ISSN 2281-4299

DIPARTIMENTO DI INGEGNERIA INFORMATICA
AUTOMATICA E GESTIONALE ANTONIO RUBERTI

SAPIENZA
UNIVERSITÀ DI ROMA

Exploiting non-linear transformations of the variables space in derivative-free global optimization

Giampaolo Liuzzi
Stefano Lucidi
Francesco Romito

Technical Report n. 1, 2021

Exploiting non-linear transformations of the variables space in derivative-free global optimization

Giampaolo Liuzzi · Stefano Lucidi · Veronica Piccialli · Francesco Romito

Abstract Solving a global optimization problem is a hard task. In this work we defined non-linear equations that are a transformations between variable spaces, allowing us to tackle the problem advantageously. The idea is to gather the information obtained during a multi-start approach and to apply a sequence of transformations in the variables space that makes the esploration easier. The aim is to expand the attraction basins of global minimizers shrinking those of local minima we have already found.

Keywords Global optimization · Derivative-free Optimization · Continuous optimization · Variable spaces transformation · Non-linear transformation

Giampaolo Liuzzi, Stefano Lucidi, Francesco Romito
Dipartimento di Ing. Informatica Automatica e Gestionale "Antonio Ruberti", "Sapienza" Università di Roma, via Ariosto 25, 00185 Roma, Italy
E-mail: liuzzi@diag.uniroma1.it, lucidi@diag.uniroma1.it, romito@diag.uniroma1.it

Veronica Piccialli
Dipartimento di Ingegneria Informatica e Ingegneria Civile, "Tor Vergata" Università di Roma, via del Politecnico 1, 00133 Roma, Italy
E-mail: veronica.piccialli@uniroma2.it

1 Introduction

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],
- to choose points at random according to a suitable distribution [5,6,7,8,9,10];

Other efficient approaches to global optimization problems are the multi-start strategies [7,11] and filled function techniques [12,13,14].
Multi-start method have two phases: the first one in which the solution is generated and the second one in which the solution is typically (but not necessarily) improved. It aims to iterate local minimization starting from points provided by probabilistic [5] or deterministic methods [11]. It is shown under suitable assumption that local minimizations can be attracted by the global minimum of both gradient related algorithm [15] and derivative-free [16,28,17].
Instead, in the filled function approaches the objective function is iteratively perturbed in order to avoid that local optimization methods getting stuck in local minima.
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.

In this work we propose a new strategy that can be useful to tackling such hard global optimization problems. The rationale behind our 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.
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 we propose two bijective non-linear transformations. In order to have a feel of the possible interest of the proposed transformations we combine it with a simple multi-start algorithms. The obtained numerical results show that the performance of the multi-start approach could improves by exploiting the transformations.
It is clear that the proposed transformations can be integrated with more complex algorithm scheme. The study of efficient global optimization algorithms with a space expansion-contraction strategy will be the object of future works.
In Sect. 2, we give some preliminary definitions that will be used throughout the paper. In Sect. 3, we describe a simple piecewise linear space transformation and outline its properties. In Sect. 4, we propose a more complex non-linear space transformation, that overcome some drawback of the piecewise linear transformation. In Sect. 6, we report a first numerical experience combining our transformations with the derivative-free local minimization technique in the multi-start framework. In Sect. 5 some preliminary considerations are made about the possibility to use these transformations as derivative-free preconditioner. Finally, we draw some conclusions 6.

2 Problem definition and Notation

Consider the following problem

$$\min_{x\in\mathcal{X}} f(x), \tag{1}$$

where

$\mathcal{X} = \{x \in \mathbb{R}^n \ : \ l \leq \ x \ \leq u\}$ with $l, u \in \mathbb{R}^n$, $l < u$ and
$f : \mathbb{R}^n \to \mathbb{R}$ is a real valued function.

In the following we assume that $\mathcal{X}$ is compact and f is continuous on $\mathcal{X}$.
Our aim is to find the global solution x^* of problem (1), namely a point such that

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

Given a vector $x \in \mathbb{R}^n$, we denote by $\|x\|_p$, with $p \in [1, +\infty)$ the ℓ_p-norm of x, i.e.

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

and by $\|x\|_\infty$ the ℓ_∞-norm of x, i.e.

$$\|x\|_\infty = \max_{i=1,\dots,n} \{|x_i|\}.$$

Definition 1 (hypercubes and hyperspheres) Given a vector $\hat{x} \in \mathbb{R}^n$ and a positive scalar r we define

$$\begin{aligned}
\mathcal{B}_\infty(\hat{x}, r) &= \left\{x \in \mathbb{R}^n : \|x - \hat{x}\|_\infty \leq r\right\} \\
\partial\mathcal{B}_\infty(\hat{x}, r) &= \left\{x \in \mathbb{R}^n : \|x - \hat{x}\|_\infty = r\right\} \\
\mathcal{B}_2(\hat{x}, r) &= \left\{x \in \mathbb{R}^n : \|x - \hat{x}\|_2 \leq r\right\} \\
\partial\mathcal{B}_2(\hat{x}, r) &= \left\{x \in \mathbb{R}^n : \|x - \hat{x}\|_2 = r\right\}
\end{aligned}$$

Definition 2 (hyperrectagles and hyperellipses) Given a vector $\hat{x} \in \mathbb{R}^n$, a positive scalar r and a square diagonal positive definite matrix

