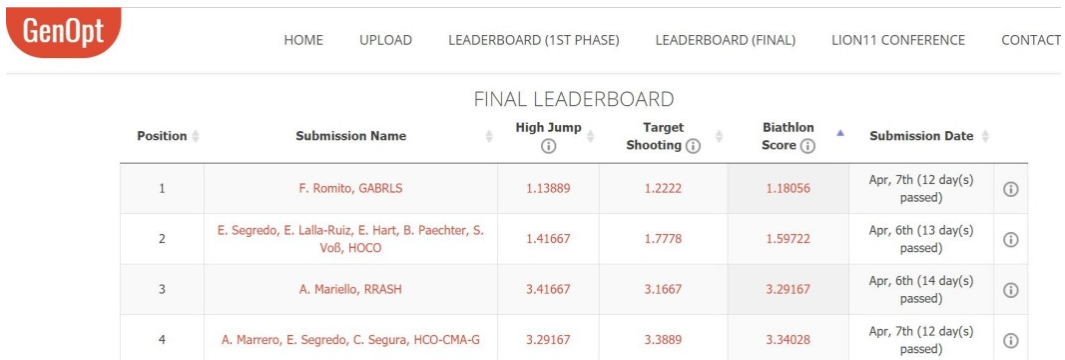
7.   Start LS from the current best point

8.   **end if**

9.   **end**

2.3 Prize and application

This section report briefly the prize obtained by GABRLS algorithm during a global optimization challenge and an application to real life problem.

2.3.1 Genopt prize

The Generalization-based Contest in Global Optimization (GENOPT, [61]) is a special session of the Learning and Intelligent Optimization Conference (LION). During LION 11 (June 19–21, 2017, Nizhny Novgorod, Russia), the GABRLS algorithm won the 1st prize in both partial categories on a test suite of 1800 multidimensional problems. Below, fig (2.4) shows the top four position of the final leaderboard and the score (less is better) on the speed of convergence (High Jump), the task solved (Target Shooting) and the overall ranking (Biathlon Score).

Figure 2.4. Final leaderboard - GENOPT competition.

GENOPT organizers provide a black-box library of 1800 problems, distinguishable in 18 classes by their dimension and other high-level characteristics. The total number of solved problem was 1605 over 1800, almost 90%. Next figure (2.5) summarize the 18 function classes and the result in term of problems solved for each class of function.

Id	Type	Type-details	Dim	GABRLS Tasks solved
0	GKLS	Non-differentiable	10	86
1			30	74
2		Differentiable	10	77
3			30	56
4		Twice differentiable	10	66
5			30	47
6	High condition	Rosenbrock	10	100
7			30	100
8		Rastrigin	10	99
9			30	100
10		Zakharov	10	100
11			30	100
12	Composite		10	100
13			30	100
14			10	100
15			30	100
16			10	100
17			30	100
Total				**1605**

Figure 2.5. Number of solved tasks per function class.

2.3.2 Case study: Design optimization of an electric motor

This subsection describes an application of the GABRLS algorithm on the optimization of the design of an electric motor. The local search integrated in the algorithm scheme is the one of [76], developed by G.Liuzzi et al., that is able to optimize over continuous and discrete variables and exploit a sequential penalty approach to handle non-linear constraints. The case study is intended to optimize the design of a Synchronous Reluctance Motor with the goal of maximizing the torque while smoothing the torque profile. This last objective means to reduce the torque ripple (instability of the torque) under an acceptable threshold. The motor model is designed in a Finite Element software e treated as a Black-Box. The strategy is to approach to the particular optimal design problem as mixed discrete constrained minimization of a suitable objective function. In particular the FE analysis can be used to evaluate the motor performance, namely to compute the objective function value and the constraints. The optimization procedure can use the information obtained by the FE program to iteratively update the set of motor parameters and try to identify an optimal motor by making a trade-off between the different parameters

of the machine.

The set of parameters x used in the optimization procedure of the SRM concern are listed in Table 2.6. Fig. (2.6) shows in details the 17 design variables (dimensions of flux barriers, tilt angles, fillet radii): they have been varied according to practical limits existing in cutting the rotor shape.

Table 2.6. Minimum and maximum ranges of design variables.

Variables		Min	Max	Type	Step
x_1. Width of rotor tooth along d-axis (mm)	$a1d$	4	8	$\mathbb{D}$	0.25
x_2. Width of barrier 1	b_1	2	4.5	$\mathbb{D}$	0.1
x_3. Width of barrier 2	b_2	2	4.5	$\mathbb{D}$	0.1
x_4. Width of barrier 3	b_3	2	4.5	$\mathbb{D}$	0.1
x_5. Distance between shaft and barrier 1	$fb1$	2	3	$\mathbb{D}$	0.1
x_6. Distance between barriers 1-2	$fb2$	3	5	$\mathbb{D}$	0.1
x_7. Distance between barriers 2-3	$fb3$	3	5	$\mathbb{D}$	0.1
x_8. Distance between barriers 3-ext.channel	$fb4$	3	5	$\mathbb{D}$	0.1
$x_9 - x_{10}$. Tilt angles barrier 1	α_1, β_1	0°	5°	$\mathbb{R}$	-
$x_{11} - x_{12}$. Tilt angles barrier 2	α_2, β_2	0°	5°	$\mathbb{R}$	-
$x_{13} - x_{14}$. Tilt angles barrier 3	α_3, β_3	0°	5°	$\mathbb{R}$	-
x_{15}. Number of conductors in slot		20	40	$\mathbb{D}$	1
$x_{16} - x_{17}$. Fillet radii (mm)	ri, re	0.5	3	$\mathbb{D}$	0.1

Moreover, several constraints have been introduced to guarantee the reliability and feasibility of the final design (Table 2.7). They are: the average torque, the
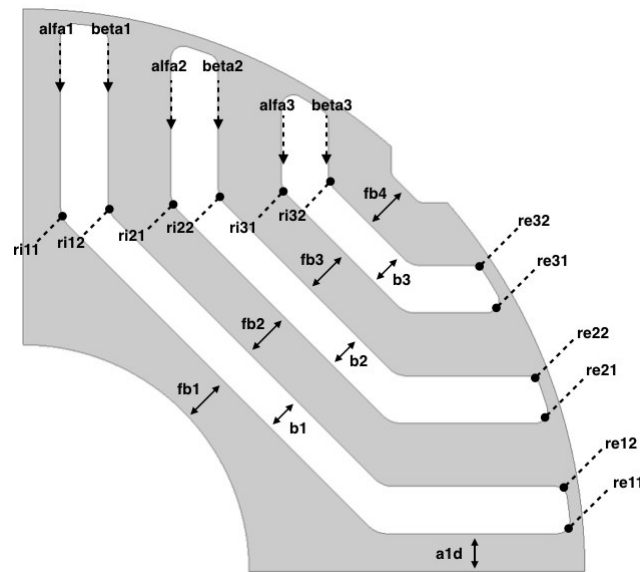
Figure 2.6. Rotor independent variables.

maximum flux densities in the stator and rotor core and the percentage of torque ripple.

Table 2.7. Constraints.

Constraints		Limits
c1. Fill factor		≤ 0.48
c2. Max flux density in the stator tooth	T	≤ 1.60
c3. Max flux density in the stator yoke	T	≤ 1.45
c4. Max flux density in the rotor tooth	T	≤ 1.50
c5. Joule losses at 90 degrees Celsius	W	≤ 240
c6. Torque ripple	%	≤ 10.0
c7. Average torque @ 1500 rpm	Nm	≥ 19.0

In particular, we deal with the following mixed discrete non-linear minimization

problem:

$$\min\ f(x)$$
$$g(x) \leq 0$$
$$lb \leq x \leq ub \qquad\qquad\qquad (2.13)$$
$$x_i \in \mathbb{D}, \quad i \in I_d$$
$$x_i \in \mathbb{R}, \quad i \in I_c = \{1, \ldots, n\} \setminus I_d,$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^m$, $lb \in \mathbb{R}^n$, $ub \in \mathbb{R}^n$, I_d is the index set of the discrete variables and I_c is the index set of the continuous variables.

The non-linear constrained problem (2.13) is approximated by a sequence of box-constrained problems of the form:

$$\min\ P(x; \epsilon_k) = f(x) + \frac{1}{\epsilon_k} \sum_{i=1}^{m} \max\{0, g_i(x)\}^p$$

$$lb \leq x \leq ub$$
$$x_i \in \mathbb{D}, \quad i \in I_d$$
$$x_i \in \mathbb{R}, \quad i \in I_c,$$

where $\epsilon_k > 0$ and $p > 1$.

The global search quickly find the most promising area, while the local search generates a sequence $\{\epsilon_k\}$ such that, for any given ϵ_k, the function $P(x; \epsilon_k)$ is reduced (with respect to x) by means of suitable derivative-free line searches along all coordinate directions. Afterwards, if the constraint violation at the new point is not sufficiently decreased, then the penalty parameter ϵ_k is updated. The single evaluation in a feasible point of the motor model in the FE software is computationally onerous even if it exploit multithread computation (up to 5 minute on 8-core desktop computer). In an unfeasible point it take less than a minute. In the following table are reported the initial non optimized motor, the result of the optimized motor with the local search only, and the optimized motor with GABRLS algorithm. The budget of function evaluation has been set to 10000 function evaluations for both algorithm. In table (2.8) the results os the comparison, whereas in figure (2.7) the torque profile in a single operating cycle.

Table 2.8. Comparison: non optimized motor, optimization with DFL and GABRLS.

	Average torque (N/m)	Standard deviation	Constr. violation
Non-Optimized	19.91	1.13	13.20
DFL	20.00	0.60	$2.43 \cdot 10^{-2}$
GABRLS	21.02	0.44	$3.88 \cdot 10^{-2}$
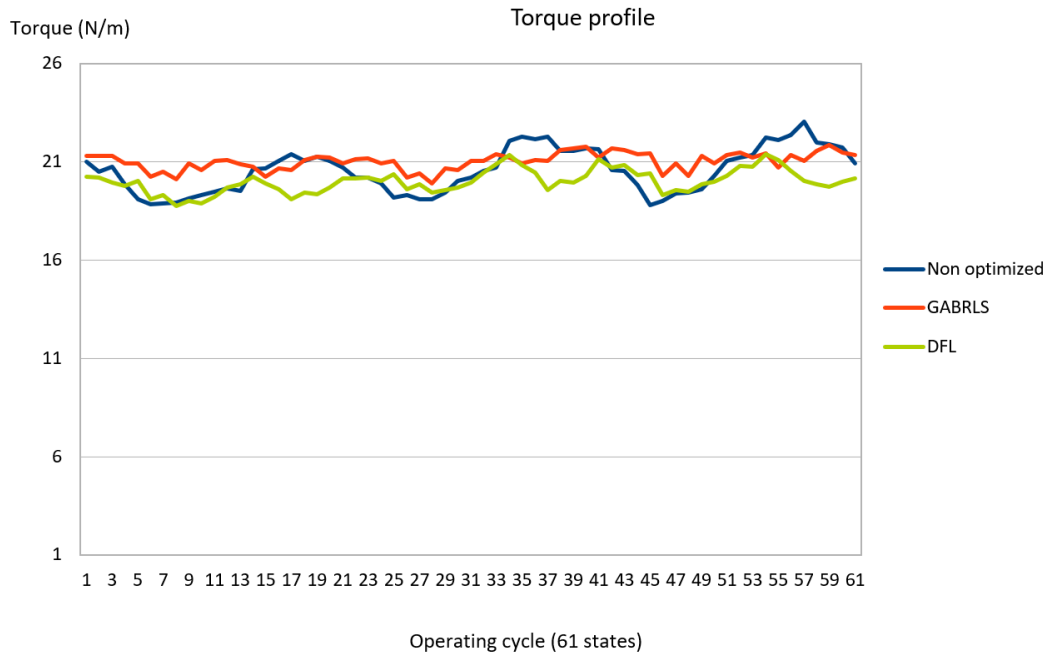
Figure 2.7. Torque profile: non optimized motor, optimized with DFL and GABRLS.

The motor optimized with GABRLS show the most high torque (21.02 Newton/meter) and also the most stable one (the value of 0.44 as standard deviation reveal the lowest torque ripple).

Chapter 3

A black box approach for the spare parts management problem

This chapter is the outcome of a joint work with prof. Laura Palagi et al., to appear on the International Journal of Information Systems and Supply Chain Management [82].
The aim of this project was to advancing the existing literature in tackling the inventory problems known as *spare parts management problems*. The idea was to develop a holistic approach that allows to directly address the non-linearities of the global optimization problem involved. Comparative results are presented on a case study of spare part management of a fleet of aircraft.

3.1 Introduction

Modern organizations are fully dependent on readily available spare parts to maximize operational capability in case of failures. Managing repairable inventories - which are spares normally characterized by high market value - represents an important managerial domain for improving operational readiness and reducing life-cycle costs for equipment. As a component downtime can be very costly, part inventories are required to keep the stock-out time as low as possible. Nevertheless, an excessive number of spare parts is to be avoided as well for reducing the cost dimension. Spare parts shall thus be thoroughly optimized to balance high system availability requirements and low cost of allocation [83].

Such repairable items are usually managed following a one-for-one replenishment policy, where the part is ordered after each substitution in lots of one. This situation is usually represented as a (S-1, S) policy where S is the optimum number of items in the inventory, and S-1 is the re-order level, i.e. the number of items below which activates the need for a re-ordering. This latter remains meaningful for parts characterized by high inventory cost and low demand, where the economic order quantity tends to a size of one [84]. In case of failure, the defective part is removed from the equipment, substituted with a functioning one. In the meantime, the original defective part is sent to a maintenance facility to be repaired.

