



# SAFIP: A STREAMING ALGORITHM FOR INVERSE PROBLEMS

Maeva Biret, Michel Broniatowski

## ► To cite this version:

Maeva Biret, Michel Broniatowski. SAFIP: A STREAMING ALGORITHM FOR INVERSE PROBLEMS. 2016. hal-01371926

**HAL Id: hal-01371926**

**<https://hal.science/hal-01371926>**

Preprint submitted on 26 Sep 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# SAFIP: A STREAMING ALGORITHM FOR INVERSE PROBLEMS

MAEVA BIRET<sup>(1)</sup> AND MICHEL BRONIATOWSKI<sup>(2,\*)</sup>

ABSTRACT. This paper presents a new algorithm which aims at the resolution of inverse problems of the form  $f(x) = 0$ , for  $x \in \mathbb{R}^d$  and  $f$  an arbitrary function with mild regularity condition. The set of solutions  $S$  may be infinite. This algorithm produces a good coverage of  $S$ , with a limited number of evaluations of the function  $f$ . It is therefore appropriate for complex problems where those evaluations are costly. Various examples are presented, with  $d$  varying from 2 to 10. Proofs of convergence and of coverage of  $S$  are presented.

## 1. INTRODUCTION

**1.1. The scope of this paper.** Assume that we are given a bounded and closed domain  $D \subset \mathbb{R}^d$ , and a continuous real-valued function  $f$  defined on  $D$ .

The aim of this paper is to present an algorithm for the solution of the problem

$$(1.1) \quad S = \{x \in D : f(x) = 0\},$$

assuming  $S \neq \emptyset$ .

Such problems have been extensively handled over the years; see [2]. The difficulty which we are confronted to lies in three main points :

- (1) the set  $S$  may contain many points, even be infinite,
- (2) the function  $f$  might be quite costly for example when defined by a simulation device,
- (3) the function  $f$  may be quite irregular; we will assume mild regularity in the neighborhood of any point in  $S$ , only.

We also provide a two-fold proof for the convergence of this algorithm, namely we first prove that any resulting sequence of points in  $D$  converges to some point in  $S$ , and secondly that any point  $x$  in  $S$  is reached asymptotically by some "good" sequence, which is a sequence starting in a suitable neighborhood of  $x$ . As usually done in random search techniques, the starting points will be defined through random sampling in  $D$ .

**1.2. Bibliographic outlook.** Most approaches to Problem (1.1) extensively use analytic properties of the function  $f$ ; dichotomy, false position, Newton, conjugate gradient, etc (see [1]) handle so called well-posed problems, when the equation  $f(x) = 0$ , for  $x \in \mathbb{R}$  and  $f$  a real-valued function, has a unique solution. The case where  $f$  is defined as a mapping from  $\mathbb{R}^d$  to  $\mathbb{R}^k$  with  $d \leq k$  is treated by singular value decomposition (see [3]), which also solves well-posed problems.

The ill-posed problems which we consider, namely the case where Problem (1.1) has multiple solutions, is usually handled through regularization techniques, which aim

---

*Date:* May 26th, 2016.

*Key words and phrases.* level set, chains, inverse problem, convergence.

at transposing (1.1) into a well-posed problem. This procedure produces a partial solution to (1.1) under appropriate knowledge on the function  $f$  (see [4]). All these techniques are out of the concern of the present work, where all solutions of  $f$  are looked for, with minimal assumption on  $f$ . We briefly present four methods, which constitute the environment of our proposal.

*Local multi-start optimization, a deterministic approach.* Looking for the value of  $x$  such that  $f(x) = 0$ , consider the function  $|f|$ ; minimizing  $|f|$  indeed produces the set  $S$ .

First we choose a local optimization technique (Newton-Raphson for example). Then consider a design, which is a grid of initial points for the local optimization. From any of those, the sequence of iterations of the local optimization algorithm may produce a limit solution in  $S$ . Obviously stationary points not in  $S$  may be produced. The initial design is of utmost importance and the method may be unstable in this respect. Furthermore the method may be very costly due to the numerous evaluations of  $f$ . A general reference for those methods is [5].

*A grid search, deterministic approach.* This method produces a sequence of grids in  $D$ . Given an initial regular grid, the function  $f$  is evaluated on each of its points. Points where  $f$  is close to 0 are selected and the grid is updated and refined in the neighborhood of those points. This method has been proposed by [6]. A serious drawback lies in its cost, when the dimension of  $D$  corresponds to real life cases. Furthermore, the stopping rule of such algorithms does not guarantee a uniform approximation of  $S$ .

*A Monte Carlo Markov Chain technique.* We assume that the function  $f$  is written as  $f(x) = g(x) + \epsilon$ .  $f$  is then a model for the real function  $g$  with an error  $\epsilon$  due to modelling. For example,  $g$  is a physical model and  $f$  a computer-based formula for  $g$ . We estimate  $S = \{x : g(x) = 0\}$ . We choose a prior distribution  $\Pi_0(x)$  on  $\mathcal{X}$  and a parametric form for the distribution of  $\epsilon$ ,  $p(\epsilon|x)$ , for fixed  $x$ . By Bayes formula, the a posteriori distribution of  $x$  given  $\epsilon$  is given by

$$(1.2) \quad \frac{p(\epsilon|x)\Pi_0(x)}{\int p(\epsilon|x)\Pi_0(x)dx}.$$

The maximum probability principle provides stochastic solutions of  $g(x) = 0$  as the maximum of (1.2) upon  $x$ , given the prior  $\Pi_0$ .

In turn it can be proved that, whenever  $\Pi_0(x) = \mathcal{N}(x_0, \sigma_0^2)$  the Gaussian distribution with mean  $x_0$  and variance  $\sigma_0^2$ , for some  $x_0 \in D$  and  $\sigma_0^2 > 0$ , solutions  $x^*$  of (1.2) can be written as

$$(1.3) \quad x^* := \operatorname{argmin}_{x \in D} \|y - g(x)\| + \frac{\sigma_\epsilon^2}{\sigma_0^2} \|x - x_0\|^2,$$

when  $\epsilon$  is assumed to follow  $\mathcal{N}(0, \sigma_\epsilon^2)$ .

In order to find the  $x^*$  solution of (1.3), MCMC routines are used. This method is described in [7].

*The MRM (Monotonous Reliability Method).* Assume that  $f : \mathbb{R}^d \rightarrow \mathbb{R}$  is a globally monotone, i. e. is monotone in each of its variables. Assume also that the set  $S$  of solutions of the equation  $f(x) = 0$  is a continuous and simply (or one) connected set.

Assume for example that  $f$  is increasing on each of its variables. At each step, choose one point  $x$  in the unexplored subset of  $D$ . When  $f(x) > 0$  then all points  $y > x$  (meaning  $y_i > x_i$  for all  $1 \leq i \leq d$ ) are discarded from the unexplored region.

In the same way, when  $f(x) < 0$ , discard all the regions  $\{y : y < x\}$ . Iteration of these steps produces an unexplored domain which shrinks to  $S$ . Various ways of choosing  $x$  in the unexplored domain define specific algorithms. See [8].

## 2. OUTLOOK OF THE SAFIP ALGORITHM

**2.1. Basic features and properties.** We start with the iteration of the equivalence

$$(2.1) \quad (f(x) = 0) \iff \left( f(x) + \frac{x}{2k} + \frac{x}{2k} = \frac{x}{k} \right),$$

which holds where  $d = 1$ , for any  $k \neq 0$ ; for sake of convenience state  $k > 0$ . We proceed defining a recurrence in the RHS in (2.1), namely define a sequence  $(z_i)_{i \in \mathbb{N}}$  with  $z_i \in D$  and such that

$$(2.2) \quad z_{i+1} = z_i + \frac{z_{i-1} - z_i}{2} + kf(z_i).$$

Defining

$$(2.3) \quad R_i = |z_i - z_{i-1}|,$$

we obtain from (2.2)

$$(2.4) \quad R_{i+1} \leq \frac{R_i}{2} + k|f(z_i)|.$$

When  $d > 1$ , we may write

$$R_i = ||z_i - z_{i-1}||.$$

Thus, any sequence  $(z_i)$  which satisfies (2.2) also satisfies (2.4). We define  $R_0 > 0$  arbitrary.

We now propose to substitute (2.2) by a random sequence  $(z_i)$  which satisfies (2.4). Also some additional conditions on  $(z_i)$  will be imposed. We will thus be able to prove the convergence of the resulting sequence  $(z_i)$  to some point in  $S$ ; reciprocally, for any  $x$  in  $S$ , when  $z_0$  is close enough to  $x$ , the limit point of  $(z_i)$  will coincides with  $x$ .

