Pareto Explorer: A Global/Local Exploration Tool for Many Objective Optimization Problems

Dr. Oliver Schütze, Cinvestav-IPN

Glasgow, Sept. 17, 2015

Outline

Introduction

- Multi- and many objective optimization
- Background

Methods for the Treatment of M(a)OPs

- Pareto Tracer for MOPs
- Pareto Explorer for MaOPs

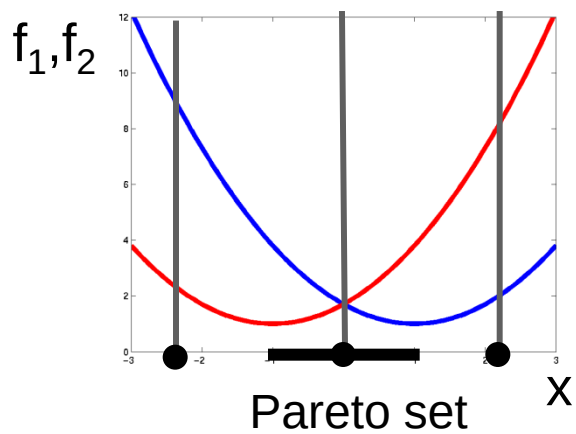
Application

- PE for laundry system

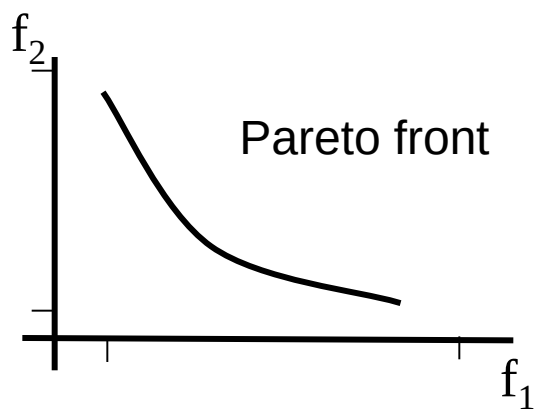
Multi-objective Optimization

Multi-objective Optimization Problem

$$\min \; F = \begin{cases} f_1 : Q \subset R^n \to R \\ \vdots \\ f_k : Q \subset R^n \to R \end{cases} \qquad \textbf{(MOP)}$$



Definition: $x \in Q$ is Pareto optimal $:\Leftrightarrow$ there exists no point $y \in Q \setminus \{x\}$ s.t. $F(y) \leq F(x)$ and $F(y) \neq F(x)$.

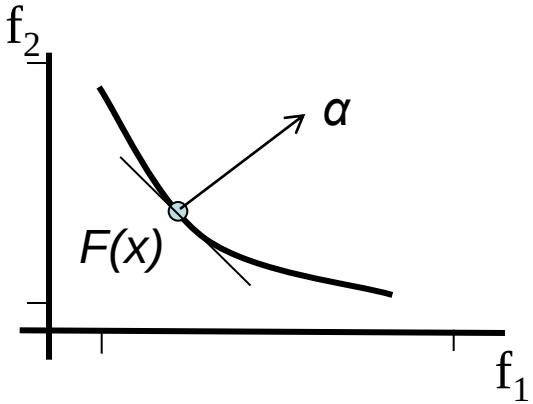


P_Q is a set of dimension (k-1)

$F(P_Q)$ = image of P_Q (***Pareto front***)

Optimality Conditions

Theorem ([Kuhn, Tucker, 1951])
Let x be a Pareto point of (MOP), then:



$$\$a_1,..,a_k \geq 0 : \sum_{i=1}^{k} a_i = 1 \text{ and } \sum_{i=1}^{k} a_i \nabla f_i(x) = 0 = \mathrm{J}^T a, \quad (C)$$

$$\text{where } \mathrm{J} = \begin{pmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_k(x)^T \end{pmatrix} \in R^{k \times n} \text{ denotes the Jacobian of } F \text{ at } x.$$

Remark α is orthogonal to the linearized Pareto front at F(x)

Definition If x∈Q satisfies *(C)*, then it is called a *Karush Kuhn Tucker* (KKT) point.

The Pareto Set is a Manifold

$$\tilde{F} : R^{n+k} \rightarrow R^{n+1}$$

$$\tilde{F}(x,\alpha) = \begin{pmatrix} \sum_{i=1}^{k} \alpha_i \nabla f_i(x) \\ \sum_{i=1}^{k} \alpha_i - 1 \end{pmatrix}$$

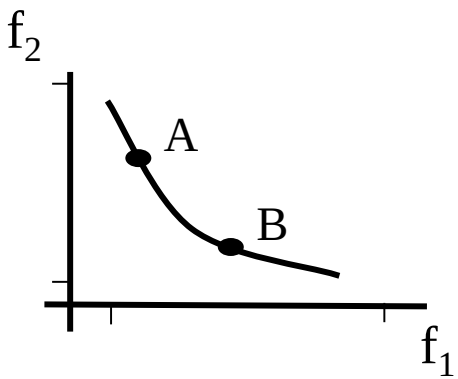
$\Rightarrow$ **[Kuhn,Tucker 1951]**

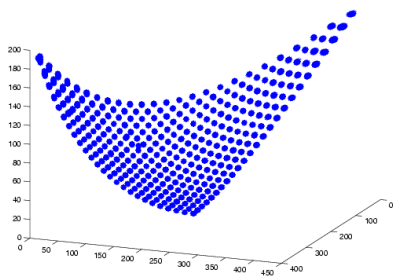
The Pareto set is contained in $\tilde{F}^{-1}(0)$