Traditionally, inventory is balanced following an item-approach process, i.e. inventory levels for each item are set independently [85], failing to give a holistic optimization. The assignment of over-simplified constraints and requirements at the item level becomess increasingly less adequate for the needs of the so-called HA-HCLDS (High Availability, High Cost and Low Demand Systems), which represent the focus of the study [86]. Starting from the original contribution of Sherbrooke [87], it is possible to set item's stock levels jointly, adopting a systemic optimization, as for the so-called system-approach. The system-approach allows a holistic perspective on the system, being fed by systemic variables (e.g., total inventory budget, overall system's availability), and supporting the identification of system-wide parameters (e.g., the budget required for an overall service level, the effect of a stock reduction on the overall system service level).

The dominant system-approach model for repairable items is the Multi-Echelon Technique for Recoverable Item Control (METRIC), which relies on Palm's theorem [87]. Since the METRIC aims to respect a systemic perspective, it has to take into account a large number of variables, e.g. for each item at each local warehouse, the demand rate, the on-site repairing time, the turn-around-time, the reparability level. The approach should be also subjected to constraints related to holding costs, and availability requirements. The corresponding algorithmic computational complexity – which is non-linearly increasing with the number of items - forces the analysts to develop and adopt approximated optimization solutions. Traditionally, METRIC approaches adopt heuristics based on the so-called marginal allocation algorithm, as proposed by [87]. The marginal allocation generates acceptable stock level solutions in a limited time interval, counting on the incremental benefits related to the placement of an additional item in stock. More specifically, the METRIC-based model aims at defining the stock level for a single site that allows minimizing the holding costs whilst satisfying availability constraints.

After reviewing the literature on optimization algorithms for the METRIC, this work explores possible enhancements for the marginal allocation heuristics in order to define an alternative optimization algorithm for solving the system-approach model for repairable items. The main contribution of the work consists thus of advancing the existing literature on operational research for multi-item inventory systems through an enhanced time-effective optimization algorithm tested in a single-echelon system.

The discussed problem can be considered a non-linear global optimization problem, where functions are not available in closed form but only as the output of a black-box system, which implies expensive evaluation and not available derivatives. Therefore, this chapter describes an original black-box derivative-free algorithm [88] for solving such a problem which fully exploits the peculiar aspects of the application. It is pointed out that the proposed derivative-free approach allows tackling the non-linearity as is, without any decomposition in subproblems and without any approximation or necessity to check the feasibility of the solution. The algorithm is inspired by pattern search algorithms [89] and it includes specific features both to exploit integrality of the variables and to locally explore promising feasible subregions by using suitable tailor-made rules. In particular, the algorithm considers an enhanced approach for the selection parameters based on the ratio between holding costs of each item at a local warehouse and the absolute value of the availability

variation associated with a change in the stock of the same item at the local warehouse. The tailored selection rules allow improving performance in terms of needs of function evaluations which represent the main computational costs in black-box optimization.

The remainder of the chapter is organized as follows. Section 3.2 provides a literature review about METRIC with a focus on optimization approaches. Section 3.3 clarifies the analytic formulation of METRIC, as applied to the case of spare management of a fleet of aircraft. Section 3.4 specifies the innovative black-box algorithm proposed for its optimization, and Section 3.5 details its application as a walk-through application in a case study, comparing the results obtained via the proposed approaches with the ones obtained via the traditional marginal allocation and some high performance direct search algorithms. Lastly, section 3.6 summarize the outcomes of the work and pave the way to future joint multi-disciplinary research combining advanced optimization algorithms and logistics.

3.2 Literature review

Multi-component systems such as aircraft fleets, nuclear power plants, oil refineries, etc., demand for a thorough analysis of system requirements before deciding how many spares should be kept in each warehouse. The analysis should consider multivariate non-linear relations, which require an analytical formulation hard to made explicitly [90]

The METRIC starts from the assumption of a Poisson-distributed demand and of independent and identically distributed repair time, characterized by any distribution with a specified mean. Both these assumptions are representative of HA-HCLDS, and in analytical terms, they allow adopting Palm's theorem. Once respected the above-mentioned assumptions, the theorem states that the steady-state probability distribution for the number of units in repair is still a Poisson distribution, where the mean can be calculated as a simple product of the mean demand (following a Poisson) and the mean repair time [91]. From a logistics perspective, the theorem represents an important milestone that can significantly ease the mathematical formulation for the allocation problem, motivating the dominant role of the METRIC as one of the most used system-approaches, especially considering its reduced degree of mathematical sophistication [91].

In terms of optimization, the METRIC traditionally adopts a marginal allocation algorithm, whose applications in literature confirm it to be a flexible method suitable for a variety of problems. The marginal allocation can be originally found in the work of Sherbrooke [87], who started promoting its advantage compared to a random trial-and-error stock assignment procedure [88, 92]. The concept of marginal analysis dates back to Gross (1956) [93], and it has been re-organized by Sherbrooke to find the optimal stock solution which maximizes the backorder reduction-versus-cost increment when marginally adding a spare individually to each item. It remained a widely used approach in several early METRIC applications; see (e.g.) the multi-echelon METRIC model assuming a compound Poisson processes for modeling the demands (Graves, 1985, [94]); or the extended MOD-METRIC for multi-indenture systems by Muckstadt (1973) [95]. For example, Kline & Bachman

(2007) [96] use the marginal allocation for their inventory optimization problem, where the volume of spares is calculated from the functioning time percentage requested for the system to work. Similarly, in the context of performance-based logistics, Nowicki et al. (2008) [97] adopt a traditional marginal analysis for inventory optimization. De Smidt Destombes et al. (2009) [98] use a marginal allocation as well for a joint optimization problem on the frequency of maintenance activities, the ability to repair and the spare level. A similar marginal analysis has been employed by Costantino et al. (2013) [99] to solve a military inventory problem in a multi-item system with non-equal maintainability certification levels, imposing a weighted availability constraint on the number of equipment at each local warehouse. The adoptions of the marginal analysis can be also confirmed in the work of Xu et al. (2015) [100], who in their METRIC approach relax the hypothesis of an infinite supplier's capacity, rather assuming a prioritized maintenance service. Even Basten & Van Houtum (2014) [101] in their review about spare parts inventory indicate the METRIC as a dominant technique for rotable spare parts management, and present a greedy algorithm, which is largely based on the traditional marginal analysis. Their paper also refers to software in place for the adoption of METRIC, or its variant VARI-METRIC.

More recently, even considering the benefits arising from the adoption of the METRIC, alternative optimization approaches have been discussed and explored in literature, mainly focused on genetic and pattern search algorithms. Kapoor et al. (2016) [102] develop a simulation approach referring to METRIC theory for a two-echelon problem in a public transport fleet counting 9000 buses. Their version of a genetic algorithm provides optimal review periods and the level of spare parts for each site. Patriarca et al. (2016) [103] create a Real Coded Genetic Algorithm, the MI-LXPM, to increase the randomness of a Mixed-Integer (MI) solution employing Laplace Crossover (LX) and Power Mutation (PM). In the same research stream, Patriarca et al. (2016) [104] adopt a similar genetic algorithm for a more computationally demanding problem, i.e. including also lateral transshipment of spares. The advantage of these MI genetic algorithms is the possibility of providing integer values as outcomes of the optimization, which is meaningful outcome for any inventory problem optimization. In general terms, an integer solution obtained by approximating the real numbers to the nearest integer values may be not feasible or produce quite different results in terms of costs. Alternatively, Duran & Perez (2014) [105] develop a hybrid Particle Swarm Optimization algorithm combined with local search to solve a multi-item problem. The explorative nature of the work is acknowledged by the authors themselves, who suggest to further consider different initialization strategies, fitness definitions, and replacement strategies. Another explorative contribution is the one developed by Costantino et al. (2014) [106], who test a Pattern Search algorithm to determine the items to stock, and their quantity in a multi-item multi-indenture problem. The solution is obtained by implementing a Generalized Pattern Search (GPS) (Audet & Dennis, 2003) [90], where no specific attention is presented for parameters' optimization. Even though not explicitly related to the METRIC, Nickel et al. (2006) [107] address a similar problem by a pattern search algorithm for spare parts allocation. In their work, the target cost function is linear, while the constraints are highly non-linear and considered as black-box functions, i.e. very difficult (and expensive) to evaluate, whose derivatives

are not available. The (resource-expensive) solution adopted in this case consists of performing a second pattern search algorithm for each solution point that is found by the initial pattern search at each iteration, to find a good integer solution. Other relevant approaches for inventory optimization can be linked to the work conducted by Topan et al. (2017) [108], and Wong et al. (2007) [109], both in multi-echelon systems. Both the researches propose variations to the standard marginal allocation approach, but with a partly different focus: the former aimed at defining the order quantities, the reorder points at the central warehouse, and the base-stock levels at the local warehouse; the latter offers a multi-echelon representation which does not necessarily hold the same results for the single-echelon proposed in this research.

Based on this review, it emerges from the literature the continuous interest in spare parts optimization over the years. Nevertheless, recent research in the area confirms an increasing interest in developing alternative optimization techniques for such complicated problems (Nowicki et al., 2012) [110], aimed at exploring the benefits of pattern search, genetic algorithms, and particle swarm optimization. These algorithms, however, are not widespread in the specific literature, partly due to the large difference in terms of ease of implementation and required computational efforts. These pieces of evidence motivate the development for a heuristic to be both effective and efficient if compared (at least) with the marginal analysis. The black-box algorithm presented in this work aims to provide near optimal solutions in a reduced computational time. This original and efficient solution remains significant to allow quick systemic parametric analyses, in order to test multiple managerial options and allocations in a non-invasive approach.

3.3   The inventory management model

This section presents a black-box optimization model for a single-echelon multi-item inventory system. It is considered a problem that is based on the METRIC and it aims at the minimization of the holding costs of items at a single site, whilst guaranteeing a required availability level.

3.3.1   The single-echelon multi-item problem and assumptions.

In Figure (3.1) is described the single-echelon multi-item system, where:

- CD (Central department)

- MD (Maintenance Department)

- LW_j (Local Warehouse $j = 1, \ldots, J$, J = total number of LW)

- LRU_i (items, Line Replaceable Unit $i = 1, \ldots, I$, I = total number of LRU)
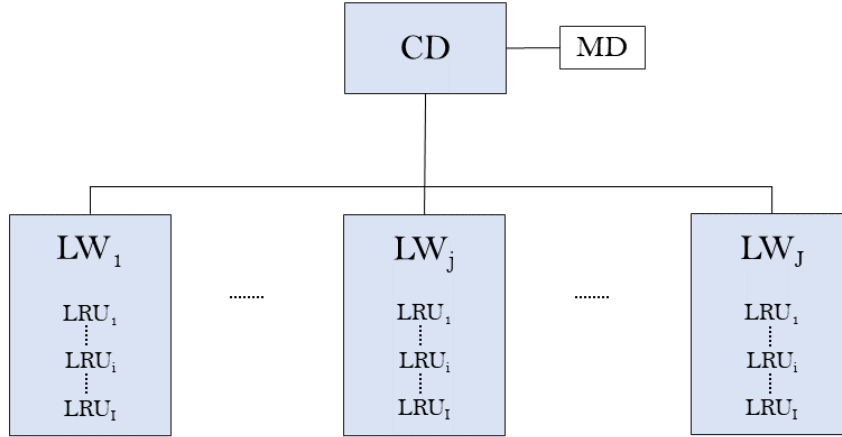
Figure 3.1. Echelons structure for multi-item LRU_i.

In Figure (3.1), it is important to clarify the assumption is that the CD has no stock levels, and it is used only as a maintenance center with infinite capacity, i.e. it always holds an appropriate resources level to execute any needed operation. Furthermore, each LW_j does not have a maintenance center. Hence each LRU_i that requires maintenance at LW_j must be sent to the CD. For these reasons, in analytical terms, the problem can be interpreted as a single-echelon optimization problem.

The proposed formulation does not consider lateral transshipment among LW_j and that the back-order queues generated by each LW_j at the CD are unrelated. Under these assumptions the authors can decompose the spare parts management problem of the LRUs on subproblems on each LW_j for all $j = 1, \ldots, J$.

The remainder of this section focuses on the definition of the optimization problem for each LW_j for the detection of an optimal stock level s_j^* such as to minimize an objective cost function $C_j(s)$ subject to nonlinear restriction which enforces the availability of vehicles at each LW_j.

3.3.2   Analytic formulation

This section describes the analytical model at the local warehouse LW_j.

As a first step, the authors define the decision variables which are the stock level $s_{i,j}$ for all LRU_i, i =1, ..., I, at the local warehouse LW_j. The goal is the minimization of the holding cost for all LRUs i at the $j - th$ local warehouse LW_j. Denoting $c_{i,j}$ the unit cost for stocking the $i - th$ item in the $j - th$ local warehouse the authors get the overall holding cost for the local warehouse LW_j.

$$C_j = \sum_{i=1}^{I} s_{i,j} \cdot c_{i,j}. \tag{3.1}$$

At any time, the stock level $s_{i,j}$ of LRU_i at LW_j can be split into the sum of three components:

$$s_{i,j} = DI_{i,j} + OH_{i,j} - BO_{i,j}, \tag{3.2}$$

where:

- $DI_{i,j}$ is the number of Due In items of LRU_i waiting for repair by the CD; Under the hypothesis of no queue at CD, they are the items being repaired in the exact moment of the calculation

- $OH_{i,j}$ is the number of On Hand items of LRU_i currently available at LW_j

- $BO_{i,j}$ is the number of Back Orders of LRU_i due to request arrived when the inventory was already out of stock at the LW_j

The feasible values of $s_{i,j}$ are thus constrained by two nonlinear constraints. The first constraint imposes a lower bound $A^{target} > 0$. In addition, the availability at site A_j^* has to verify the target availability constraint:

$$A_j^* \geq A^{target}. \tag{3.3}$$

The availability at site refers to the availability of the fleet planned for the site. In this case, the authors consider a fleet of machines, so that each site can be modelled as a passive redundancy system constituted by N_j machines, when M_j machines are active, the remaining $N_j - M_j$ are put in cold stand-by ready to substitute the active machine in case of failure.