Define  $z_0$  and  $z_1$  uniformly in  $D$  and  $R_1 = ||z_1 - z_0||$ .

For  $i \geq 1$  compare  $f(z_i)$  and  $f(z_{i-1})$ . Let  $C \in [\frac{1}{2}, 1]$ . If

$$(2.5) \quad |f(z_i)| \leq C|f(z_{i-1})|,$$

then obtain  $z_{i+1}$  by

$$(2.6) \quad z_{i+1} := z_i + u_i,$$

where  $u_i$  is randomly drawn on  $\mathcal{B}(\underline{0}, \frac{R_i}{2} + k|f(z_i)|)$ , where  $\mathcal{B}(\omega, r)$  is the ball with center  $\omega$  and radius  $r$ .

If (2.5) is not fulfilled then the sequence  $(z_j)_{j \in \mathbb{N}}$  stops. Draw then  $z_0$  and  $z_1$  again. At this point we state

**Theorem 1.** *Any infinite sequence  $(z_i)$  defined as above converges a. s. with limit in  $S$ .*

We now add a number of conditions on the function  $f$  which entail that any point in  $S$  is reached asymptotically.

Let  $x \in S$  and set  $z_0 \in \mathcal{B}(x, \epsilon_0) = \{z : \|z - x\| \leq \epsilon_0\}$  for some  $\epsilon_0 > 0$ . Define further

$$(2.7) \quad E_0 := B \cap \{z : \|z - z_0\| > k_1 |f(z_0)|\},$$

with  $0 < k_1 < k$  and such that  $k_1 |f(z_0)| < 2\epsilon_0$ ;  $B$  is the ball with center  $z_0$  and radius  $\frac{R_0}{2} + k |f(z_0)|$ . Therefore,  $E_0$  is an annulus around  $z_0$ .

Let

$$(2.8) \quad A_1 = \text{int}\{\mathcal{B}(x, \epsilon_0) \cap B\}.$$

By its very definition, the set  $A_1$  is not void.

Assume that  $f$  satisfies the following regularity conditions

- (1) For all  $x \in S$ , there exists some  $\epsilon_0(x) > 0$  such that if  $z_0, z_1 \in \mathcal{B}(x, \epsilon_0)$  and  $\|x - z_1\| \leq \|x - z_0\|$  then

$$\{z : |f(z)| \leq |f(z_1)|\} \subsetneq \{z : |f(z)| \leq |f(z_0)|\}.$$

- (2) There exists  $0 < m < \frac{1}{4\epsilon_0}$  such that for all  $x \in S$ , for all  $z_0 \in \mathcal{B}(x, \epsilon_0)$  for all  $0 < \epsilon < k/2$ , for all  $z \in E_0 \cap A_1$ ,

$$|f(z_0)| - |f(z)| \geq m \|z - z_0\|.$$

By condition (1), the LHS in this inequality is non negative.

We then have

**Theorem 2.** *Let  $x \in S$  and  $\epsilon_0 > 0$  such that (1) and (2) hold. When  $z_0 \in \mathcal{B}(x, \epsilon_0)$ , the sequence  $(z_i)$  is infinite and satisfies Theorem 1. Furthermore  $\lim z_i = x$  a. s.*

In order to cover all  $S$  by the limiting points of such sequences we also propose to add a step where we randomly select  $p$  points uniformly in  $D$ . These points are initial points of new sequences; this allows to obtain a good covering of  $S$  by the limits of all these sequences.

Obviously this latest step does not substitute the entire algorithm; clearly a huge number of such points will approximate  $S$  from the start, the most inefficient Monte-Carlo random search method.

The stopping rule is defined through the definition of an accuracy index call  $tol$ . Define  $N$  the number of points to be reached in  $S$ . We may decide to stop the algorithm when  $N$  sequences  $(z_i)$  are such that the extremities are in  $S$  up to the accuracy, denoted  $tol$  in the sequel.

**2.2. Enhanced algorithm.** In order to improve the coverage of  $S$ , keeping the same set of points  $z_0$ , we propose to modify the choice of  $z_{i+1}$  as given in (2.5) and (2.6) as follows. From  $z_0, \dots, z_i$  we build indeed  $i$  chains, each one starting from  $z_j, 1 \leq j \leq i$ . Obviously the sequence starting at  $z_i$  is as described above; the new  $i - 1$  ones spread and develop in all directions. Any of these chains inherit of the properties mentioned in Theorem 1. Also, any  $x$  in  $S$  is asymptotically reached by one of those sequences, as  $i$  increases.

The sequences defined by an algorithm may be finite; indeed condition (2.5) may not hold for  $(z_{i-1}, z_i)$  and therefore  $z_{i+1}$  cannot be simulated. Thus no point  $z_{i+1}$  will be simulated since his father would be higher than his grandfather.

However his grandfather  $z_{i-1}$  is indeed lower than his grand-grandfather; therefore

his grandfather may have offspring. This grandfather is the root of a new generation, hence a new  $z_i$  which may satisfy (2.5). In the same way all ancestors of  $z_{i-1}$  satisfy (2.5) and are eligible for fatherhood.

We call a step of the algorithm the generation of all the offspring of the eligible points in the existing population of points. Such a step is followed by the generation of  $p$  uniformly distributed points in  $D$  as done in the basic algorithm.

In the sequel, we focus on the basic algorithm described in Section 2.2.

**2.3. Reducing the computational cost tuning the parameters.** Firstly this algorithm makes use of very few parameters. Furthermore those can be tuned easily according to the complexity of the problem at hand. Indeed these parameters can be interpreted in connection with the computational burden. In some cases the function  $f$  is very costly and running an algorithm for a long time, without evaluating  $f$  often, may be of great advantage. Sometimes the function  $f$  is easy to calculate and the need is to get a quick description of  $S$ . Tuning  $k, C$  and  $m$ , together with the number of initiating points, makes use for those choices.

The following examples illustrate the role of each of the parameters, all the other ones being kept fixed.

The number of solutions which we require in the tolerance zone around  $S$  is fixed to 1000, but in the last example where the algorithm is evaluated with respect to this number.

Examples are presented in dimension 2. Higher dimension examples are presented in Section 2.4. Red points are couples  $(x_1, x_2)$  such that  $f(x_1, x_2) > 0$ . Points with negative values of  $f$  are blue. Black points are all blue or red ones whose  $f$  value belongs to  $[-tol, tol]$ .

Each example is summarized by three indicators. The first one is the runtime. The second one is the efficiency coefficient (EC) which is the ratio between the total number of evaluations of  $f$  and the number of solutions, which equals 1000 in all but the last example. This indicator is a measure of the number of calls to  $f$  which are required in order to obtain one solution to the equation  $f(x) = 0$ . The third indicator is of visual nature; in all those examples which are in dimension 2, the quality of the coverage of  $S$  can be considered qualitatively.

**Remark 1.** *The most important indicator is EC, since in all industrial applications, what really matters is the cost in evaluating  $f$ .*

*The initialization step.* Call  $n$  the number of initiating points  $z_0$ , randomly selected on  $D$ . This is the initial cost of the method since the function  $f$  will be evaluated  $n$  times. Due to section 2.2,  $n$  should not be too large.

