

Multi-objective Bayesian optimization with reference point – sequential and parallel versions

David Gaudrie^{1,2}, Rodolphe Le Riche¹, Victor Picheny³,
Benoît Enaux², Vincent Herbert²

¹ CNRS LIMOS at Mines Saint-Etienne, France
² PSA ³ Prowler.io

20 March 2019

GDR Mascot-Num annual conference
IFPEN, Rueil-Malmaison, France

Goal: multi-objective optimization (1)

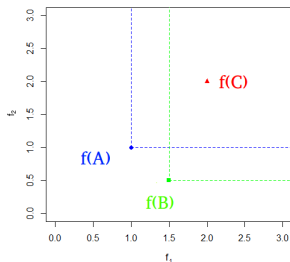
In general there are many (even an infinite number of) trade-off solutions to

$$\min_{\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^d} (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$$

called the Pareto set (in \mathcal{X}) or front (in \mathcal{Y}).

It is composed of Non-Dominated points,

$$\{\mathbf{x} \in \mathcal{X} : \nexists \mathbf{x}' \neq \mathbf{x} \in \mathcal{X}, \forall i f_i(\mathbf{x}') \leq f_i(\mathbf{x}) \ \& \ \exists j f_j(\mathbf{x}') < f_j(\mathbf{x})\}.$$



C is dominated,
 A and B non-dominated

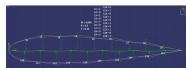
Notation:

$A \prec C, B \prec C$

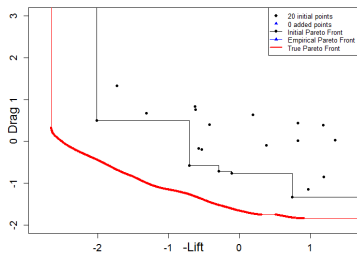
Goal: multi-objective optimization (2)

True Pareto front vs. empirical Pareto front.

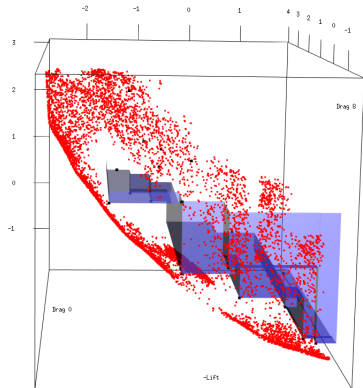
Examples from the metaNACA test bed [6], $\mathbf{x} :=$



$m = 2$ objectives



$m = 3$ objectives

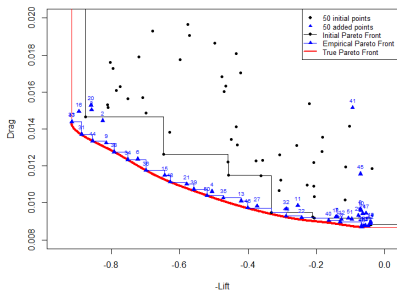


The Pareto fronts can have holes.

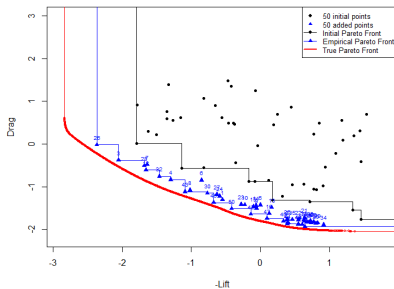
Curse of dimensionality: number of variables

At a given budget, optimization performance degrades with the number of variables:

$d = 3$



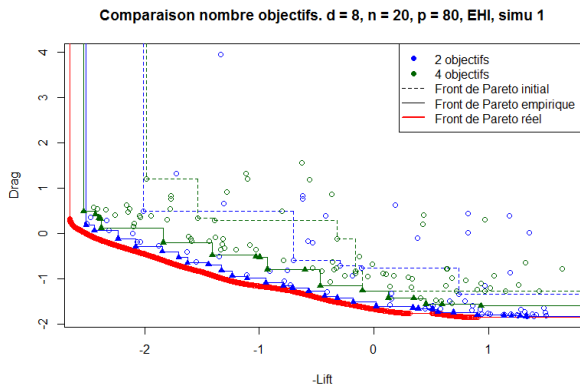
$d = 22$



(optimization algorithm: EHI – Emmerich et al. [3] – with GPareto – Binois and Picheny [1])

Curse of dimensionality: number of objectives (1)

At a given budget, optimization performance degrades with the number of objectives:



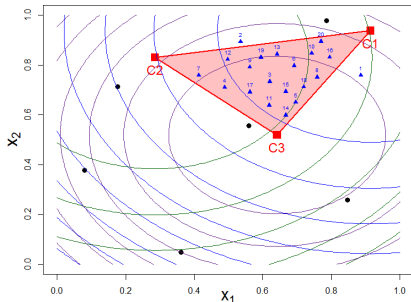
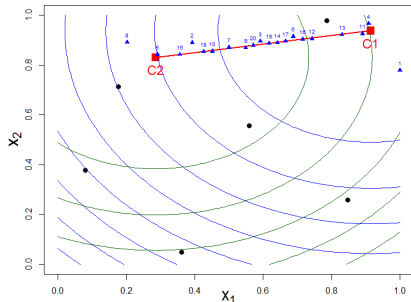
(optimization algorithm: EHI – Emmerich et al. [3] – with GPareto – Binois and Picheny [1])

Curse of dimensionality: number of objectives (2)

As the number of objectives increases, a larger part of \mathcal{X} becomes Pareto optimal:

$m = 2$

$m = 3$



Ex: sphere functions centered on C1, C2, C3. Pareto sets (in red) are all convex combinations of the C's. Blue triangles: points sampled by MO Bayesian algorithm (GPareto). With 4 objectives at the corners of \mathcal{X} , every point could be a Pareto solution. As the Pareto set becomes larger, the optimization algorithm degenerates in a space filling algorithm. Give up the utopian search for all of the Pareto set.

Restricting ambitions in MOO of costly functions

Metamodels of costly functions does not solve everything: they still need to be learned. Recently, we have explored ways to proportionate ambitions to search budget (about 100 functions evaluations):

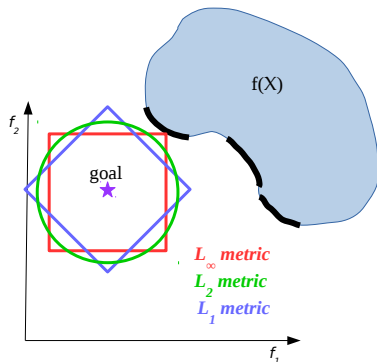
- **Today's talk:** how to focus on specific regions of the Pareto front. The R-mEI algorithm explained step by step:
 - 1 Finding one Pareto optimal point
 - 2 Widening the search
 - 3 A q -points batch versionDetails and proofs in [6, 7].
- **See also:** variable dimension reduction:
 - variables selection for optimization ($m = 1$). Work with Adrien Spagnol and Sébastien Da Veiga, Safran Tech, cf. [16].
 - problem reformulation (eigenshapes), cf. other presentation by David Gaudrie here.