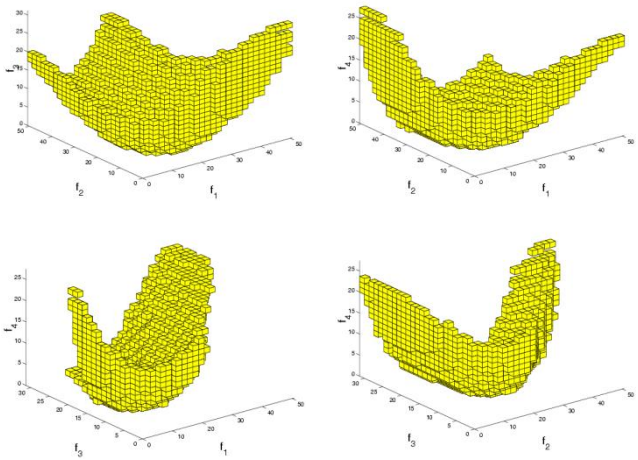
$\Rightarrow$ **[Hillermeier 2001]**

Pareto set: (*k-1*)-dim manifold

Impact on Decision Making and Algorithm Design





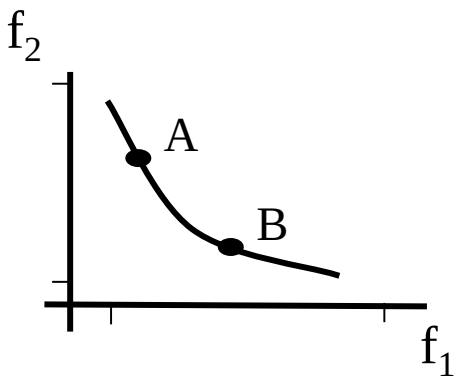


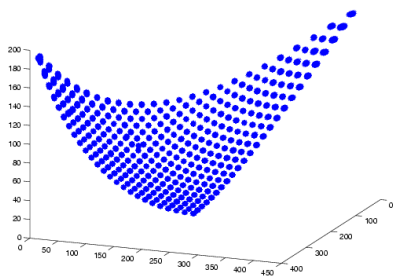
k=2 k=3 k=4

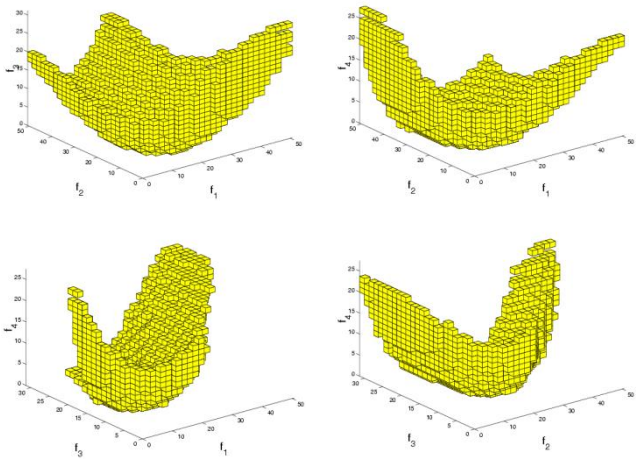
Finite size approximation of the entire Pareto set/front: exponential growth in k ($N=m^{k-1}$)

Example: 3 points per dimension, k=14 → $N=3^{13}\approx 1{,}600{,}000$

Impact on Decision Making and Algorithm Design







k=2 k=3 k=4

Definition:

- $k \leq 3$ multi-objective optimization problem (MOP)
- $k > 3$ many objective optimization problem (MaOP)

Steering the Search in MOO

Given point $x \in R^n$, direction $v \in R^n$, step size $t \in R_+$.
New candidate found via line search

$$x_{new} = \mathrm{x} + \mathrm{t}v$$

Then the according movement in objective space (for infinitesimal step sizes) is given by

$$\mathrm{J}v$$

"Proof"

$$\lim_{t \to \infty} \frac{f_i(x + tv) - f_i(x)}{t} = < \nabla f_i(x), v > = (Jv)_i$$

Methods for the Treatment of M(a)OPs

Mathematical Programming Techniques

- Scalarization methods
- Descent methods
- Interactive methods

Set based methods

- Evolutionary strategies
- Subdivision techniques
- Cell mapping techniques

Continuation Methods

- Perform a movement along the Pareto set/front (done by MP and SON methods)

Scalarization Methods

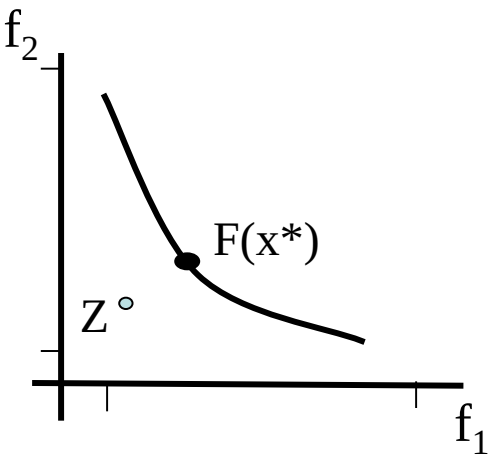
Scalarization methods transform the given MOP into a 'suitable' SOP
→ *one* solution expected, according to the given setting

Examles:

- Weighted sum method
- ε-Constraint method
- Reference point methods
- Normal boundary intersection (NBI)
- Pascoletti-Serafini
- …

Example: Reference Point Methods

If the Decision Maker has a rough idea about his/her product, one can look for the solution which is 'closest' to a given reference point Z (defined in objective space).



$$\min_{x} d(F(x), Z) \quad (*)$$

But: does the solution of (*) give us (exactly) what we need?
Note that the solution depends on

Similar problems with all scalarizing functions (e.g., ASF)

- The shape of the Pareto front.