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

we define

$$\begin{aligned}
\mathcal{E}_\infty(\hat{x}, D, r) &= \left\{x \in \mathbb{R}^n : \ \|D(x - \hat{x})\|_\infty \leq r\right\} \\
\partial\mathcal{E}_\infty(\hat{x}, D, r) &= \left\{x \in \mathbb{R}^n : \ \|D(x - \hat{x})\|_\infty = r\right\} \\
\mathcal{E}_2(\hat{x}, D, r) &= \left\{x \in \mathbb{R}^n : \ \|D(x - \hat{x})\|_2 \leq r\right\} \\
\partial\mathcal{E}_2(\hat{x}, D, r) &= \left\{x \in \mathbb{R}^n : \ \|D(x - \hat{x})\|_2 = r\right\}
\end{aligned}$$

Definition 3 Given a point $\hat{x} \in \mathcal{X}$, we denote by

$$y = T_{\hat{x}}(x) = \bar{y}(x),$$

a mapping $\mathcal{X} \mapsto \mathcal{X}$. The point $\hat{x}$ is the so called "centroid" of the transformation. Conversely,

$$x = T_{\hat{x}}^{-1}(y) = \bar{x}(y)$$

denotes the inverse mapping. Hence, by definition, for every $x \in \mathcal{X}$ it holds that

$$x = T_{\hat{x}}^{-1} \circ T_{\hat{x}}(x)$$

Given the inverse mapping $T_{\hat{x}}^{-1}(y)$, problem (1) becomes the *transformed* optimization problem

$$\min_{y \in \mathcal{Y}} f(T_{\hat{x}}^{-1}(y)) \tag{2}$$

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 (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.

3 Piecewise linear mappings

In this section we introduce a special kind of mapping $T_{\hat{x}}$, namely the piecewise linear mapping. This mapping shall be piecewise linear with respect to any single variable. Figure (1) illustrates the mapping in $\mathbb{R}^2$.
For the sake of simplicity, let us consider $\hat{y} = \mathbf{0}_n \in \mathcal{X}$ as the reference point. This point is set as the centroid of the hypercube $\mathcal{B}_\infty(\hat{y}, r_y)$ in which the mapping impact range is confined. The point $\hat{y}$ is also set as centroid of the smaller hypercube $\mathcal{B}_\infty(\hat{y}, \epsilon_y)$ in which the mapping impact has the main focus. The centroid $\hat{y}$ is a fixed point of the transformation, i.e. $\hat{y} = \hat{x}$, where $\hat{x}$ is the centroid of the corresponding hypercube $\mathcal{B}_\infty(\hat{x}, r_x)$.
For each dimension $i = 1, 2, ..., n$ the piecewise linear transformation (blue line in fig. (1)) propagates outward from the centroid in the relative space.
The larger the angle between the bisector (green line in fig. (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

- $l_i \leq \hat{x}_i - r_x$;
- $\hat{x}_i + r_x \leq u_i$;
- $l_i \leq \hat{y}_i - r_y$;
- $\hat{y}_i + r_y \leq u_i$.

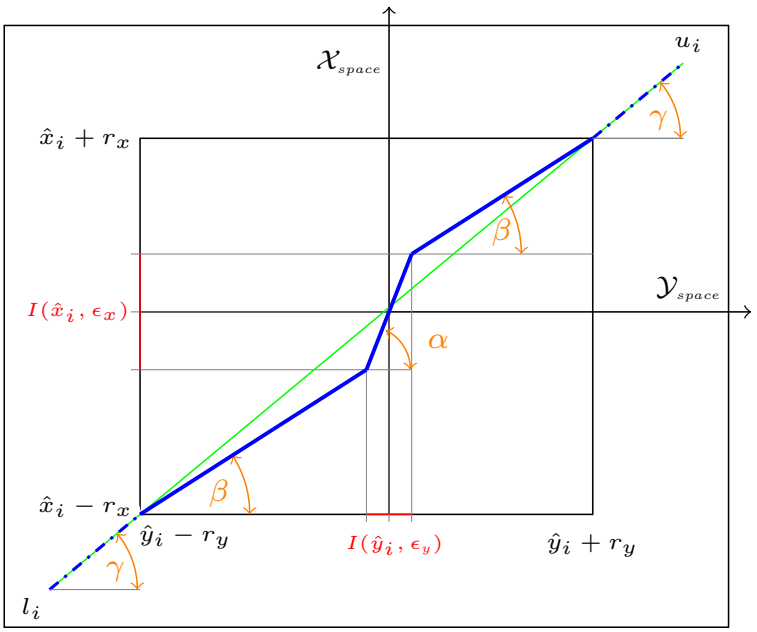

Fig. 1 Graph of the piecewise linear transformation in the i^{th} dimension.

Definition 4 (Piecewise linear mapping) Let $\hat{x}, \hat{y} \in \mathbb{R}^n$, and $\epsilon_x,\ r_x,\ \theta_1,\ \theta_2,\ \theta_3$ be positive scalar quantities, with $\epsilon_x \leq r_x$. The piecewise linear mapping $y = T_{\hat{x}}(x)$ is defined as foolows. For each $i = 1, \ldots, n$:

$$\bar{y}(x)_i = \hat{y}_i + \theta_1 (x_i - \hat{x}_i), \quad |x_i - \hat{x}_i| \leq \epsilon_x \tag{3}$$

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} (\theta_1 \epsilon_x + \theta_2(|x_i - \hat{x}_i| - \epsilon_x)), \quad \epsilon_x \leq |x_i - \hat{x}_i| \leq r_x \tag{4}$$

$$\bar{y}(x)_i = \hat{y}_i + \frac{(x_i - \hat{x}_i)}{|x_i - \hat{x}_i|} (\theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x) + \theta_3 (|x_i - \hat{x}_i| - r_x)), \quad |x_i - \hat{x}_i| \geq r_x \tag{5}$$

The following proposition reports properties of the transformation and necessary coupling condition to preserve the continuity of the transformation.

