

Targeting Well-Balanced Solutions in Multi-Objective Bayesian Optimization under a Restricted Budget

David Gaudrie^{1,2,3}, Rodolphe Le Riche², Victor Picheny³,
Benoît Enaux¹, Vincent Herbert¹

¹Groupe PSA

²CNRS LIMOS, École Nationale Supérieure des Mines de Saint-Étienne

³Institut National de la Recherche Agronomique, MIAT

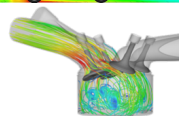
LION12 Conference, Kalamata (Greece), June 11th 2018

Industrial context

- Multi-objective optimization (m up to 6-8 objectives) of high dimensional systems (d up to 40)

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

- Complex systems and physics \Rightarrow use of computationally expensive CFD codes (12-24 hours per simulation) \Rightarrow optimization under restricted budget (≈ 100 evaluations)
- Multi-criteria decision-aid: choice among the optimal solutions made by a Decision Maker (DM)
- How to obtain several *optimal* trade-off solutions in spite of the extremely parsimonious use of the computer code?



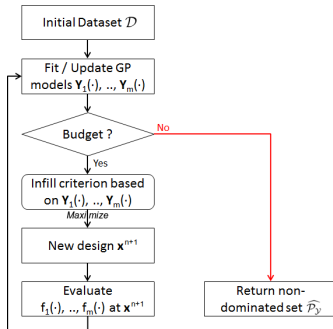
Bayesian multi-objective optimization

$$\min_{\mathbf{x} \in X \subset \mathbb{R}^d} (f_1(\mathbf{x}), \dots, f_m(\mathbf{x})), f_i(\cdot) \text{ expensive black-box}$$

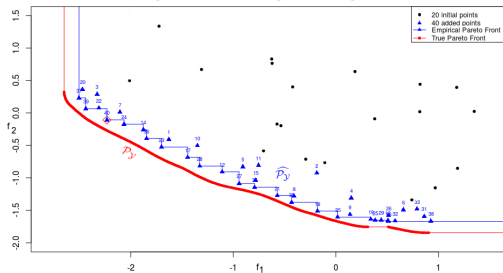
- Multi-objective extension to EGO [2]
- Fit m independent Gaussian Processes (GP) $Y_1(\cdot), \dots, Y_m(\cdot)$ for the objectives to an initial design of experiments
 $\mathcal{D}_n = \{(\mathbf{x}^1, \mathbf{y}^1), \dots, (\mathbf{x}^n, \mathbf{y}^n)\}; \mathbf{f}(\mathbf{x}^k) = \mathbf{y}^k = (y_1^k, \dots, y_m^k)^T$
- Computable Kriging mean predictor $\hat{y}_i(\mathbf{x})$ and variance $s_i^2(\mathbf{x}) \forall \mathbf{x} \in X$

Bayesian multi-objective optimization

- Use a Bayesian multi-objective infill criterion ("multi-objective version" of EI [3], e.g. EHI [1], EMI [6], SMS [5], SUR [4], ...) relying only on $\mathbf{Y}(\cdot)$ to determine \mathbf{x}^{n+1}
- Evaluate $f_i(\cdot), i = 1, \dots, m$ at $\mathbf{x}^{n+1} \Rightarrow \mathbf{y}^{n+1}$
- Update the GP metamodels



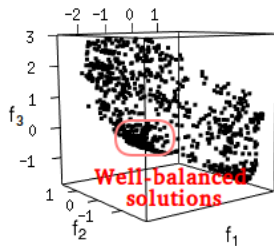
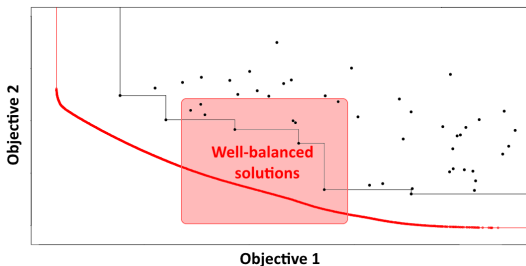
Example of Multi-Objective Optimization



