Convergence of Stochastic Search Algorithms to Gap-Free Pareto Front Approximations

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

Oliver Schütze, INRIA Futurs
Marco Laumanns, ETH Zurich
Emilia Tantar, INRIA Futurs
Carlos A. Coello Coello, CINVESTAV
El-Ghazali Talbi, INRIA Futurs

Outline

- **Introduction**
  - Background
  - Need for Suitable Archivers for EMO Algorithms
  - Need for Gap-Free Pareto Front Approximations
- **Archiver ArchiveUpdateTight1**
  - The Object
  - The Strategy
  - Theoretical Investigation (Convergence, Bounds)
  - Numerical Results

INRIA

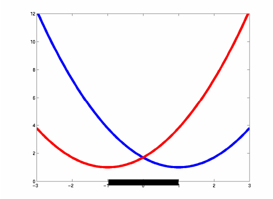
Background

(MOP) $\min_x F : R^n \to R^k$

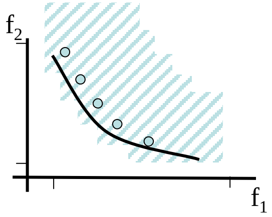
F continuous

Definition (*ε-dominance)*
Let $\varepsilon \in R^k_+$ and $x,y \in R^n$. x ε-dominates y ($x <_\varepsilon y$) if
$F(x) - \varepsilon \leq_p F(y)$ and $F(x) - \varepsilon \neq F(y)$

Definition (*ε-approximate Pareto set*):
A set $A \subset R^n$ is called an *ε-approximate Pareto set* of (MOP), if every point $x \in R^n$ is ε-dominated by at least one $a \in A$.
→ finite approximation possible



Pareto set



Generic Stochastic Search Algorithm

Algorithm GSSA

1: ▶*Initialization:*
2: $P_0 \subset Q$ drawn at random
3: A_0:= ArchiveUpdate(P_0,∅)
4: **for** $j=0,1,2,\ldots$ **do**
5: P_{j+1} = Generate(P_j)
6: A_{j+1} = ArchiveUpdate(P_{j+1},A_j)
7: **end for**

Given:

$\min F{:}Q \subset R^n \to R^k$ (**MOP**)

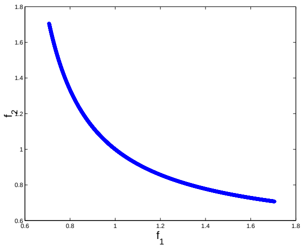
where
- Q domain (compact)
- P_l population at step l
- A_l archive at step l

Generator: extensively studied since several years.

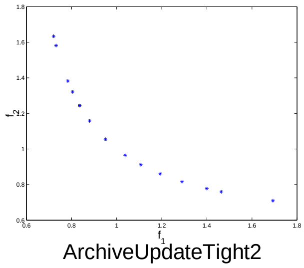
Archiver: investigation is scarce
Rudolph, *Laumanns*, *Corne & Knowles*: discrete models
Hanne: continuous models

Example 1: Effect of Archiver

Setting: both archivers are fed with N= 500,000 randomly generated test points (same sequence for both archivers)



'classical' archiver



ArchiveUpdateTight2
based on ε-dom.

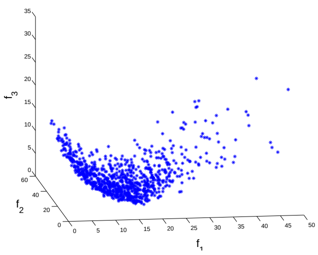
Time required to update the archive (averaged)

T = 3730 sec.

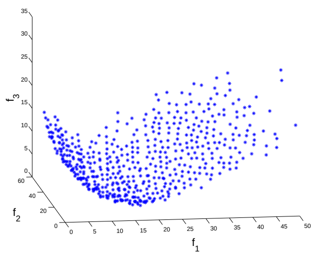
T = 2.8 sec.

Example 2: Fixed Running Time (3 Obj.)

Setting: fixed running time for simple EMO algorithm (5 minutes, same sequence of points), different archivers.