**Example 1.** *Let  $f$  be a bivariate function defined by*

$$(x_1, x_2) \mapsto f(x_1, x_2) = x_1^2 + x_2^2 - 0.5$$

*The aim is to find  $N = 1000$  pairs  $(x_1, x_2)$  such that  $|f(x_1, x_2)| \leq tol$  where  $tol$  is the accuracy. All parameters but  $n$  are fixed. The tolerance is 0.01; the value of  $C$  is fixed being 0.75; the value of  $k$  is 1; the number  $p$  of supplementary points at each step of the algorithm is 1.*

*The solutions are close to  $S = \{(x_1, x_2), f(x_1, x_2) = 0\}$ , the circle with center  $(0, 0)$  and radius  $\sqrt{0.5}$ . In Figure 1(a), the function  $f$  is intersected by the horizontal*

| $n$ | $tol$ | $N$  | $C$  | $k$ | $p$ | $Time$ | $EC$ | $Coverage$ |
|-----|-------|------|------|-----|-----|--------|------|------------|
| 5   | 0.01  | 1000 | 0.75 | 1   | 1   | 0.32s  | 4.33 | -          |
| 100 | 0.01  | 1000 | 0.75 | 1   | 1   | 0.60s  | 6.32 | +          |
| 300 | 0.01  | 1000 | 0.75 | 1   | 1   | 1.54s  | 9.14 | ++         |

TABLE 1. Results for Example 1 with different values of  $n$ 

plane  $z = 0$ . The Figure 1(b) represents the intersection in the variables frame. The circle is then clearly visible. In Figures 2 (a), (b), (c), we have considered

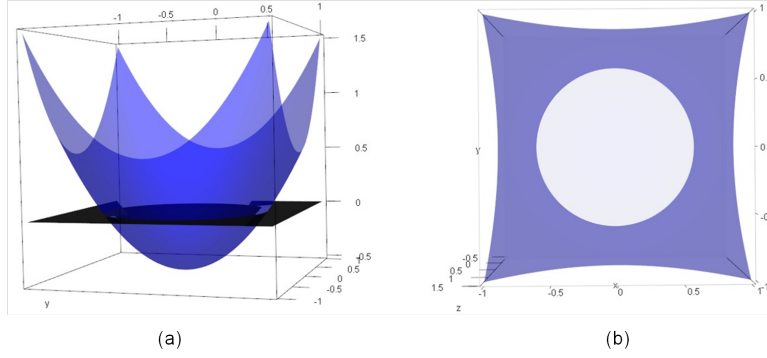
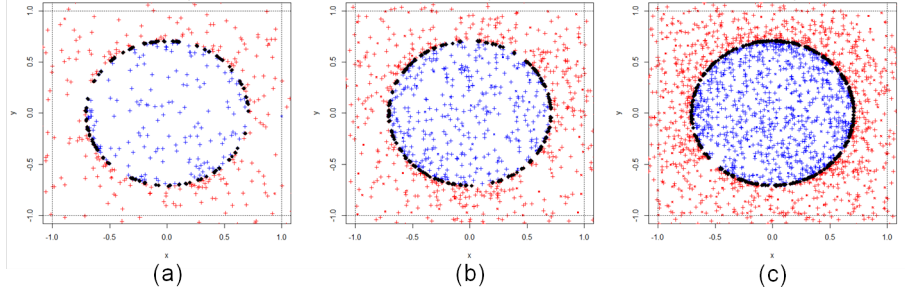


FIGURE 1. Representations of the quadratic function

respectively  $n = 5$ ,  $n = 100$  and  $n = 300$ . Clearly the more numerous the initial

FIGURE 2. Solving quadratic equation using SAFIP for three values of  $n$ 

points, the more the number of chains, and therefore the more numerous the points where the function  $f$  is evaluated; so the algorithm is costly as  $n$  increases. At the contrary, the better the coverage of  $S$ . Results are gathered in Table 1.

*The rate of convergence.* The value of  $C$  pertains to the rate of convergence of the algorithm. Assume  $C$  small ( $C$  close to  $1/2$ ); thus condition (2.5) is rarely satisfied. The selected points will define chains with a fast convergence to  $S$ . However in order to satisfy (2.5), many simulations in the ball  $B$  are required, leading to an increased runtime.

**Example 2.** Let  $f$  be a bivariate function defined by

$$(x_1, x_2) \mapsto f(x_1, x_2) = x_1^4 + x_2^3 - 0.5$$

The aim is to find  $N = 1000$  pairs  $(x_1, x_2)$  such that  $|f(x_1, x_2)| \leq \text{tol}$  where  $\text{tol}$  is the accuracy. All parameters but  $C$  are fixed. The number of initial points is 10; the tolerance is 0.015; the value of  $k$  is 1; the number  $p$  of supplementary points at each step of the algorithm is 1.

In Figure 3(a), the function  $f$  is intersected by the horizontal plane  $z = 0$ . The Figure 3(b) represents the intersection in the variables frame. In Figures 4 (a), (b),

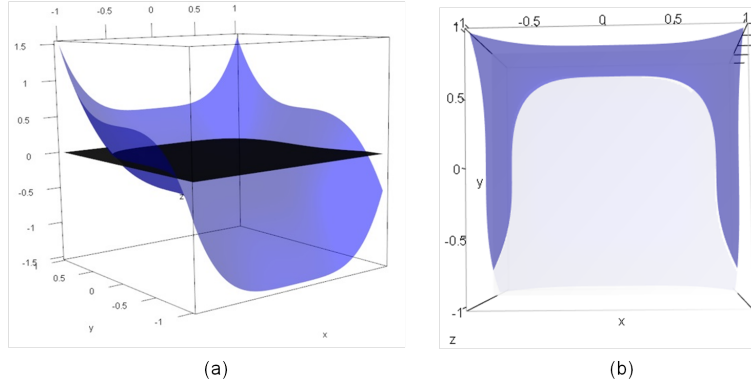


FIGURE 3. Representations of the function with a chair shape

(c), we have considered respectively  $C = 0.55$ ,  $C = 0.75$  and  $C = 0.95$ . The greater

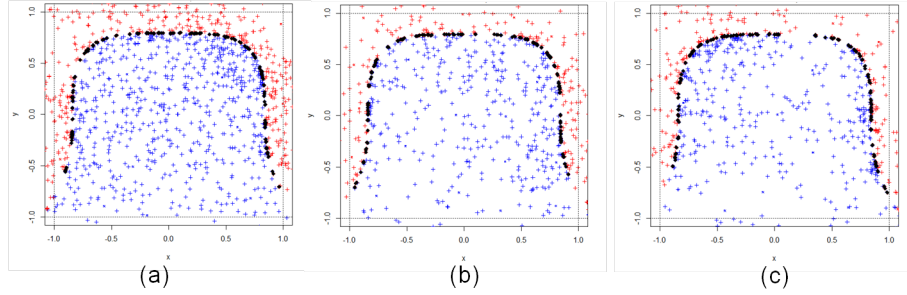


FIGURE 4. Solving equation for the function with a chair shape using SAFIP for three values of  $C$

$C$ , the less the number of evaluations of  $f$ ; furthermore the runtime decreases as  $C$  increases. Results are gathered in Table 2.

*The role of  $k$ .* The parameter  $k$  is crucial for the simulation around  $z_i$ . In order to give some insight on the value of  $k$ , suppose that  $z$  belongs to  $[-1, 1]^2$ , and that the mean value of  $|f(z)|$  is  $\bar{f} = 10$ . The current radius of the ball  $B$  is  $\frac{R}{2} + k|f(z)|$ , with  $R$  the distance between two points in the chain. Thus  $k$  should be at most of order  $\frac{1}{\bar{f}}$ ; in this way the ball  $B$  lays in  $[-1, 1]^2$ , roughly. This appears clearly in Example 3.

| $n$ | $tol$ | $N$  | $C$  | $k$ | $p$ | $Time$ | $EC$ | $Coverage$ |
|-----|-------|------|------|-----|-----|--------|------|------------|
| 10  | 0.015 | 1000 | 0.55 | 1   | 1   | 0.62s  | 8.36 | +          |
| 10  | 0.015 | 1000 | 0.75 | 1   | 1   | 0.44s  | 5.33 | +          |
| 10  | 0.015 | 1000 | 0.95 | 1   | 1   | 0.42s  | 5.05 | +          |

TABLE 2. Results for Example 2 with different values of  $C$ 

**Example 3.** Let  $f$  be a bivariate function defined by

$$(x_1, x_2) \mapsto f(x_1, x_2) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2 - 50$$

The aim is to find  $N = 1000$  pairs  $(x_1, x_2)$  such that  $|f(x_1, x_2)| \leq tol$  where  $tol$  is the accuracy. All parameters but  $k$  are fixed. The number of initial points is 10; the tolerance is 3; the value of  $C$  is 0.75; the number  $p$  of supplementary points at each step of the algorithm is 1.