Output: Pareto front approximation $\hat{\mathcal{P}}_Y$

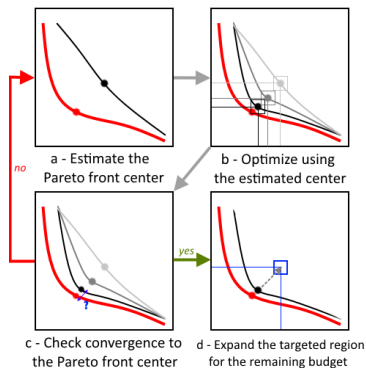
Targeting: motivations

- Restricted budget, many objectives \Rightarrow impossible for classical MO-EGO approaches to find the Pareto front (growing size of \mathcal{P}_Y with m)
- \Rightarrow Uncovering the whole Pareto front in a "region of interest"
- Targeting solutions in objective space
 - Shrink the search to a smaller subset \Rightarrow faster convergence
 - Emphasize solutions that *equilibrate* the objectives: (unknown) central part of the PF \Rightarrow interesting solutions for the DM



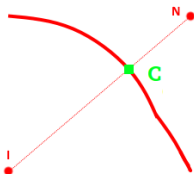
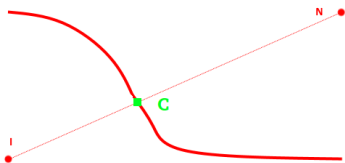
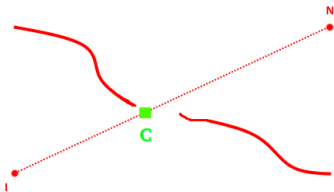
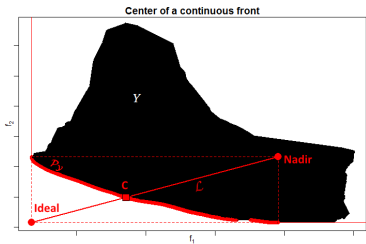
Proposed methodology

C-EHI (Centered Expected Hypervolume Improvement) algorithm



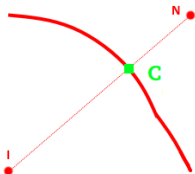
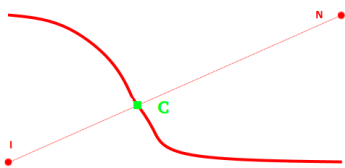
Center of the Pareto front: definition

- **C**: Intersection between \mathcal{P}_Y and Ideal-Nadir line \mathcal{L} (**IN**), if existing
- Closest point of \mathcal{L} to \mathcal{P}_Y otherwise (projection on \mathcal{L})



Center of the Pareto front: properties

- Invariant to a linear scaling when \mathcal{P}_y intersects \mathcal{L}
- No intersection: closest point $\mathbf{y} \in \mathcal{P}_y$ to \mathcal{L} remains the same after linear scaling in bi-objective problems
- Low sensitivity¹ to \mathbf{I} and \mathbf{N} : $|\frac{\partial C_i}{\partial I_j}|$ and $|\frac{\partial C_i}{\partial N_j}| < 1, i, j = 1, \dots, m.$



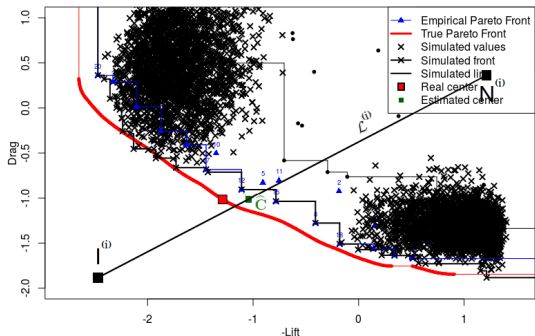
- In Game Theory: particular Kalai-Smorodinsky solution (disagreement point $\mathbf{d} \equiv \mathbf{N}$)

¹Under some regularity assumptions

Center of the Pareto front: estimation

- \mathbf{C} located on $\mathcal{L} \Rightarrow \mathbf{I}$ and \mathbf{N} required
- Use GP simulations to produce estimates $\hat{\mathbf{I}}$ and $\hat{\mathbf{N}} \Rightarrow \hat{\mathcal{L}}$

GP simulations directed towards the extremes of the Pareto front

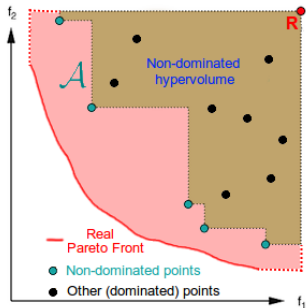


- \triangle Choice of \mathbf{x} 's where $\mathbf{Y}(\cdot)$ is simulated is critical!
- Choose \mathbf{x} 's according to probability to lead to \mathbf{I} or \mathbf{N} through probability of being extreme and non-dominated