MOO: related work (1)

- Evolutionary Multi-objective Optimization (EMO): a field in itself (10 EMO international conferences, Deb's book [12]), expensive without model of the function.
- Model-based multi-objective optimization:
 - gradient on Gaussian Process (GP) mean (Zerbinati et al. [19]),
 - the family of Bayesian MOO (EHI, SMS, SUR, EMI – GPareto [1], Wagner et al. [17] –), constrained EHI (Feliot et al. [4]).
 - They target the entire front, cheaper than EMO but still expensive for us (curse of criteria dimensionality, see earlier).

MOO: related work (2)

- User preference: scalarize the MOO by minimizing a distance to a user given goal, $\min_{\mathbf{x} \in \mathcal{X}} \text{dist}(\mathbf{f}(\mathbf{x}), \text{goal})$ (Miettinen [15]). But choice of the metric, choice of the goal.

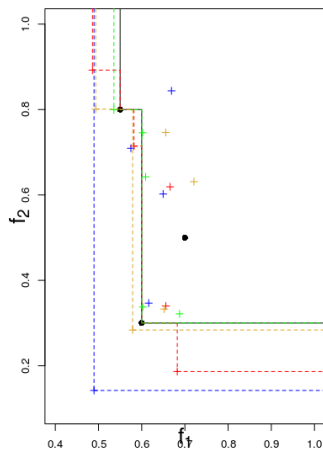
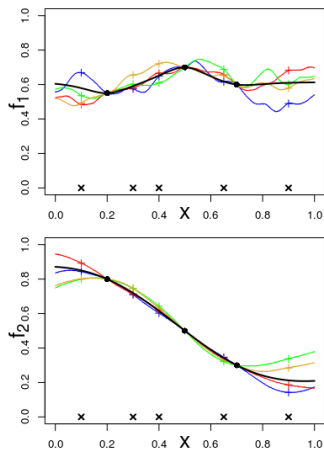


- User preference in Bayesian optimization: weighted EHI (Feliot et al. [5]), truncated EHI (Yang et al. [18]). But: need to specify the weight or the truncation region.
- ⇒ we now propose a Bayesian MOO with or without user preference (reference point and its default), R/C-mEI.

1. FINDING ONE PARETO-OPTIMAL POINT

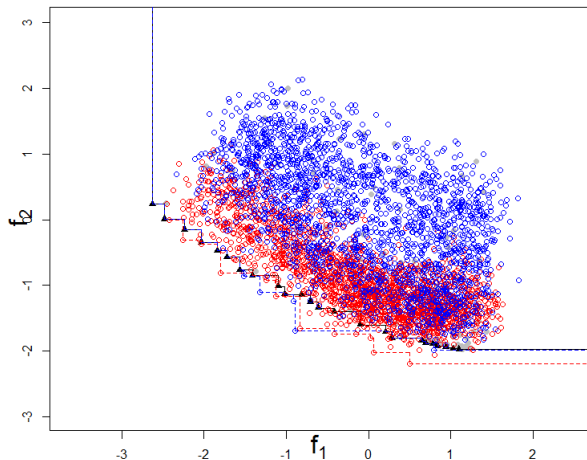
Bayesian multi-objective optimization (1)

Equipped with observations of the true functions and GPs, we can simulate possible Pareto fronts at given \mathbf{x} 's:



Simulations points for the Pareto front (1)

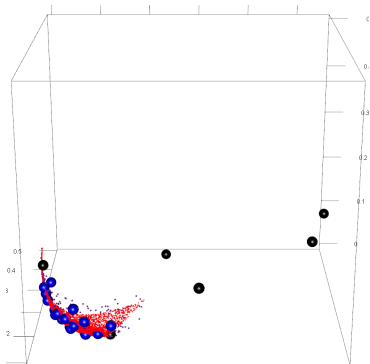
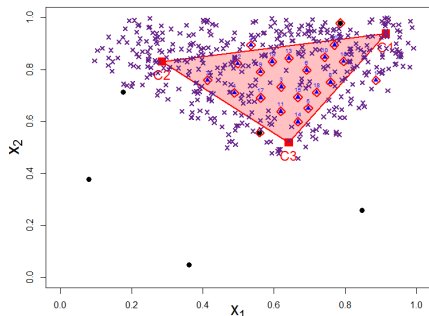
The choice of the \mathbf{x} 's where the simulations are performed matters. Below, blue points are random, red points selected proportionally to their probability of being non dominated by the empirical front $\widehat{\mathcal{P}}_y$:



($d = 22$)

Simulations points for the Pareto front (2)

Choose \mathbf{x} 's with a probability proportional to $\mathbb{P}(\widehat{\mathcal{P}}_{\mathbf{Y}} \not\subseteq \mathbf{Y}(\mathbf{x}))$.
In the quadratic case, $d = 2$, $m = 3$, after 20 iterations

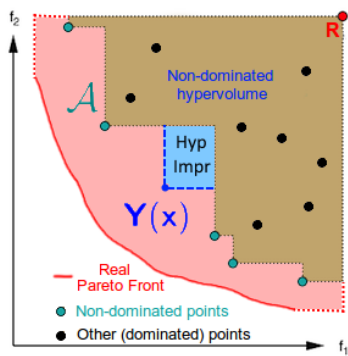


DoE \cdot & \bullet & \bullet , Non Dominated \diamond , selected point \times , sampled \mathbf{Y} ND \cdot .

Simulation points are uniformly distributed near the Pareto set.

Bayesian multi-objective optimization (1)

Where to put the next point, \mathbf{x}^{n+1} , where to call the costly \mathbf{f} ? At the point that maximizes, on the average of the $\mathbf{Y}(\mathbf{x})$ samples, the *Hypervolume Improvement* (over the empirical Pareto front $\widehat{\mathcal{P}}_{\mathbf{y}}$):



EHI implementation from [14]

$$H(\mathcal{A}; \mathbf{R}) = \bigcup_{\mathbf{y} \in \mathcal{A}} \int_{\mathbf{y} \preceq \mathbf{z} \preceq \mathbf{R}} dz$$

$$I_H(\mathbf{f}; \mathbf{R}) = H(\widehat{\mathcal{P}}_{\mathbf{y}} \cup \{\mathbf{f}\}; \mathbf{R}) - H(\widehat{\mathcal{P}}_{\mathbf{y}}; \mathbf{R})$$

$$\text{EHI}(\mathbf{x}; \mathbf{R}) = \mathbb{E}(I_H(\mathbf{Y}(\mathbf{x}); \mathbf{R}))$$

EHI favors $\mathbf{Y}(\mathbf{x})$ dominating the empirical Pareto front and far from already observed $\mathbf{f}(\mathbf{x}^i)$ s.

Notice: the omnipresence of \mathbf{R} ; $\forall \mathbf{R} \succeq \widehat{\mathcal{P}}_{\mathbf{y}}$, EHI generalizes the EI criterion of EGO (Jones [10]).