'classical' archiver



ArchiveUpdateTight1

Number of points evaluated during run of the algorithm

6.6e5 function calls

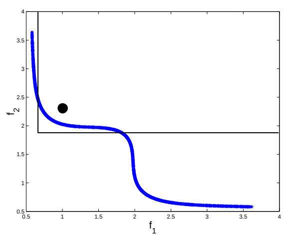
3.5e7 function calls

Possible Tasks for the Archiver

- **Pareto Set Approximation**
  - all efficient points
  - all ε-efficient points
  - finite size Hausdorff approximation ($d_H(A,P) < \varepsilon$)
- **Pareto Front Approximation**
  - ε-approximate Pareto set, ε-Pareto set
  - gap-free approximation
  - entire front
- **Limited Archive ($|A| \leq N$)**
  - maximal spread
  - ‘hot spots‘ (e.g., knee)

INRIA

Need for Gap-Free Approximation



‘Problem’ when only using ε-dominance:
Gaps can occur where the Pareto front is ‘flat’ (e.g., when dents occur).

These gaps can be unwanted in certain situations:

Local Search / Memetic Strategies
Search algorithms which are able to search along the Pareto set (continuation methods, Predator-Prey models).
→ ‘**Barrier**’ induced by ε-dominance could lead to false termination of the algorithm.

Optimal Control of Mechatronical Systems
Setting: Online selection of ‘optimal‘ Pareto points (computed offline).
Abrupt jumps in the objective values have to be avoided → **Instabilities**.

INRIA

The Object

Definition: A set $A \subset R^n$ is called a ***(Δ_1,Δ_2)-tight ε-approximate Pareto set*** of (MOP) if

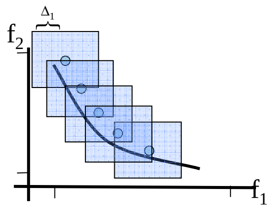
(a) A is an ε-approx. Pareto set,

(b) $\sup_{p \in P} \inf_{a \in A} d_\infty(F(p), F(a)) \le \Delta_1$ (*)

(Δ_1: tightness)

(c) $d_\infty(F(a_1), F(a_2)) \ge \Delta_2, \quad \forall a_1, a_2 \in A, a_1 \ne a_2$

(Δ_2: uniformity level)



Remark: (*) is equivalent to $F(P) \subset C_{A,\Delta} := \bigcup_{a \in A} B_\Delta^\infty(F(a))$

Algorithm ArchiveUdateTight1

Task: obtain an (Δ,ε)-tight ε–approx. Pareto set in the limit

Given:
$\varepsilon \in R^k_+$, $\Delta \in R_+$
$\Theta \in (0,1)$, $\Delta^* < \Delta$

Remark:
in practise: Θ=1, Δ^*=Δ

A:= ArchiveUpdateTight1 (P, A_0)

A:= A_0
for all $p \in P$ **do**
  if $(\exists a \in A : a \prec p)$ or
    $(\exists a_1 \in A : a_1 \prec_{\Theta\varepsilon} p$ and
    $\exists a_2 \in A : d_\infty(F(a_2), F(p)) \le \Delta^*)$ then
      CONTINUE
  end if
  A:= A $\cup$ {p}
  for all $a \in A$ **do**
    if p<a **then** A:=A\{a}
  end for
end for

Convergence Result

Theorem 1: Let *(MOP)* be given, where *F* is continuous, let $Q \subset R^n$ be a compact set and $\varepsilon \in R^k_+$. Let $\Theta \in (1,0)$ and $\varepsilon_m := min_{i=1...k} \varepsilon_i$ and $\varepsilon_M := max_{i=1..k}\, \varepsilon_i$, let Δ, $\Delta^* \in R_+$ such that $\varepsilon_M < \Delta^* < \Delta$, and

$$\forall x \in Q \text{ and } \forall \delta > 0: \qquad P(\exists l \in N : P_l \cap B_\delta(x) \cap Q \neq \emptyset) = 1 \quad \text{(A)}$$

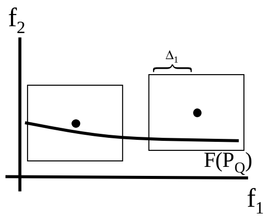
Then an application of Algorithm 1, where *ArchiveUpdateTight1* is used to update the archive, leads to a sequence of archives, such that

(a) there exists with probability one an $l_0 \in N$ such that A_l is an **ε-approximate Pareto set** for all $l \geq l_0$.

(b) there exists with probability one an $l_1 \in N$ such that A_l is a **$(\Delta, \Theta\varepsilon_m)$-tight ε–approximate Pareto set** for all $l \geq l_1$.