- Estimated center $\hat{\mathbf{C}}$: closest point to $\hat{\mathcal{P}}_y$ on $\hat{\mathcal{L}}$

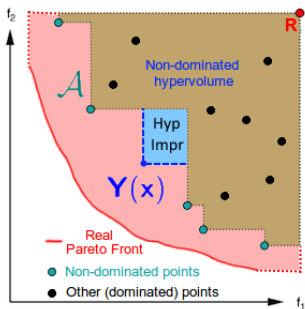
EHI: a Bayesian multi-objective infill criterion

- Hypervolume Indicator $H(\mathcal{A}; \mathbf{R}) = \text{Vol} \left(\bigcup_{\mathbf{y} \in \mathcal{A}} \{\mathbf{z} : \mathbf{y} \preceq \mathbf{z} \preceq \mathbf{R}\} \right)$,
 \mathbf{R} reference point [7]



EHI: a Bayesian multi-objective infill criterion

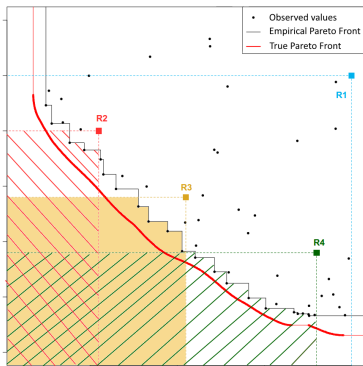
- Hypervolume Indicator $H(\mathcal{A}; \mathbf{R}) = \text{Vol} \left(\bigcup_{\mathbf{y} \in \mathcal{A}} \{\mathbf{z} : \mathbf{y} \preceq \mathbf{z} \preceq \mathbf{R}\} \right)$, \mathbf{R} reference point [7]



- $\text{EHI}(\mathbf{x}; \mathbf{R}) = \mathbb{E}[H(\widehat{\mathcal{P}}_{\mathbf{y}} \cup \{\mathbf{Y}(\mathbf{x})\}; \mathbf{R}) - H(\widehat{\mathcal{P}}_{\mathbf{y}}; \mathbf{R})]$: Expected **Hypervolume Improvement** (relatively to \mathbf{R}) [1], if adding design \mathbf{x}
- Maximal Hypervolume \Rightarrow Uncover the whole Pareto front

Targeting the center via EHI

- In other works, \mathbf{R} chosen to cover the whole Pareto front (**R1**)
- Here: use \mathbf{R} to target *central* solutions by setting $\mathbf{R} = \hat{\mathbf{C}}$
⇒ Null hypervolume improvement for solutions that do not dominate $\hat{\mathbf{C}}$: $H(\hat{\mathcal{P}}_y \cup \{\mathbf{y}\}; \hat{\mathbf{C}}) = 0 \quad \forall \mathbf{y} \not\prec \hat{\mathbf{C}}$



EHI-mEI equivalence

Definition

$mEI(\cdot; \mathbf{R}) = \prod_{i=1}^m EI_i(\cdot; R_i)$ with $EI_i(\cdot; R_i)$ EI considering R_i as the current minimum in objective i .

- Closed form, cheap to compute (even for large number of objectives)

Proposition

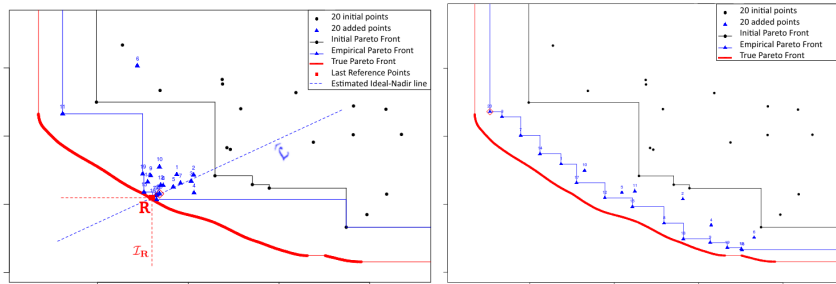
If $Y_1(\cdot), \dots, Y_m(\cdot)$ are independent GPs, and $\widehat{\mathcal{P}}_y \not\subseteq \mathbf{R}$,
 $EHI(\cdot; \mathbf{R}) = mEI(\cdot; \mathbf{R})$