Bayesian multi-objective optimization (2)

Algorithm 1 Multi-objective EHI Bayesian optimizer

Require: DoE = $\{(\mathbf{x}^1, \mathbf{f}(\mathbf{x}^1)), \dots, (\mathbf{x}^n, \mathbf{f}(\mathbf{x}^n))\}$, \mathbf{R} , n^{\max}

- 1: **while** $n < n^{\max}$ **do**
 - 2: Build m independent GPs, $\mathbf{Y}() = (Y_1(), \dots, Y_m())$, from current DoE
 - 3: Find next iterate by solving $\mathbf{x}^{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \text{EHI}(\mathbf{x}; \mathbf{R})$
{internal optimization problem, no call to $\mathbf{f}()$ }
 - 4: Calculate $\mathbf{f}(\mathbf{x}^{n+1})$
 - 5: $n \leftarrow n + 1$
 - 6: **end while**
-

Targeting improvement regions with EHI

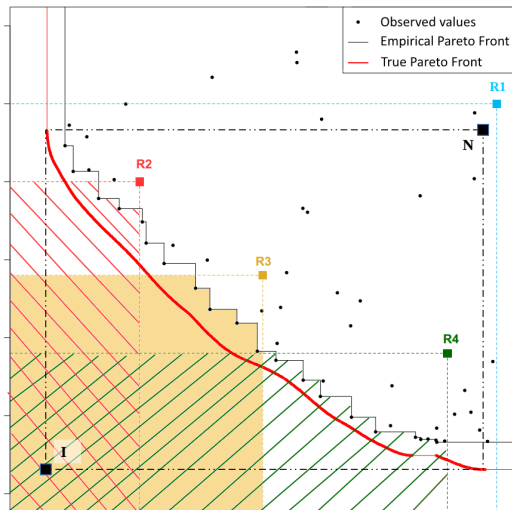
To find the entire Pareto front, \mathbf{R} must be dominated by the Nadir point, \mathbf{N} : **R1** is the default in the literature.

But the entire Pareto front is *i)* too large to be described *ii)* not interesting in general (e.g., extreme solutions).

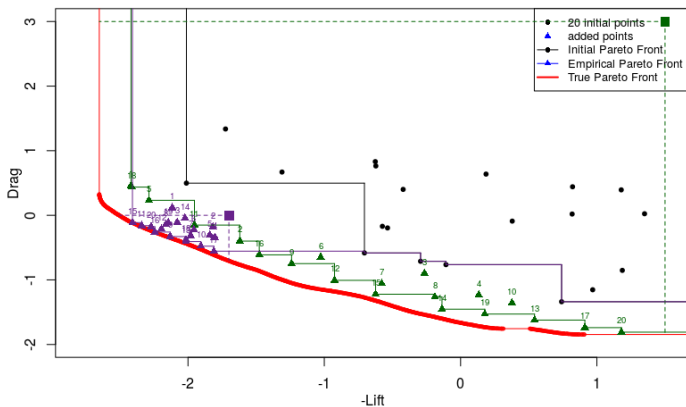
\Rightarrow move \mathbf{R} and control the *improvement region*,

$$\mathcal{I}_{\mathbf{R}} := \{\mathbf{y} \in \mathcal{Y} : \mathbf{y} \preceq \mathbf{R}\}$$

(keeps the Pareto rank for non comparable functions)



Example: targeted EHI versus EHI



Violet: mEI convergence with \mathbf{R} at ■.

Green: EHI convergence with \mathbf{R} at ■.

Note the more local and accurate convergence with mEI. ($d = 8$)

mEI, a computationally efficient proxy to EHI

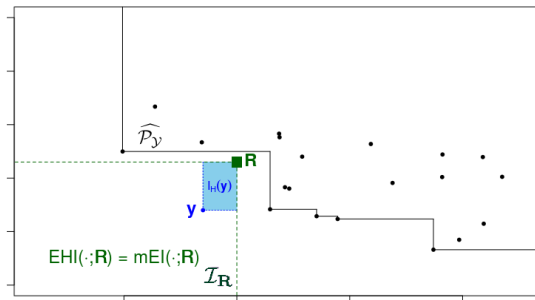
Once \mathbf{R} is freed from $\widehat{\mathcal{P}}_y$, a new acquisition criterion is possible.

Definition:

$$\text{mEI}(\mathbf{x}; \mathbf{R}) := \prod_{j=1}^m \text{EI}_j(\mathbf{x}; R_j) \stackrel{Y\text{'s indep.}}{=} \mathbb{E} \prod_{j=1}^m \max(0, R_j - Y_j(\mathbf{x}))$$

Property:

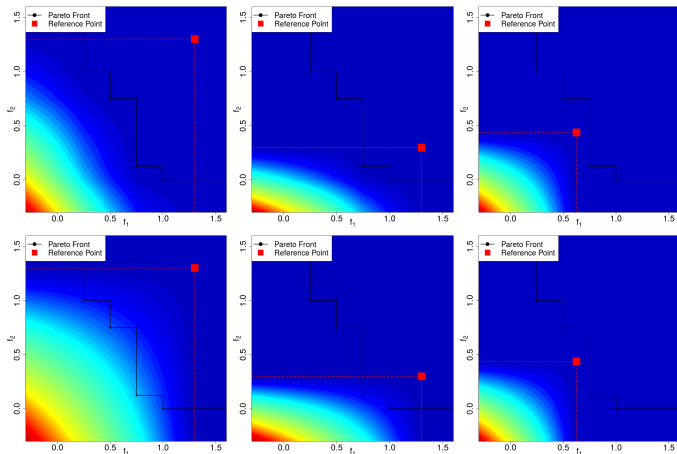
If $\widehat{\mathcal{P}}_y \not\subseteq \mathbf{R}$, $\text{EHI}(\cdot; \mathbf{R}) = \text{mEI}(\cdot; \mathbf{R})$.



$\text{mEI}(\mathbf{x}; \mathbf{R})$ is analytical in $m_i(\mathbf{x})$ and $C_i(\mathbf{x}, \mathbf{x})$, computationally much more efficient than EHI which involves Monte Carlo simulations when $m > 2$ (ms vs min).

Illustration: mEI versus EHI

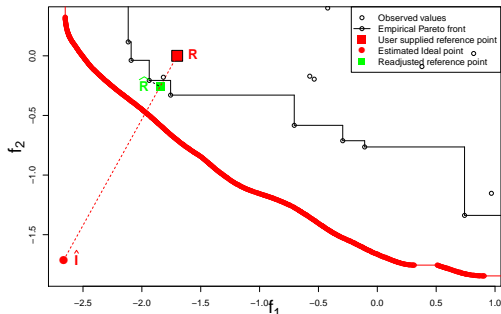
EHI top row, mEI bottom row



Reference point updating: principle

- $\hat{\mathbf{R}}$ too ambitious (e.g., near Ideal): the algorithm degenerates into a space filling (variance driven).
- $\hat{\mathbf{R}}$ easy to reach: favors already known high achievers.