The total availability A_j^*, is described by the following formula (Costantino et al., 2013 [99]):

$$A_j^* = A_j^{M_j} \sum_{k=0}^{N_j - M_j} \frac{[-M_j \ ln(A_j)]^k}{k!}. \tag{3.4}$$

The availability of the single machine A_j depends on the availability of the LRUs that affect it, so that a machine is available when all the items that contribute to form it are available. The mathematical relation is here simplistic described as a series:

$$A_j = \prod_{i=1}^{I} A_{i,j}, \tag{3.5}$$

where $A_{i,j}$ is the availability of the single item LRU_i at site j and it is expressed by:

$$A_{i,j} = 1 - \frac{E[BO_{i,j}]}{N_j}, \tag{3.6}$$

where $E[BO_{i,j}]$ is the expected value of the Back-Order corresponding to the stock level s_{ij} and the ratio between $E[BO_{i,j}]$ and N_j is the unavailability level.

Since $s_{i,j}$ can only take non-negative integer values, at least one of $BO_{i,j}$ and $OH_{i,j}$ is zero, the Expected Back Order is the positive part of the delta between the Due In $DI_{i,j}$ and the stock level $s_{i,j}$ (Costantino et al., 2018):

$$EBO_{i,j}(s_{i,j}) = E[BO_{i,j}] = E[(DI_{i,j} - s_{i,j})^+]. \tag{3.7}$$

Hence the expected value of backorders $BO_{ij}$is obtained, following the assumption of Poisson distribution, as:

$$E[BO_{i,j}] = \sum_{x=s_{i,j}+1}^{\infty} (x - s_{i,j})f_{i,j}(x), \tag{3.8}$$

where $f_{i,j}$ represents the fraction of demand due to item i at site j (Sherbrooke, 2004):

$$f_{i,j} = \frac{m_{i,j}(1 - r_{i,j})}{m_{i,0}}. \tag{3.9}$$

In the last relation $m_{i,j}$ is the share of LRU_i sent from LW_j to the CD with probability $1 - r_{i,j}$ with respect to the total demand $m_{i,0}$ of $LRUs$ of all local warehouses. Since it is assumed that the local warehouses have no maintenance center the probability of a repair is $r_{i,j} = 0$.

Since the availability of the single item $A_{i,j}$ is non-negative, from the formula (3.6), a second non-linear constraint is needed:

$$E[BO_{i,j}] \leq N_j. \tag{3.10}$$

The whole formulation of the problem has to deal with is the following integer nonlinear constrained minimization for all local warehouse $j = 1, ..., J$:

$$\min_{s} \quad C_j = \sum_{i=1}^{I} s_{i,j} \cdot c_{i,j}, \tag{3.1}$$

$$\text{s.t.} \quad A_j^* \geq A^{target}, \tag{3.3}$$

$$E[BO_{i,j}(s_{i,j})] \leq N_j, \tag{3.10}$$

$$lb \leq s_{i,j} \leq ub, \tag{3.11}$$

$$s_{i,j} \in Z. \tag{3.12}$$

The lower bound and the upper bound (3.11) limit the values that can be assumed by the inventory level. The lower bound lb is usually set to 0 to avoid those solutions that present one or more negative - unrealistic - values. The upper bound, on the other hand, can model different requirements such as a limitation in the space of storage or a precise restriction about the quantities in stock for one or more items. Since indivisible goods are considered, the last constraint (3.12) ensure that the stock levels are integers.

Notice that constraints (3.3) and (3.10) embed the values of the stock levels $s_{i,j}$, and are tackled explicitly in our algorithm scheme so that feasibility is retained at every trial solution.

Table (3.1) summarize the main parameters involved in the definition of the mathematical model.

Table 3.1. Variables and parameters of the mathematical model.

I	Number of items (LRU)
J	Number of Local Warehouse (LW)
i	Item index (LRU index) $i = 1 : I$
j	Warehouse index of the LW $j = 1 : J$
$m_{i,0}$	Annual average demand of LRU_i at CD
$m_{i,j}$	Annual average demand of LRU_i at LW_j
$s_{i,j}$	Stock level of LRU_i at LW_j
s_j	$[s_{1,j}, s_{2,j}, \ldots, s_{I,j}]$ vector of stock levels of LRU_i
$r_{i,j}$	Repair rate of LRU_i at LW_j
N_j	Number of vehicles at LW_j
M_j	Number of active vehicles at LW_j
c_{ij}	Unit inventory holding cost of $LRUi$ at LW_j
A_j^*	Total fleet availability at LW_j
A_j	Aircraft availability at LW_j
$A_{i,j}$	Availability of LRU_i at LW_j
A^{target}	Target value of availability for each LW

$BO_{i,j}$	Back Order level for LRU_i at LW_j
lb, ub	Lower and Upper bounds on stock level of LRU_i at LW_j
Z	The set of integers
T_{max}	Number of items which may be considered all together during the optimization. It may limit the computational effort that is machine-dependent.

3.3.3 Optimization using the Marginal Analysis

At this step, it is possible to solve the above-described problem via marginal analysis, which represents the most common approach for METRIC-like problems. The following 7 steps summarize the main phases of a marginal analysis approach applied to inventory optimization:

1. Set the index of the iteration count $k = 0$ and choose the starting point (zero - vector in our case).

2. Calculate the expected backorder

$$E[BO_{i,j}].$$

3. Calculate the rating vector

$$R_j = \{\rho_{i,j} = \frac{E[BO_{i,j}]}{c_{i,j}} \; : \; i = 1, ..I\},$$

where $\rho_{i,j}$ is a variable created to prioritize the solution.

4. Select the index $\hat{i}$ which returns the $\max_i \; \rho_{i,j}$.

5. Calculate s_j^{k+1} as $s_{ij}^{k+1} = \begin{cases} s_{\hat{i}j}^k + 1 & : & for \; i = \hat{i}, \\ s_{ij}^k & : & \text{otherwise.} \end{cases}$

6. Evaluate the availability A_j^* in S_j^{k+1}.

7. Stopping criterion:

   If $\quad A_j^* < A_j^{target} \quad$ Set $k = k + 1$ and repeat from 2,
   else $A_j^* \geq A_j^{target} \quad$ then evaluate the objective. function and ends.

The marginal analysis provides an incremental solution that does not consider complex re-combinations of different items: once an item is assigned, it cannot be removed from the optimal solution.

Nevertheless, the solution provided by the marginal analysis can be used as a good starting point for more accurate and complex optimization algorithms, i.e. to define a feasible point. In this regard, one can note that the service level constraint (3.3) and the upper bound on the back order expected value (3.10) are highly non-linear constraints involving all the variables $s_{i,j}$ which represent the hard constraints in the spare management problem.

3.4   The Deterministic Black Box Integer Feasible Optimization

In this section, it is presented the Derivative Free Optimization (DFO) method, which is an innovative deterministic feasible algorithm to deal with the integer black-box constrained problem represented by the inventory problem at hand. It is well known that the complexity of finding an optimal solution of black-box optimization problems increases exponentially with the number of variables (Vavasis, 1995 [111]). Indeed, an exhaustive search is not conceivable because of the time-consuming evaluation of the objective function and of the constraints due to the simulations involved. In addition to the Pattern Search type approach already mentioned, examples of recent efficient and effective algorithms for black-box optimization, that later we would take in consideration for the comparison of our approach, are:

- the NonMonotone Black-Box Optimization Algorithm (NM-BBOA) developed by Liuzzi et al. (2018) [112] that is suited for managing only integer variable and exploit primitive direction to perform a dense neighborhood search. Even if the algorithm is conceived for local searches the authors explore the effect on the global search by increasing a parameter β that allows to explore larger and larger neighborhoods.

- the GABRLS algorithm presented in chapter 2 that is a modified genetic algorithm, integrated with a Bounding Restart technique and derivative free Local Searchers (Romito, 2017 [113]).

The algorithm proposed in this research is inspired by the pattern search framework and it includes specific features both to exploit integrality of the variables and to locally explore promising areas of the feasible region by using tailor-made rules. A generic DFO pattern search algorithm starts from a first feasible solution and at each iteration produces points that lie on a rational lattice. Elementary directions are combined and scaled with a step length parameter on a finer and finer grid to meet, where possible, convergence requirements. Elementary displacements are movement along one direction with unit step-length. The set of used directions must constitute a positive spanning set [114] such as in any positive basis method [115] or a dense set of directions in case of integer variable [116]. The proposed algorithm is able to explore suitable vector combinations to speed up the search in the most promising neighborhood but it is also able to explore densely a neighborhood by enumeration if allowed by the budget of function evaluation.

Key elements that influence the performance of a DFO algorithm are the rules for selecting the i-th elementary displacement and the number of trial points along the selected direction.

The proposed contribution stays in the definition of tailor-made rules for selecting the mesh of trial points that exploit the relationship among constraints and objective function. The tailored selection rules allow improving performance in terms of needs of function and constraint evaluations which represent the main computational costs in black-box optimization.

3.4.1 DeB^{2}IFO (Deterministic Black Box Integer Feasible Optimization)

As previously discussed, the main innovative aspect of the proposed algorithm is to explore the feasible integer mesh as best as possible using tailor-made rules. Indeed, the aim is to avoid expensive function computations at points where the expected value of the objective function is locally worse than the current best solution. Thus, the it is introduced a parametric selection strategy for choosing subsets of items with the best expected improvement in the objective function. The selection strategy of an item i is based on the trade-off between its holding cost and the percentage availability variation due to a unit variation of the stock s_i. The idea of using similar ratios has been also explored in [109] within a greedy procedure to find a feasible solution.

Formally it is introduced the selection indicator (3.13) for each item $i - th$ and the local warehouse LW_j

$$k_{i,j} = \frac{c_{i,j}}{\Delta A_{i,j}} \tag{3.13}$$

which is the ratio between the storing cost of the item i at LW_j and the value of the change in availability $\Delta A_{i,j}$ due to a change of the stock $\Delta s_{i,j}$ of the item i at the same LW_j. It is necessary to discuss some issues in order to explain the role played by $k_{i,j}$ in the construction of the new trial point.

Two scenarios may occur following a unit variation of stock $\Delta s_{i,j}$:

- when $\Delta s_{i,j} = +1$, namely the stock increases, the availability $A_{i,j}$ increases too so that the difference $\Delta A_{i,j}$ is positive and the cost of storage increases by $c_{i,j}$ (the unit storage price of the item i-th in LW_j);

- when $\Delta s_{i,j} = -1$, namely the stock decreases, the availability of the item $A_{i,j}$ is reduced, so the difference $\Delta A_{i,j}$ is negative and the cost of storage is reduced by $c_{i,j}$.

Hence for each LW_j two vectors are constructed

- $k_j^+ = [k_{i,j}]_{i:k_{i,j}>0}$;

- $k_j^- = [k_{i,j}]_{i:k_{i,j}<0}$.

corresponding respectively to positive or negative variations on all the components $\Delta s_{i,j}$ sorted in ascending order, namely such that

$$(k_j^+)_h < (k_j^+)_{h+1} \quad \text{and} \quad (k_j^-)_h < (k_j^-)_{h+1}. \tag{3.14}$$

It is possible to select heuristically combinations of items that seem to be more promising to provide a decrease in the cost in order to obtain an improvement of the objective function. It is selected the items corresponding to the first $T_{\max}$ components of k_j^+ and k_j^-, which are linked respectively to a positive and negative unit variation of availability, and therefore to an elementary displacement on the integer mesh. The algorithm moves along a grid defined by selecting a bunch of

items of cardinality $2T_{\max}$. It is denoted with k_j the vector made up of the $2T_{\max}$ selected items.

$$k_j = [k_j^+, k_j^-] \tag{3.15}$$

Note that, at first glance, using items $i-th$ corresponding to the values in k_j^+ may appear to be useless. In fact, if the stock increases, the cost function increases too, thus deteriorating the objective function. However, it is essential to consider the positive variations due to the presence of the service level constraint. Merely decreasing the stock could lead to a reduction in availability, with a possible violation of the corresponding constraint. What is of interest now are $A_{i,j}^+$ and $A_{i,j}^-$ which represent respectively the increase and reduction in availability due to a change in stock of items in k_j^+, k_j^-. These availability factors are combined in order to maintain feasibility. The core idea of this approach consists of choosing items corresponding either to low reduction in availability at a high cost or high increase in availability at a low cost, thus reducing the total cost whilst satisfying the service level constraint.

The parameter $T_{\max}$ plays a crucial role in the computational effort needed by the algorithm. It is explored the neighbourhood of feasible points combining in all possible feasible ways the changes in the stock of at most $2T_{\max}$ items corresponding to different components of vectors k_j^+, k_j^-. It makes no sense to consider combinations related to the same item which will lead either to a null displacement or a double displacement which is not allowed in the algorithm. The number of combinations is exponential in $T_{\max}$. In principle it is used an iterative incremental strategy to select the value $T_{\max}$ starting from a minimum number $T_{\min}$ of two items, in order to limit the computational effort at the first iterations when it is far from the optimal solution. All search points are therefore determined for the current iteration as the most promising subset of the mesh points in a neighborhood of s_j with radius $T_{\min}$.

Now it is proceeded to the last step, the Direct Search. Consider a starting point s_j for the current iteration with value of the objective function C_j. This value of the objective function is the target for the considered iteration. The algorithm plans to move to the search points earlier identified. For each of them, indicated by s_j^{trial}, the value of the objective function $C_j(s_j^{trial})$ is calculated. Only if

$$C_j(s_j^{trial}) \leq C_j,$$

the correspondence of the point s_j^{trial} to the bounds and to the constraints is verified, while if

$$C_j(s_j^{trial}) > C_j,$$

the point is discarded.