Predictor Corrector (PC) Methods

Given : $H : R^{N+1} \to R^{N}, \quad H(x) \overset{!}{=} 0$

$x_0 \in R^{N+1}$ s.t. $H(x_0) = 0, \; rk\left(H'(x_0)\right) = N$

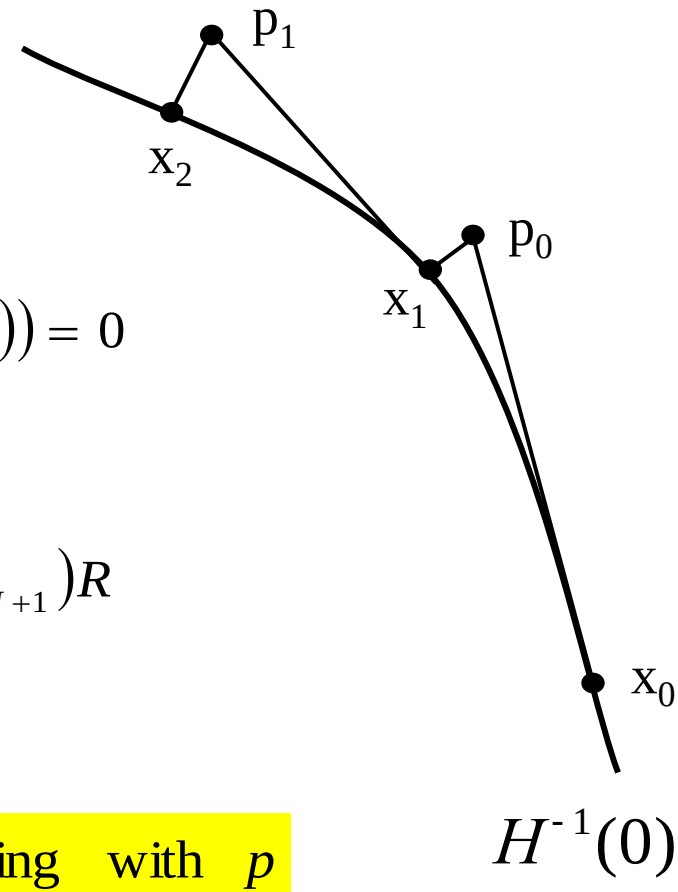
IFT : $\exists \; c : (-\varepsilon, \varepsilon) \to R^{N+1} \quad c(0) = x_0 + H(c(s)) = 0$

Differentiation : $H'(c(s)) \cdot c'(s) = 0$

Orthogonalization : $H'(x_0) = QR = (q_1 \cdots q_{N+1})R$

Predictor : $p = x_0 + t_0 q_{N+1}$

Corrector : Newton method on $H(x) = 0$ starting with p

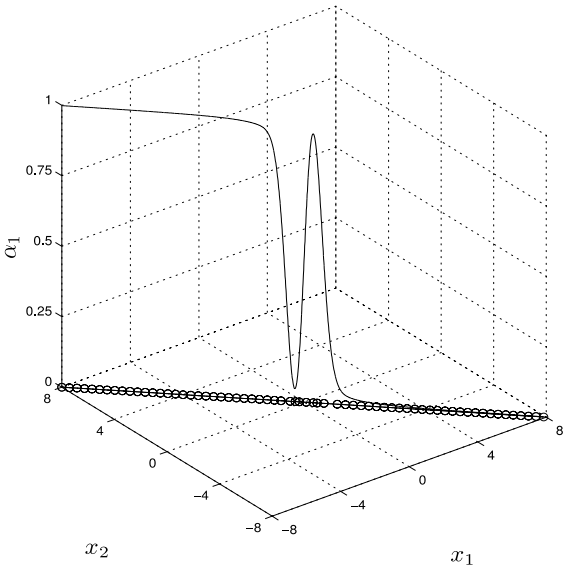


PC Method of Hillermeier

[Hillermeier, 2001] Use PC method on

$$\tilde{F} : R^{n+k} \to R^{n+1}$$

$$\tilde{F}(x,\alpha) = \begin{pmatrix} \sum_{i=1}^{k} \alpha_i \nabla f_i(x) \\ \sum_{i=1}^{k} \alpha_i - 1 \end{pmatrix}$$



Possible drawbacks

- Consideration of compound (x,α) space may increase nonlinearity
- No treatment of inequality constraints discussed

Pareto Tracer [Martin, Sch., 2014]

Basic Idea:
Use PC method as above, but separate decision (x) and weight (α) space whenever possible

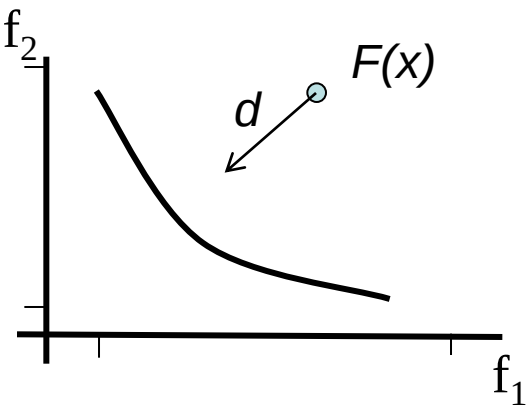
P Directly compute kernel vectors of F-tilde and utilize ‘steering’ in objective space
C Use Newton method for MOPs by **[Fliege & Svaiter, 2009]**

Pareto Tracer: Predictor (1/3)

Kernel vector of $\tilde{F}'$

$$\tilde{F}'(x,\alpha)\begin{pmatrix}\upsilon\\ \mu\end{pmatrix}=\begin{pmatrix}H_\alpha & J^T\\ 0 & 1\cdots 1\end{pmatrix}\begin{pmatrix}\upsilon\\ \mu\end{pmatrix}\overset{!}{=}\begin{pmatrix}0\\ 0\end{pmatrix}$$

$$\Rightarrow \begin{cases}\sum_{i=1}^{k}\mu_i=0 & \text{(equation II)}\\ H_a\nu=-J^T\mu & \text{(equation I)}\end{cases}$$



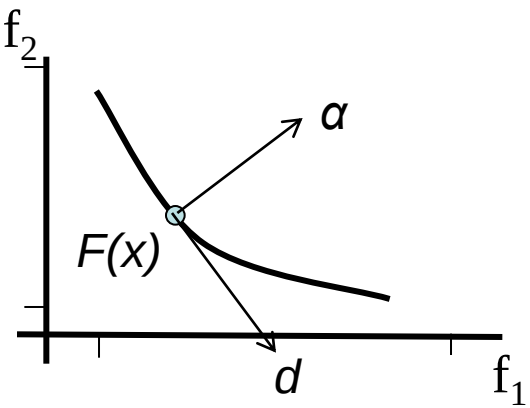
Steer the search in direction $d\in R^k$: compute $\nu\in R^n$ s.t. $Ju=d$

$$\begin{matrix} I: u_d=-JH_a^{-1}J^T\mu_d \\ II: \quad \sum\mu_i=0 \end{matrix}\Rightarrow \begin{pmatrix}-JH_a^{-1}J^T & \\ 1\cdots 1 & \end{pmatrix}\begin{pmatrix}\mu_d\\ \end{pmatrix}=\begin{pmatrix}d\\ 0\end{pmatrix}$$

$$\Rightarrow u_d=-JH_a^{-1}J^T\mu_d$$

Pareto Tracer: Predictor (2/3)