$\Rightarrow \hat{\mathbf{R}}$ near the Empirical Pareto front as the right amount of ambitions (exploration vs intensification).

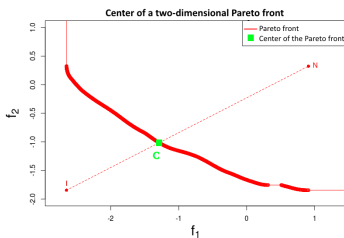
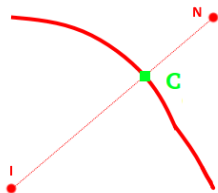
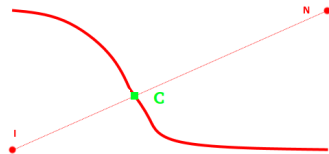
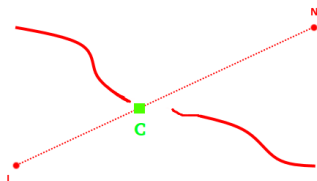


No user preference (\mathbf{R})? Default with the *Pareto front center* (next).

The Pareto front center

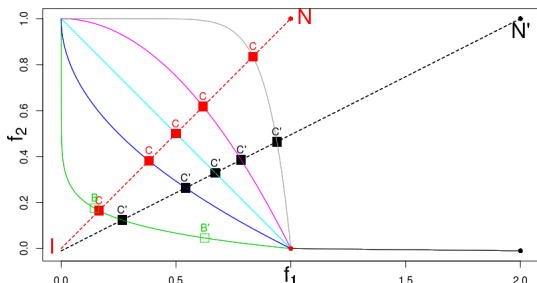
Which point should be targeted through **R**? By default, the point where objectives are “balanced”.

Definition: The center **C** is the point of the Ideal-Nadir line the closest in Euclidean distance to the Pareto front.



Properties of the Pareto front center

- The Pareto front center is equivalent, in game theory, to the Kalai-Smordinsky solution with a disagreement point at the Nadir [11].
- The Pareto front center is invariant w.r.t. a linear scaling of the objectives either when the Pareto front intersects the Ideal-Nadir line, or when $m = 2$ (not true in general though).
- The Pareto front center is stable w.r.t. perturbations in Ideal and Nadir: $\|\Delta \mathbf{C}\|_2 < \|\Delta \mathbf{N}\|_2$ and $\|\Delta \mathbf{C}\|_2 < \|\Delta \mathbf{I}\|_2$.



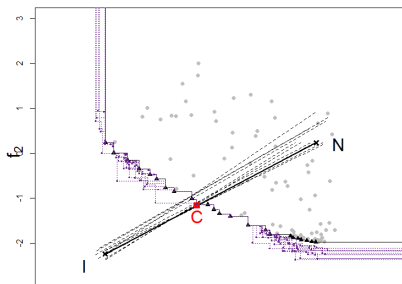
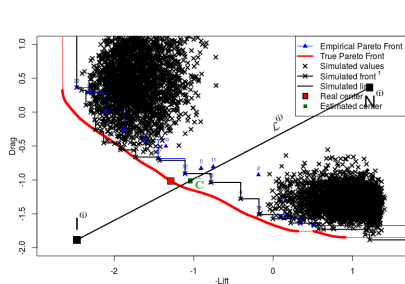
Estimating the Pareto front center

Crude estimators:

$$\hat{\mathbf{I}} = \left(\min_{\mathbf{y} \in \text{DoE}} (y_1), \dots, \min_{\mathbf{y} \in \text{DoE}} (y_m) \right),$$

$$\hat{\mathbf{N}} = \left(\max_{\mathbf{y} \in \widehat{\mathcal{P}}_y} (y_1), \dots, \max_{\mathbf{y} \in \widehat{\mathcal{P}}_y} (y_m) \right),$$

but they may be misleading early in the search. Take advantage of the GPs uncertainties \Rightarrow estimate them from Pareto front simulations (at carefully selected \mathbf{x} 's, see next slides) and take their median.



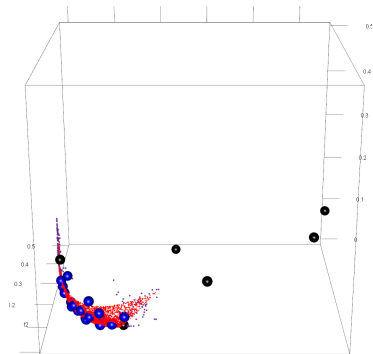
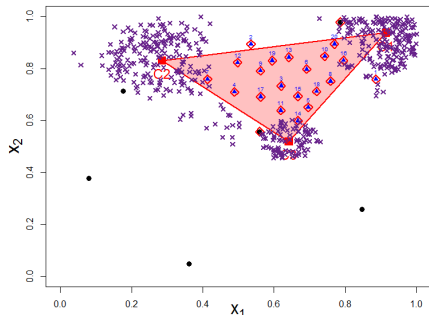
Simulation points for the Ideal and the Nadir (1)

- (For the Pareto front, choose \mathbf{x} 's with a probability proportional to $\mathbb{P}\left(\widehat{\mathcal{P}}_y \not\subseteq \mathbf{Y}(\mathbf{x})\right)$.) \leftarrow see earlier
- For the Ideal, choose \mathbf{x} 's with a probability proportional to $\mathbb{P}\left(Y_i(\mathbf{x}) \leq \min_j f_i^j\right)$, $j = 1, n$, $i = 1, m$ (analytical).
- For the Nadir, choose \mathbf{x} 's with a probability proportional to $\mathbb{P}\left(Y_i(\mathbf{x}) > \widehat{\mathbf{N}}_i, \mathbf{Y}(\mathbf{x}) \text{ non dominated}\right) + \mathbb{P}\left(\mathbf{Y}(\mathbf{x}) \preceq \arg \widehat{\mathbf{N}}_i\right)$, $i = 1, m$

More details in [6]

Simulation points for the Ideal and the Nadir (2)

In the quadratic case, $d = 2$, $m = 3$, after 20 iterations



DoE \cdot & \bullet & \bullet , Non Dominated \diamond , selected point \times , sampled \mathbf{Y} ND \cdot .

Simulation points are grouped around the centers which make the Ideal and Nadir.