- $\widehat{\mathcal{P}}_y \not\subseteq \widehat{\mathbf{C}}^1 \Rightarrow EHI(\cdot; \widehat{\mathbf{C}})$ equivalent to $mEI(\cdot; \widehat{\mathbf{C}})$

¹Occurs almost always

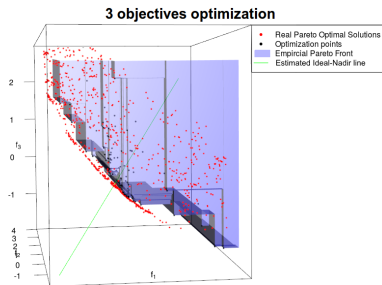
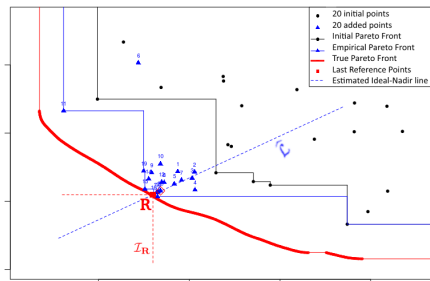
Example

- Estimated center $\hat{\mathbf{C}} + mEI$
- Budget of 20 calls to the expensive function: center-targeting (left) and classical approach (right)



Example

- 2 objectives (left) and 3 objectives (right)



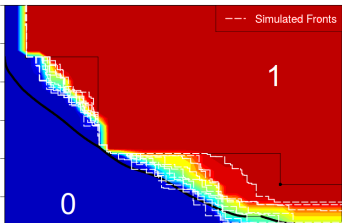
- Local convergence to the Pareto front achieved
- \mathcal{I}_R unreachable (\mathbf{R} on \mathcal{P}_y) \Rightarrow waste of resources
- Convergence to \mathbf{C} has to be verified: stopping criterion

Convergence criterion using the probability of domination

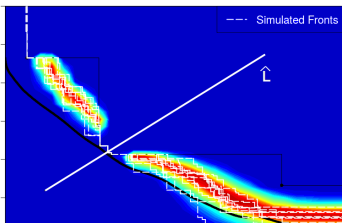
- Probability of domination $p(\mathbf{y})$: probability that objective vector $\mathbf{y} \in \mathbb{R}^m$ can be dominated by some $(f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$, $\mathbf{x} \in X$
- Estimated using simulated Pareto fronts $\widetilde{\mathcal{P}}_y^{(k)}$:

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

Probability of domination in Y space

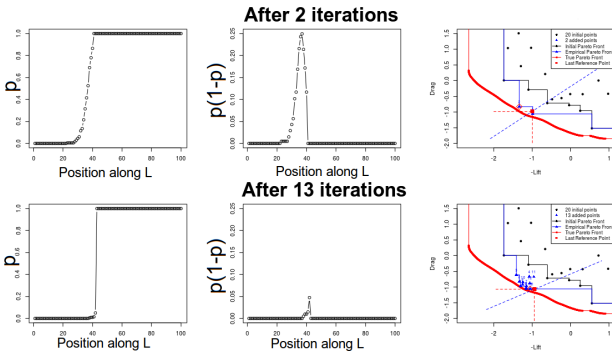


$p(1-p)$ in Y space



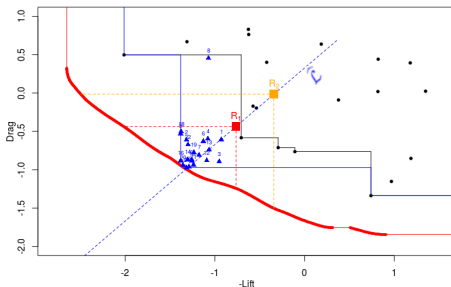
Convergence criterion using the probability of domination

- Information about uncertainty and (local) convergence towards the PF: convergence in areas where p goes quickly from 0 to 1 / where $p(1 - p)$ equals 0
- Assume local convergence to the center of the PF when $\int_{\hat{L}} p(\mathbf{y})(1 - p(\mathbf{y})) d\mathbf{y} \leq \varepsilon$