Proof (Tightness)

Proof not constructive. Key: replacements



Proof (Tightness)

Proof not constructive. Key: replacements

F is continuous on a compact domain Q:

$m_i := \min_{x \in Q} f_i(x), \qquad M_i := \max_{x \in Q} f_i(x)$

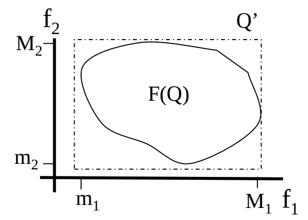
$F(Q) \subset Q' := [m_1 - \Theta\varepsilon_1, M_1] \times \ldots \times [m_k - \Theta\varepsilon_k, M_k]$

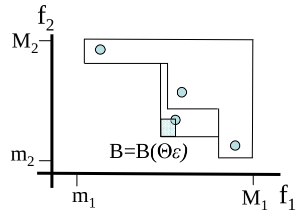
$D(A,\varepsilon) := \{ y \in Q' \mid \exists a \in A : F(a) - \Theta\varepsilon \leq_p y \}$

Insertion :

(I1) $\neg \exists a \in A : a \prec_{\Theta\varepsilon} \mathrm{p}$, or

(I2) $\neg \exists a \in A : a \prec \mathrm{p}$ and $\forall \mathrm{a} \in A : d_\infty(\mathrm{F(a)}, \mathrm{F(p)}) > \Delta^*$





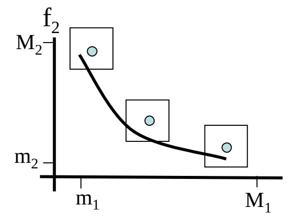
→Volume increases by every insertion by Vol(B)>0

→only **finitely many insertions** possible since Vol(D(A,ε))≤(Vol(Q')<∞

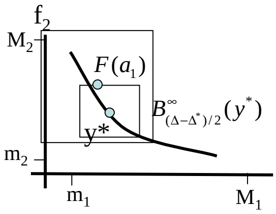
Proof (Tightness)

Assume claim (b) is wrong, that is

$$\exists y_i \in F(P_Q) \setminus C_{A_i,\Delta} \quad \forall i \in N \qquad (1)$$



Proof (Tightness)



Assume claim (b) is wrong, that is

$$\exists y_i \in F(P_Q) \setminus C_{A_i,\Delta} \quad \forall i \in N \qquad (1)$$

Since F(Q) is bounded there exists a subsequence

$$y_{i_j} \to y^* \in F(P_Q) \qquad (2)$$

Let $U := B^{\infty}_{(\Delta-\Delta^*)/2}(y^*)$. By assumption there exists with probability one an $l_1 \in N$ such that there exists an $x1 \in P_{l1}$ with $y_1=F(x) \in U$, and it holds (simplified!)

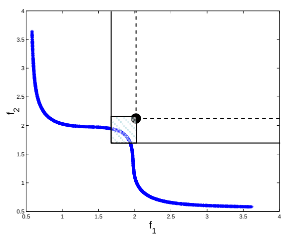
$$\exists a_1 \in A: \quad d_\infty(F(a_1), \tilde{y}) < \Delta \quad \forall \tilde{y} \in U$$

By (1) + (2) it follows that there exists an $l_2 > l_1 \in N$ and an $y_{l2} \in (F(P_Q) \cap U) \setminus C_{Al2,\Delta}$.
This is only possible via a replacement of a1.
Analogue: infinitely many replacements → contradiction! → claim (b)

INRIA

Proof (Uniformity Level)

Uniformity level (minimal distance between two entries in the archive): given by $\Theta\varepsilon_m$ ■



Remark: this 'exclusion strategy'

+ makes it possible to obtain the following upper bound on the archive size:

$$|A_l| \le \sum_{\substack{i_1,\dots,i_{k-1}=1 \\ i_1>\dots>i_{k-1}}}^{k} \prod_{j=1}^{k-1} \left\lceil \frac{M_{i_j} - m_{i_j}}{\theta \varepsilon_{i_j}} \right\rceil$$

– prevents convergence of the entries of the archive toward the Pareto set.
→ *ArchiveUpdateTight2, see* ***[Sch., Laumanns, Tantar, Coello, Talbi; 2007]***

INRIA

Bounds on the Archive Sizes

Theorem 2: Let $|A_0|=1$ and $m_i=\min_{x\in Q} f_i(x)$, $M_i=\max_{x\in Q} f_i(x)$. Then, when using ArchiveUpdateTight1, the archive size maintained by GSSA for all $l\in N$ is bounded as

$$|A_l| \le \sum_{\substack{i_1,\dots,i_{k-1}=1 \\ i_1>\dots>i_{k-1}}}^{k} \prod_{j=1}^{k-1} \left\lceil \frac{M_{i_j}-m_{i_j}}{\theta\varepsilon_{i_j}} \right\rceil$$