Task: compute d duch that a movement along the *linearized* Pareto front is performed.



x KKT point $\$a \in R^k$ s.t. $J^T a = 0$

$$a = QR = (q_1 \cdots q_k)\begin{pmatrix} r_{11} & \cdots \\ \vdots & \vdots \\ 0 & \cdots \end{pmatrix}$$

$a = r_{11} q_1 \in span\{q_1\}$ $\Rightarrow$ $\{q_2, \ldots, q_k\}$ forms an orthonormal basis (ONB) of $a^\perp$

$\Rightarrow$ choose directions $d_i := q_{i+1},\ i = 1, .., k-1$

and use u_{d_i} as predictors

Pareto Tracer: Predictor (3/3)

Theorem: Let x be a KKT point of *(MOP)* with weight vector $\alpha>0$. Further, let $\nabla f_i(x)\neq 0$, i=1,..,k, rank(J)=k-1 and H_α be regular. Let d_i and μ_d as above, then

$$u_{\eta_i} = -JH_\alpha^{-1}J^T \eta_{d_i}, \quad i=1,..,k-1$$

point along the linearized Pareto set at x and the vectors

$$Ju_{\eta_i}, \quad i=1,..,k-1$$

form an ONB of the linearized Pareto front at F(x).

Pareto Tracer: Corrector

[Fliege & Svaiter, 2009]

Idea: given x and (MOP), compute the Newton direction υ via solving

$$\max_{u} \min_{i=1,..,k} \nabla f_i(x)^T u + \frac{1}{2} u^T \nabla^2 f_i(x) u$$

Remark: necessary condition for k=1: $\nabla^2 f(x) u + \nabla f(x) = 0$

$$\min_{u,t} \quad t$$

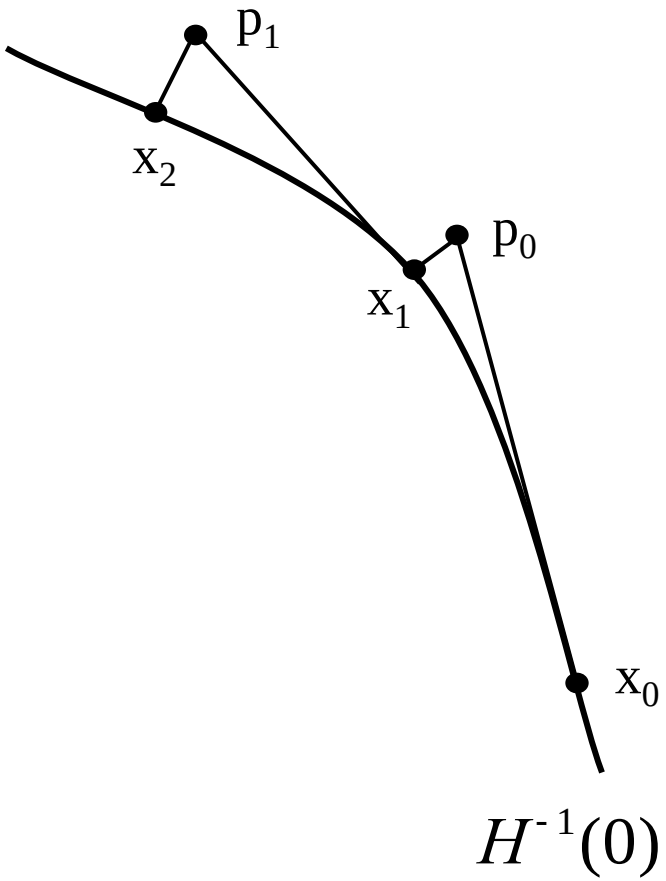
$$\text{s.t.} \quad \nabla f_i(x)^T u + \frac{1}{2} u^T \nabla^2 f_i(x) u \leq 0, \quad i = 1,..,k$$

Hessian Free Realization

Successive update of the Hessians via quasi Newton strategies, i.e., approximation of H_α by

$$B_\alpha = \sum_{i=1}^{k} \alpha_i B_i, \text{ where } B_i \approx \nabla^2 f_i(x)$$

→ Only gradient information required

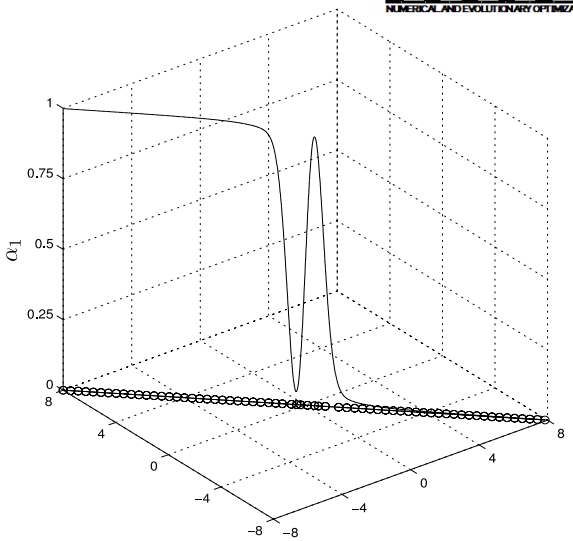


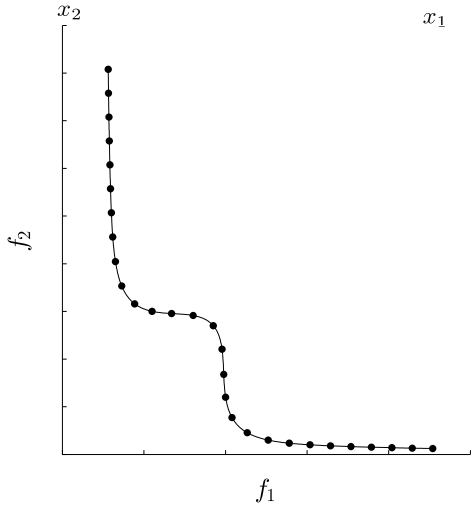
Example PT1: Unconstr. MOP

$$f_1(x) = \frac{1}{2}\left(\sqrt{1+(x_1+x_2)^2} + \sqrt{1+(x_1-x_2)^2} + x_1 - x_2\right) + l\, e^{-(x_1-x_2)^2}$$

$$f_2(x) = \frac{1}{2}\left(\sqrt{1+(x_1+x_2)^2} + \sqrt{1+(x_1-x_2)^2} - x_1 + x_2\right) + l\, e^{-(x_1-x_2)^2}$$