Ongoing work: expansion of the PF approximation after convergence

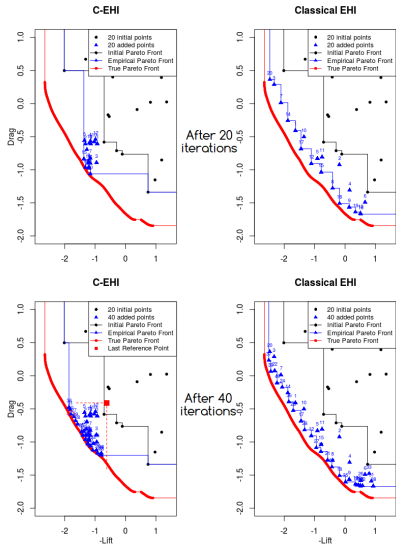
- Convergence detected: how to use the remaining budget? Target a *broader* central part of the PF
- Use $\text{EHI}(\cdot; \mathbf{R})$ with \mathbf{R} shifted backwards on $\hat{\mathcal{L}} \Rightarrow$ larger but still central targeted region



- Targeted region for the last iterations: largest area where accurate enough convergence is forecasted within the remaining budget

Comparison with standard EHI

- Budget of 40 iterations: better convergence of C-EHI to \mathcal{P}_y in its central part
- Convergence detected at 22nd iteration \Rightarrow broader area targeted for the last 18 iterations (red square in bottom left figure)



References



Michael TM Emmerich, Kyriakos C Giannakoglou, and Boris Naujoks.

Single- and multiobjective evolutionary optimization assisted by gaussian random field metamodels. [IEEE Transactions on Evolutionary Computation](#), 10(4):421–439, 2006.



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.



J Mockus.

On bayesian methods for seeking the extremum.

In [Optimization Techniques IFIP Technical Conference](#), pages 400–404. Springer, 1975.



Victor Picheny.

Multiobjective optimization using gaussian process emulators via stepwise uncertainty reduction. [Statistics and Computing](#), 25(6):1265–1280, 2015.



Wolfgang Ponweiser, Tobias Wagner, Dirk Biermann, and Markus Vincze.

Multiobjective optimization on a limited budget of evaluations using model-assisted s-metric selection. In [International Conference on Parallel Problem Solving from Nature](#), pages 784–794. Springer, 2008.



Joshua D Svenson.

[Computer experiments: Multiobjective optimization and sensitivity analysis](#). PhD thesis, The Ohio State University, 2011.



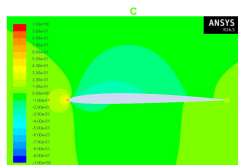
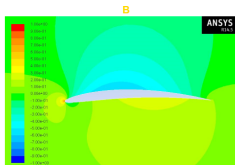
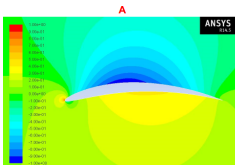
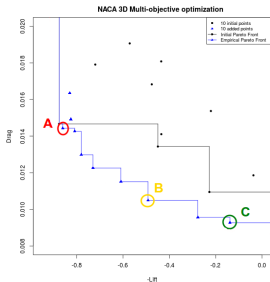
Eckart Zitzler.

Evolutionary algorithms for multiobjective optimization: Methods and applications. 1999.

Thank you for your attention,

Do you have any question?

Attractive solutions: central solutions



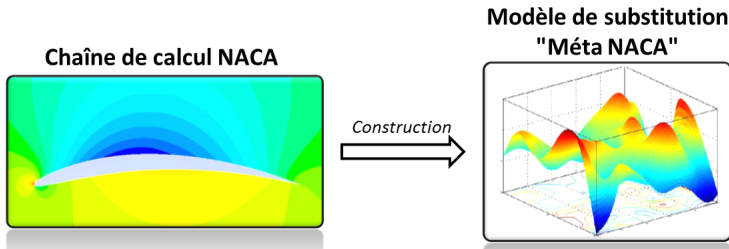
← +

Drag (minimize)

Lift (maximize)

Benchmark: "Meta NACA"

- For each test case (dimension $d = 3, 8$ or 22), creation of a surrogate to the computer code using
 - 1000 points (complete factorial design) in 3D
 - 1200 points (LHS-maximin design + refinement in areas of compromise) in 8D and 22D



- Variable number of objectives: $m \in \{2, 3, 4\}$

Ideal-Nadir line of the empirical Pareto front

- Center of a Pareto front \Rightarrow Ideal-Nadir line
- Not robust when using \mathbf{I} and \mathbf{N} from the approximation front $\widehat{\mathcal{P}}_y$

Ideal-Nadir line of the empirical Pareto Front: bad estimate of the center