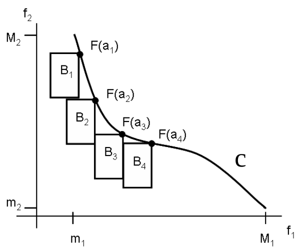
INRIA

Proof (k=2)

Points are located on a 'virtual' curve

$$c:[m_1,M_1]\to R^2$$
$$u \mapsto (u, f(u))$$

where $f:[m_1,M_1]\to[m_2,M_2]$ is strictly monotonically decreasing. The length of c can be bounded by:



$$L(c)=\int_{m_1}^{M_1}\|c'(u)\|_2\,du=\int_{m_1}^{M_1}\sqrt{|1|^2+|f'(u)|^2}\,du \le \int_{m_1}^{M_1}1du+\int_{m_1}^{M_1}|f'(u)|\,du=\int_{m_1}^{M_1}1du-\int_{m_1}^{M_1}f'(u)du$$
$$\le (M_1-m_1)+(M_2-m_2)$$

→ Bound for k=2: $|A_l| \le \left\lceil\frac{M_1-m_1}{\Theta\varepsilon_1}\right\rceil+\left\lceil\frac{M_2-m_2}{\Theta\varepsilon_2}\right\rceil$

INRIA

Proof (k≥2)

Proof analogue (integration in R^k)

$$\Phi : K \to R^k, \quad \Phi(u_1,\dots,u_{k-1}) \mapsto (u_1,\dots,u_{k-1}, f(u_1,\dots,u_{k-1}))$$

$$Vol_k(\Phi) = \int_K \sqrt{\|\nabla f\|^2 + 1}\,du = \int_K \sqrt{\left(\frac{\partial f}{\partial u_1}\right)^2 + \dots + \left(\frac{\partial f}{\partial u_{k-1}}\right)^2 + 1}\,du$$

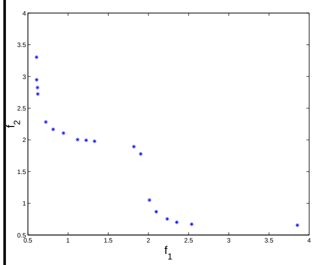
$$\le \int_K \left|\frac{\partial f}{\partial u_1}\right| du + \dots + \int_K \left|\frac{\partial f}{\partial u_{k-1}}\right| du + \int_K 1 du$$

$$= \sum_{i=1}^{k-1} \left(\int_{K_{(i)}} \left(\int_{m_i}^{M_i} \left|\frac{\partial f}{\partial u_i}\right| du_i \right) du_{(i)} \right) + \int_K 1 du = \sum_{i=1}^{k-1} \left(\int_{K_{(i)}} \left(-\int_{m_i}^{M_i} \frac{\partial f}{\partial u_i} du_i \right) du_{(i)} \right) + \int_K 1 du$$

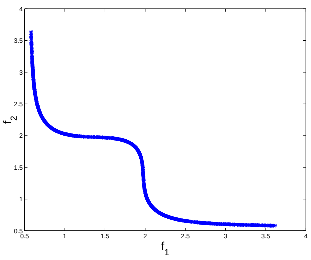
$$\le \sum_{\substack{i_1,\dots,i_{k-1}=1 \\ i_1 > \dots > i_{k-1}}}^{k} \prod_{j=1}^{k-1} (M_{i_j} - m_{i_j})$$

Example 2: Pareto Front with Dent

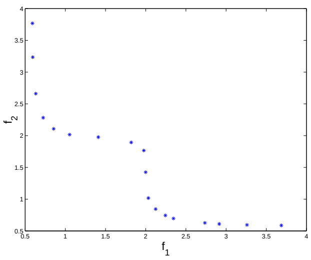
Generator: random search, *N=200,000* points within *Q=[-1,1]²*



UpdateEps1
T = 0.28 sec.



UpdateND
T = 134.03 sec.

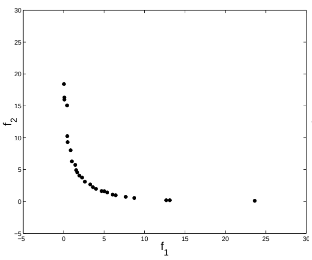


UpdateTight1
T = 0.39 sec.