In the case s_j^{trial} satisfies both the constraints and the bounds, the value of the objective function $C_j(s_j^{trial})$ is stored in a matrix, called Cost Matrix, while s_j^{trial} is put in a second matrix, called Stock Matrix.

Once this procedure has been repeated for all the selected displacements, the stock level is chosen from the Stock Matrix related to the respective lowest value in the Cost Matrix. This level of inventory is imposed as the starting point of the

next iteration and the corresponding value of the objective function becomes the new target. The number of elements taken for the combinations must be increased in those cases where it would not be possible to determine a point better than the starting point, i.e. one of the following situations does not hold:

i) the value of the objective function $C_j(s_j^{trial})$ is lower or equal than the target value C_j^0 of the current iteration;

ii) s_j^{trial} satisfies the bounds;

iii) s_j^{trial} satisfies the constraints.

The algorithm stops when the number of elements taken to construct the combinations is equal to the number of items $T_{\max}$ previously identified as reasonable choice for computational resource limitation. It is pointed out that by setting $T_{\max}$ equal the number of items, the algorithm is able to perform an exhaustive search in the feasible domain.

The evaluation of the cost function is the first step because the cost function is much less onerous, in terms of computational time, than the constraint functions. This allows to avoid the calculation of the latter for those points which, in any case, will not be chosen at the end of the iteration, since they present a value of the cost function greater than the target one. The following section shows formally the algorithm steps for each LW_j.

3.4.2 Optimization using the DeB²IFO

In agreement with the described black-box approaches, the following steps detail the algorithm as applied to the inventory optimization problem at hand. This subsection presents the approach intended to enhance the traditional optimization based on the marginal analysis.

1. Set $l = 1$ as the counter of inner the iterations of the algorithm and consider l_{max} as max number of iterations. Choose $\hat{n} \leq n$ as the number of items to consider in the selection rule (3.13). Choose $T_{\min} <= T_{max}$ and set $T = T_{\min}$ for the starting number of items to be optimized jointly. Find the starting point s_j through marginal analysis and evaluate its availability

$$
\begin{aligned}
s_j &= \{s_{i,j} : i = 1, \ldots, n\}, \\
A_j &= \{A_{i,j}(s_{i,j}) : i = 1, \ldots, n\}.
\end{aligned}
$$

2. Perturb s_j and find

$$
\begin{aligned}
s_j^+ &= s_j + 1, \\
s_j^- &= s_j - 1.
\end{aligned}
$$

3. Calculate the corresponding availability A_j^+ and A_j^-

$$A_j^+ = \{A_{i,j}(s_{i,j}^+) : i = 1, \ldots, n\},$$

$$A_j^- = \{A_{i,j}(s_{i,j}^-) : i = 1, \ldots, n\}.$$

4. Calculate the delta availability w.r.t. the base scenario ΔA_j^+ and ΔA_j^-

$$\Delta A_j^+ = A_j^+ - A_j,$$

$$\Delta A_j^- = A_j^- - A_j.$$

5. Calculate the ratio k_j^+ and k_j^- to rank the solutions for all the items

$$k_j^+ = \{k_{i,j}^+ : k_{i,j}^+ = \frac{c_{i,j}}{\Delta A_{i,j}^+}, \quad i = 1, \ldots, n\},$$

$$k_j^- = \{k_{i,j}^- : k_{i,j}^- = \frac{c_{i,j}}{\Delta A_{i,j}^-}, \quad i = 1, \ldots, n\}.$$

6. Sort k_j^+ and k_j^- in ascending order to facilitate the ranking.

7. Create vector $\hat{k}_j$ by selecting a subset of items, i.e. the first $\hat{n} \leq n$ values of k_j^+ and k_j^-

$$\hat{k}_j = [(k_j^+)_1, \ldots, (k_j^+)_h, \ldots, (k_j^+)_{\hat{n}}, (k_j^-)_1, \ldots, (k_j^-)_h, \ldots, (k_j^-)_{\hat{n}}].$$

8. Create all the possible combinations $\binom{\hat{n}}{T}$ of values of $(A_j^+)_h$ and $(A_j^-)_h$ corresponding to the combination of the components of $\hat{k}_j$. Notice that each components of $\hat{k}_j$ is linked to a specific item i, so in practice for each A_j^+ or A_j^- we have a specific combination of items.

9. Eliminate combinations of items involving null displacements or multiple displacements along the same component.

10. Create the expected availability vector for each remaining combination $p = 1, \ldots, P$

$$A_j^p = \begin{cases} A_{i,j}^p : A_{i,j}^p = A_{i,j}(s_{i,j}^+), \forall i \in I, \ i \text{ linked to } h \to (k_j^+)_h \in \hat{k}_j \\ A_{i,j}^p : A_{i,j}^p = A_{i,j}(s_{i,j}^-), \forall \ i \in I, \ i \text{ linked to } h \to (k_j^-)_h \in \hat{k}_j \ . \\ A_{i,j}^p : A_{i,j}^p = A_{i,j}(s_{i,j}), \forall i \in I, \ i \text{ linked to } h \to (k_j)_h \notin \hat{k}_j \end{cases}$$

11. Evaluate the total expected availability A_j^{*p} (from 3.4) and deletes those less than the target value A^{target}.

12. Consider the item indices used for the remaining P combination and construct the points s_j^p , $p = 1, \ldots, P$.

13. Construct complex trial displacements $s_j^{trial,p} = s_j + s_j^p$, $p = 1, \ldots, P$.

14. Evaluate all the trial mesh points $s_j^{trial,p}$ and compare it to the current best solution s_j.

15. If $\exists p$ such that $Cj(s_j^{trial,p}) \leq C_j(s_j)$ and $(s_j^{trial,p}$ is feasible) put

$$s_j = s_j^{trial,p}, \quad l = l + 1.$$

16. Stopping criterion

    Execute instruction 2-15 until $l < l_{\max}$ or at the current iteration l there is no further trial point to evaluate. In the last case
    if $T < T_{\max}$ set

    $$l = 1, \qquad T = T + 1, \qquad \text{repeats from 2.}$$

    else

    $$\text{return an optimal (local) solution } s_j^* = s_j.$$

    end

    Any suitable criterion can be added, such as time limit or a max number of function evaluations.

3.5 Case study: Spare part management of a fleet of aircraft

In this section, it is illustrated an application of the DeB^{2}IFO for a single-site case study for a fleet of aircraft.

3.5.1 Scenario description

The system under investigation is a multi-item system constituted by one CD and 3 LWs. This problem becomes a single-echelon scenario where each *LW* is responsible for a different fleet of aircraft. The case study is intended to test the proposed optimization process for a subset of items (23 items) which constitute the main LRU for an aircraft flight system. The items considered in this explorative research are linked to an aircraft hydraulic plant. Each site has the possibility to activate cold stand-by aircraft. This situation is representative of several civil aviation real operating contexts, in which the stand-by aircraft can be used to guarantee fleet operability and deal with unexpected failures generating the so-called AoG (aircraft on ground), i.e. a problem serious enough to prevent an aircraft from flying. For each LW, the efficient number of aircraft must satisfy an availability target equal to 0,96. In operating contexts, the availability level is usually set by the decision-maker based on the company scheduled service level and it depends on the market competitiveness, as well as customers' expectations.

The sites are characterized by total vehicles N_j and active vehicles M_j, $j = 1, 2, 3$ respectively

- $N_1 = 97, \quad M_1 = 96;$

- $N_2 = 23, \quad M_2 = 22;$

- $N_3 = 202, \quad M_3 = 201;$

while the holding cost are distributed between 104 €/piece and 5705 €/piece, and an average value of about 1300 €/item.

Note that the proposed scenario remains representative also of an MRO company (Maintenance, Repair, Overhaul) network. In this case, the company remains responsible for a subset of maintenance interventions for an airline, or a pool of airlines, and has to manage the spare parts to be located in 3 *LWs* intended for maintenance operations.

The aim of the computation experiment presented in this case study is threefold without the ambition of being exhaustive. First, the DeB^{2}IFO was compared with the traditional marginal analysis; secondly, the Pattern Search, that is one of the most used in literature metaheuristic is chosen, to assess the quality of the proposed tailored grid search in the neighborhood of feasible solutions. Moreover, two recent efficient and effective algorithm approaches are included in the comparison to have a feel of the strength to exploration of DeB^{2}IFO.

3.5.2 Results

The starting solution benchmark is the result of marginal analysis. It is compared with three direct search alternatives and the proposed approach on 23 demand

scenarios for 3 local warehouse. The used software environment is MATLAB. All the search algorithms must be stopped if they exceed the timelimit of 1 hour for solving all 69 problems. To do so each of them has no more than 20000 function evaluation to solve an instance (1 scenarios of 1 local warehouse), according to the single evaluation time of the objective function and constraints. The algorithms are compared in terms of the overall total cost of all problems. The search algorithms are tuned as follows:

- Marginal analysis doesn't need any tuning;

- DeB^{2}IFO adopt $T_{\min} = T_{\max} = 4$ (i.e. the maximum number of items to be considered jointly during the optimization), $l_{max} = 1000$ as max number of iteration, and $\hat{n} = 10$ as the number of best performer items to consider in the selection rule.

- NM-BBOA has the β parameter that specifies the neighborhood size is set to 50 according to the global search analysis reported in [112];

- GABRLS has the initial set of point (population) and the number of iteration (generations) as two main feature to tune. They are set to 100 and 70 respectively according to the dimension of the problems.

The Pattern Search algorithms is the most common direct search alternative to marginal analysis. Since the Pattern Search algorithms integrated into the MATLAB optimization toolbox allow several parametrizations, it is made a preliminary analysis to select the most robust tuning. The fundamental three parameters tuned concern the search directions:

- Poll method (Generalized Pattern Search, Generating Set Search, Mesh Adaptive Direct Search (Audet & Dennis, 2006) combined to positive basis 2N);

- Polling order (random, success, consecutive);

- Complete search (on/off).

To ensure an optimal setting of parameters, twelve different configurations have been considered. The best configuration corresponds to MADS algorithm combined to a positive basis 2N as reported in table (3.2). This configuration is used in the following comparisons.

Table 3.2. Selected Configuration.

Poll Method	Polling Order	Complete Search
MADS 2N	Random	Off

In table (3.2) it is compared the results of all the algorithms.

Fig. (3.2) shows the total amount of the costs for the 23 demand scenarios per each LW, while table (3.3) summarize the total costs on all *LW* and the saving

w.r.t. the best solution found. It is interesting to observe how the DeB^{2}IFO ensures a cost reduction of approximately 3.6% w.r.t. MADS 2N, or about 6.6% w.r.t. a traditional marginal analysis. A comparative table with detailed results on the 23 demand scenarios is provided in the appendix B.

	LW1	LW2	LW3
■ DeB2IFO	463905.84	525985.99	873468.88
■ NM-BBOA	465605.3319	532931.7551	876225.6797
■ GABRLS	465311.5168	535724.7214	883278.7618
■ MADS 2N	481485.61	552757.54	895537.55
■ Marginal Analisys	499224.41	567825.64	918946.28
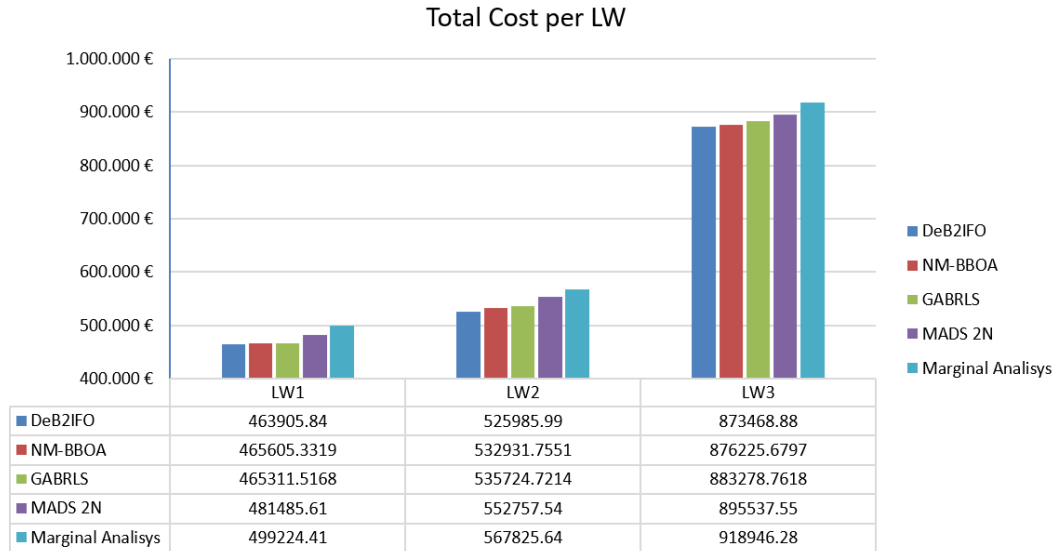
Figure 3.2. Total costs for each *LW* (23 items).

Table 3.3. Total Item's holding costs in € and computational time in seconds.

	DeB^{2}IFO	NM BBOA	GABRLS	MADS 2N	Marginal Analysis
Total €	**1863361**	**1874763**	**1884315**	**1929781**	**1985996**
Δ€	–	11402	20954	66540	122635
% Δ€	–	% 0,61	% 1,12	% 3,57	% 6,59
Total time (sec.)	3427	3481	2893	1832	281
Δ time (sec.)	3146	3200	2612	1832	281
% Δ time	% 11,20	% 11,39	% 9,30	% 5,52	-

All these results are highly satisfactory and confirm the goodness of the optimization procedure and the effectiveness of the proposed approach.