First phase of \mathbf{R} estimations

Require: DoE =

$$\{(\mathbf{x}^1, \mathbf{f}(\mathbf{x}^1)), \dots, (\mathbf{x}^n, \mathbf{f}(\mathbf{x}^n))\}, \\ n^{\max}$$

1: Build the m independent GPs;

2: **repeat**

3: **if** no \mathbf{R} **then**

4: estimate Nadir $\hat{\mathbf{N}}$;
 $\mathbf{R} \leftarrow \hat{\mathbf{N}}$;

5: **end if**

6: estimate Ideal $\hat{\mathbf{I}}$;

7: $\hat{\mathbf{R}} \leftarrow$ Project on $\hat{\mathbf{I}\mathbf{R}}$ the
closest point of $\hat{\mathcal{P}}_y$ to $\hat{\mathbf{I}\mathbf{R}}$;

8: $\mathbf{x}^{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \text{mEI}(\mathbf{x}; \hat{\mathbf{R}})$;

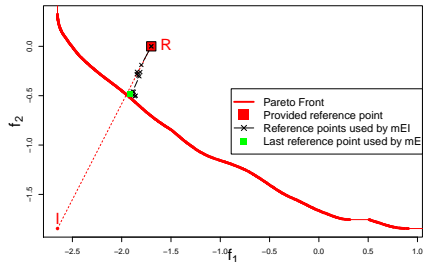
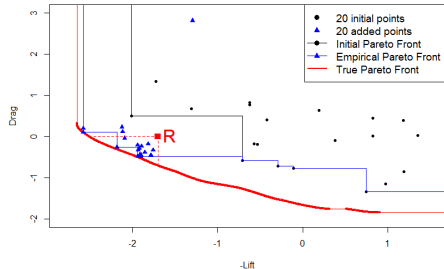
9: evaluate $\mathbf{f}(\mathbf{x}^{n+1})$ and
update the GPs;

10: $n \leftarrow n+1$;

11: **until** $n > n^{\max}$

Often $\hat{\mathbf{R}}$ is at the true Pareto front before the end. Cannot be further improved. Waste of computation.

Example of convergence to one Pareto-optimal point



Need a **stopping criterion**.

Uncertainty in center location (1)

Need a stopping criterion. mEI and EHI are too unstable: depend on f_i 's scales and \mathbf{R} .

Define the domination probability,

$$p(\mathbf{y}) := \mathbb{P}(\exists \mathbf{x} \in \mathcal{X} : \mathbf{Y}(\mathbf{x}) \preceq \mathbf{y})$$

Estimation: simulate n_{sim} Pareto fronts (at well-chosen \mathbf{x} 's), $\widetilde{\mathcal{P}}_y^{(i)}$, and

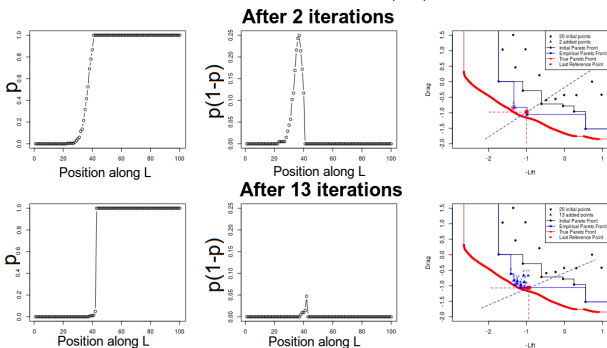
$$\widehat{p}(\mathbf{y}) = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \mathbb{1}(\widetilde{\mathcal{P}}_y^{(i)} \preceq \mathbf{y})$$

Uncertainty in center location (2)

If $\widehat{p(\mathbf{y})}$ is near 1 or 0, we are quite sure that \mathbf{y} is dominated or not. The uncertainty is $p(\mathbf{y})(1 - p(\mathbf{y}))$, the variance of the Bernoulli variable $D(\mathbf{y}) = \mathbb{1}(\mathcal{P}_{\mathbf{Y}(\cdot)} \preceq \mathbf{y})$.

Define the uncertainty in center location as

$$U(\widehat{\mathcal{L}}) := \frac{1}{|\widehat{\mathcal{L}}|} \int_{\widehat{\mathcal{L}}} p(\mathbf{y})(1 - p(\mathbf{y})) d\mathbf{y}.$$



($d = 8$)

Finding one Pareto-optimal point

Algorithm 2 First phase of the R/C-mEI algorithm

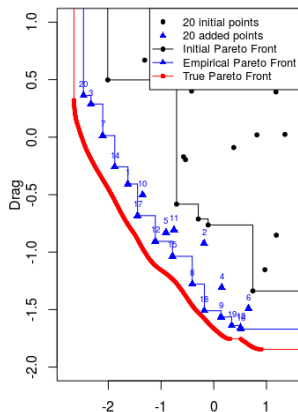
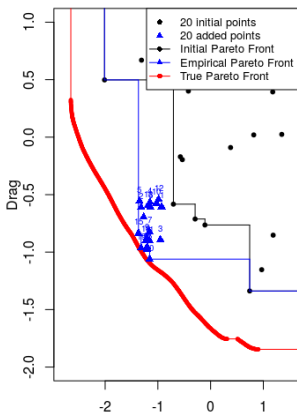
Require: DoE = $\{(\mathbf{x}^1, \mathbf{f}(\mathbf{x}^1)), \dots, (\mathbf{x}^n, \mathbf{f}(\mathbf{x}^n))\}$, ε_1 , n^{\max}

- 1: Build the m independent GPs;
 - 2: **repeat**
 - 3: estimate $\hat{\mathbf{R}}$ (i.e., $\hat{\mathbf{I}}$, and $\hat{\mathbf{N}}$ if no user reference);
 - 4: $\mathbf{x}^{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \text{mEI}(\mathbf{x}; \hat{\mathbf{R}})$;
 - 5: evaluate $\mathbf{f}(\mathbf{x}^{n+1})$ and update the GPs;
 - 6: compute $U(\hat{\mathbf{I}}\hat{\mathbf{R}})$;
 - 7: $n \leftarrow n+1$;
 - 8: **until** $U(\hat{\mathcal{L}}) \leq \varepsilon_1$ **or** $n > n^{\max}$
-

(If no \mathbf{R} , \mathbf{R} defaults to $\hat{\mathbf{N}}$ and $\hat{\mathbf{R}}$ is $\hat{\mathbf{C}}$)

Example of targeted MO Bayesian opt. vs EHI

MetaNACA, $n^{\max} = 40$



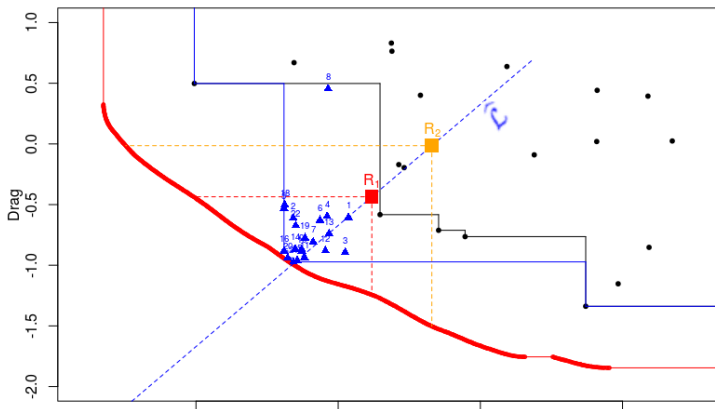
(Statistically significant results can be found in [6])

($d = 8$)

2. WIDENING THE SEARCH

Remaining budget: second phase

What if convergence to the Pareto front occurs before n^{\max} ?
 \Rightarrow widen the search around the last $\hat{\mathbf{R}}$ (or $\hat{\mathbf{C}}$) by moving $\hat{\mathbf{R}}$ along $\hat{\mathbf{I}}\mathbf{R}$ away from the Ideal by a distance that is *compatible with the remaining budget*, $n^{\max} - n$.



