

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

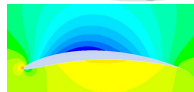
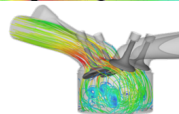
²École Nationale Supérieure des Mines de Saint-Étienne, CNRS LIMOS

³Institut National de la Recherche Agronomique, MIAT

PGMO Days, Saclay, November 14th 2017

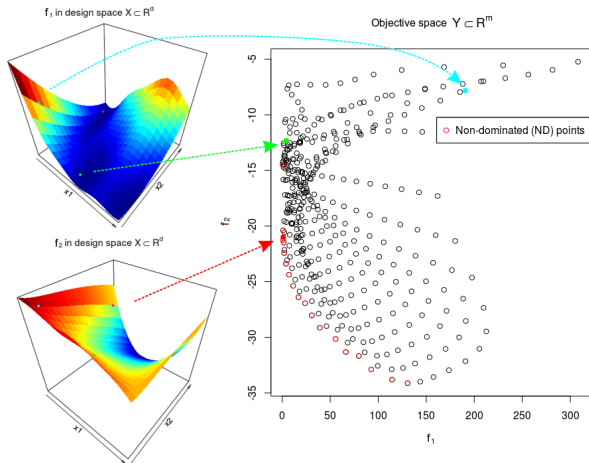
Industrial context

- Multi-objective optimization (up to 6-8 objectives) of high dimensional systems (d up to 40)
- Complex systems and physics \Rightarrow use of computationally expensive CFD codes (12-24 hours per simulation) \Rightarrow optimization under (very) restricted budget (≈ 200 evaluations)
- Test case : aerodynamic optimization of a NACA airfoil (3, 8 or 22 parameters, 2 to 4 objectives)
- Multi-criteria decision-aid : choice among the optimal solutions made by a Decision Maker
- How to obtain several *optimal* trade-off solutions in spite of an extremely parsimonious use of the computer code ?



Multi-objective optimization (MOO)

- Multi-objective problem : $\min_{\mathbf{x} \in X} (f_1(\mathbf{x}), \dots, f_m(\mathbf{x}))$



- Aim of MOO : find an approximation to the Pareto Front, set of all ND values in Y

Gaussian Processes for Multi-Objective Optimization

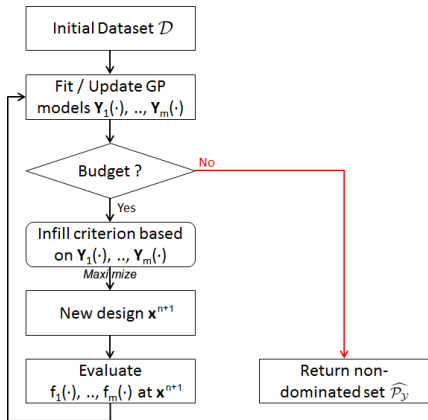
- Limited knowledge of the objective functions \Rightarrow metamodeling
- Initial dataset $\mathcal{D} = (\mathbf{x}_{1:n}, \mathbf{y}_{1:n}) = ((\mathbf{x}^{(1)}, \mathbf{y}^{(1)}), \dots, (\mathbf{x}^{(n)}, \mathbf{y}^{(n)}))$

$$\text{with } \mathbf{y}^{(k)} = \begin{pmatrix} y_1^{(k)} \\ \vdots \\ y_m^{(k)} \end{pmatrix}$$

- Fit m independent GP models $\mathbf{Y}_1(\cdot), \dots, \mathbf{Y}_m(\cdot)$ for the objectives
- Prediction : $\hat{y}_i(\cdot)$, Variance : $s_i^2(\cdot)$
- Use Bayesian Multi-Objective infill criterion for the choice of \mathbf{x}^{n+1}
- Evaluate $f_i(\cdot), i = 1, \dots, m$ at $\mathbf{x}^{(n+1)}$ as $\mathbf{y}^{(n+1)}$
- Update the GP models

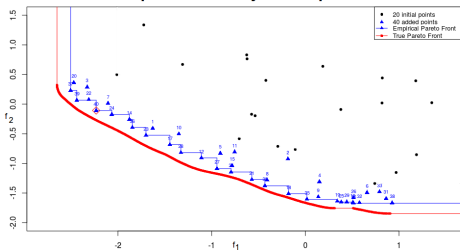
Gaussian Processes for Multi-Objective Optimization