3.6 Conclusion

A novel strategy to reduce the costs associated with spare parts management has been presented. The proposed approach started by modeling the spare parts management problem as an integer constrained minimization of a linear objective function. Then the focus of the chapter was devoted to the introduction of the DeB^2IFO algorithm, a deterministic method that takes into account the relationships between different items and a robust selection strategy for subsets of items with the most promising impact in reducing the total cost function. The advantages of DeB^2IFO are the ability to handle integer solutions always satisfying feasibility and the tailor-made criterion of choosing only subsets of promising directions that avoid costly evaluations of the objective function and constraints. The results of comparison with the traditional marginal analysis and the others black-box algorithms were satisfactory both in terms of quality of the solutions and computational time. Nevertheless, the focus of this paper was mainly devoted to the definition of the analytical formulation of the algorithm itself and as such, there are several possibilities for further research. Firstly, the algorithm might be tested in more complex logistic networks, with other logistic solutions (e.g.) multi-echelon, lateral transshipment, cannibalization. It could be also relevant to expand the proposed solution enhancing the objective function via other ordering costs related to transportation or administrative aspects. The proposed algorithm could be also used for optimization related to additional METRIC-like solutions referred to the management of performance-based contract, such as the PBC-METRIC (Patriarca et. al., [103]). Moreover, with respect to multi-echelon scenarios, it will be relevant to compare, and possibly integrate the results of this research with other algorithms developed as a variation of the greedy algorithm, or Dantzig-Wolfe decomposition and Lagrangian heuristics (Wong et al., 2007 [109]; Topan et al., 2017 [108]). There might be also possible to test the algorithm in multi-indenture systems: systems whose LRUs are made up of multiple SRUs (Shop Replaceable Units), i.e. items at a lower level of the bill of material. In this case, it is expected to further increase the effects of reduced computational efforts for the proposed solution. Lastly, as for the general formulation, the DeB^2IFO remains conceptually suitable for a wide range of spare parts optimization problems, where items are subject to 1 by 1 replenishment policy (S-1, S) to guarantee high service levels.

Appendix A

Properties and special cases of the inverse transformations

In the following sections are reported the properties and the special cases of the inverse equations for all the geometries of the piecewise linear and non-linear transformations.

A.1 Piece Wise Linear inverse transformation

A.1.1 Properties of the inverse PLT - hypercube

Here below are listed the properties and the necessary coupling condition to preserve the continuity of the inverse transformation between the original space $\mathcal{X}$ and the transformed space $\mathcal{Y}$.

i) The reference point $\hat{y} \in \mathcal{Y}$ is transformed in the corresponding one $\hat{x}$ in the original space $\mathcal{X}$:

$$\hat{x} = T_{\hat{x}}^{-1}(\hat{y}).$$

ii) The region $\{\, y \in \mathbb{R}^n \;:\; |y_i - \hat{y}_i| \leq \epsilon_y, \; i = 1, \ldots, n \,\}$ is transformed in

$$\{\, x \in \mathbb{R}^n \;:\; |x_i - \hat{x}_i| \leq \epsilon_x, \; i = 1, \ldots, n \,\},$$

where the coupling condition at the boundary is

$$\epsilon_x = \frac{\epsilon_y}{\theta_1}. \tag{A.1}$$

iii) The region $\{\, y \in \mathbb{R}^n \;:\; \epsilon_y \leq |y_i - \hat{y}_i| \leq r_y, \; i = 1, \ldots, n \,\}$ is transformed in

$$\{\, x \in \mathbb{R}^n \;:\; \epsilon_x \leq |x_i - \hat{x}_i| \leq r_x, \; i = 1, \ldots, n \,\},$$

where the coupling condition at the boundary is

$$r_x = \epsilon_x + \frac{1}{\theta_2}(r_y - \epsilon_y). \tag{A.2}$$

iv) The region $\{\, y \in \mathbb{R}^n \;:\; |y_i - \hat{y}_i| \geq r_y \;,\; i = 1, \ldots, n \,\}$ is transformed in

$$\{\, x \in \mathbb{R}^n \;:\; |x_i - \hat{x}_i| \geq \; r_x \;,\; i = 1, \ldots, n \,\},$$

where from (1.14)

$$\bar{x}\,(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|} \left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(r_y - \epsilon_y) + \frac{1}{\theta_3}(|y_i - \hat{y}_i| - r_y) \right),$$

substituting (A.1) and (A.2)

$$\bar{x}\,(y)_i = \hat{x}_i + r_x + \frac{1}{\theta_3}(|y_i - \hat{y}_i| - r_y). \tag{A.3}$$

As long as the (A.3) refers to the outermost region there is no need to further coupling condition. It means that θ_3 is free of choice.

A.1.2   Special case of the inverse PLT - hypercube

Consider a transformation that has the following features:

- $\hat{y} = \hat{x}$;

- $\theta_1 = \frac{\epsilon_y}{\epsilon_x} \neq 1$;

- $r_x = r_y = r$;

- $\theta_2 = \frac{r - \epsilon_y}{r - \epsilon_x}$;

- $\theta_3 = 1$.

From this choices, the equation (1.12), for $y \in \mathbb{R}^n : |y_i - \hat{y}_i| \leq \epsilon_y, \ i = 1, \ldots, n$,

$$\bar{x}(y)_i = \hat{x}_i + \frac{1}{\theta_1}(y_i - \hat{y}_i),$$

becomes

$$\bar{x}(y)_i = \hat{x}_i + \frac{\epsilon_x}{\epsilon_y}(y_i - \hat{y}_i).$$

The equation (1.13), for $y \in \mathbb{R}^n : \epsilon_y \leq |y_i - \hat{y}_i| \leq r, \ i = 1, \ldots, n$,

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(|y_i - \hat{y}_i| - \epsilon_y)\right),$$

becomes

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\epsilon_x + \frac{r - \epsilon_x}{r - \epsilon_y}(|y_i - \hat{y}_i| - \epsilon_y)\right).$$

The equation (1.14), for $y \in \mathbb{R}^n : |y_i - \hat{y}_i| \geq r, \ i = 1, \ldots, n$,

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(r_y - \epsilon_y) + \frac{1}{\theta_3}(|y_i - \hat{y}_i| - r_y)\right),$$

becomes

$$\bar{x}(y)_i = y_i.$$

In summary the inverse piecewise linear transformation in the this special case:

- For $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \leq \epsilon_y, \; i = 1, \ldots, n,$

$$\bar{x}\,(y)_i = \hat{x}_i + \frac{\epsilon_x}{\epsilon_y}\,(y_i - \hat{y}_i)\,. \qquad (A.4)$$

- For $y \in \mathbb{R}^n$: $\epsilon_y \leq |y_i - \hat{y}_i| \leq r, \; i = 1, \ldots, n,$

$$\bar{x}\,(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\epsilon_x + \frac{r - \epsilon_x}{r - \epsilon_y}\,(|y_i - \hat{y}_i| - \epsilon_y)\right)\,. \qquad (A.5)$$

- For $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \geq r, \; i = 1, \ldots, n,$

$$\bar{x}\,(y)_i = y_i\,. \qquad (A.6)$$

A.1.3 Properties of the inverse PLT - hyperrectangle

Here below are listed the properties and the necessary coupling condition to preserve the continuity of the inverse transformation between the original space $\mathcal{X}$ and the transformed space $\mathcal{Y}$.

i) The reference point $\hat{y} \in \mathcal{Y}$ is transformed in the corresponding one $\hat{x}$ in the original space $\mathcal{X}$:

$$\hat{x} = T_{\hat{x}}^{-1}(\hat{y}).$$

ii) The region $\left\{ y \in \mathbb{R}^n \; : \; |y_i - \hat{y}_i| \leq \dfrac{\epsilon_y}{\nu_i}, \; i = 1, \ldots, n \right\}$ is transformed in

$$\left\{ x \in \mathbb{R}^n \; : \; |x_i - \hat{x}_i| \leq \dfrac{\epsilon_x}{\sigma_i}, \; i = 1, \ldots, n \right\},$$

where the coupling condition at the boundary is

$$\epsilon_x = \frac{\epsilon_y}{\theta_1}. \tag{A.7}$$

iii) The region $\left\{ y \in \mathbb{R}^n \; : \; \dfrac{\epsilon_y}{\nu_i} \leq |y_i - \hat{y}_i| \leq \dfrac{r_y}{\nu_i}, \; i = 1, \ldots, n \right\}$ is transformed in

$$\left\{ x \in \mathbb{R}^n \; : \; \dfrac{\epsilon_x}{\sigma_i} \leq |x_i - \hat{x}_i| \leq \dfrac{r_x}{\sigma_i}, \; i = 1, \ldots, n \right\},$$

where the coupling condition at the boundary is

$$r_x = \epsilon_x + \frac{1}{\theta_2}(r_y - \epsilon_y). \tag{A.8}$$

iv) The region $\left\{ y \in \mathbb{R}^n \; : \; |y_i - \hat{y}_i| \geq \dfrac{r_y}{\nu_i} \; , \; i = 1, \ldots, n \right\}$ is transformed in

$$\left\{ x \in \mathbb{R}^n \; : \; |x_i - \hat{x}_i| \geq \dfrac{r_x}{\sigma_i} \; , \; i = 1, \ldots, n \right\},$$

where from (1.26)

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|} \left(\frac{\nu_i}{\sigma_i} \right) \left(\frac{1}{\theta_1} \left(\frac{\epsilon_y}{\nu_i} \right) + \frac{1}{\theta_2} \left(\frac{r_y - \epsilon_y}{\nu_i} \right) + \frac{1}{\theta_3}(|y_i - \hat{y}_i| - r_y) \right);$$

substituting (A.7) and (A.8)

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|} \left(\frac{\nu_i}{\sigma_i} \right) \left(\frac{r_x}{\nu_i} + \theta_3 \left(|y_i - \hat{y}_i| + \frac{r_y}{\nu_i} \right) \right). \tag{A.9}$$

As long as the (A.9) refers to the outermost region there is no need to further coupling condition. It means that θ_3 is free of choice.

A.1.4   Special case of the inverse PLT - hyperrectangle

Consider a transformation that has the following features:

- $\hat{x} = \hat{y}$;

- $D_x = D_y$ and so $\sigma_i = \nu_i$ for $i = 1, \ldots, n$;

- $\theta_1 = \frac{\epsilon_y}{\epsilon_x} \neq 1$;

- $r_x = r_y = r$;

- $\theta_2 = \frac{r - \epsilon_y}{r - \epsilon_x}$;

- $\theta_3 = 1$.

From this choices, the equation (1.24), for $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \leq \frac{\epsilon_y}{\nu_i}$, $i = 1, \ldots, n$,

$$\bar{x}(y)_i = \hat{x}_i + \frac{1}{\theta_1} \left(\frac{\nu_i}{\sigma_i}\right)(y_i - \hat{y}_i),$$

becomes

$$\bar{x}(y)_i = \hat{x}_i + \frac{\epsilon_x}{\epsilon_y}(y_i - \hat{y}_i).$$

The equation (1.25), for $y \in \mathbb{R}^n$: $\frac{\epsilon_y}{\nu_i} \leq |y_i - \hat{y}_i| \leq \frac{r}{\nu_i}$, $i = 1, \ldots, n$,

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\frac{\nu_i}{\sigma_i}\right)\left(\frac{1}{\theta_1}\left(\frac{\epsilon_y}{\nu_i}\right) + \frac{1}{\theta_2}\left(|y_i - \hat{y}_i| - \frac{\epsilon_y}{\nu_i}\right)\right),$$

becomes

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\left(\frac{\epsilon_x}{\nu_i}\right) + \frac{r - \epsilon_y}{\nu_i}\left(|y_i - \hat{y}_i| - \frac{\epsilon_y}{\nu_i}\right)\right).$$

The equation (1.26), for $x \in \mathbb{R}^n$: $|x_i - \hat{x}_i| \geq r$, $i = 1, \ldots, n$,

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\frac{\nu_i}{\sigma_i}\right)\left(\frac{1}{\theta_1}\left(\frac{\epsilon_y}{\nu_i}\right) + \frac{1}{\theta_2}\left(\frac{r_y - \epsilon_y}{\nu_i}\right) + \frac{1}{\theta_3}(|y_i - \hat{y}_i| - r_y)\right).$$

becomes

$$\bar{x}(y)_i = y_i.$$

In summary the inverse piecewise linear transformation in this special case:

- For $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \leq \frac{\epsilon_y}{\nu_i}$, $i = 1, \ldots, n$,

$$\bar{x}(y)_i = \hat{x}_i + \frac{1}{\theta_1}(y_i - \hat{y}_i). \tag{A.10}$$

- For $y \in \mathbb{R}^n$: $\frac{\epsilon_y}{\nu_i} \leq |y_i - \hat{y}_i| \leq \frac{r}{\nu_i}$, $i = 1, \ldots, n$,

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\left(\frac{\epsilon_x}{\nu_i}\right) + \frac{r - \epsilon_y}{\nu_i}\left(|y_i - \hat{y}_i| - \frac{\epsilon_y}{\nu_i}\right)\right). \tag{A.11}$$

- For $y \in \mathbb{R}^n$: $|y_i - \hat{y}_i| \geq \frac{r}{\nu_i}$, $i = 1, \ldots, n$,

$$\bar{x}(y)_i = y_i. \tag{A.12}$$

A.2   Non-Linear inverse transformation

A.2.1   Properties of the inverse NLT - hypersphere

Here below are listed the properties and the necessary coupling condition to preserve the continuity of the inverse transformation between the original space $\mathcal{X}$ and the transformed space $\mathcal{Y}$.

  i)   The reference point $\hat{y} \in \mathcal{Y}$ is transformed in the corresponding one $\hat{x}$ in the original space $\mathcal{X}$:

$$\hat{x} = T_{\hat{x}}^{-1}(\hat{y}).$$

  ii)   The region $\{\, y \in \mathbb{R}^n \;:\; \|y - \hat{y}\|_2 \le \epsilon_y \,\}$ is transformed in

$$\{\, x \in \mathbb{R}^n \;:\; \|x - \hat{x}\|_2 \le \epsilon_x \,\},$$

where the coupling condition at the boundary is

$$\epsilon_x = \frac{\epsilon_y}{\theta_1}. \tag{A.13}$$

  iii)   The region $\{\, y \in \mathbb{R}^n \;:\; \epsilon_y \le \|y - \hat{y}\|_2 \le r_y \,\}$ is transformed in

$$\{\, x \in \mathbb{R}^n \;:\; \epsilon_x \le \|x - \hat{x}\|_2 \le r_x \,\},$$

where the coupling condition at the boundary is

$$r_x = \epsilon_x + \frac{1}{\theta_2}(r_y - \epsilon_y). \tag{A.14}$$

  iv)   The region $\{\, \|y - \hat{y}\|_2 \ge r_y \,\}$ is transformed in

$$\{\, x \in \mathbb{R}^n \;:\; \|x - \hat{x}\|_2 \ge r_x \,\}$$

where from (1.47)

$$\|x - \hat{x}\|_2 - r_x = \frac{1}{\theta_3}(\|y - \hat{y}\|_2 - r_y),$$

substituting (A.13) and (A.14)

$$\|x - \hat{x}\|_2 - r_x = \frac{1}{\theta_3}(\|y - \hat{y}\|_2 - r_y). \tag{A.15}$$

As long as the (A.15) refers to the outermost region there is no need to further coupling condition. It means that θ_3 is free of choice.