Proposition 1 *Given $x \in \mathbb{R}^n$, $y = T_{\hat{x}}(x)$ is such that:*

i) The reference point $\hat{x} \in \mathcal{X}$ is mapped to $\hat{y}$, i.e. $\hat{y} = T_{\hat{x}}(\hat{x})$.
ii) For every index i, the function $(T_{\hat{x}}(x))_i$ is continuous.
iii) for every index i such that $|x_i - \hat{x}_i| \leq \epsilon_x$, y_i is such that

$$|y_i - \hat{y}_i| \leq \theta_1 \epsilon_x.$$

iv) For every index i such that $\epsilon_x \leq |x_i - \hat{x}_i| \leq r_x$, y_i is such that

$$\theta_1 \epsilon_x \leq |y_i - \hat{y}_i| \leq \theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x).$$

v) For every index i such that $|x_i - \hat{x}_i| \geq r_x$, y_i is such that

$$|y_i - \hat{y}_i| \geq \theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x).$$

Proof Point i) is trivially true. Point ii). For every index $i = 1, \ldots, n$, the function $(T_{\hat{x}}(x))_i$ is obviously continuous when x_i is such that $|x_i - \hat{x}_i| \neq \epsilon_x$ and $|x_i - \hat{x}_i| \neq r_x$. Then, we have to only worry about x_i such that either $|x_i - \hat{x}_i| = \epsilon_x$ or $|x_i - \hat{x}_i| = r_x$. Let us first consider the case $|x_i - \hat{x}_i| = \epsilon_x$. According to the definition of $T_{\hat{x}}$, continuity is guaranteed if

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

and since $|x_i - \hat{x}_i| = \epsilon_x$, we have

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

which proves continuity. Now, let us consider x_i such that $|x_i - \hat{x}_i| = r_x$. In this case, continuity is guaranteed provided that

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

When $|x_i - \hat{x}_i| = r_x$, the above can be written as

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

which is always true.
Point iii). Let i be an index such that

$$|x_i - \hat{x}_i| \leq \epsilon_x.$$

Then, by definition of $T_{\hat{x}}$, we have

$$y_i = (T_{\hat{x}}(x))_i = \hat{y}_i + \theta_1(x_i - \hat{x}_i).$$

Hence,

$$|y_i - \hat{y}_i| = \theta_1 |x_i - \hat{x}_i| \leq \theta_1 \epsilon_x,$$

which proves the result.
Point iv). Let i be such that $\epsilon_x \leq |x_i - \hat{x}_i| \leq r_x$. By definition of $T_{\hat{x}}$, we have

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

which, by considering that $|x_i - \hat{x}_i| - \epsilon_x \geq 0$, gives

$$|y_i - \hat{y}_i| = \theta_1 \epsilon_x + \theta_2(|x_i - \hat{x}_i| - \epsilon_x).$$

Then, recalling that $|x_i - \hat{x}_i| \leq r_x$, we can write

$$|y_i - \hat{y}_i| \leq \theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x). \tag{6}$$

On the other hand, recalling that $|x_i - \hat{x}_i| \geq \epsilon_x$

$$|y_i - \hat{y}_i| \geq \theta_1 \epsilon_x. \tag{7}$$

Hence, by (6) and (7), we have that

$$\theta_1 \epsilon_x \leq |y_i - \hat{y}_i| \leq \theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x),$$

from which the result of point iii) follows.
Point v). Let now i be an index such that $|x_i - \hat{x}_i| \geq r_x$. By definition of $T_{\hat{x}}$, we can write

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

Then, recalling that $|x_i - \hat{x}_i| \geq r_x$,

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

and the result follows, concluding the proof. □

Proposition 2 (inverse of the PLT) *Let $\hat{x}, \hat{y} \in \mathbb{R}^n$, and $\epsilon_y = \theta_1 \epsilon_x$, $r_y = \theta_1 \epsilon_x + \theta_2(r_x - \epsilon_x)$, $\theta_1, \theta_2, \theta_3$ be positive scalar quantities. The following equations define the inverse piecewise linear mapping. For each $i = 1, \ldots, n$,*

$$\bar{x}(y)_i = \hat{x}_i + \frac{1}{\theta_1}(y_i - \hat{y}_i), \quad |y_i - \hat{y}_i| \leq \epsilon_y \tag{8}$$

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(|y_i - \hat{y}_i| - \epsilon_y)\right), \quad \epsilon_y \leq |y_i - \hat{y}_i| \leq r_y \tag{9}$$

$$\bar{x}(y)_i = \hat{x}_i + \frac{(y_i - \hat{y}_i)}{|y_i - \hat{y}_i|}\left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(r_y - \epsilon_y) + \frac{1}{\theta_3}(|y_i - \hat{y}_i| - r_y)\right), \quad |y_i - \hat{y}_i| \geq r_y \tag{10}$$