In Figure 5(a), the function  $f$  is intersected by the horizontal plane  $z = 0$ . Figure 5(b) represents the intersection in the variables frame. The mean value of  $f$  is

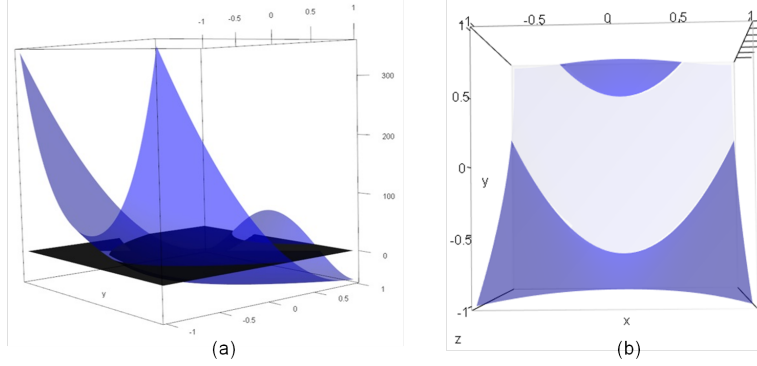
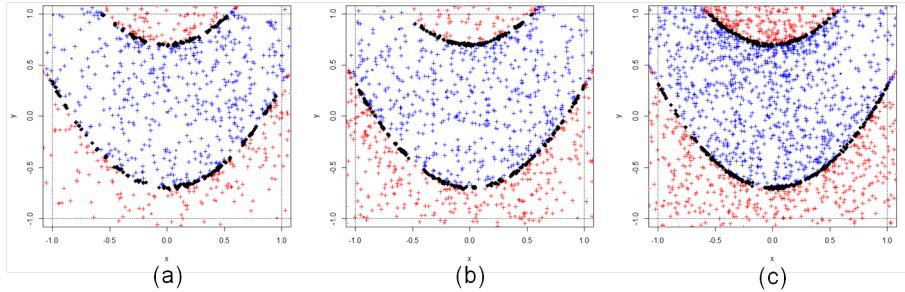


FIGURE 5. Representations of the Rosenbrock function

200 and its variations belong to  $[-50, 350]$ . In Figures 6 (a), (b), (c), we have considered respectively  $k = 1/200$ ,  $k = 10/200$  and  $k = 50/200$ . As  $k$  increases, the

FIGURE 6. Solving equation for the Rosenbrock function using SAFIP for three values of  $k$ 

runtime also increases as does the number of evaluations of  $f$  in order to obtain

| $n$ | $tol$ | $N$  | $C$  | $k$   | $p$ | $Time$ | $EC$  | $Coverage$ |
|-----|-------|------|------|-------|-----|--------|-------|------------|
| 10  | 3     | 1000 | 0.75 | 0.005 | 1   | 0.76s  | 10.69 | +          |
| 10  | 3     | 1000 | 0.75 | 0.05  | 1   | 2.76s  | 18.71 | +          |
| 10  | 3     | 1000 | 0.75 | 0.25  | 1   | 4.16s  | 48.49 | ++         |

TABLE 3. Results for Example 3 with different values of  $k$ 

one solution, and also the coverage of  $S$  improves. When  $f$  is costly,  $k$  should be chosen small. Results are gathered in Table 3.

*The role of  $p$ .* The number of intermediate points is important since it allows to explore new points of  $D$  in quest for  $S$ . This number should be chosen small with respect to the number  $n$  of initializing points. The following example shows that very small values of  $p$  may be good choices.

**Example 4.** Let  $f$  be a bivariate function defined by

$$(x_1, x_2) \mapsto f(x_1, x_2) = (x_1 - 0.5)^2 + 3x_1x_2 - x_2^3 - 2.25$$

The aim is to find  $N = 1000$  pairs  $(x_1, x_2)$  such that  $|f(x_1, x_2)| \leq tol$  where  $tol$  is the accuracy. All parameters but  $p$  are fixed. The number of initial points is 10; the tolerance is 0.04; the value of  $C$  is 0.75; the number  $k$  is 0.25.

In Figure 7(a), the function  $f$  is intersected by the horizontal plane  $z = 0$ . Figure 7(b) represents the intersection in the variables frame.  $p$  is chosen as 1, 3 and

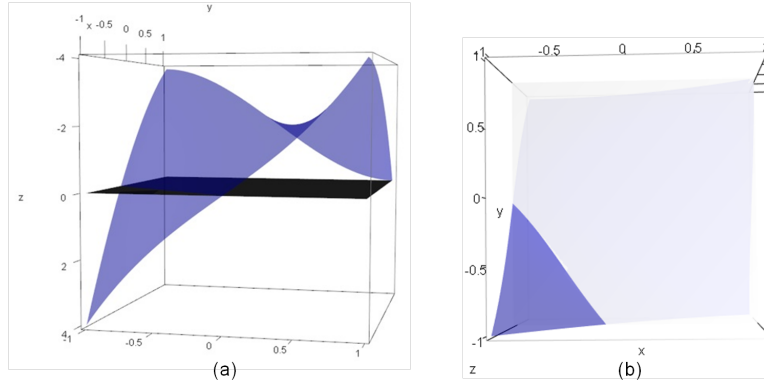


FIGURE 7. Representations of the polynomial function

5. In Figures 8 (a), (b), (c), we see that the algorithm has produced some insight to elements in  $S$  at the north-east region; however, the 1000 solutions have been obtained on the south-west component of  $S$ . Having asked for more solutions, we would have obtained the north-east component. Increasing  $p$  to 3 or 5, the coefficient  $EC$  increases noticeably and the coverage of  $S$  clearly increases. Results are gathered in Table 4.

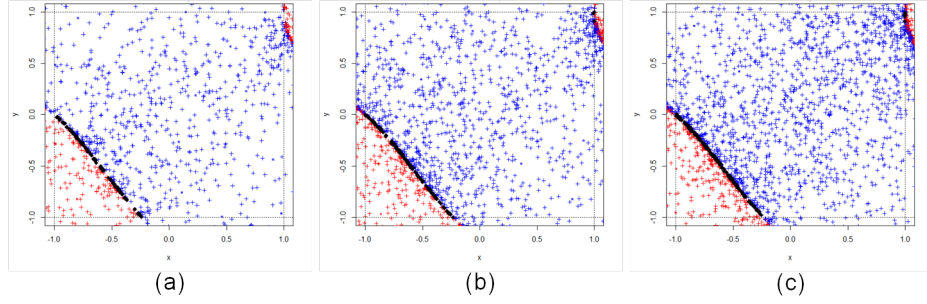


FIGURE 8. Solving equation for the polynomial function using SAFIP for three values of  $p$

| $n$ | $tol$ | $N$  | $C$  | $k$  | $p$ | $Time$ | $EC$  | $Coverage$ |
|-----|-------|------|------|------|-----|--------|-------|------------|
| 10  | 0.04  | 1000 | 0.75 | 0.25 | 1   | 2.12s  | 15.94 | +          |
| 10  | 0.04  | 1000 | 0.75 | 0.25 | 3   | 3.24s  | 14.58 | +          |
| 10  | 0.04  | 1000 | 0.75 | 0.25 | 5   | 4.96s  | 17.01 | ++         |

TABLE 4. Results for Example 4 with different values of  $p$

| $n$ | $tol$ | $N$  | $C$  | $k$  | $p$ | $Time$ | $EC$  | $Coverage$ |
|-----|-------|------|------|------|-----|--------|-------|------------|
| 10  | 0.15  | 1000 | 0.75 | 0.25 | 1   | 2.6s   | 43.47 | -          |
| 10  | 0.75  | 1000 | 0.75 | 0.25 | 1   | 1.68s  | 32.2  | -          |
| 10  | 1.5   | 1000 | 0.75 | 0.25 | 1   | 1.3s   | 22.85 | -          |

TABLE 5. Results for Example 5 with different values of  $p$

The tolerance factor  $tol$ . The strongest the tolerance (i. e. when  $tol$  is small), the highest the number of evaluations of  $f$ , and the longest the runtime.

**Example 5.** Let  $f$  be a bivariate function defined by

$$\begin{aligned}
 (x_1, x_2) \mapsto f(x_1, x_2) = & 8 \sin(7(x_1 - 0.9)^2)^2 + 6 \sin((14(x_1 - 0.9)^2)^2) + (x_1 - 0.9)^2 \\
 & + 8 \sin((7(x_2 - 0.9)^2)^2) + 6 \sin((14(x_2 - 0.9)^2)^2) \\
 & + (x_2 - 0.9)^2 - 15
 \end{aligned}$$

The aim is to find  $N = 1000$  pairs  $(x_1, x_2)$  such that  $|f(x_1, x_2)| \leq tol$  where  $tol$  is the accuracy. All parameters but  $tol$  are fixed. The number of initial points is 10; the value of  $C$  is 0.75; the number  $k$  is 0.08; the number  $p$  of supplementary points at each step of the algorithm is 1.