	Hillermeier	PT-N	PT-QN	PT-SD
Solutions	65.0000	63.0000	63.0000	63.0000
Avg. corrector iterations	2.3594	0.0000	0.0000	0.0000
Function evaluations	65.0000	63.0000	63.0000	63.0000
Jacobian evaluations	216.0000	63.0000	63.0000	63.0000
Hessian evaluations	432.0000	126.0000	-	-

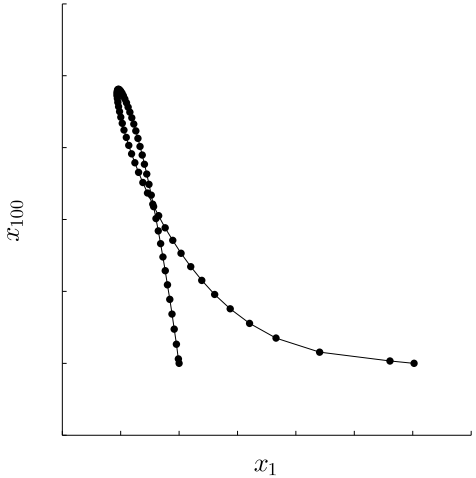




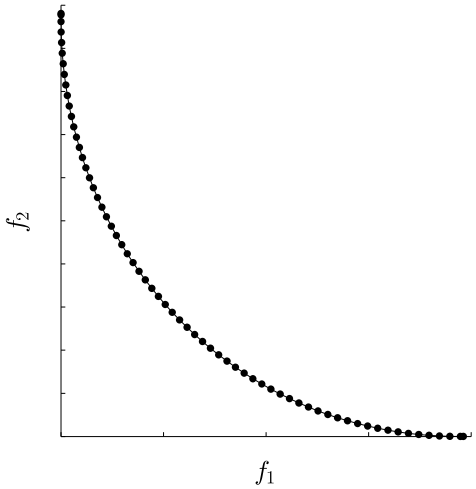
PT2: Eq. Constr.MOP, n=100

$$f_j(x) = \sum_{i=1, i \neq j}^{100} (x_i - a_i^j)^2 + (x_j - a_i^j)^4, \quad j = 1, 2$$

$$s.t. \qquad \|x - c\|^2 = r^2$$



	Hillermeier	PT-N	PT-QN	PT-SD
Solutions	63.0000	64.0000	64.0000	63.0000
Avg. corrector iterations	3.1746	1.8254	1.9688	4.7778
Function evaluations	64.0000	179.0000	192.0000	365.0000
Jacobian evaluations	264.0000	179.0000	191.0000	365.0000
Hessian evaluations	528.0000	358.0000	-	-

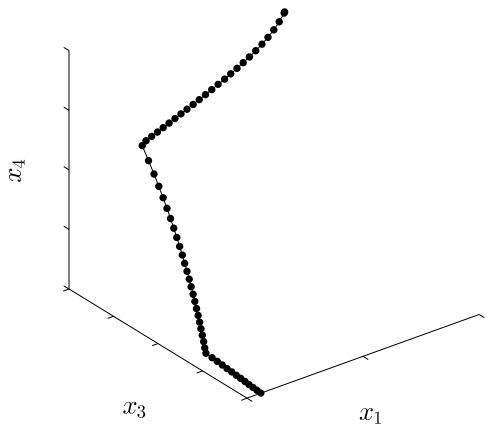


PT3: Eq. + Box Constr. MOP

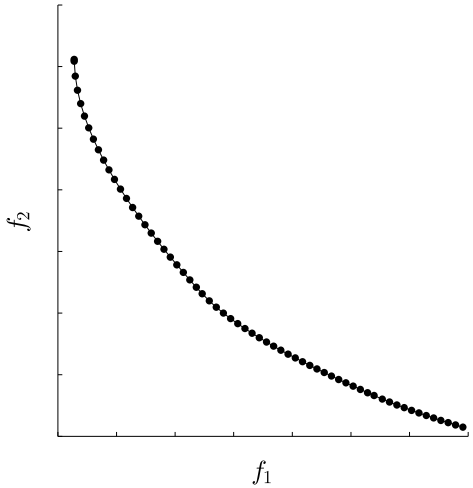
$$f_1(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2$$

$$f_2(x) = 3x_1 + 3x_2 - \frac{x_3}{3} + 0.01(x_4 - x_5)^3$$

$$\begin{aligned} s.t. \quad & x_1 + 2x_2 - x_3 - 0.5x_4 + x_5 = 2 \\ & 4x_1 - 2x_2 + 0.8x_3 + 0.6x_4 + 0.5x_5^2 = 0 \\ & x_i \hat{\mathrm{I}} \ [-2, 2], i = 1, \ldots, 5 \end{aligned}$$