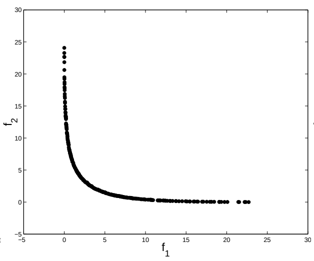
`http://paradiseo.gforge.inria.fr`

Example 3: Pareto Front with 'Flat' Parts

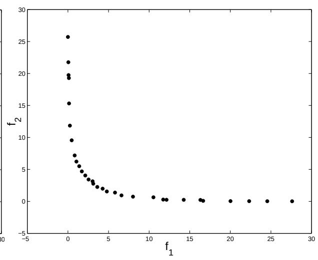
Generator: random search, *N=200,000* points within *Q=[-1.5,1.5]³*



UpdateEps1
T = 0.29 sec.



UpdateND
T = 36.46 sec.



UpdateTight1
T = 0.36 sec.

INRIA

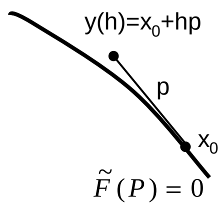
Outlook Memetic Strategies: Hybridization with Continuation Methods

Given: $x_0 \in Q$ (Pareto candidate)

$p \in R^n$ with $\|p\|_\infty = 1$, F L-cont., ε

Task: find $h \in R_+$ such that for $y(h)=x_0+hp$ it holds

$\|F(x_0)-F(y(h))\|_\infty \approx \varepsilon$



$$\|F(x_0) - F(y(h))\|_\infty \leq L\|x_0 - y(h)\|_\infty \quad \forall h \in (-\delta, \delta)$$

$$\overset{!}{=} \varepsilon \qquad L \approx \|DF(x_0)\|_\infty = \max_{i=1,\dots,k} \|\nabla f_i(x_0)\|_1 =: L_{x_0}$$

$$\|x_0 - y(h)\|_\infty = \|x_0 - x_0 - hp\|_\infty = h$$

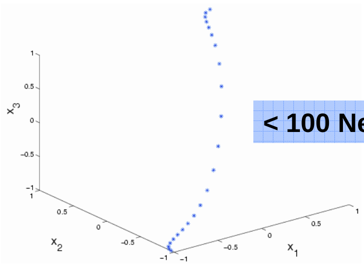
$$h(x_0) \approx \frac{\theta\varepsilon}{L_{x_0}}, \quad \theta \in (0,1)$$

INRIA

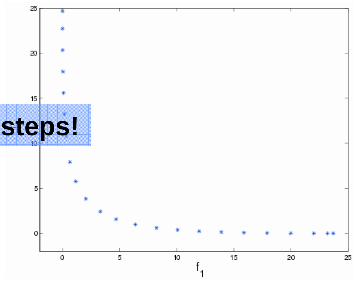
Example (n=3)

$$F : R^n \rightarrow R^2 \qquad F = (f_1, f_2)$$

$$f_i(x) = \sum_{\substack{j=1 \\ j \neq i}}^{n} (x_j - a_j^i)^2 + (x_i - a_i^i)^4$$



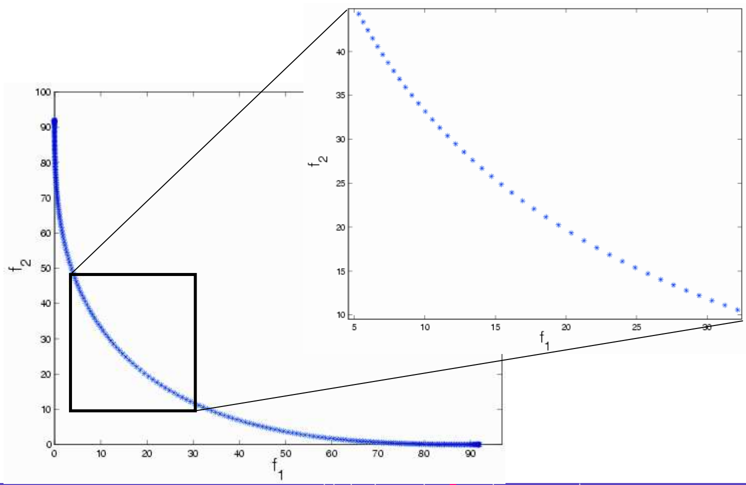
< 100 Newton steps!



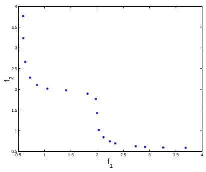
Parameter Space

Image Space

Example (n=20)



Conclusion



Task: archiver for EMO algorithms to maintain finite size gap-free Pareto set approximations.