In Figure 9(a), the function  $f$  is intersected by the horizontal plane  $z = 0$ . Figure 9(b) represents the intersection in the variables frame. The function oscillates between -15 and 15. In Figures 10 (a), (b), (c), algorithm results are illustrated for three values of  $tol$  : 0.15, 0.75 and 1.5 . When  $tol$  varies from 0.15 to 1.5, the coefficient  $EC$  gets divided by 2. Results are gathered in Table 5.

Due to the complexity of the function and of the set  $S$ , coverage is mild whatever  $tol$ ; it depends upon the required number of solutions only.

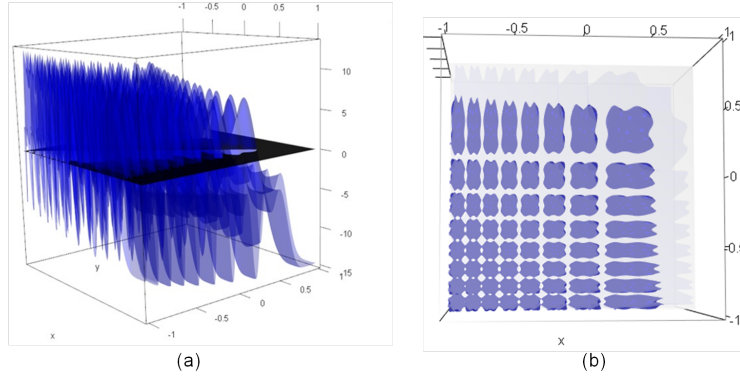
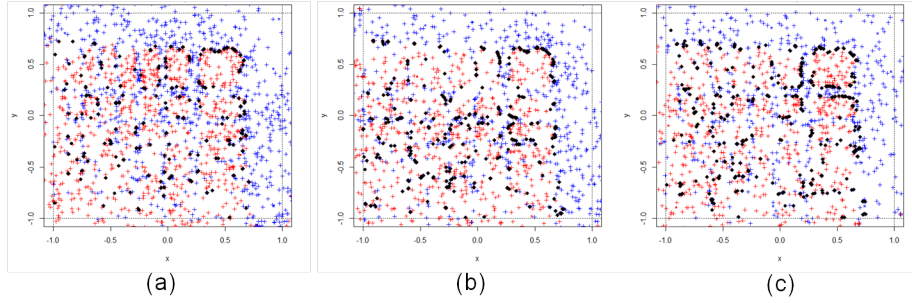


FIGURE 9. Representations of the trigonometric function

FIGURE 10. Solving equation for the trigonometric function using SAFIP for three values of  $tol$ 

*The role of  $N$ , the required number of solutions.* The same function as in Example 4 is used in order to focus on the role of the number of solutions. When we ask for 15000 points in  $S$ , then the runtime remains quite satisfactory; the EC coefficient is 76, due to a choice of  $n = 1000$ . The coverage of  $S$  is quite fair. Clearly the quality of the solutions improves with the required number of solutions. Not only do we get more solutions, but the coverage of  $S$  improves noticeably.

**Example 6.** Let  $f$  be a bivariate function defined by

$$(x_1, x_2) \mapsto f(x_1, x_2) = 20 + x_1^2 - 10 \cos(2\pi x_1) + x_2^2 - 10 \cos(2\pi x_2) - 60$$

*The aim is to find  $N$  pairs  $(x_1, x_2)$  such that  $|f(x_1, x_2)| \leq tol$  where  $tol$  is the accuracy. All parameters but  $N$  are fixed. The number of initial points is 10;  $tol$  is fixed to 0.4; the value of  $C$  is 0.75; the number  $k$  is 0.025; the number  $p$  of supplementary points at each step of the algorithm is 1.*

*In Figure 12(a), the function  $f$  is intersected by the horizontal plane  $z = 0$ . Figure 12(b) represents the intersection in the variables frame. In Figures 13 (a), (b), (c), algorithm results are illustrated for three values of  $N$  : 100, 1000 and 2000. When  $N$  is small, the important feature of the result is that  $S$  is covered equally. So no cluster of solutions seems to appear; this is important for exploratory analysis. Results are gathered in Table 6.*

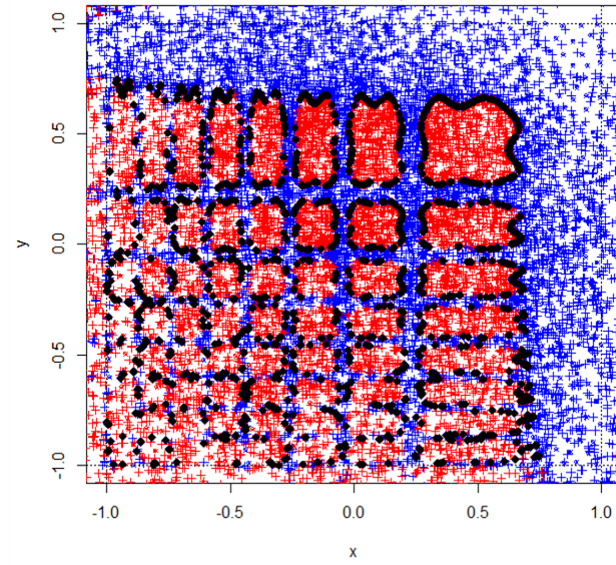


FIGURE 11. Solving equation for the trigonometric function using SAFIP for a bigger number of required final points and a tolerance of 0.15

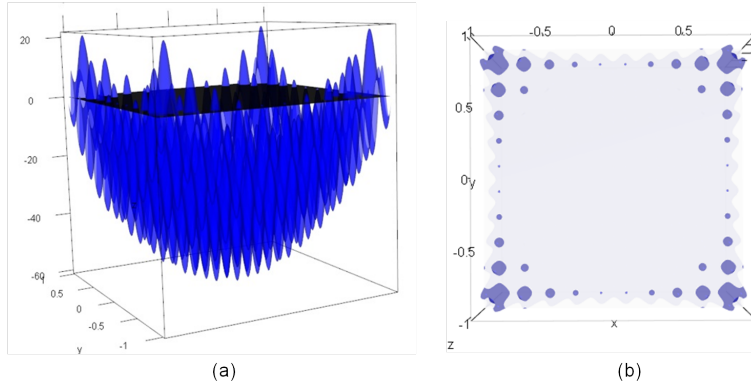


FIGURE 12. Representations of the Rastrigin function

| $n$ | $tol$ | $N$  | $C$  | $k$   | $p$ | $Time$ | $EC$  | $Coverage$ |
|-----|-------|------|------|-------|-----|--------|-------|------------|
| 10  | 0.4   | 1000 | 0.75 | 0.025 | 1   | 0.48s  | 55.33 | -          |
| 10  | 0.4   | 1000 | 0.75 | 0.025 | 1   | 3.96s  | 60.64 | -          |
| 10  | 0.4   | 1000 | 0.75 | 0.025 | 1   | 8.6s   | 83.82 | -          |

TABLE 6. Results for Example 6 with different values of  $N$

**2.4. Increasing the dimension.** We consider a collection of functions which mimic Example 1, increasing the dimension. The required number of solutions is kept as  $N = 500$  in all cases.

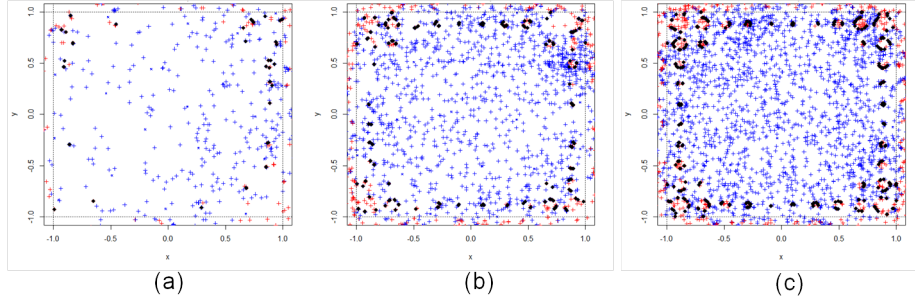


FIGURE 13. Solving equation for the Rastrigin function using SAFIP for three values of  $N$

We firstly consider the case in dimension 3, namely we look at points situated in

$$(2.9) \quad S := \{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 - 0.5 = 0\},$$

with  $-1 \leq x_i \leq 1$  for  $i = 1, 2, 3$ . The result appears in Figure 14. We also have

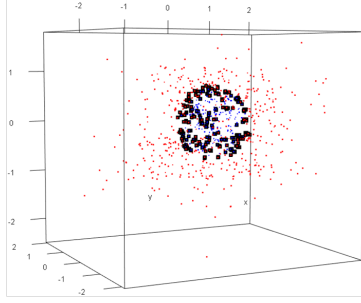


FIGURE 14. Results for spheres in dimension 3

considered the set

$$(2.10) \quad S := \{(x_1, x_2, x_3) : \max(x_1, x_2, x_3) - 0.5 = 0\};$$

See Figure 15.