Proof When $|y_i - \hat{y}_i| \leq \epsilon_y$, we have to show that $\bar{x}(y)_i$ is such that $|\bar{x}(y)_i - \hat{x}_i| \leq \epsilon_x$. In fact, in this case $T_{\hat{x}}(\bar{x}(y))_i = y_i$. From (8), we get that

$$|\bar{x}\,(y)_i - \hat{x}_i| = \frac{1}{\theta_1}|y_i - \hat{y}_i|$$

and, since $|y_i - \hat{y}_i| \leq \epsilon_y = \theta_1 \epsilon_x$, we obtain

$$|\bar{x}\,(y)_i - \hat{x}_i| \leq \epsilon_x.$$

Now, let us suppose that $\epsilon_y \leq |y_i - \hat{y}_i| \leq r_y$. In this case, we have to show that $\epsilon_x \leq |\bar{x}(y)_i - \hat{x}_i| \leq r_x$. From (9) we get

$$|\bar{x}\,(y)_i - \hat{x}_i| = \left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(|y_i - \hat{y}_i| - \epsilon_y)\right). \tag{11}$$

Since $\epsilon_y \leq |y_i - \hat{y}_i|$, (11) yields

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

On the other hand and since $|y_i - \hat{y}_i| \leq r_y$, again from (11) we obtain

$$|\bar{x}\,(y)_i - \hat{x}_i| \leq \left(\epsilon_x + \frac{1}{\theta_2}(r_y - \epsilon_y)\right).$$

This, recalling the expressions of ϵ_y and r_y, gives

$$|\bar{x}\,(y)_i - \hat{x}_i| \leq \left(\epsilon_x + \frac{1}{\theta_2}(\theta_1\epsilon_x + \theta_2(r_x - \epsilon_x) - \theta_1\epsilon_x)\right).$$

from wihch we get $|\bar{x}\,(y)_i - \hat{x}_i| \leq r_x$.
Finally, let us suppose that $|y_i - \hat{y}_i| \geq r_y$. In this last case, we have to show that $|\bar{x}(y)_i - \hat{x}_i| \geq r_x$. From (10) we can write

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

and, considering that $|y_i - \hat{y}_i| \geq r_y$,

$$|\bar{x}\,(y)_i - \hat{x}_i| \geq \left(\frac{\epsilon_y}{\theta_1} + \frac{1}{\theta_2}(r_y - \epsilon_y)\right)$$

which, again recalling the expressions of ϵ_y and r_y, gives $|\bar{x}\,(y)_i - \hat{x}_i| \geq r_x$, thus concluding the proof. □

The piecewise linear transformation is well suited for all those global optimization approaches where the feasible region is partitioned in sets of hyper-rectangles.
For instance, deterministic algorithms such as DIRECT-type approach (DIviding RECTangle [2] and its variants [20,21]). In the modifications of the DIRECT, there are hybridization with local minimization algorithm [11] 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.

4 Piecewise nonlinear mapping

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 effects are confined to the set $\mathcal{B}_2(\hat{x}, r_x)$
In particular, we introduce a piecewise nonlinear mapping $y = T_{\hat{x}}(x)$. This latter mapping is not component-wise, even though it is still piecewise.
For the sake of simplicity, let us consider $\hat{y} = \mathbf{0}_n$ as the reference point. The mapping impact range is confined to the set $\mathcal{B}_2(\hat{y}, r_y)$. The mapping has the main focus within the set $\mathcal{B}_2(\hat{y}, \epsilon_y)$. We point out that the centroid $\hat{y}$ is a fixed point of the mapping, i.e. $\hat{y} = \hat{x}$.

Definition 5 (Piecewise nonlinear mapping) Let $\hat{x}, \hat{y} \in \mathbb{R}^n$, and $\epsilon_x,\ r_x,\ \theta_1,\ \theta_2,\ \theta_3$ be positive scalar quantities, with $\epsilon_x \leq r_x$. Then, for each $x \in \mathbb{R}^n$, let

$$\bar{y}(x) = \hat{y} + \theta_1 (x - \hat{x}), \quad \|x - \hat{x}\|_2 \leq \epsilon_x, \tag{12}$$

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\theta_1 \epsilon_x + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x\right)\right), \quad \epsilon_x \leq \|x - \hat{x}\|_2 \leq r_x, \tag{13}$$

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\theta_1 \epsilon_x + \theta_2 (r_x - \epsilon_x) + \theta_3 \left(\|x - \hat{x}\|_2 - r_x\right)\right), \quad \|x - \hat{x}\|_2 \geq r_x. \tag{14}$$

Below we report the properties of the piecewise nonlinear mapping.

Proposition 3 *i) The reference point $\hat{x} \in \mathcal{X}$ is transformed in $\hat{y}$, i.e. $\hat{y} = T_{\hat{x}}(\hat{x})$.*
ii) The region $\mathcal{B}_2(\hat{x}, \epsilon_x)$ is transformed in $\mathcal{B}_2(\hat{y}, \epsilon_y)$
iii) The region $\left\{x \in \mathbb{R}^n : \epsilon_x \leq \|x - \hat{x}\|_2 \leq r_x\right\}$ is transformed in

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

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

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

Proof Points i) and ii) are trivial.
Point iii). From (13) we have

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

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

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

that is

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

and by substituting

$$\epsilon_y = \theta_1 \epsilon_x. \tag{15}$$

it impose at the boundary the following coupling condition

$$r_y = \epsilon_y + \theta_2 (r_x - \epsilon_x). \tag{16}$$

Point iv). From (14) we have

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

substituting (15) and (16)

$$\bar{y}(x) = \hat{y} + \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(r_y + \theta_3 \left(\|x - \hat{x}\|_2 - r_x \right) \right). \tag{17}$$

As long as the (17) refers to the outermost region there is no need to further coupling condition. It means that θ_3 is free of choice. □

We are going to prove that the piecewise nonlinear mapping is invertible.