Optimal final search region

- For a given $\widehat{\mathbf{R}}$, anticipate the future space filling of the algorithm by virtual iterates (Kriging Believer, [8]) $\Rightarrow \mathbf{Y}^{KB}(\mathbf{x})$ built from $\{(\mathbf{x}^1, \mathbf{f}(\mathbf{x}^1)), \dots, (\mathbf{x}^n, \mathbf{f}(\mathbf{x}^n))\} \cup \{(\mathbf{x}^{n+1}, \boldsymbol{\mu}(\mathbf{x}^{n+1})), \dots, (\mathbf{x}^{n^{\max}}, \boldsymbol{\mu}(\mathbf{x}^{n^{\max}}))\}$
- Measure the remaining uncertainty in Pareto domination

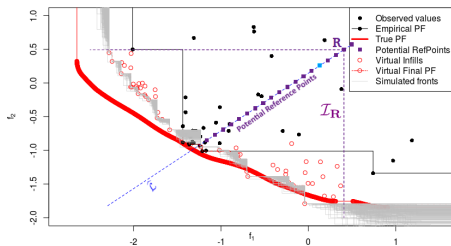
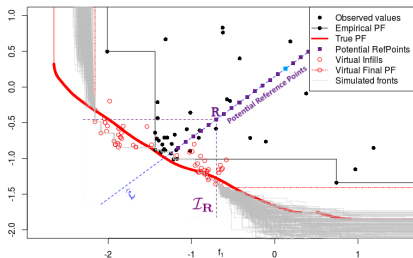
$$U(\widehat{\mathbf{R}}, \mathbf{Y}) := \frac{1}{\text{Vol}(\widehat{\mathbf{I}}, \widehat{\mathbf{R}})} \int_{\widehat{\mathbf{I}} \preceq \mathbf{y} \preceq \widehat{\mathbf{R}}} p(\mathbf{y})(1 - p(\mathbf{y})) d\mathbf{y} .$$

- Second phase optimal reference point defined through

$$\mathbf{R}^* = \arg \max_{\widehat{\mathbf{R}} \in \widehat{\mathbf{I}}\mathbf{R}} \|\widehat{\mathbf{R}} - \widehat{\mathbf{I}}\| \quad \text{such that} \quad U(\widehat{\mathbf{R}}; \mathbf{Y}^{KB}) \leq \varepsilon_2 \quad (1)$$

by enumeration.

Optimal final search region: illustration



The remaining uncertainty in Pareto domination can be seen by the sampled fronts roaming (in grey). It is small enough on the left, too large on the right.

R^* is in blue. $d = 8$.

Budgeted and Targeted MO Bayesian Optimization

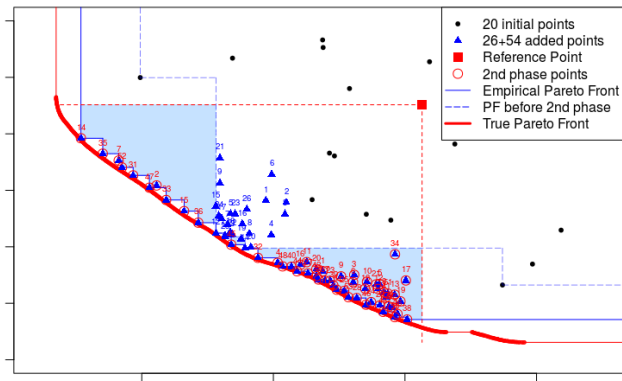
Algorithm 3 The R-mEI algorithm

Require: DoE = $\{(\mathbf{x}^1, \mathbf{f}(\mathbf{x}^1)), \dots, (\mathbf{x}^n, \mathbf{f}(\mathbf{x}^n))\}$, ε_1 , ε_2 , n^{\max}

- 1: Build the m independent GPs;
- 2: **repeat**
- 3: estimate $\widehat{\mathbf{R}}$ (i.e., $\widehat{\mathbf{I}}$, and $\widehat{\mathbf{N}}$ if no user reference);
- 4: $\mathbf{x}^{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \text{mEI}(\mathbf{x}; \widehat{\mathbf{R}})$;
- 5: evaluate $\mathbf{f}(\mathbf{x}^{n+1})$ and update the GPs;
- 6: compute $U(\widehat{\mathbf{I}}\mathbf{R})$;
- 7: $n \leftarrow n+1$;
- 8: **until** $U(\widehat{\mathbf{I}}\mathbf{R}) \leq \varepsilon_1$ **or** $n > n^{\max}$
- 9: **if** $n < n^{\max}$ **then**
- 10: Calculate \mathbf{R}^* solution of Eq. (1); # needs ε_2
- 11: **end if**
- 12: **while** $n < n^{\max}$ **do**
- 13: $\mathbf{x}^{n+1} = \arg \max_{\mathbf{x} \in \mathcal{X}} \text{EHI}(\mathbf{x}; \mathbf{R}^*)$;
- 14: evaluate $f_i(\mathbf{x}^{(t+1)})$ and update the GPs;
- 15: $n = n + 1$;
- 16: **end while**
- 17: **return** final DoE, final GPs, and approximation front $\widehat{\mathcal{P}}_y$

C-mEI: illustration of the 2nd phase

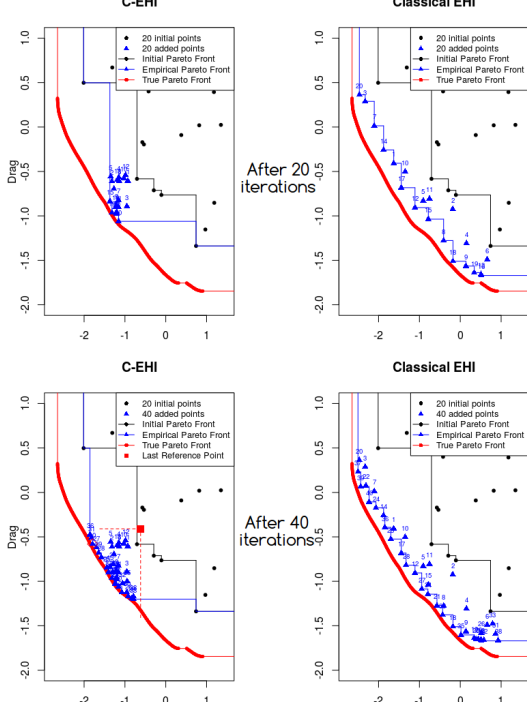
(video demo)



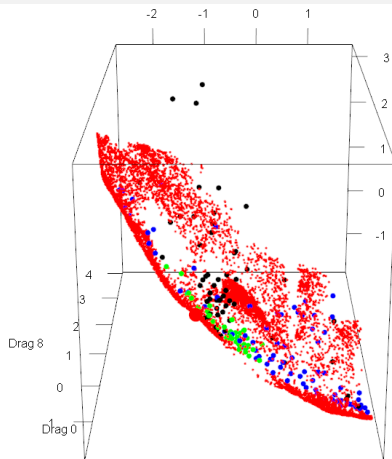
The objective values added during the 2nd phase are circled in red. Compared to the initial front obtained when searching for the center, the last approximation front is expanded as highlighted by the blue hypervolumes. $d = 8$.