Looking at similar examples as (2.9), we consider  $d = 4$  and  $d = 10$ ; the results

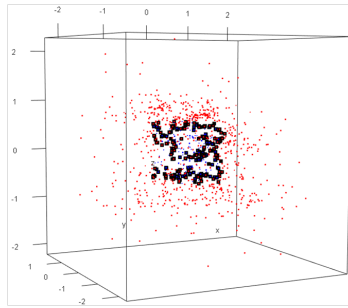


FIGURE 15. Results for cubes in dimension 3

| Dim | n    | tol | N   | C    | k | p | Time  | EC   |
|-----|------|-----|-----|------|---|---|-------|------|
| 2   | 5    | 0.1 | 500 | 0.75 | 1 | 1 | 0.22s | 4.81 |
| 3   | 25   | 0.1 | 500 | 0.75 | 1 | 1 | 4.72s | 6.64 |
| 4   | 75   | 0.1 | 500 | 0.75 | 1 | 1 | 0.4s  | 9.7  |
| 10  | 1000 | 0.1 | 500 | 0.75 | 1 | 1 | 53s   | 449  |

TABLE 7. Results for spheres in different dimensions

| Dim | n    | tol | N   | C    | k | p | Time  | EC   |
|-----|------|-----|-----|------|---|---|-------|------|
| 2   | 5    | 0.1 | 500 | 0.75 | 1 | 1 | 0.16s | 4    |
| 3   | 25   | 0.1 | 500 | 0.75 | 1 | 1 | 4.2s  | 5.04 |
| 4   | 75   | 0.1 | 500 | 0.75 | 1 | 1 | 0.72s | 8    |
| 10  | 1000 | 0.1 | 500 | 0.75 | 1 | 1 | 51s   | 614  |

TABLE 8. Results for cubes in different dimensions

comparing three dimensions are in Table 7. The same is available for (2.10) in Table 9. The number of initializing points has been chosen accordingly:  $n = 75$  for  $d = 4$ , and  $n = 1000$  for  $d = 10$ ; a coherent choice for  $n$  would have been  $n = 5^9$  for  $d = 10$ , an impracticable choice.

Obviously the indicator EC increases with  $n$ . However, choosing  $n = 5^9$  and  $N = 500$ , the value of EC exceeds 2000, which proves that  $n$  should be kept low, growing slowly with respect to  $d$ .

### 3. SIMULTANEOUS INVERSE PROBLEMS

**3.1. Algorithm.** Let  $f$  and  $g$  denote two functions defined on  $D$ ; each of these functions  $f$  and  $g$  is assumed to satisfy hypothesis (2.5) together with conditions (1) and (2). We will make use of constants  $C$ ,  $k$ ,  $n$  and  $p$  defined in Section 2.2; these constants will play a similar role in the present on  $f$  and  $g$ . The number of common solutions to the system

$$(3.1) \quad \begin{cases} f(x) &= 0 \\ g(x) &= 0 \end{cases}$$

is denoted  $N$ .

Also the present section considers simultaneous inverse problems pertaining to two functions; quantization to a given number of functions is straightforward.

The algorithm is as follows with similar notation as in Section 2.2, it holds

$$(3.2) \quad \begin{cases} f(x) &= 0 \\ g(x) &= 0 \end{cases} \Leftrightarrow \begin{cases} f(x) + \frac{x}{2k} + \frac{x}{2k} &= \frac{x}{2} \\ g(x) + \frac{x}{2k} + \frac{x}{2k} &= \frac{x}{2} \end{cases}$$

which yields to define

$$(3.3) \quad z_{i+1} = z_i + \frac{z_{i-1} - z_i}{2} + k \max(|f(z_i)|, |g(z_i)|).$$

Inequality (2.4) is substituted by

$$(3.4) \quad R_{i+1} \leq \frac{R_i}{2} + k \max(|f(z_i)|, |g(z_i)|).$$

Similarly as in (2.6), the choice of  $z_{i+1}$  follows the rule

$$(3.5) \quad z_{i+1} = z_i + u_i$$

where  $u_i$  is drawn randomly on  $\mathcal{B}(0, \frac{R_i}{2} + k \max(|f(z_i)|, |g(x_i)|))$ .

With those changes, denoting  $S = \{x : f(x) = 0, g(x) = 0\}$ , it holds

**Theorem 3.** *Any sequence  $(z_i)$  defined as above converges a. s. with limit in  $S$ .*

and

**Theorem 4.** *For any  $x \in S$  and  $\epsilon_0 > 0$  such that (1) and (2) hold simultaneously for  $f$  and  $g$ , and when  $z_0 \in \mathcal{B}(x, \epsilon_0)$ , thus the sequence  $(z_n)$  is infinite and converges to  $x$ .*

**3.2. Examples.** Due to (3.5), the point  $z_{i+1}$  is randomly chosen in a ball  $B$  centered at  $z_i$  when both  $|f(z_i)|$  and  $|g(z_i)|$  share a common mesural order of magnitude. The best case is when  $B$  has a moderate radius; it is therefore useful to normalize  $f$  and  $g$  on  $D$ ; this preliminary procedure obviously does not modify the set  $S$ . We present three examples of simultaneous inversion, based on the functions presented on Section 2.2. In all examples the parameters are  $n = 20$ ,  $p = 1$ ,  $tol = 0.01$ ,  $C = 0.75$ ,  $k = 1$ .  $N$  equals 10 in Example 7, it equals 100 in Example 8 and Example 9.

**Example 7** (A regular case). *We choose  $f$  as in Example 2 and  $g(x) = f(x-a)$ ,  $a = (0.2, -0.2)$ . Therefore  $f(x) = 0$  is as in Example 2 and  $g(x) = 0$  is a circle with same radius and center  $a$ .*

*Figures 16(a) and (b) show the graphs of  $f$  and  $g$  together with the intersection of the plane  $z = 0$ . The set  $S$  consists in the two points shown in Figure 16(b). Those*

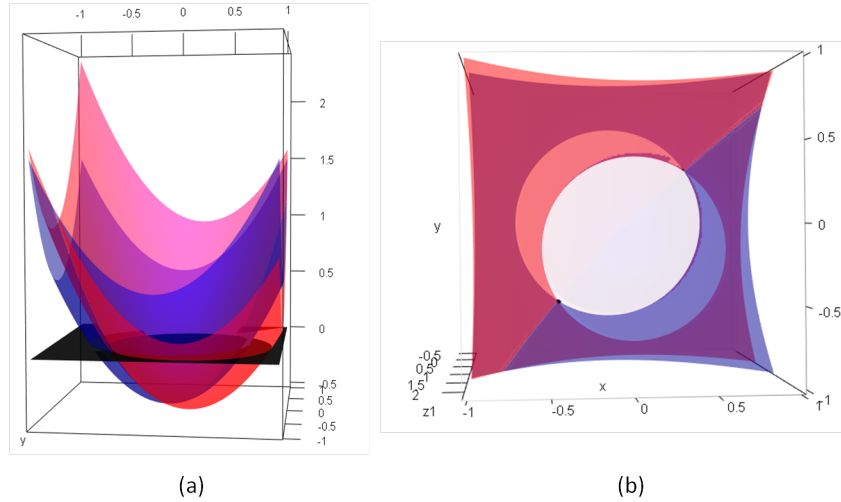


FIGURE 16. Representations of  $f$ ,  $g$  and  $S$

*points are indeed well estimated by the present algorithm, as seen in Figure 17. The runtime is 0.62s and the efficiency coefficient is 516.*