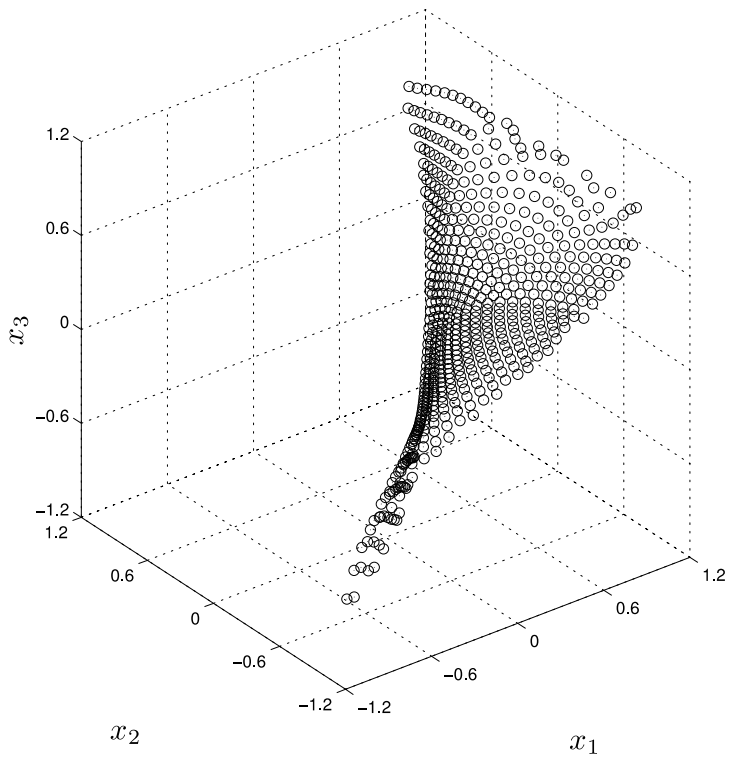


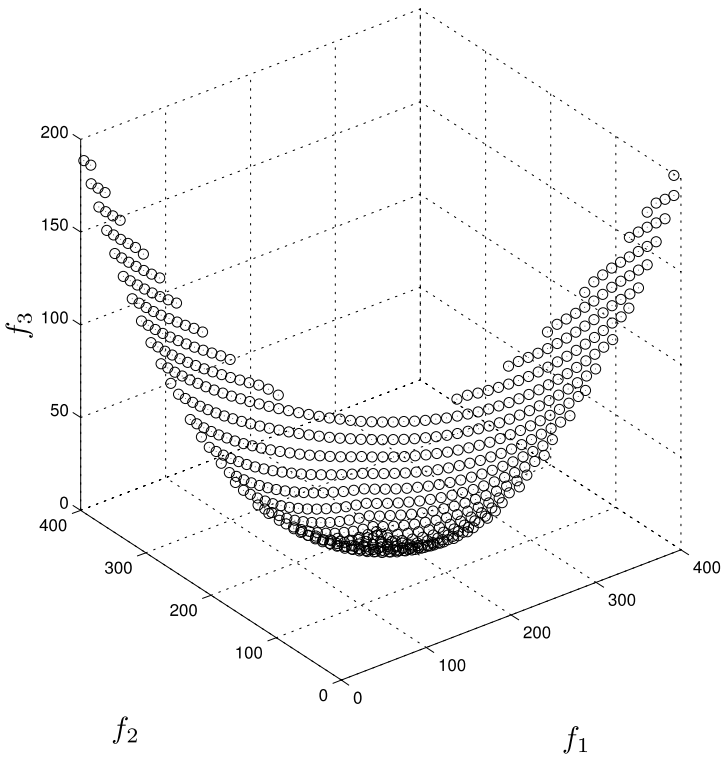
	PT-N	PT-QN	PT-SD
Solutions	62.0000	62.0000	62.0000
Avg. corrector iterations	1.0323	1.0323	2.3968
Function evaluations	128.0000	132.0000	219.0000
Jacobian evaluations	127.0000	127.0000	215.0000
Hessian evaluations	254.0000	-	-



PT4: Unconstr. MOP, k=3

$$f_j(x) = \sum_{i=1, i \neq j}^{100} (x_i - a_i^j)^2 + (x_j - a_i^j)^4, \quad j = 1, 2, 3$$



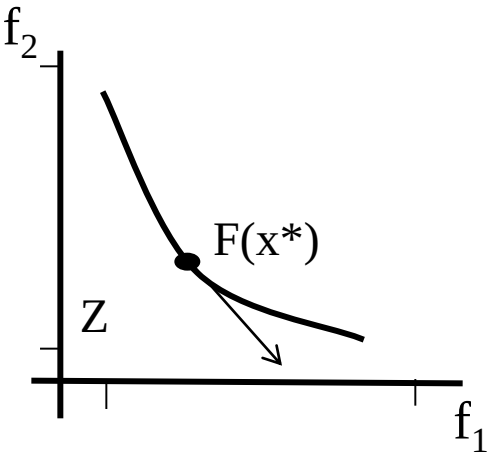


PT-QN: 4109 F, 2504 J

Pareto Explorer

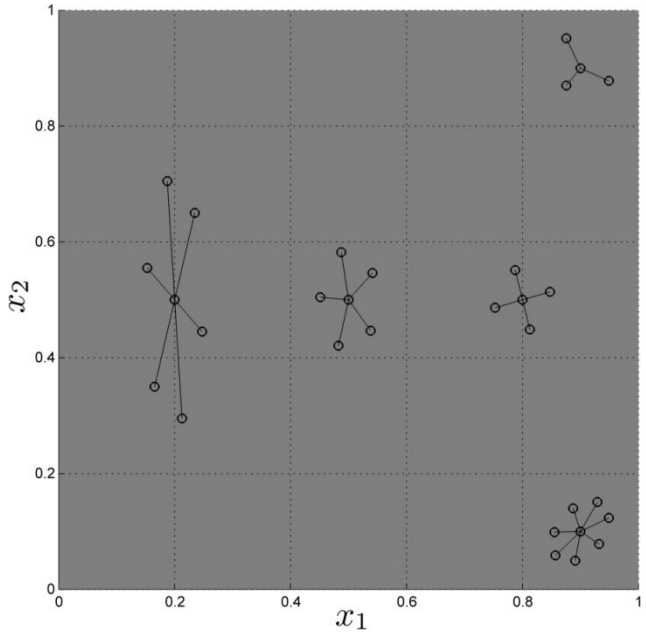
Basic Idea:
Global/local exploration of the performance landscape via