A.2.2 Special case of the inverse NLT - hypersphere

It is considered a transformation that has the following features:

- $\hat{y} = \hat{x}$;

- $\theta_1 = \frac{\epsilon_y}{\epsilon_x} \neq 1$;

- $r_x = r_y = r$;

- $\theta_2 = \frac{r - \epsilon_y}{r - \epsilon_x}$;

- $\theta_3 = 1$.

From this choices, the equation (1.45), for $y \in \mathbb{R}^n \; : \|y - \hat{y}\|_2 \leq \epsilon_y$,

$$\bar{x}(y) = \hat{x} + \frac{1}{\theta_1} (y - \hat{y}),$$

becomes

$$\bar{x}(y) = \hat{x} + \frac{\epsilon_x}{\epsilon_y} (y - \hat{y}).$$

The equation (1.46), for $y \in \mathbb{R}^n \; : \; \epsilon_y \leq \|y - \hat{y}\|_2 \leq r_y$,

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \left(\frac{1}{\theta_1} \epsilon_y + \frac{1}{\theta_2} (\|y - \hat{y}\|_2 - \epsilon_y) \right),$$

becomes

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \left(\epsilon_x + \frac{r - \epsilon_x}{r - \epsilon_y} (\|y - \hat{y}\|_2 - \epsilon_y) \right).$$

The equation (1.47), for $x \in \mathbb{R}^n \; : \; \|y - \hat{y}\|_2 \geq r_y$,

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \left(\frac{1}{\theta_1} \epsilon_y + \frac{1}{\theta_2} (r_y - \epsilon_y) + \frac{1}{\theta_3} (\|y - \hat{y}\|_2 - r_y) \right),$$

becomes

$$\bar{x}(y) = y.$$

In summary the inverse nonlinear transformation in this special case:

- For $y \in \mathbb{R}^n\ :\ \|y - \hat{y}\|_2 \leq \epsilon_y$,

$$\bar{x}\left(y\right) = \hat{x} + \frac{\epsilon_x}{\epsilon_y}\left(y - \hat{y}\right). \tag{A.16}$$

- For $y \in \mathbb{R}^n\ :\ \epsilon_y \leq \|y - \hat{y}\|_2 \leq r_y$,

$$\bar{x}\left(y\right) = \hat{x} + \frac{\left(y - \hat{y}\right)}{\|y - \hat{y}\|_2}\left(\epsilon_x + \frac{r - \epsilon_x}{r - \epsilon_y}\left(\|y - \hat{y}\|_2 - \epsilon_y\right)\right). \tag{A.17}$$

- For $x \in \mathbb{R}^n\ :\ \|y - \hat{y}\|_2 \geq r_y$,

$$\bar{x}\left(y\right) = y. \tag{A.18}$$

A.2.3 Properties of the inverse NLT - hyperellipse

Here below are listed the properties and the necessary coupling condition to preserve the continuity of the inverse transformation between the original space $\mathcal{X}$ and the transformed space $\mathcal{Y}$.

i) The reference point $\hat{y} \in \mathcal{Y}$ is transformed in the corresponding one $\hat{x}$ in the original space $\mathcal{X}$:

$$\hat{x} = T_{\hat{x}}^{-1}(\hat{y}).$$

ii) The region $\left\{ y \in \mathbb{R}^n \; : \; \|D_y(y - \hat{y})\|_2 \le \epsilon_y \right\}$ is transformed in

$$\left\{ x \in \mathbb{R}^n \; : \; \|D_x(x - \hat{x})\|_2 \le \epsilon_x \right\},$$

where the coupling condition at the boundary is

$$\epsilon_x = \frac{\epsilon_y}{\theta_1}. \tag{A.19}$$

iii) The region $\left\{ y \in \mathbb{R}^n \; : \epsilon_y \le \|D_y(y - \hat{y})\|_2 \le r_y \right\}$ is transformed in

$$\left\{ x \in \mathbb{R}^n \; : \epsilon_x \le \|D_x(x - \hat{x})\|_2 \le r_x \right\},$$

where the coupling condition at the boundary is

$$r_x = \epsilon_x + \frac{1}{\theta_2}(r_y - \epsilon_y). \tag{A.20}$$

iv) The region $\left\{ \|D_y(y - \hat{y})\|_2 \ge r_y \right\}$ is transformed in

$$\left\{ x \in \mathbb{R}^n \; : \; \|D_x(x - \hat{x})\|_2 \ge r_x \right\}$$

where from (1.59)

$$\bar{x}(y) = \hat{x} + D_x^{-1} \frac{D_y(y - \hat{y})}{\|D_y(y - \hat{y})\|_2} \left(\frac{1}{\theta_1}\epsilon_y + \frac{1}{\theta_2}(r_y - \epsilon_y) + \frac{1}{\theta_3}\left(\|D_y(y - \hat{y})\|_2 - r_y\right) \right),$$

substituting (A.19) and (A.20)

$$\bar{x}(y) = \hat{x} + D_x^{-1} \frac{D_y(y - \hat{y})}{\|D_y(y - \hat{y})\|_2} \left(r_y + \frac{1}{\theta_3}\left(\|D_y(y - \hat{y})\|_2 - r_y\right) \right). \tag{A.21}$$

As long as the (A.21) refers to the outermost region there is no need to further coupling condition. It means that θ_3 is free of choice.

A.2.4   Special case of the inverse NLT - hyperellipse

Let us consider a transformation that has the following features:

- $\hat{x} = \hat{y}$;

- $D_x = D_y$ and so $\sigma_i = \nu_i$ for $i = 1, \ldots, n$;

- $\theta_1 = \frac{\epsilon_y}{\epsilon_x} \neq 1$;

- $r_x = r_y = r$;

- $\theta_2 = \frac{r - \epsilon_y}{r - \epsilon_x}$;

- $\theta_3 = 1$.

From this choices, the equation (1.57), for $y \in \mathbb{R}^n \; : \; \|D_y(y - \hat{y})\|_2 \leq \epsilon_y$,

$$\bar{x}\,(y) = \hat{x} + \frac{1}{\theta_1} D_x^{-1} D_y \,(y - \hat{y})\,,$$

becomes

$$\bar{x}\,(y) = \hat{x} + \frac{\epsilon_x}{\epsilon_y} D_x^{-1} D_y \,(y - \hat{y})\,.$$

The equation (1.58), for $y \in \mathbb{R}^n \; : \; \epsilon_y \leq \|D_y(y - \hat{y})\|_2 \leq r_y$,

$$\bar{x}\,(y) = \hat{x} + D_x^{-1} \frac{D_y(y - \hat{y})}{\|D_y(y - \hat{y})\|_2} \left(\frac{1}{\theta_1} \epsilon_y + \frac{1}{\theta_2} \left(\|D_y(y - \hat{y})\|_2 - \epsilon_y \right) \right),$$

becomes

$$\bar{x}\,(y) = \hat{x} + D_x^{-1} \frac{D_y(y - \hat{y})}{\|D_y(y - \hat{y})\|_2} \left(\frac{\epsilon_x}{\epsilon_y} + \frac{r - \epsilon_x}{r - \epsilon_y} \left(\|D_y(y - \hat{y})\|_2 - \epsilon_y \right) \right).$$

The equation (1.59), for $x \in \mathbb{R}^n \; : \; \|D_y(y - \hat{y})\|_2 \geq r_y$,

$$\bar{x}\,(y) = \hat{x} + D_x^{-1} \frac{D_y(y - \hat{y})}{\|D_y(y - \hat{y})\|_2} \left(\frac{1}{\theta_1} \epsilon_y + \frac{1}{\theta_2}(r_y - \epsilon_y) + \frac{1}{\theta_3} \left(\|y - \hat{y}\|_2 - r_y \right) \right),$$

becomes

$$\bar{x}\,(y) = y.$$

In summary the inverse nonlinear transformation in this special case:

- For $y \in \mathbb{R}^n$: $\|D_y(y - \hat{y})\|_2 \leq \epsilon_y$,

$$\bar{x}\left(y\right) = \hat{x} + \frac{\epsilon_x}{\epsilon_y} D_x^{-1} D_y \left(y - \hat{y}\right). \tag{A.22}$$

- For $y \in \mathbb{R}^n$: $\epsilon_y \leq \|D_y(y - \hat{y})\|_2 \leq r_y$,

$$\bar{x}\left(y\right) = \hat{x} + D_x^{-1} \frac{D_y(y - \hat{y})}{\|D_y(y - \hat{y})\|_2} \left(\epsilon_x + \frac{r - \epsilon_x}{r - \epsilon_y} \left(\|D_y(y - \hat{y})\|_2 - \epsilon_y\right)\right). \tag{A.23}$$

- For $x \in \mathbb{R}^n$: $\|D_y(y - \hat{y})\|_2 \geq r_y$,

$$\bar{x}\left(y\right) = y. \tag{A.24}$$

Appendix B

Report on item's holding costs

The following table provide item's holding costs in € per demand scenario and per Local Warehouse.

Table B.1. Item's holding costs in € for each demand and at each LW.

Demand	LW	DeB²IFO	NM-BBOA	GABRLS	MADS 2N	Marginal Analysis
1	LW1	23566,89	23566,89	23566,89	24634,95	28714,21
	LW2	22251,05	22251,05	22251,05	22381,09	22881,35
	LW3	40547,17	40835,06	40835,06	41303,13	42436,93
	Total	86365,11	86652,99	86652,99	88319,17	94032,48
2	LW1	18478,07	18542,34	18478,07	19250,1	20324,77
	LW2	19648,67	20461,29	20461,29	21950,27	22216,94
	LW3	40354,59	40409,27	41299,89	41201,48	44397,73
	Total	78481,33	79412,91	80239,25	82401,86	86939,44
3	LW1	23566,89	23566,89	23566,89	27804,65	28714,21

	LW2	22251,05	22251,05	22251,05	22323,77	22881,35
	LW3	40547,17	40835,06	40835,06	41303,13	42436,93
	Total	86365,11	86652,99	86652,99	91431,55	94032,48
4	LW1	18478,07	18542,34	18478,07	18668,77	20324,77
	LW2	19648,67	20461,29	20461,29	21586,64	22216,94
	LW3	40354,59	40409,27	41299,89	42238,63	44397,73
	Total	78481,33	79412,91	80239,25	82494,03	86939,44
5	LW1	19533,26	19533,26	19589,11	19786,78	19786,78
	LW2	25537,68	25537,68	27257,45	28269,57	29388,45
	LW3	35999,3	35950,45	35999,30	36175,46	36175,46
	Total	81070,25	81021,39	82845,86	84231,81	85350,69
6	LW1	19888,64	19888,64	19888,64	19993,07	19993,07
	LW2	23223,02	23223,02	23223,02	25071,17	26092,98
	LW3	36661,16	36789,17	36957,79	36652,83	37771,71
	Total	79772,82	79900,83	80069,45	81717,07	83857,76
7	LW1	17026,2	17026,20	17026,20	17026,2	17026,2
	LW2	25289,88	26017,64	25911,35	26761,21	28052,14
	LW3	41502,06	41502,06	41502,06	41502,06	41502,06

	Total	83818,14	84545,91	84439,61	85289,47	86580,41
8	LW1	23566,89	23566,89	23566,89	24036,82	28714,21
	LW2	22251,05	22251,05	22251,05	22323,77	22881,35
	LW3	40547,17	40835,06	40835,06	42436,93	42436,93
	Total	86365,11	86652,99	86652,99	88797,52	94032,48
9	LW1	18478,07	18542,34	18478,07	20108,45	20324,77
	LW2	19648,67	20461,29	20461,29	21229,59	22216,94
	LW3	40354,59	40409,27	41299,89	42181,94	44397,73
	Total	78481,33	79412,91	80239,25	83519,98	86939,44
10	LW1	19896,23	19995,07	19970,83	20082,71	20082,71
	LW2	23463,68	24422,22	24461,38	24177,38	25252,05
	LW3	29185,68	29321,32	29334,37	29523,72	29804,3
	Total	72545,59	73738,61	73766,58	73783,81	75139,07
11	LW1	23566,89	23566,89	23566,89	25838,95	28714,21
	LW2	22251,05	22251,05	22251,05	22705,19	22881,35
	LW3	40547,17	40835,06	40835,06	41929,7	42436,93
	Total	86365,11	86652,99	86652,99	90473,84	94032,48
12	LW1	18478,07	18542,34	18478,07	20058,1	20324,77