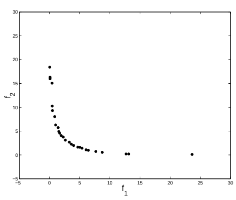
For this we

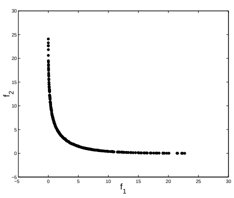
- proposed novel objects according to the problem (e.g., $(\Delta,\Theta\varepsilon_m)$-tight ε–approximate Pareto sets)
- designed two novel archivers
- investigated the limit set (convergence, bounds on the archive) under certain assumptions on the generator
- investigated the practical relevance of the novel archiver (numerical results)

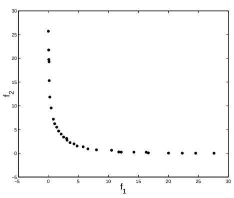
INRIA

Thank You!

Questions?







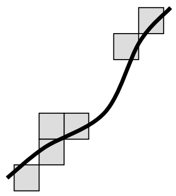
More information:

`http://paradiseo.gforge.inria.fr`

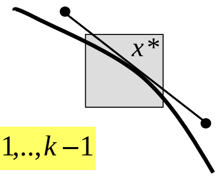
INRIA

Recovering (Smooth MOPs)

Recover-Algorithm: predictor predictor variant



$$\tilde{F}(P) = 0$$



$$x_{start} = c \pm h_i q_i \quad i = 1,..,k-1$$

$$span\{q_1,..,q_{k-1}\} = T(\tilde{F}(x^*,\alpha^*))$$

$$< q_i, q_j > = \delta_{ij}$$

[Rheinboldt; 89], [Hillermeier: 01]
[Sch.;04], [Sch., Dell'Aere, Dellnitz;05]

INRIA

Example 1: 3 Objectives

$$a^1, a^2, a^3 \in R^n$$

$$f_i : R^n \to R$$

$$f_i(x) = \sum_{\substack{j=1 \\ j \neq i}}^{n} (x_j - a_j^i)^2 + (x_i - a_i^i)^4$$

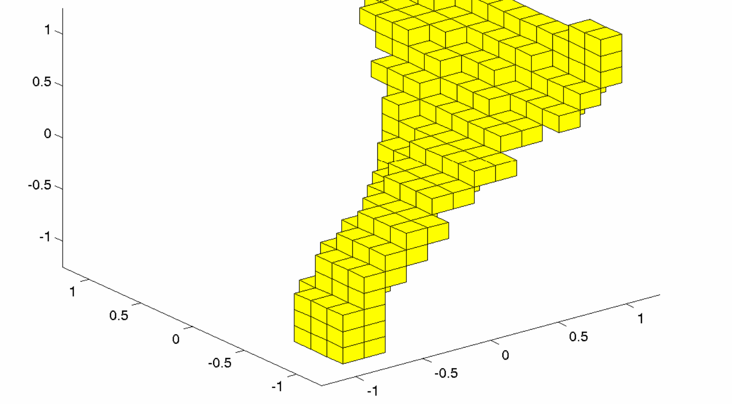
$$F : R^n \to R^3 \qquad F = (f_1, f_2, f_3)$$

INRIA

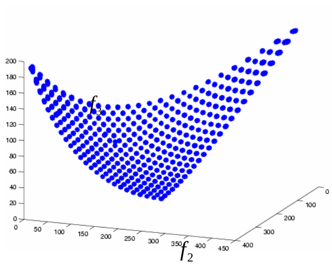
Example 1: n=3

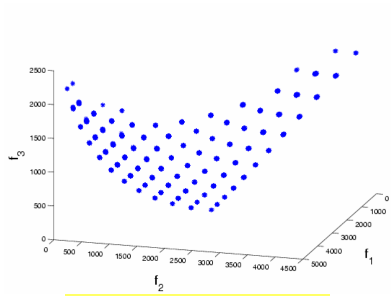
$\tilde{F}^{-1}(0)$

$\tilde{F}: R^6 \rightarrow R^4$



Example 1: n = 100/1000





$\tilde{F}: R^{103} \rightarrow R^{101}$ $\tilde{F}: R^{1003} \rightarrow R^{1001}$

Computational Time:

$\approx$ 1min $\approx$ 5h

N=100	$\approx$ 1.5 sec
N=1000	$\approx$ 15 sec
N=30.000	$\approx$ 11mins

INRIA