Proposition 4 *The piecewise nonlinear mapping $T_{\hat{x}}$ is invertible.*

Proof Let us consider (13):

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

substituting (15)

$$y - \hat{y} = \frac{(x - \hat{x})}{\|x - \hat{x}\|_2} \left(\epsilon_y + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right) \right), \tag{18}$$

taking the ℓ_2-norm and recalling 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

$$\|y - \hat{y}\|_2 = \epsilon_y + \theta_2 \left(\|x - \hat{x}\|_2 - \epsilon_x \right), \tag{19}$$

by inverting the last expression we have

$$\|x - \hat{x}\|_2 = \epsilon_x + \frac{1}{\theta_2} (\|y - \hat{y}\|_2 - \epsilon_y). \tag{20}$$

Substituting (19) in (18)

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

by inverting the last expression

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

and substituting (20) we have

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

that is

$$\bar{x}\left(y\right)=\hat{x}+\frac{\left(y-\hat{y}\right)}{\left\|y-\hat{y}\right\|_2}\left(\epsilon_x+\frac{1}{\theta_2}(\left\|y-\hat{y}\right\|_2-\epsilon_y)\right). \tag{21}$$

Now consider the equation (14):

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

substituting (15) and (16)

$$y-\hat{y}=\frac{\left(x-\hat{x}\right)}{\left\|x-\hat{x}\right\|_2}\left(r_y+\theta_3\left(\left\|x-\hat{x}\right\|_2-r_x\right)\right), \tag{22}$$

taking the ℓ_2-norm operator and recalling that $\left\|x-\hat{x}\right\|_2\geq r_x$

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

and so

$$\left\|y-\hat{y}\right\|_2=r_y+\theta_3\left(\left\|x-\hat{x}\right\|_2-r_x\right), \tag{23}$$

by inverting the last expression we have

$$\left\|x-\hat{x}\right\|_2=r_x+\frac{1}{\theta_3}(\left\|y-\hat{y}\right\|_2-r_y). \tag{24}$$

Substituting (23) in (22)

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

by inverting the last expression

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

and substituting (24) we have

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

that is

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \left(r_x + \frac{1}{\theta_3} (\|y - \hat{y}\|_2 - r_y) \right). \tag{25}$$

This concludes the proof. □

We can now state the formal definition of the inverse piecewise nonlinear mapping.

Definition 6 Let $\hat{y} \in \mathbb{R}^n$, ϵ_y, r_y, θ_1, θ_2, θ_3 be positive scalar quantities, with $\epsilon_y \leq r_y$.

$$\bar{x}(y) = \hat{x} + \frac{1}{\theta_1}(y - \hat{y}), \quad \|y - \hat{y}\|_2 \leq \epsilon_y, \tag{26}$$

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \left(\frac{1}{\theta_1} \epsilon_y + \frac{1}{\theta_2} (\|y - \hat{y}\|_2 - \epsilon_y) \right), \quad \epsilon_y \leq \|y - \hat{y}\|_2 \leq r_y, \tag{27}$$

$$\bar{x}(y) = \hat{x} + \frac{(y - \hat{y})}{\|y - \hat{y}\|_2} \left(\frac{1}{\theta_1} \epsilon_y + \frac{1}{\theta_2} (r_y - \epsilon_y) + \frac{1}{\theta_3} (\|y - \hat{y}\|_2 - r_y) \right), \quad \|y - \hat{y}\|_2 \geq r_y. \tag{28}$$

5 Scaling and Preconditioning

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 (2) 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 have been set as follows:

$$\begin{array}{llll} \epsilon_x = 2.0 & r_x = 2.5 & \theta_1 = 1 & \theta_3 = 1.2 \\ \epsilon_y = 2.0 & r_y = 3.0 & \theta_2 = 0.5 & \hat{x} = \hat{y} = (0, 0)^\top \end{array}$$

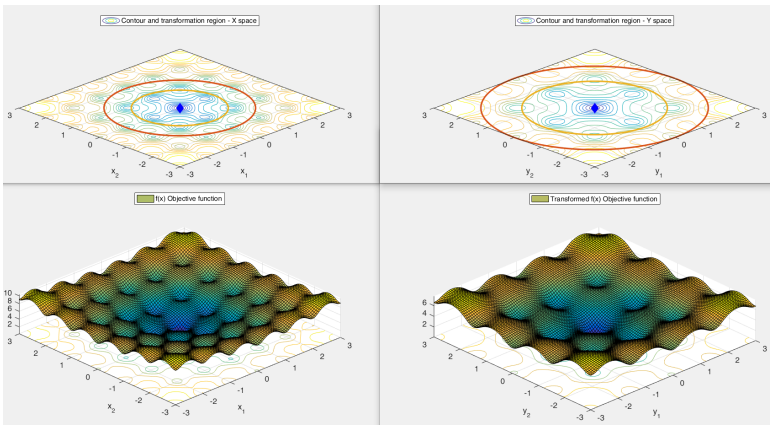

Fig. 2 A spherical neighborhood of $\hat{x}$ has been scaled to an expanded one in the transformed space.