	LW2	19648,67	20461,29	20461,29	22056,57	22216,94
	LW3	40354,59	40409,27	41299,89	41304,58	44397,73
	Total	78481,33	79412,91	80239,25	83419,25	86939,44
13	LW1	20941,92	21251,17	21251,17	21467,49	21907,11
	LW2	23791,91	23743,65	23791,91	23829,2	23829,2
	LW3	38853,24	38853,24	38853,24	38853,24	38853,24
	Total	83587,07	83848,05	83896,31	84149,93	84589,55
14	LW1	23462,76	24546,16	23987,38	25741,03	25741,03
	LW2	25987,32	25987,32	25987,32	25987,32	25987,32
	LW3	35194,15	36048,96	36152,67	36935,04	37132,71
	Total	84644,23	86582,43	86127,37	88663,39	88861,06
15	LW1	18458,65	18354,22	18354,22	18709,53	18885,69
	LW2	25279,49	25957,44	25957,44	27236,69	27236,69
	LW3	34999,58	34999,58	35537,85	36151,27	36943,91
	Total	78737,72	79311,24	79849,50	82097,49	83066,29
16	LW1	19497,57	19497,57	19497,57	19497,57	19497,57
	LW2	23458,77	24045,43	24045,43	26143,33	26334,03
	LW3	42952,91	42952,91	43070,86	43070,86	43070,86

	Total	85909,25	86495,91	86613,86	88711,76	88902,46
17	LW1	24420,26	24420,26	24420,26	24596,42	24890,19
	LW2	21168,16	21168,16	21168,16	21168,16	21168,16
	LW3	34985,39	34985,39	35569,12	35440,78	37627,36
	Total	80573,81	80573,81	81157,55	81205,36	83685,71
18	LW1	17382,66	17382,66	17382,66	17382,66	17382,66
	LW2	22690,73	22690,73	22903,86	23282,41	25316,91
	LW3	33593,27	33593,27	33593,27	34093,53	34093,53
	Total	73666,65	73666,65	73879,79	74758,6	76793,1
19	LW1	17977,21	17977,21	17977,21	18137,58	18137,58
	LW2	23985,26	24189,88	24071,04	26654,01	27528,09
	LW3	40660,19	40660,19	40660,19	40660,19	40660,19
	Total	82622,65	82827,28	82708,44	85451,78	86325,87
20	LW1	17539,25	17539,25	17992,39	17992,39	18096,82
	LW2	20521,95	21109,98	21523,08	22440,21	25397,01
	LW3	36946,9	36920,53	36946,90	38673,65	41996,76
	Total	75008,1	75569,76	76462,38	79106,25	85490,59
21	LW1	18972,43	18972,43	18972,43	19244,69	19442,36

	LW2	25488,18	25488,18	26072,86	26233,23	26233,23
	LW3	37843,51	38250,57	38250,57	38481,3	39038,88
	Total	82304,11	82711,18	83295,86	83959,22	84714,47
22	LW1	24296,02	24351,40	24351,40	24993,76	25568,12
	LW2	23999,27	23999,27	23999,27	24444,95	24556,84
	LW3	31173,85	31110,05	32391,69	34168,79	35387
	Total	79469,14	79460,72	80742,36	83607,5	85511,96
23	LW1	16432,94	16432,94	16470,23	16432,94	16630,6
	LW2	24501,81	24501,81	24501,81	24501,81	25059,38
	LW3	39310,65	39310,65	39919,07	41255,31	41549,67
	Total	80245,39	80245,39	80891,11	82190,05	83239,66

Table B.2. Total Item's holding costs in € for all LW and all demands.

	DeB^{2}IFO	NM BBOA	GABRLS	MADS 2N	Marginal Analysis
Total €	**1863361**	**1874763**	**1884315**	**1929781**	**1985996**
Δ€	–	11402	20954	66540	122635
% Δ	–	% 0,61	% 1,12	% 3,57	% 6,59

Bibliography

[1] Sergeyev, Ya D., Kvasov, D.E.: Deterministic Global Optimization: An Introduction to the Diagonal Approach. Springer, New York (2017) https://doi.org/10.1007/978-1-4939-7199-2.

[2] Jones, D.R., Perttunen, C.D., Stuckman, B.E.: Lipschitzian optimization without the Lipschitz constant. J. Optim. Theory Appl. 79(1), 157–181 (1993) https://doi.org/10.1007/BF00941892.

[3] Lera, D., Sergeyev, Y.D.: GOSH: derivative-free global optimization using multi-dimensional space-filling curves. J. Glob Optim (2017) https://doi.org/10.1007/s10898-017-0589-7.

[4] Lera, D., Sergeyev, Y.D.: Deterministic global optimization using space-filling curves and multiple estimates of Lipschitz and Hölder constants. Commun. Nonlinear Sci. Numer. Simul. 23(1–3), 328–342 (2015).

[5] Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P.: Optimization by simulated annealing. Science 220, 671–680 (1983).

[6] Francisco J. Solis and Roger J.-B. Wets.: Minimization by Random Search Techniques. Mathematics of Operations Research, 6(1):19–30, ISSN 0364-765X (1981). doi: 10.1287/moor.6.1.19.

[7] Dixon, L. C. W., Szego, G. P. (eds.): Towards global optimization 1: North-Holland, Amsterdam (1975).

[8] Dixon, L. C. W., Szego, G. P. (eds.): Towards global optimization 2: North-Holland, Amsterdam (1978).

[9] Greenhalgh D., Marshall, S.: Convergence criteria for genetic algorithms, SIAM J. Comput. 30(1), 269–282 (2000) https://doi.org/10.1137/S009753979732565X.

[10] Hartl, R. F., Belew, R. K.: A global convergence proof for a class of genetic algorithms. Tech. rep. University of Technology, Vienna (1990).

[11] Rudolph, G.: Convergence of evolutionary algorithms in general search spaces. Proceedings of the Third IEEE Conference on Evolutionary Computation, pp. 50–54, IEEE Press, Piscataway (NJ) (1996).

[12] Liuzzi, G., Lucidi, S., Piccialli, V.: Exploiting derivative-free local searches in direct-type algorithms for global optimization. Comput. Optim. Appl. 65, 449–475 (2016) https://doi:10.1007/s10589-015-9741-9.

[13] Lucidi, S., Piccialli, V. New Classes of Globally Convexized Filled Functions for Global Optimization. Journal of Global Optimization 24, 219–236 (2002) https://doi.org/10.1023/A:1020243720794.

[14] Xu, Z., Huang, H., Pardalos, P.M. et al.: Filled functions for unconstrained global optimization. Journal of Global Optimization 20, 49–65 (2001) https://doi.org/10.1023/A:1011207512894.

[15] Wu, Z.Y., Lee, H.W.J., Zhang, L.S. et al.: A Novel Filled Function Method and Quasi-Filled Function Method for Global Optimization. Comput Optim Applic 34, 249–272 (2006). https://doi.org/10.1007/s10589-005-3077-9.

[16] Bertsekas, D.P.: Constrained Optimization and Lagrange Multipliers Methods. Academic Press, New York (1982) https://doi.org/10.1002/net.3230150112.

[17] Di Pillo, G., Lucidi, S., Rinaldi, F.: A Derivative-Free Algorithm for Constrained Global Optimization Based on Exact Penalty Functions. J Optim Theory Appl 164, 862–882 (2015) https://doi.org/10.1007/s10957-013-0487-1.

[18] Liuzzi, G., Lucidi, S., Piccialli, V., Sotgiu, A.: A magnetic resonance device designed via global optimization techniques. Math. Program. 101(2), 339–364 (2004).

[19] Bertolazzi, P., Guerra, C., Liuzzi, G.: A global optimization algorithm for protein surface alignment. BMC Bioinform. 11, 488–498 (2010).

[20] Locatelli, M., Schoen, F.: Efficient algorithms for large scale global optimization: Lennard-Jones clusters. Comput. Optim. Appl. 26(2), 173–190 (2003).

[21] Jones, D.R., Martins, J.R.R.A.: The DIRECT algorithm: 25 years Later. J Glob Optim (2020). https://doi.org/10.1007/s10898-020-00952-6.

[22] Liuzzi, G., Lucidi, S., Piccialli, V.: A DIRECT-based approach exploiting local minimizations for the solution of large-scale global optimization problems. Comput. Optim. Appl. 45, 353–375 (2010).

[23] Sergeyev, Ya D., Kvasov, D.E.: Deterministic Global Optimization: An Introduction to the Diagonal Approach. Springer, New York (2017) https://doi.org/10.1007/978-1-4939-7199-2.

[24] J. J. Liang, Q. B. Y., and S. P. N.: "Problem Definitions and Evaluation Criteria for the CEC 2013 Special Session and Competition on Real-Parameter Optimization," Computational Intelligence Laboratory, Zhengzhou University, Zhengzhou China And Nanyang Technological University, Singapore, Technical Report 201212, (2013).

[25] Liuzzi, G., Lucidi, S., Piccialli, V.: A partition-based global optimization algorithm. J Global Optim 48(1), 113–128 (2010) https://doi.org/10.1007/s10898-009-9515-y.

[26] Goldberg, D.E., Kalyanmoy, D.: A comparative analysis of selection schemes used in genetic algorithms. Found. Genetic Algorithms 1, 69–93 (1991).

[27] Al-Baali, M., Caliciotti, A., Fasano, G., Roma, M.: A Class of Approximate Inverse Preconditioners Based on Krylov-Subspace Methods for Large-Scale Nonconvex Optimization in SIAM JOURNAL ON OPTIMIZATION, vol. 30, pp. 1954-1979 (ISSN 1052-6234) (2020)

[28] Caliciotti, A., Fasano, G., Roma, M.: Preconditioned Nonlinear Conjugate Gradient methods based on a modified secant equation in APPLIED MATHEMATICS AND COMPUTATION, vol. 318, pp. 196-214 (ISSN 0096-3003) (2018)

[29] Fasano, G., Roma, M.: Preconditioning Newton-Krylov Methods in Non-Convex Large Scale Optimization in COMPUTATIONAL OPTIMIZATION AND APPLICATIONS, vol. 56, pp. 253-290 (ISSN 0926-6003) (2013)

[30] Morales JL, Nocedal J. Automatic preconditioning by limited memory Quasi-Newton updating. SIAM J Optim. 10(4): 1079–1096 (2000)

[31] Brown, P.N., Walker, H.F., Wasyk, R., Woodward, C.S.: On using approximate finite-differences in matrix-free Newton–Krylov methods. SIAM J. Numer. Anal. 46, 1892–1911 (2008)

[32] Dixon, L. C. W., Szego, G. P.: The global optimization problem: an introduction. Towards global optimization, 2, 1-15 (1978)

[33] Picheny, V., Wagner, T., Ginsbourger, D.: A benchmark of kriging-based infill criteria for noisy optimization (2012)

[34] Lucidi, S., Sciandrone, M.: A Derivative-Free Algorithm for Bound Constrained Optimization, Computational Optimization and Applications, 21(2): 119-142 (2002) https://doi.org/10.1023/A:1013735414984

[35] Yun-Wei Shang and Yu-Huang Qiu. A note on the extended rosenbrock function. Evol. Comput., 14(1):119–126 (2006)

[36] Matsumoto, M., Nishimura, T.: Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul., 8(1), 3–30 (1998) https://doi.org/10.1145/272991.272995

[37] Horst, R. & Pardalos, P. M. (eds): Handbook of Global Optimization, vol. 1, Kluwer Academic Publishers, Dordrecht (1995)

[38] Pintér, J. D. Global Optimization in Action (Continuous and Lipschitz Optimization: Algorithms, Implementations and Applications). Kluwer Academic Publishers, Dordrecht (1996)

[39] Price, K., Storn, R. M., Lampinen, J. A. Differential Evolution: A Practical Approach to Global Optimization. Natural Computing Series. Springer, New York (2005)

[40] Sergeyev, Y. D., Strongin, R. G., Lera, D.: Introduction to Global Optimization Exploiting Space-Filling Curves. (Springer, New York, 2013).

[41] Holland, J. H.: Adaptation in Natural and Artifcial Systems: an Introductory Analysis with Applications to Biology, Control, and Artifcial Intelligence. University of Michigan Press (1975)

[42] Strongin, R. G., Sergeyev, Y. D.: Global Optimization with Non-Convex Constraints: Sequential and Parallel Algorithms. Kluwer Academic Publishers, Dordrecht (2000)

[43] Paulavičius R., Žilinskas J.: Simplicial Global Optimization. SpringerBriefs in Optimization. Springer-Verlag, New York (2014).

[44] Sergeyev, Y., D., & Kvasov, D., E.: Deterministic Global Optimization. Springer (2017).

[45] Pardalos, P. M. & Romeijn, H. E. (eds): Handbook of Global Optimization, volume 2. Kluwer Academic Publishers, Dordrecht, (2002).

[46] Journal of Global Optimization https://www.springer.com/journal/10898.

[47] Grippo, L., Sciandrone, M.: Metodi di Ottimizzazione non Vincolata. Springer (2011).

[48] Schoen, F.: Two-phase methods for global optimization, Handbook of global optimization, Vol. 2, Kluwer Acad. Publ., Dordrecht, pp. 151-177 (2002).

[49] Conn, A. R., Scheinberg, K., Vicente, L. N.: Introduction to Derivative-Free Optimization. SIAM, Philadelphia, USA, (2009).

[50] Jones, D.R. A Taxonomy of Global Optimization Methods Based on Response Surfaces. Journal of Global Optimization 21, 345–383 (2001) https://doi.org/10.1023/A:1012771025575.

[51] Hu J., Wang Y., Zhou E., Fu M.C., Marcus S.I.: A Survey of Some Model-Based Methods for Global Optimization. In: Hernández-Hernández D., Minjárez-Sosa J. (eds) Optimization, Control, and Applications of Stochastic Systems. Systems & Control: Foundations & Applications. Birkhäuser, Boston (2012).