C-mEI vs. EHI: illustration $m = 2$

C-mEI (left) vs. EHI (right), top after 20 iterations, bottom after 40 iterations. C-mEI local convergence has occurred at 22 iterations, a wider optimal improvement region (under the red square) is targeted for the 18 remaining iterations. Compared to the standard EHI, C-mEI searches in a smaller balanced part of the objective space, at the advantage of a better convergence. $d = 8$



C-mEI vs. EHI: illustration $m = 3$



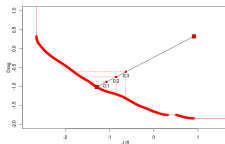
-Lift

($d = 8$)

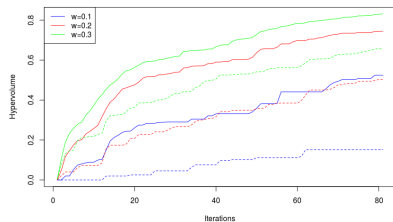
green, C-mEI; blue, EHI; black, initial front; red, true front, ●, true center.

C-mEI vs. EHI: tests

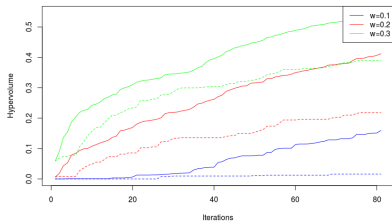
Hypervolumes of the C-mEI (continuous line) and EHI (dashed) averaged over 10 runs. Initial DoE of size 20, 80 iterations. Blue, red and green correspond to the improvement regions $\mathcal{I}_{0.1}$, $\mathcal{I}_{0.2}$ and $\mathcal{I}_{0.3}$, respectively. $d = 8$.



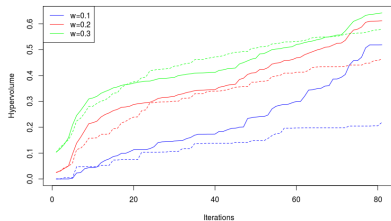
$m = 2$



$m = 3$



$m = 4$



C-mEI $>$ EHI, except when $m = 4$ and $\mathbf{R}_{0.3}$ because it is a large region.

3. A q -POINTS BATCH VERSION

Parallel MOO: related work

Three existing ways to obtain a batch of q points to parallelize the function evaluations in MOO (Horn et al. [9]):

- parallel execution of q searches with q different goals (Deb and Sundar [2]),
- select q points from an approximation to the Pareto front set,
- perform q sequential steps of a Bayesian MOO with a Kriging Believer strategy.

But it is not theoretically clear in which way these strategies are optimal \Rightarrow a batch criterion for MOO (details in [7]).

Batch mEI is q -mEI

In the same spirit as the q -EI criterion for single objective, we introduce a batch version of the mEI for MOO.

1 objective

$$\text{EI}(\mathbf{x}) = \mathbb{E} (f_{\min} - Y(\mathbf{x}))_+ \quad (\cdot)_+ := \max(0, \cdot)$$

$$q\text{-EI}(\mathbf{x}^1, \dots, \mathbf{x}^q) = \mathbb{E} \max_{i=1, \dots, q} (f_{\min} - Y(\mathbf{x}^i))_+$$

m objectives

$$\text{mEI}(\mathbf{x}; \mathbf{R}) = \mathbb{E} \prod_{j=1}^m (R_j - Y_j(\mathbf{x}))_+$$

$$q\text{-mEI}(\mathbf{x}^1, \dots, \mathbf{x}^q; \mathbf{R}) = \mathbb{E} \max_{i=1, \dots, q} \prod_{j=1}^m (R_j - Y_j(\mathbf{x}^i))_+$$

average the max of the hyper-rectangles areas

q - m El but not m - q El (1)

The correct batch m El is

$$q\text{-}m\text{El}(\mathbf{x}^1, \dots, \mathbf{x}^q; \mathbf{R}) = \mathbb{E} \max_{i=1, \dots, q} \prod_{j=1}^m (R_j - Y_j(\mathbf{x}^i))_+$$

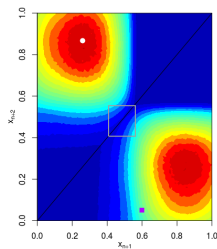
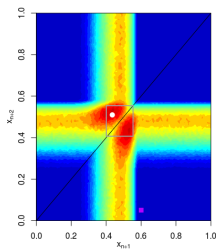
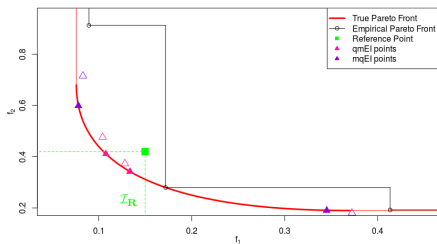
but the product of q El's is not correct

$$m\text{-}q\text{El}(\mathbf{x}^1, \dots, \mathbf{x}^q; \mathbf{R}) = \prod_{j=1}^m \mathbb{E} \max_{i=1, \dots, q} (R_j - Y_j(\mathbf{x}^i))_+$$

because when $q \geq m$ the maximum is obtained for each \mathbf{x}^i
maximizing one of the El's independently from the other objectives
 \Rightarrow no longer solves the MO problem.

q -mEI but not m - q EI (2)

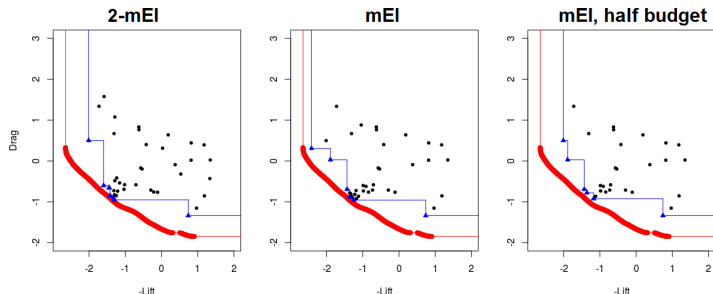
q -mEI (left) vs m - q EI (right) for $d = 1$, $m = 2$, $q = 2$.



The targeted region \mathcal{I}_R is attained inside the gray rectangles.
The purple square is an example of training point where q -mEI is null
but m - q EI is not.