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 (3) 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 have been set as follows:

$$\begin{array}{llll} \epsilon_x = 2.0 & r_x = 3.0 & \theta_1 = 1 & \theta_3 = 0.83 \\ \epsilon_y = 2.0 & r_y = 2.5 & \theta_2 = 2 & \hat{x} = \hat{y} = (0,0)^\top \end{array}$$

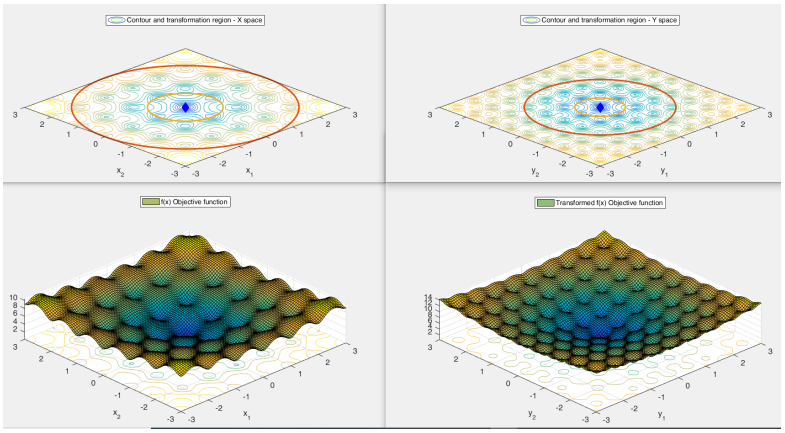

Fig. 3 An entire spherical neighborhood of $\hat{x}$ has been scaled to an expanded one in the transformed space. The innermost region of radius $\epsilon_x = \epsilon_x$ remain unchanged.

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 $\mathcal{B}(\hat{y}, \epsilon_y)$ where the transformation equation involved is

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

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 $\mathcal{B}(\hat{y}, \epsilon_y)$, corresponds an identical hypersphere in the original space $\mathcal{B}(\hat{x}, \epsilon_x)$
The transformation parameters have been set as follows:

$$\begin{array}{llll} \epsilon_x = 2.0 & r_x = 2.5 & \theta_1 = 1 & \theta_3 = 1.2 \\ \epsilon_y = 2.0 & r_y = 3.0 & \theta_2 = 0.5 & \hat{x} = \hat{y} = (0,0)^\top \end{array}$$

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 (5) 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 have been set as follows:

$$\begin{array}{llll} \epsilon_x = 2.0 & r_x = 3.0 & \theta_1 = 1 & \theta_3 = 0.83 \\ \epsilon_y = 2.0 & r_y = 2.5 & \theta_2 = 2 & \hat{x} = \hat{y} = (0,0)^\top \end{array}$$

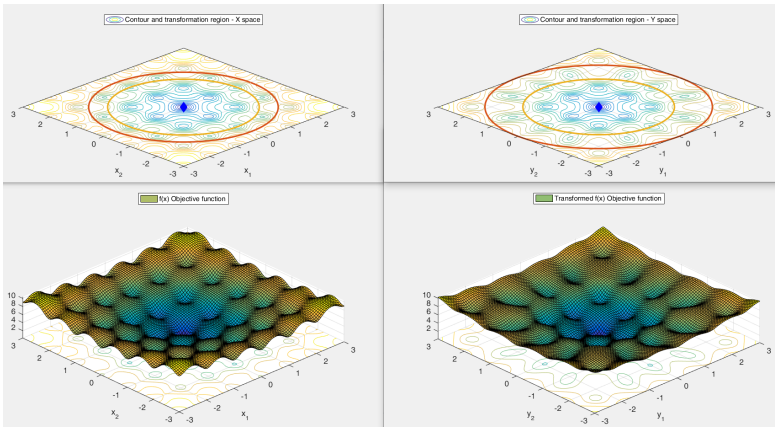

Fig. 4 A spherical neighborhood of $\hat{x}$ has been scaled to an expanded one in the transformed space. The innermost region of radius $\epsilon_x = \epsilon_y$ remains unchanged.

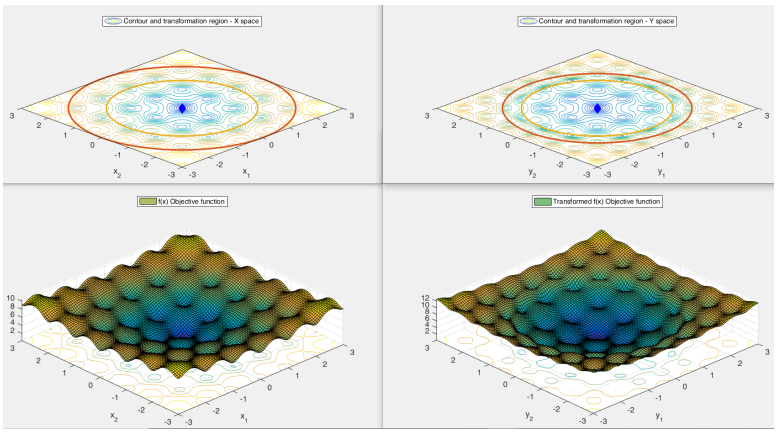

Fig. 5 A spherical neighborhood of $\hat{x}$ has been scaled to a contracted one in the transformed space. The innermost region of radius $\epsilon_x = \epsilon_y$ remains unchanged.