• The algorithm



• Pareto Front approximation $\widehat{\mathcal{P}}_y$

Example of Multi-Objective Optimization

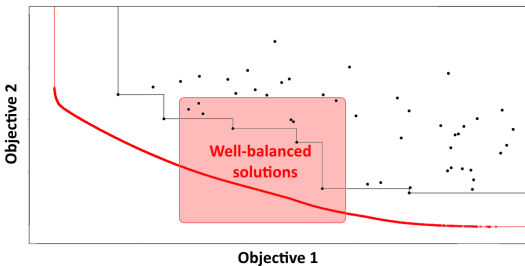


Targeting : motivations

- Restricted budget \Rightarrow Uncovering the whole Pareto Front in a "region of interest"
- Targeting solutions in objective space
 - Provide only "interesting" solutions to the Decision Maker
 - Shrink the search to a smaller subset \Rightarrow faster convergence ?
- Growing size of \mathcal{P}_y with m

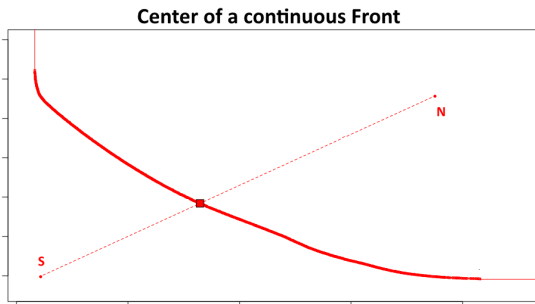
Attractive solutions

- Multi-objective optimization : *compromise* between objectives
- Favour solutions that balance these objectives : central part of the Pareto Front. "Preference-Based Optimization" towards the center



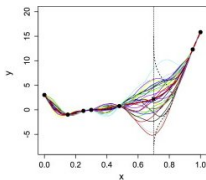
Center of a Pareto Front

- Used definition
 - Intersection between continuous prolongation of \mathcal{P}_Y and Nadir-Shadow line \mathcal{L} , if existing
 - Closest point of \mathcal{P}_Y to \mathcal{L} otherwise

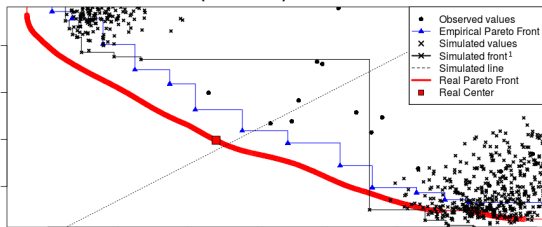


Conditional simulations for estimating the center of a Pareto Front

- Real Pareto Front, Nadir and Shadow are unknown \Rightarrow GP conditional simulations to learn about these points
- Use extreme values of simulated front(s) \Rightarrow estimated NS line $\hat{\mathcal{L}}$ \Rightarrow estimation of the central region of the Pareto Front
- Slightly discard extreme values : extreme quantiles of the simulated Front (e.g. 5%-95% quantiles line)



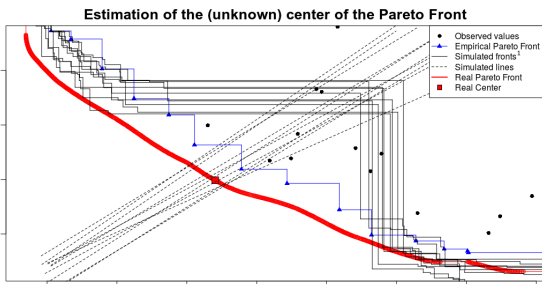
Estimation of the (unknown) center of the Pareto Front



¹ Fronts that are meant to be only accurate at the extremes of $\widehat{\mathcal{P}}_y$

Conditional simulations for estimating the center of a Pareto Front

- Real Pareto Front, Nadir and Shadow are unknown \Rightarrow GP conditional simulations to learn about these points
- Use extreme values of simulated front(s) \Rightarrow estimated NS line $\hat{\mathcal{L}}$ \Rightarrow estimation of the central region of the Pareto Front



- Provides a direction to follow during optimization

¹Fronts that are meant to be only accurate at the extremes of $\widehat{\mathcal{P}}_y$