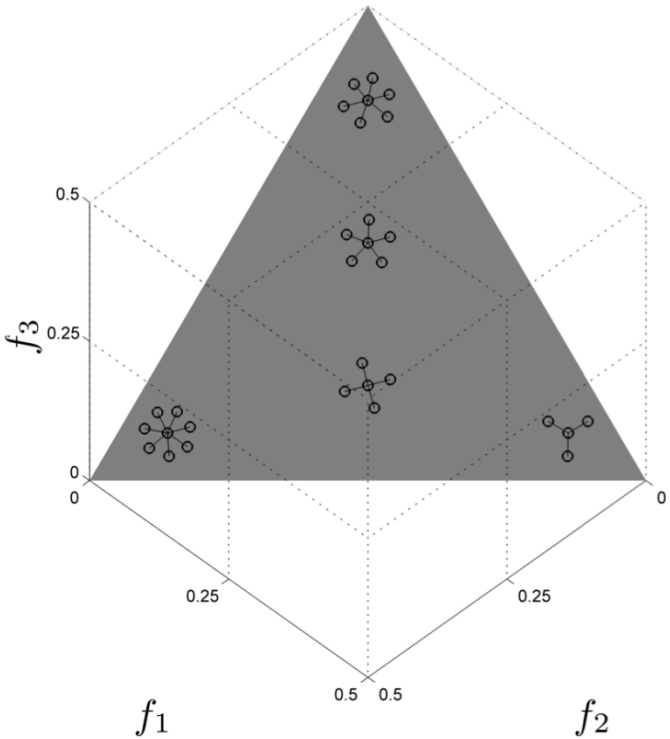
1. Compute set of initial candidate solutions (via global method)
2. Explore solution manifold locally: starting from a given (optimal) solution, steer the search along the Pareto set/front into given user specified directions (decision space, objective space, weight space, dynamic reference points)



Neighborhood Exloration

Example: bias free neighborhood exploration in objective space for DTLZ2

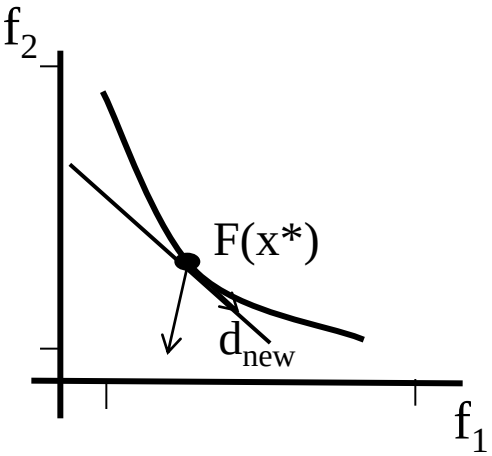




PE: Steering in Objective Space

Given: x KKT point, $a \in R^k$ s.t. $J^T a = 0$

d_y : user specified direction in obj. space



$$a = QR = \begin{pmatrix} q_1 & \cdots & q_k \end{pmatrix} \begin{pmatrix} r_{11} & \cdots \\ 0 & \cdots \end{pmatrix}$$

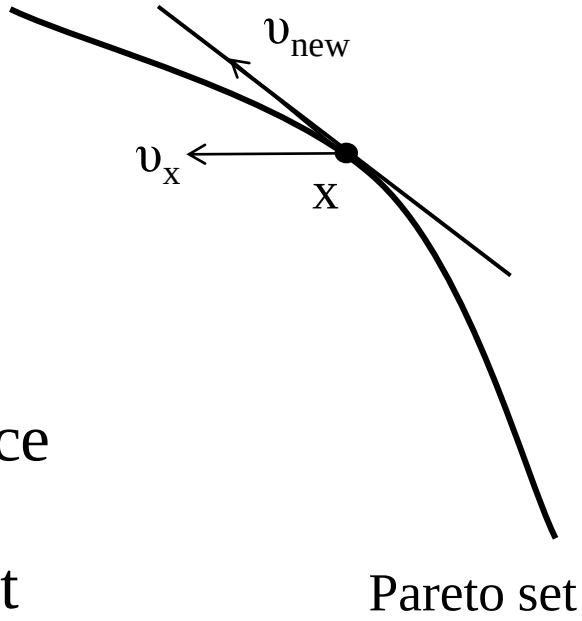
$$Q_2 := \begin{pmatrix} q_2 & \cdots & q_k \end{pmatrix}$$

$d_{new} = Q_2 Q_2^T d_y$ orthogonal projection onto linearized PF

→ apply Pareto Tracer using d_{new}

PE: Steering in Decision Space (1)

Example: $x = \begin{pmatrix} 100 \\ 1 \\ 2 \end{pmatrix} \Rightarrow$ reduce x_1



Given: x KKT point

u_x : user specified direction in decision space

$Q_x = \begin{pmatrix} u_1 & \cdots & u_{k-1} \end{pmatrix}$ ONB of lin. Pareto set

$u_{new} = Q_x Q_x^T u_x$ orth. proj. onto lin. Pareto set

→ apply Pareto Tracer using υ_{new}

PE: Steering in Decision Space (2)

Setting Decision Maker is satisfied with objective value F(x)

Possible solution find "best" direction v under the constraint that the movement in objective space is minimal ($Jv = 0$)

→ Predictor direction is solution of

$$\min_{v} \; -< v, d_x > \\ s.t. \quad Jv = 0 \\ \quad ||v|| = 1$$

Example: Laundry System

Simplified model for a laundry system

Objectives:

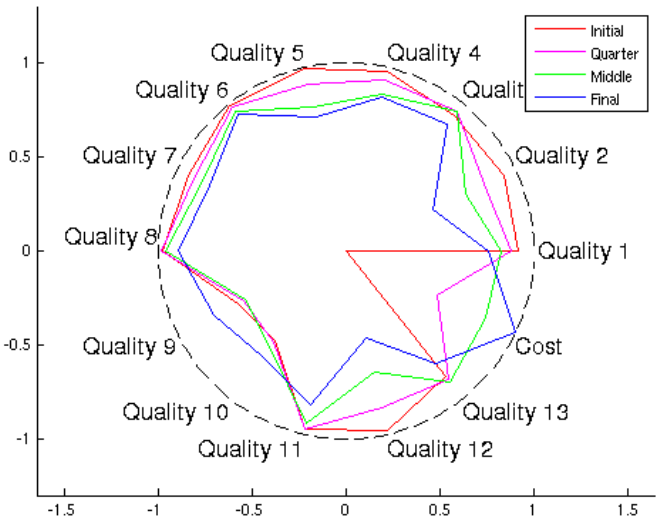
1. Wool grease type A
2. Wool grease type B
3. Red wine
4. Sebum type A
5. Sebum type B
6. Curry
7. Motor oil
8. Petroleum
9. Blood
10. Egg
11. Starch
12. Vegetable fat
13. Cocoa
14. Cost