2-mEI vs mEI, example on MetaNACA

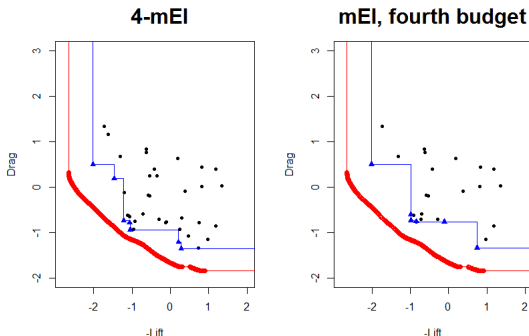
In all tests, q -mEI estimated with 10,000 Monte Carlo samples.
10 iterations of 2-mEI (left) vs 20 iterations of mEI (center) vs 10 iterations of mEI (right) for $d = 8$, $m = 2$.



The performance of 2-mEI is barely degraded wrt mEI at the same number of function evaluations, but the wall-clock time is half. At constant wall-clock time (iterations), 2-mEI outperforms mEI.

4-mEI vs mEI, example on MetaNACA

Constant wall-clock time comparison: 5 iterations of 4-mEI (left) vs 5 iterations of mEI (right) for $d = 8$, $m = 2$.



The performance of 4-mEI is degraded wrt mEI at the same number of function evaluations, it outperforms mEI at the same wall-clock time.

q -mEI tests on MetaNACA

10 independent runs, average (*std. dev.*) of hypervolumes in $\mathcal{I}_{0.3}$ after 20 and 50 additional evaluations in $d = 8, 22$, respectively.

Criterion	2-mEI	mEI	mEI, half budget
$d = 8$	0.234 (0.022)	0.265 (0.035)	0.209 (0.067)
$d = 22$	0.327 (0.045)	0.353 (0.048)	0.318 (0.048)

\Rightarrow q -mEI slightly less efficient than its sequential counterpart at the same number of evaluations, but better (and more stable) at same number of iterations (wall-clock time).

Conclusions

Summary The R-mEI algorithm

- allows to tackle multi-objective problems without assumptions on the functions beyond a bounded Pareto front when the budget is very small,
- has no arbitrary user settings (metrics, goals) and preserves objectives incommensurability,
- targets a specific region of improvement (as a default the center of the front),
- searches for a part of the Pareto front adapted to the budget.

Perspectives

- Account for couplings between the objectives.
- Calculate the gradient of q -mEI because optimization in increased dimension, $q \times d$ (cf. Marmin et al. [13]).

References I

- [1] Mickael Binois and Victor Picheny.
GPareto: An R package for Gaussian-process based multi-objective optimization and analysis.
2018.
- [2] Kalyanmoy Deb and J Sundar.
Reference point based multi-objective optimization using evolutionary algorithms.
In Proceedings of the 8th annual conference on Genetic and evolutionary computation, pages 635–642. ACM, 2006.
- [3] Michael Emmerich, André Deutz, and Jan Willem Klinkenberg.
Hypervolume-based expected improvement: Monotonicity properties and exact computation.
In Evolutionary Computation (CEC), 2011 IEEE Congress on, pages 2147–2154. IEEE, 2011.
- [4] Paul Feliot, Julien Bect, and Emmanuel Vazquez.
A Bayesian approach to constrained single- and multi-objective optimization.
Journal of Global Optimization, 67(1):97–133, January 2017.

References II

- [5] Paul Feliot, Julien Bect, and Emmanuel Vazquez.
User preferences in bayesian multi-objective optimization: the expected weighted hypervolume improvement criterion.
arXiv preprint arXiv:1809.05450, 2018.
- [6] David Gaudrie, Rodolphe Le Riche, Victor Picheny, Benoit Enaux, and Vincent Herbert.
Budgeted multi-objective optimization with a focus on the central part of the pareto front – extended version.
preprint arXiv:1809.10482, 2018.
- [7] David Gaudrie, Rodolphe Le Riche, Victor Picheny, Benoit Enaux, and Vincent Herbert.
Targeting solutions in bayesian multi-objective optimization: Sequential and parallel versions.
preprint arXiv:1811.03862, 2018.
- [8] David Ginsbourger, Rodolphe Le Riche, and Laurent Carraro.
Kriging is well-suited to parallelize optimization.
In *Computational Intelligence in Expensive Optimization Problems*, pages 131–162. Springer, 2010.

References III

- [9] Daniel Horn, Tobias Wagner, Dirk Biermann, Claus Weihs, and Bernd Bischl.
Model-based multi-objective optimization: taxonomy, multi-point proposal, toolbox and benchmark.
In International Conference on Evolutionary Multi-Criterion Optimization, pages 64–78.
Springer, 2015.
- [10] Donald R Jones, Matthias Schonlau, and William J Welch.
Efficient Global Optimization of expensive black-box functions.
Journal of Global optimization, 13(4):455–492, 1998.
- [11] Ehud Kalai and Meir Smorodinsky.
Other solutions to Nash's bargaining problem.
Econometrica: Journal of the Econometric Society, pages 513–518, 1975.
- [12] Deb Kalyanmoy et al.
Multi objective optimization using evolutionary algorithms.
John Wiley and Sons, 2001.
- [13] Sébastien Marmin, Clément Chevalier, and David Ginsbourger.
Differentiating the multipoint expected improvement for optimal batch design.
In International Workshop on Machine Learning, Optimization and Big Data, pages 37–48.
Springer, 2015.

References IV

- [14] Olaf Mersmann.
EMOA: Evolutionary multiobjective optimization algorithms.
R package version 0.5-0, 2012.
- [15] Kaisa Miettinen.
Nonlinear multiobjective optimization, volume 12.
Springer Science & Business Media, 1998.
- [16] Adrien Spagnol, Rodolphe Le Riche, and Sebastien Da Veiga.
Global sensitivity analysis for optimization with variable selection.
SIAM/ASA Journal on Uncertainty Quantification, 2019.
to appear.
- [17] Tobias Wagner, Michael Emmerich, André Deutz, and Wolfgang Ponweiser.
On expected-improvement criteria for model-based multi-objective optimization.
In *International Conference on Parallel Problem Solving from Nature*, pages 718–727.
Springer, 2010.
- [18] Kaifeng Yang, Andre Deutz, Zhiwei Yang, Thomas Back, and Michael Emmerich.
Truncated expected hypervolume improvement: Exact computation and application.
In *Evolutionary Computation (CEC), 2016 IEEE Congress on*, pages 4350–4357. IEEE, 2016.

References V

- [19] Adrien Zerbinati, Jean-Antoine Désidéri, and Régis Duvigneau.
Application of metamodel-assisted multiple-gradient descent algorithm (MGDA) to air-cooling duct shape optimization.
In *ECCOMAS-European Congress on Computational Methods in Applied Sciences and Engineering-2012*, 2012.

Comparisons of estimations for the Pareto front center

Example, $d = 8$:

Ideal-Nadir line of the empirical PF: a less robust estimator for the center of the Pareto front