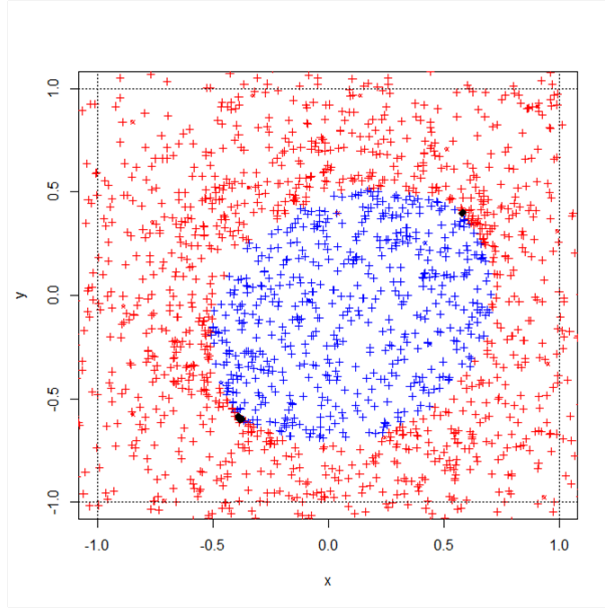


FIGURE 17. Solutions obtained with SAFIP algorithm

| $C$  | $EC$ | $Temps$ |
|------|------|---------|
| 0.55 | 905  | 4.72s   |
| 0.75 | 469  | 1.66s   |
| 0.95 | 311  | 1.24s   |
| $k$  | $EC$ | $Temps$ |
| 1    | 546  | 5.02s   |
| 10   | 1963 | 8.8s    |
| 50   | 6372 | 32.04s  |
| $n$  | $EC$ | $Temps$ |
| 10   | 577  | 2.54s   |
| 100  | 622  | 3.36s   |
| 300  | 708  | 3.36s   |

TABLE 9. Results for cubes in different dimensions

**Example 8** (Mixing a regular function and an irregular one). We choose  $f(x)$  as defined in Example 6, a regular function, and  $g(x)$  the Rastrigin function of Example 13. The Figure 18(a) shows the two functions, and Figure 18(b) provides the set  $S$ , which is defined as the intersection of the frontier points of the red domains (the solutions to  $g(x) = 0$ ) with the frontier points of the blue domains (the solutions to  $f(x) = 0$ ). There are 29 points in  $S$ . The algorithm provides solutions as shown in Figure 19, with runtime 14s and efficiency coefficient 375. Table ?? provides results for different values of  $C$ ,  $k$  and  $n$ .

As  $C$  increases,  $EC$  decreases; as  $k$  or  $n$  increases,  $EC$  increases too.

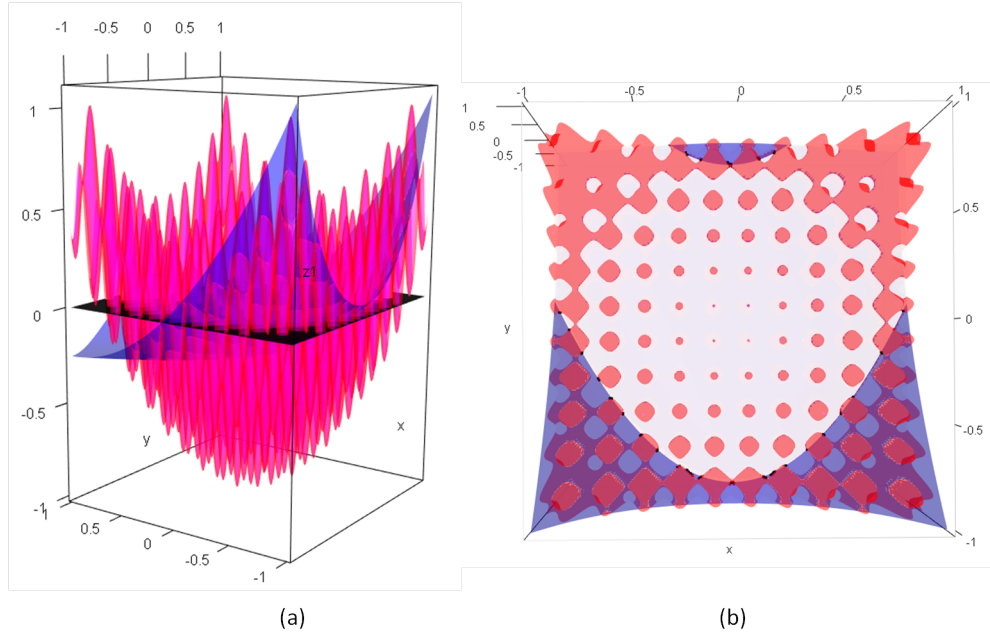
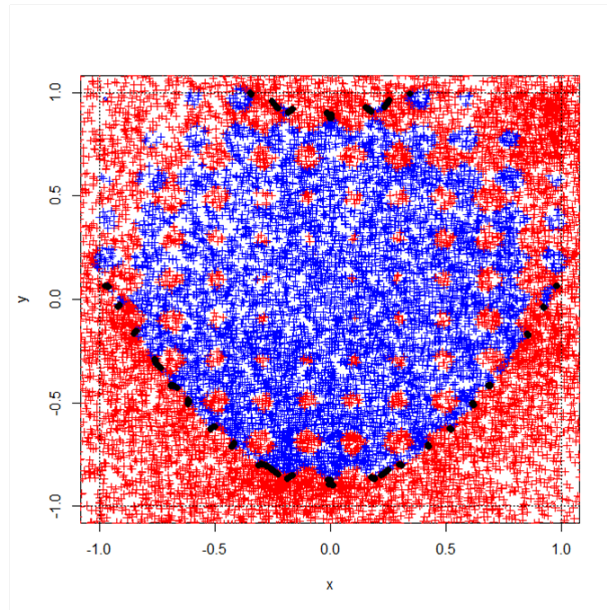
FIGURE 18. Representations of  $f$ ,  $g$  and  $S$ 

FIGURE 19. Solutions obtained with SAFIP algorithm

A clear feature in Figure 19 is that all the 29 points in  $S$  are obtained a limiting points of SAFIP.

**Example 9** (A last example). We choose  $f(x)$  as in Example 2 and  $g(x)$  the trigonometric function of Example 10. Figure 20(a) shows the functions  $f$  and  $g$ ; Figure 20 (b) shows the intersection set  $S$  which contains 33 points. We asked for

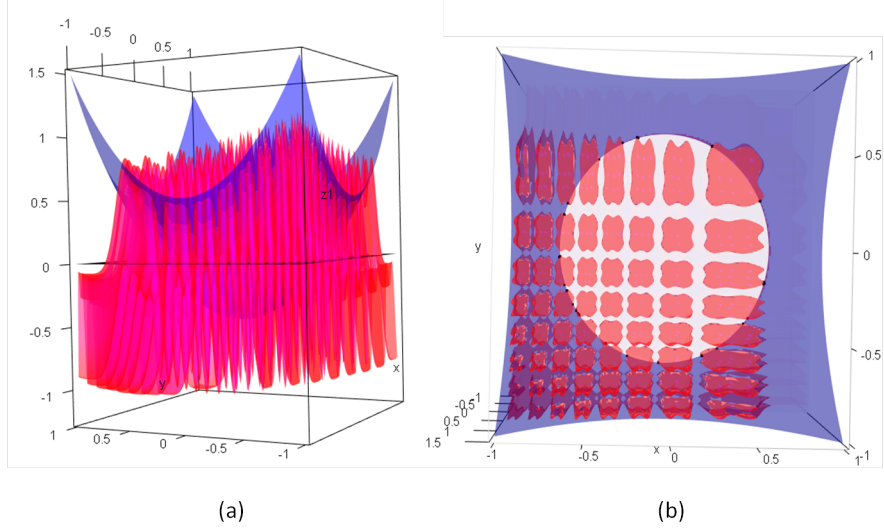


FIGURE 20. Representations of  $f$ ,  $g$  and  $S$

$N = 100$  solutions; the set  $S$  is not totally covered (we obtain 26 points in  $S$  as it can be seen on Figure 21); a larger value of  $N$  would provide all solutions. The

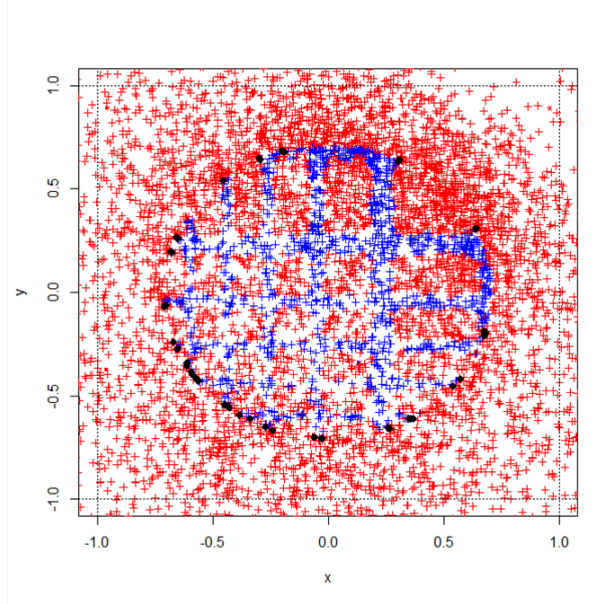


FIGURE 21. Solutions obtained with SAFIP algorithm  
runtime is 4.1s and EC is 1102.