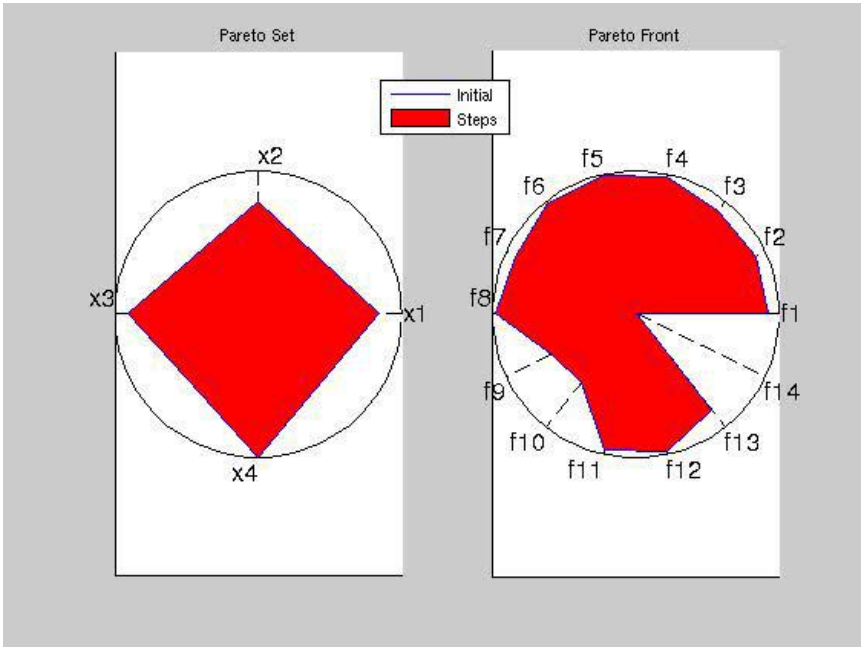
Parameters:

1. Temperature
2. Amount of cleaner
3. Washing time
4. Frequency of the rotating unit

Laundry System: Steer in Obj. Space

Start: high quality solution (high cost)
Steering: reduce cost ($=f_{14}$) → $d_y=-e_{14}$



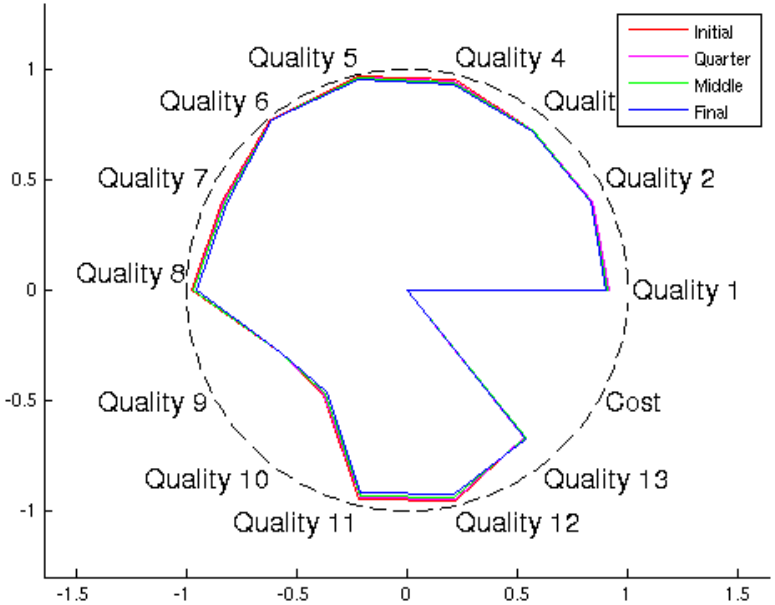


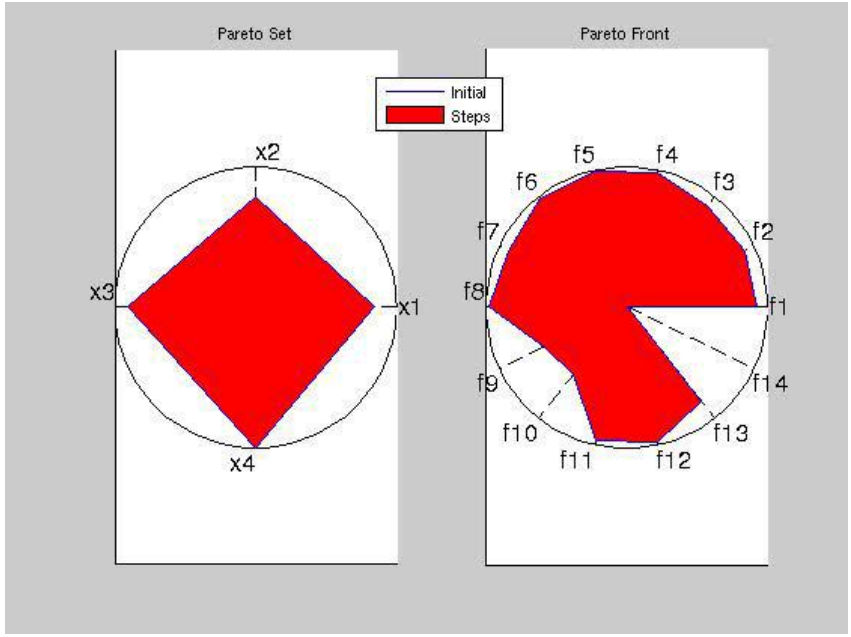
Laundry System: Steer in Dec. Space

Interpretation:

$x_3(t_0=0)=1.45$, $x_3(t_{final})=0.73$

→ Washing time can be reduced by approx. 50% without (significantly) changing the washing quality





Conclusions and Future Work

Tools

- MOPs: Pareto Tracer
- MaOPs: Pareto Explorer

Future Work

Questions?

- hybrids of EAs and continuation methods
- evolutionary continuation-like strategies
- applications to problems related to SMD!