5.1 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 ($f(x) = \frac{1}{2}x^T Hx$ 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 [22,23,24] or an approximation of them, such as finite differences [25,26].

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

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

where $\lambda\in\{1,\dots,10^{10}\}$. This function has a global minimum at the point $x^*=(1,1,\dots,1)$ and $f(x^*)=0$.
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 [35,37]. Moreover for large enough λ and n, has one local minimum close to $x=[-1,1,...,1]$, see also [32]. In figure (6) 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.

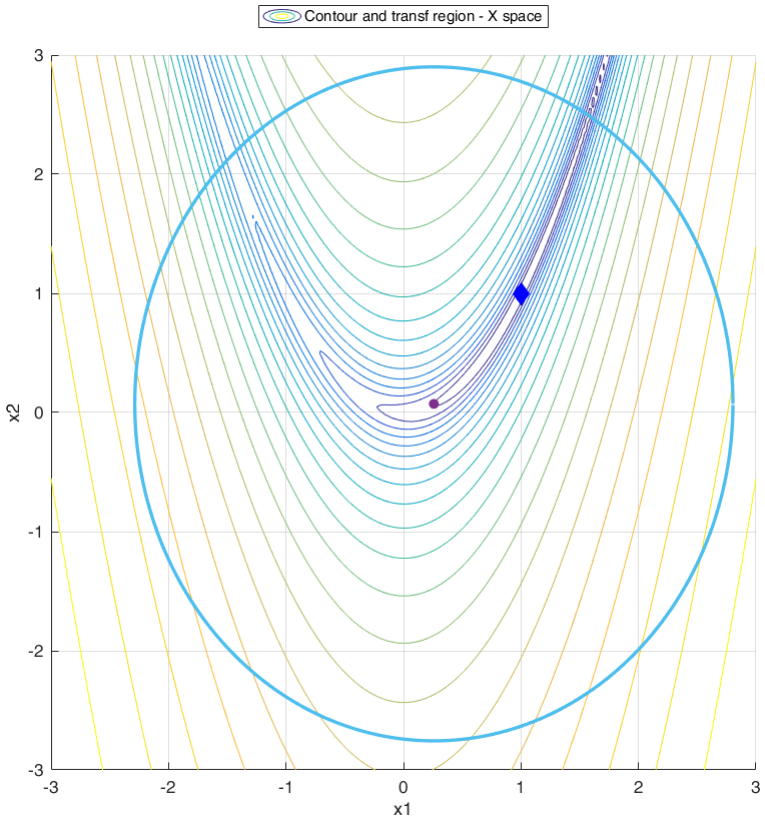

Fig. 6 Contour plot of Rosenbrock function in 2D

.

It considers a derivative-free algorithm DFL such as [33] (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 in such a way as to prevent the ill-conditioning. It can be done by a suitable choice of the diagonal entries of the matrices D_x and D_y. In particular a possible choice is

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

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

In figure (7), 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 $\max\limits_i(\alpha_i) < 10^{-6}$ is satisfied, without the use of any transformation.

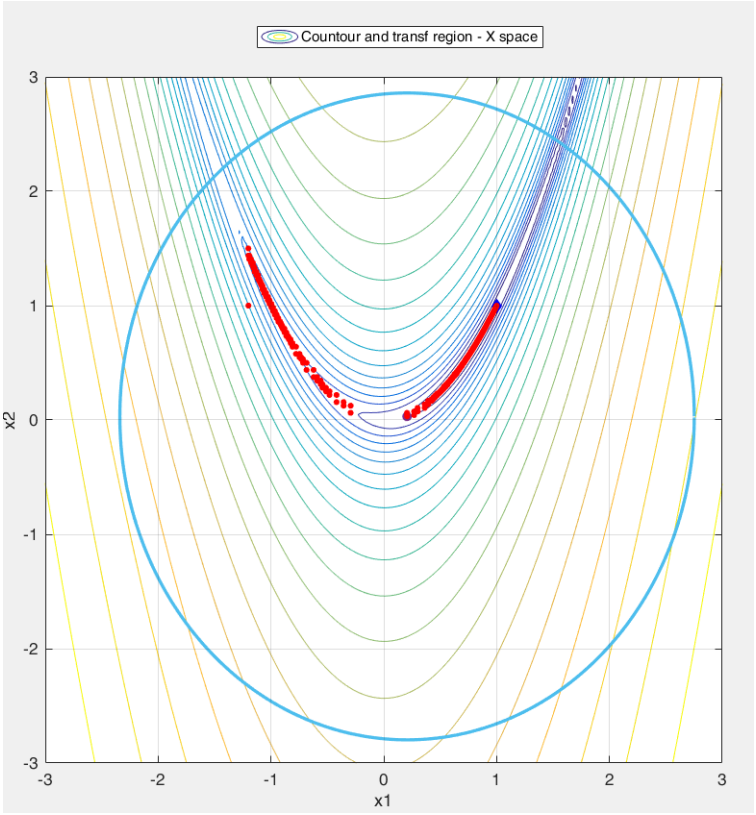

Fig. 7 Trace of the improving points explored by the algorithm DFL

.

In figure (8), it is drawn the trace of the 1246 improving points explored by DFL algorithm with the use of the non-linear transformation.