[52] Bertsimas, D., Tsitsiklis, J.: Simulated annealing. Statistical Science 8(1), 10–15 (1993).

[53] Wah, B., Chen, Y., Wang, T.: Simulated annealing with asymptotic convergence for nonlinear constrained global optimization. Journal of Global Optimization 39, 153–162 (2002).

[54] Reeves, C.: Genetic algorithms for the operations researcher. INFORMS Journal on Computing 9, 231–250 (1997).

[55] Greenhalgh D., Marshall, S.: Convergence criteria for genetic algorithms, SIAM J. Comput. 30(1), 269–282 (2000).

[56] Rudolph, G.: Convergence of evolutionary algorithms in general search spaces. Proceedings of the Third IEEE Conference on Evolutionary Computation, pp. 50–54, IEEE Press, Piscataway (NJ), (1996).

[57] Glover, F.: Tabu search: A tutorial. Interfaces 20, 74–94 (1990). DOI 10.1287/inte.20.4.74

[58] C. A. Floudas and P. M. Pardalos, editors. Encyclopedia of Optimiza-tion (6 Volumes). Springer, 2nd edition, (2009).

[59] Serani, A., Fasano, G., Liuzzi, G., Lucidi, S., Iemma, U., Campana, E.F., Stern, F., Diez, M.: Ship hydrodynamic optimization by local hybridization of deterministic derivative-free global algorithms. Appl. Ocean Res. 59, 115–128 (2016).

[60] Di Pillo, G., Facchinei, F.: Exact barrier function methods for Lipschitz programs. Appl. Math. Optim. 32, 1–31 (1995) https://doi.org/10.1007/BF01189901.

[61] Battiti R., Sergeyev Y.D., Brunato M., Kvasov D.E. GENOPT 2016: Design of a generali-zation-based challenge in global optimization. In Sergeyev Y.D., Kvasov D.E., Dell'Accio F., Mukhametzhanov. M. S. (Eds.), AIP Conference Proceedings, Vol. 1776. No. 060005. AIP Publishing (2016).

[62] Regis, R. G.: On the properties of positive spanning sets and positive bases. Optim. Eng. 17(1) 229-262 (2016) https://doi.org/10.1007/s11081-015-9286-x.

[63] Lewis, R.M., Torczon V.: Rank ordering and positive bases in pattern search algorithms. Technical Report TR96-71, ICASE, NASA Langley Research Center (1999).

[64] Abramson, M.A., Audet, C., Dennis, J.E.Jr., Le Digabel, S.: OrthoMADS: a deterministic mads instance with orthogonal directions. SIAM Journal on Optimization 20(2), 948-966. (2009) https://doi.org/10.1137/080716980.

[65] Kelley, C.T.: Implicit Filtering (Software, Environments and Tools). SIAM (2011) https://doi.org/10.1137/1.9781611971903.

[66] Lucidi, S., Sciandrone, M.: On the global convergence of derivative-free methods for unconstrained optimization, SIAM J. Optim. 13, 97–116 (2002) https://doi.org/10.1137/S1052623497330392.

[67] Conn, A.R., Gould, N.I.M., Toint,P.L.: Trust-Region Methods, MPS/SIAM Ser. Optim. 1, SIAM, Philadelphia, (2000) https://doi.org/10.1137/1.9780898719857.

[68] Coope, I.D., Price, C.J.: On the convergence of grid-based methods for unconstrained optimization. SIAM Journal on Optimization 11, 859-869 (2001) https://doi.org/10.1137/S1052623499354989.

[69] Lewis, R.M., Torczon, V.: Pattern search algorithms for bound constrained optimization. SIAM Journal on Optimization 9(4), 1082-1099, (1999).

[70] Coope, I.D., Price, C.J.: Positive bases in numerical optimization. Computational Optimization and Applications 21(2), 169-175 (2002).

[71] Coope, I.D., Price, C.J.: Frames and grids in unconstrained and linearly constrained optimization: A nonsmooth approach, SIAM J. Optim. 14, 415–438 (2003).

[72] Goldberg, D.E., Kalyanmoy, D.: A comparative analysis of selection schemes used in genetic algorithms. Found. Genetic Algorithms 1, 69–93 (1991).

[73] Halton, J.: On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numerische Mathematik 2, 84–90 (1960).

[74] Di Pillo, G., Grippo, L.: Exact penalty functions in constrained optimization. Siam J. Control Optim. 27, 1333–1360 (1989).

[75] Liuzzi, G., Lucidi, S., Rinaldi, F.: Derivative-free methods for bound constrained mixed-integer optimization. Computational Optimization and Applications - COMPUT OPTIM APPL 53 (2012). DOI https://doi.org/10.1007/s10589-011-9405-3.

[76] G.Liuzzi, S.Lucidi, F.Rinaldi. Derivative-Free Methods for Mixed-Integer Constrained Optimization Problems, Journal of Optimization Theory and Applications, 164(3): 933-965 (2015) https://doi.org/10.1007/s10957-014-0617-4.

[77] Lewis, R., Torczon, V.: A globally convergent augmented lagrangian pattern search algorithm for optimization with general constraints and simple bounds. SIAM J. Optim. 12, 1075–1089 (2002).

[78] Lin, C.J., Lucidi, S., Palagi, L., Risi, A.: Decomposition algorithm model for singly linearly-constrained problems subject to lower and upper bounds. Journal of Optimization Theory and Applications 141, 107–126 (2009). https://doi.org/10.1007/s10957-008-9489-9.

[79] Liuzzi, G., Lucidi, S., Sciandrone, M.: Sequential penalty derivative-free methods for nonlinear constrained optimization. SIAM J. Optim. 20, 2614–2635 (2010).

[80] Sergeyev, Y.D., Kvasov, D.E.: Global search based on efficient diagonal partitions and a set of Lipschitz constants. SIAM J. Optim. 16(3), 910–937 (2006).

[81] Nævdal, G. Positive bases with maximal cosine measure. Optim Lett 13, 1381–1388 (2019). https://doi.org/10.1007/s11590-018-1334-y

[82] Bernabei, G., Costantino, F., Palagi, L., Patriarca, R., Romito, F.: An Integer Black-Box Optimization Model for Repairable Spare Parts Management. International Journal of Information Systems and Supply Chain Management (IJISSCM), 14(2), 46-68 (2021) https://doi.org/10.4018/IJISSCM.2021040103

[83] Sleptchenko, A., Hanbali, A. Al, Zijm, H.: Joint planning of Service Engineers and Spare Parts. Eur. J. Oper. Res. 271, 97–108 (2018) https://doi.org/10.1016/j.ejor.2018.05.014.

[84] Díaz, A., Fu, M.C.: Models for multi-echelon repairable item inventory systems with limited repair capacity. Eur. J. Oper. Res. 97, 480–492 (1997) https://doi.org/10.1016/S0377-2217(96)00279-2.

[85] Wong, H., Van Houtum, G.J., Cattrysse, D., Van Oudheusden, D.: Multi-item spare parts systems with lateral transshipments and waiting time constraints. Eur. J. Oper. Res. 171, 1071–1093 (2006) https://doi.org/10.1016/j.ejor.2005.01.018.

[86] Costantino F., Di Gravio G., Patriarca R., Petrella L.: Spare parts management for irregular demand items. Omega (United Kingdom), 81, 57-66 (2018) https://doi:10.1016/j.omega.2017.09.009.

[87] Sherbrooke, C. C.: METRIC: A Multi-Echelon Technique for Recoverable Item Control. Operations Research 16, 122-141 (1968).

[88] Sherbrooke, C. C.: VARI-METRIC: Improved Approximations for Multi-Indenture, Multi-Echelon Availability Models. Operations Research 34, 311-319 (1986).

[89] Audet, C., Hare, W.: Derivative-free and blackbox optimization. Berlin: Springer International Publishing (2017) https://10.1007/978-3-319-68913-5.

[90] Audet, C. and Dennis, Jr. J. E.: Analysis of Generalized Pattern Searches. SIAM Journal on Optimization. Volume 13, Number 3, pp. 889–903 (2003).

[91] Marseguerra, M., Zio, E., Podofillini, L., 2005. Multiobjective spare part allocation by means of genetic algorithms and Monte Carlo simulation. Reliab. Eng. Syst. Saf. 87, 325–335. https://doi.org/10.1016/j.ress.2004.06.002.

[92] Sherbrooke, C.C.: Optimal Inventory Modeling of Systems: Multi-Echelon Techniques (2004).

[93] Gross, O.: A Class of Discrete-Type Minimization Problems. RAND Corporation, RM-1655-PR, Santa Monica, CA. (1956).

[94] Graves, S.C.: Multi-echelon inventory model for a repairable item with one-for-one replenishment. Management Science 31 (10), 1247 - 1256 (1985).

[95] Muckstadt, J.A.: A model for a multi-item, multi-echelon, multi-indenture inventory system. Management Science 20 (4), 472 - 481 (1973).

[96] Kline, R.C., Bachman, T.C.: Estimating Spare Parts Requirements with Commonality and Redundancy. J. Spacecr. Rockets 44, 977–984 (2007). https://doi.org/10.2514/1.28072.

[97] Nowicki, D., Kumar, U.D., Steudel, H.J., Verma, D.: Spares provisioning under performance-based logistics contract: profit-centric approach. Journal of the Operational Research Society 59 (3), 342–352 (2008).

[98] De Smidt Destombes, K., van der Heijden, M., van Harten, A.: Joint optimization of spare part inventory maintenance frequency. International Journal of Production Economics, 118(1), 260-268 (2009).

[99] Costantino, F., Di Gravio, G., Tronci, M.: Multi-echelon, multi-indenture spare parts inventory control subject to system availability and budget constraints. Reliability Engineering and System Safety 119, 95–101 (2013).

[100] Xu, L., Li, Q., Li, H.: Inventory control of multi-echelon maintenance supply system with multiple repair priorities. Hangkong Xuebao/Acta Aeronautica et Astronautica Sinica, 36(4), 1185-1194 (2015).

[101] Basten, R.J.I., Van Houtum, G.J.: System-oriented inventory models for spare parts. Surveys in Operations Research and Management Science, 19(1), 34-55 (2014) https://doi.org/10.1016/j.sorms.2014.05.002.

[102] Kapoor, R., Shah, B. J., Shah, N. H.: A simulation- and genetic algorithm-based optimisation of closed-loop multi-echelon inventory system. International Journal of Mathematics in Operational Research, 8(1), 28–59 (2016) https://doi.org/10.1504/IJMOR.2016.073278.

[103] Patriarca, R., Costantino, F., Di Gravio, G., Tronci, M., 2016 (b). Inventory optimization for a customer airline in a Performance Based Contract. J. Air Transp. Manag., 57, 206–216 (2016) https://doi.org/10.1016/j.jairtraman.2016.08.005 .

[104] Patriarca, R., Costantino, F., Di Gravio, G., 2016 (a). Inventory model for a multi-echelon system with unidirectional lateral transhipment. Expert Syst. Appl. 65, 372–382 (2016) https://doi.org/10.1016/j.eswa.2016.09.001.

[105] Duran, O., Perez, L.: Optimization of the multi echelon system for repairable spare parts using swarm intelligence combined with a local search strategy. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Volume 8583 LNCS, Issue PART 5, 2014, Pages 747-761 (2014).

[106] Costantino, F., Di Gravio, G., Patriarca, R., Tronci, M.: Spare parts inventory control model for the aeronautical industry. Proceedings of the Summer School Francesco Turco Volume 09-12-September-2014, 2014, Pages 120-125 (2014).

[107] Nickel, R., Mikolic Torreira, I., Tolle, J.: Computing Aviation Sparing Policies: Solving a Large Nonlinear Integer Program. Computational Optimization and Applications, 35, 109–126 (2006).

[108] Topan, E., Bayındır, Z.P., Tan T.: Heuristics for multi-item two-echelon spare parts inventory control subject to aggregate and individual service measures. European Journal of Operational Research, 256 (1), 126-138 (2017).

[109] Wong, H., Kranenburg, B., van Houtum, G., Cattrysse, D.: Efficient heuristics for two-echelon spare parts inventory systems with an aggregate mean waiting time constraint per local warehouse. OR Spectrum, 29 , 699–722 (2007).

[110] Nowicki DR, Randall WS, Ramirez-Marquez JE.: Improving the computational efficiency of metric-based spares algorithms. Eur J Oper Res. 219(2):324-334 (2012) https://doi:10.1016/j.ejor.2011.12.033.

[111] Vavasis S.A.: Complexity Issues in Global Optimization: A Survey. In: Horst R., Pardalos P.M. (eds) Handbook of Global Optimization. Nonconvex Optimization and Its Applications, vol 2. Springer, Boston, MA (1995) https://doi.org/10.1007/978-1-4615-2025-2_2.

[112] Liuzzi, G., Lucidi, S., Rinaldi, F.: An algorithmic framework based on primitive directions and nonmonotone line searches for black-box optimization problems with integer variables. Math. Prog. Comp. 12, 673–702 (2020) https://doi.org/10.1007/s12532-020-00182-7.

[113] Romito, F.: Hybridization and Discretization Techniques to Speed Up Genetic Algorithm and Solve GENOPT Problems. In: Battiti R., Kvasov D., Sergeyev Y. (eds) Learning and Intelligent Optimization. LION, 2017. Lecture Notes in Computer Science, vol 10556. Springer, Cham (2017) https://doi.org/10.1007/978-3-319-69404-7_20.

[114] Regis, R. G.: On the properties of positive spanning sets and positive bases. Optim. Eng. 17(1) 229-262 (2016) https://doi.org/10.1007/s11081-015-9286-x.

[115] Coope, I. D., Price, C. J.: Positive bases in numerical optimization. Computational Optimization and Applications 21(2), 169-175 (2002).

[116] Audet, C., Dennis Jr., J.E.: Pattern search algorithms for mixed variable programming. SIAM J. Optim. 11(3), 573–594 (2001).