## 4. APPENDIX

*Proof of Theorem 1.*

*Step 1.* We prove that the sequence  $(R_i)_{i \in \mathbb{N}}$  converges to 0 a. s.

Denote  $a := |f(z_0)| > 0$ . By (2.5),

$$|f(z_i)| \leq aC^i,$$

hence  $R_{i+1} \leq \frac{R_i}{2} + akC^i$ .

The sequence  $(R_i)_{i \in \mathbb{N}}$  is now compared to the sequence  $(x_i)_{i \in \mathbb{N}}$  defined by

$$x_{i+1} = \frac{x_i}{2} + akC^i.$$

It holds

$$\begin{aligned} x_n &= \frac{x_0}{2^n} + \frac{ak}{2^{n-1}} + \frac{akC}{2^{n-2}} + \frac{akC^2}{2^{n-3}} + \dots + \frac{akC^{n-2}}{2^1} + akC^{n-1} \\ (4.1) \quad &= \frac{x_0}{2^n} + akC^{n-1} \sum_{j=0}^{n-1} \left( \frac{1}{2C} \right)^j. \end{aligned}$$

When  $C > 1/2$ , it follows that  $x_n$  given in (4.1) tends to 0 as  $n \rightarrow \infty$ .

Since the generic term of  $(R_n)_{n \in \mathbb{N}}$  satisfies

$$(4.2) \quad R_n \leq \frac{R_0}{2^n} + akC^{n-1} \sum_{j=0}^{n-1} \left( \frac{1}{2C} \right)^j,$$

where the RHS is  $x_n$ , it follows that  $R_n$  tends to 0 as  $n \rightarrow \infty$ .

*Step 2.* Assume at present that  $(z_n)_{n \in \mathbb{N}}$  is an a. s. convergent sequence, and denote  $l$  its limit. We prove that  $l$  belongs to  $S$ . Indeed by (2.2), writing  $u_n = v_n(\frac{R_n}{2} + k|f(z_n)|)$  for  $v_n$  uniformly distributed on  $\mathcal{B}(0, 1)$ , the unit ball in  $\mathbb{R}^d$ . Going to the limit in (2.2),  $l = l + \lim_{n \rightarrow \infty} u_n$ . It follows that  $\lim_{n \rightarrow \infty} \frac{R_n}{2} + k|f(z_n)| = 0$ . Since  $\lim_{n \rightarrow \infty} R_n = 0$ , it holds

$$\lim_{n \rightarrow \infty} |f(z_n)| = 0 \text{ a. s.}$$

By continuity of  $f$ , it follows that  $\lim_{n \rightarrow \infty} |f(z_n)| = f(l)$  and then  $f(l) = 0$ . We have proved that  $l \in S$ .

It remains to prove that  $(z_n)_{n \in \mathbb{N}}$  converges, showing that it is a Cauchy sequence. Let  $(m, n) \in \mathbb{N}^2, m > n$ . Then

$$\begin{aligned} \sup_{m > n} \|z_m - z_n\| &\leq \sup_{m > n} \sum_{j=n+1}^m \|z_j - z_{j-1}\| \\ &\leq \sup_{m > n} \sum_{j=n+1}^m r_j. \end{aligned}$$

By (4.2),

$$\begin{aligned} \sup_{m>n} \|z_m - z_n\| &\leq \sup_{m>n} \sum_{j=n+1}^m \left( \frac{r_0}{2^j} + akC^{j-1} \left( \frac{2C - (\frac{1}{2C})^{j-1}}{2C-1} \right) \right) \\ &\leq \sup_{m>n} \left( \frac{r_0 \left(1 - (\frac{1}{2})^{m-n}\right)}{2^n} + \frac{2akC^{n+1}}{2C-1} \times \frac{1 - C^{m-n}}{1-C} - \frac{ak}{(2C-1)2^n} \times \frac{1 - (\frac{1}{2})^{m-n}}{\frac{1}{2}} \right), \end{aligned}$$

with  $0 < 2C - 1 < 1$ . Since  $m > n$  and  $C < 1$

$$\sup_{m>n} \|z_m - z_n\| \leq \frac{r_0}{2^{n+1}} + \frac{2akC^{n+1}}{(2C-1)(1-C)} - \frac{ak}{(2C-1)2^{n-1}}$$

and therefore

$$(4.3) \quad \lim_{n \rightarrow \infty} \sup_{m>n} \|z_m - z_n\| = 0,$$

which proves the claim. ■

*Proof of Theorem 2.* By (2.7), we have  $E_0 = \{z : k_1|f(z_0)| \leq \|z - z_0\| \leq \frac{R_0}{2} + k|f(z_0)|\}$ , with  $\epsilon_0 = \|x - z_0\|$ . We have to prove that  $E_0 \cap A_1 \neq \emptyset$ .

By (2.8) and since  $E_0 \subset B$ , this is equivalent to prove that  $\mathcal{B}(x, \epsilon_0) \cap E_0 \neq \emptyset$ . By the definition of  $E_0$  which is an annulus centred on  $z_0$  with a minimal radius of  $2\epsilon_0$  and since  $z_0 \in \partial\mathcal{B}(x, \epsilon_0)$  according to the definition of  $\epsilon_0$ ,  $\mathcal{B}(x, \epsilon_0) \cap E_0 \neq \emptyset$  and so  $E_0 \cap E_1 \neq \emptyset$ .

Let  $z_1 \in A_1 \cap E_0$ . we prove that  $z_1$  satisfies (2.5).

By condition 2, it follows

$$|f(z_0)| - |f(z_1)| \geq mk_1|f(z_0)|,$$

since  $z_1 \in E_0$ . This is equivalent to

$$|f(z_1)| \leq (1 - mk_1)|f(z_0)|$$

With an arbitrary  $k_1$  close to 0 such that  $0 < mk_1 < \frac{1}{2}$ . Getting  $C = (1 - mk_1) \in [\frac{1}{2}, 1]$ , we have  $|f(z_1)| \leq C|f(z_0)|$  for  $z_1 \in E_0$ . Thus  $z_1 \in A_1 \cap \{z_1, |f(z_1)| \leq C|f(z_0)|\}$  and  $z_0$  can have an offspring.

Iterating the above argument we can construct a sequence of balls  $\mathcal{B}(x, \epsilon_i)$  with lower bounded and decreasing sequence of radius. Thus this sequence converges to some limit. By Theorem 1,  $\lim_{i \rightarrow \infty} z_i = x^* \in S$ .

We show that  $x^* = x$  by contradiction.

If  $x^* \neq x$ , thus there exists  $i \in \mathbb{N}$  such that  $x \notin \mathcal{B}(x^*, \|x^* - z_i\|)$ . But  $z$  is simulated around  $x$  with decreasing radius to 0. Hence is the contradiction. Thus  $x^* = x$  and we have proved Theorem 2. ■

## REFERENCES

- [1] Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409, 1990.
- [2] Gene H. Golub and Charles F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.
- [3] András György and Levente Kocsis. Efficient multi-start strategies for local search algorithms. *Journal of Artificial Intelligence Research*, pages 407–444, 2011.
- [4] Mohamed Achibi Maëva Biret and Michel Broniatowski. Recherche des ensembles de niveaux d'une fonction multi variée à valeurs réelles sous conditions de monotonie. *I-Revues CNRS, Actes du Congrès Lambda-Mu 19*, 2014.

- [5] Curtis Miller. Search for level sets of functions using computer experiments. 2005.
- [6] Gen Nakamura and Roland Potthast. *Inverse Modeling*. 2053-2563. IOP Publishing, 2015.
- [7] Endre Süli and David F. Mayers. *An introduction to numerical analysis*. Cambridge university press, 2003.
- [8] A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and Anatoly G. Yagola. *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media, 2013.

*Current address:* <sup>(1)</sup>Safran Aircraft Engine, Moissy-Cramayel

*Current address:* <sup>(2)</sup>Université Pierre et Marie Curie, Paris

*E-mail address:* <sup>(\*)</sup> **Corresponding author:** `michel.broniatowski@upmc.fr`