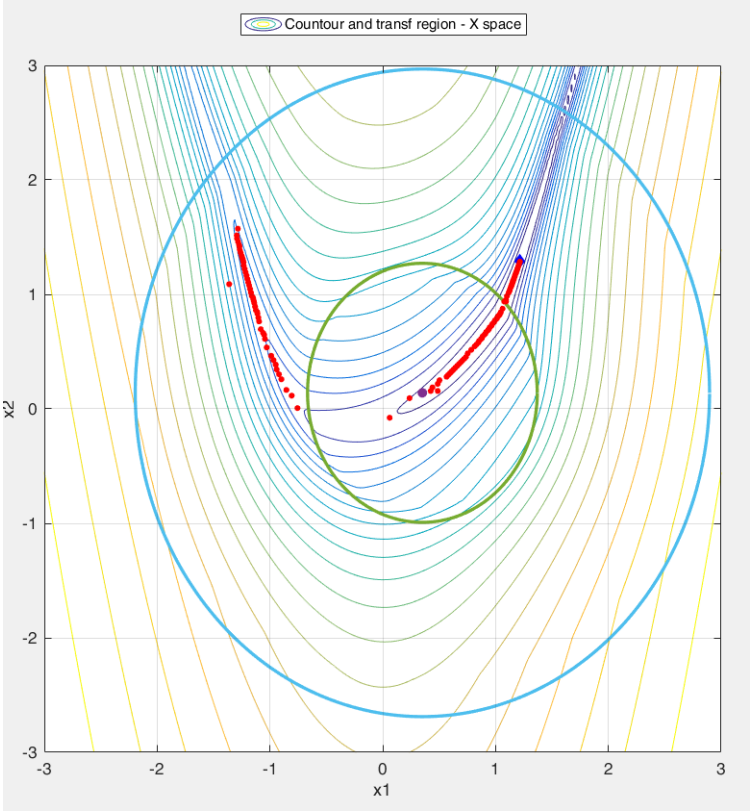

Fig. 8 Trace of the improving points explored by the DFL with transformation.

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, 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 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 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.

6 Preliminary numerical experience

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:

$$\frac{f(\bar{x}) - f^*}{max\,(1, |f^*|)} \leq 10^{-5} \tag{29}$$

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 [36] 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 (9) 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 [27]).

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:

	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: $[-100,100]^D$			

Fig. 9 CEC13 problems. https://www.ntu.edu.sg/CEC2013/CEC2013.htm

- 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.

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 3 Hyperrectangle expansion in 2D. Number 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 4 Hyperrectangle expansion in 5D. Number 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 5 Hyperrectangle expansion in 10D. Number 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 6 Hyperrectangle shrinking in 2D. Number 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 7 Hyperrectangle shrinking in 5D. Number 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 8 Hyperrectangle shrinking in 10D. Number 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 9 Ellipsoidal expansion in 2D. Number 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 10 Ellipsoidal expansion in 5D. Number 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 11 Ellipsoidal expansion in 10D. Number 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 12 Ellipsoidal shrinking in 2D. Number 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 13 Ellipsoidal shrinking in 5D. Number 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 14 Ellipsoidal shrinking in 10D. Number 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

7 Conclusions

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.

References

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, February 1981. ISSN 0364-765X. doi: 10.1287/moor.6.1.19.
7. Towards global optimization 2: L.C.W. DIXON and G.P. SZEGÖ (eds.) North-Holland, Amsterdam, (1978).
8. Greenhalgh D., Marshall, S.: Convergence criteria for genetic algorithms, SIAM J. Comput. 30(1), 269–282. (2000). https://doi.org/10.1137/S009753979732565X
9. Hartl, R. F., Belew, R. K.: A global convergence proof for a class of genetic algorithms. Tech. rep. University of Technology, Vienna. (1990).
10. 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).
11. 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) doi:10.1007/s10589-015-9741-9.
12. 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
13. 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

14. 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
15. Bertsekas, D.P.: Constrained Optimization and Lagrange Multipliers Methods. Academic Press, New York (1982) https://doi.org/10.1002/net.3230150112
16. 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
17. Liuzzi, G., Lucidi, S., Piccialli, V., Sotgiu, A.: A magnetic resonance device designed via global optimization techniques. Math. Program. 101(2), 339–364 (2004)
18. Bertolazzi, P., Guerra, C., Liuzzi, G.: A global optimization algorithm for protein surface alignment. BMC Bioinform. 11, 488–498 (2010)
19. Locatelli, M., Schoen, F.: Efficient algorithms for large scale global optimization: Lennard-Jones clusters. Comput. Optim. Appl. 26(2), 173–190 (2003)
20. 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.
21. 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).
22. 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)
23. 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)
24. 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)
25. Morales JL, Nocedal J. Automatic preconditioning by limited memory Quasi-Newton updating. SIAM J Optim. 10(4): 1079–1096 (2000)
26. 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)
27. 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
28. 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)
29. Paulavičius R., Žilinskas J.: Simplicial Global Optimization. SpringerBriefs in Optimization. Springer-Verlag, New York (2014)
30. 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)
31. 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
32. 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
33. 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
34. Goldberg, D.E., Kalyanmoy, D.: A comparative analysis of selection schemes used in genetic algorithms. Found. Genetic Algorithms 1, 69–93 (1991)
35. Dixon, L. C. W., Szego, G. P.: The global optimization problem: an introduction. Towards global optimization, 2, 1-15 (1978)
36. 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).
37. Picheny, V., Wagner, T., Ginsbourger, D.: A benchmark of kriging-based infill criteria for noisy optimization (2012)