Infill Criteria in Bayesian Optimization

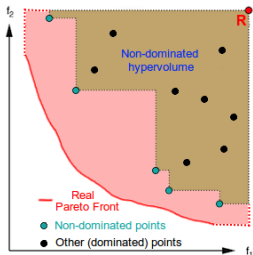
- Based on the metamodel(s)
- $IC = f(Y_1(\cdot), \dots, Y_m(\cdot); \mathbf{x}; \Theta)$ to be maximized w.r.t \mathbf{x} \Rightarrow directs the search towards attractive new designs \mathbf{x}^* , given the hyper-parameters (of the IC) Θ
- Modify existing infill criteria through Θ to direct the search towards the estimated central area of \mathcal{P}_y
- Three infill criteria will be considered : *EHI* (Expected Hypervolume Improvement), *PI* (Probability of Improvement) and *EI* (Expected Improvement)

Expected Hypervolume Improvement

- Non-dominated Hypervolume Indicator

$$H(\mathcal{A}; \mathbf{R}) = \Lambda \left(\bigcup_{\mathbf{y} \in \mathcal{A}} \{ \mathbf{z} : \mathbf{y} \preceq \mathbf{z} \preceq \mathbf{R} \} \right)$$

with $\Lambda(\mathcal{E})$ Lebesgue measure of non-dominated set \mathcal{E}

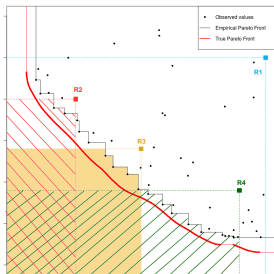


E. Zitzler and L. Thiele,

Multiobjective Optimization Using Evolutionary Algorithms - A Comparative Case Study
Conference on Parallel Problem Solving from Nature (PPSN V), pages 292–301 (1998)

Expected Hypervolume Improvement

- $EHI(\mathbf{x}; \mathbf{R})$: expected improvement of the non-dominated hypervolume (relatively to \mathbf{R}), if adding design \mathbf{x}
- $EHI(\mathbf{x}; \mathbf{R}) = \mathbb{E}[H(\widehat{\mathcal{P}}_y \cup \{\mathbf{Y}(\mathbf{x})\}; \mathbf{R}) - H(\widehat{\mathcal{P}}_y; \mathbf{R})]$
- Improvement : exclusively in areas dominating \mathbf{R}



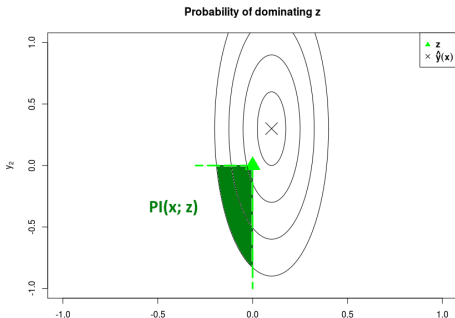
M. Emmerich, K. Giannakoglou, B. Naujoks,

Single and Multiobjective Evolutionary Optimization Assisted by Gaussian Random Field Metamodels
 IEEE Transactions on Evolutionary Computation, 10 (2006).

Probability of Improvement

- $\mathbf{Y}(\cdot) = (Y_1(\cdot), \dots, Y_m(\cdot))$ with $Y_i(\cdot), i = 1, \dots, m$ independent GPs
- $\forall \mathbf{x} \in X, \forall \mathbf{z} \in Y, PI(\mathbf{x}; \mathbf{z})$: probability that $\mathbf{Y}(\mathbf{x})$ dominates a given (target) vector \mathbf{z} in objective space

$$PI(\mathbf{x}; \mathbf{z}) = \mathbb{P}(\mathbf{Y}(\mathbf{x}) \preceq \mathbf{z}) = \prod_{i=1}^m \Phi\left(\frac{z_i - \hat{y}_i(\mathbf{x})}{s_i(\mathbf{x})}\right)$$



Expected Improvement

- Mono-objective Expected Improvement :

$$EI(\mathbf{x}) = (y_{min} - \hat{y}(\mathbf{x}))_+ \Phi \left(\frac{y_{min} - \hat{y}(\mathbf{x})}{s(\mathbf{x})} \right) + s(\mathbf{x}) \phi \left(\frac{y_{min} - \hat{y}(\mathbf{x})}{s(\mathbf{x})} \right) \quad (1)$$

- Multi-objective optimization : no straightforward extension
- Product of Expected Improvement w.r.t. a certain (possibly non-dominated) point $\mathbf{z} \in Y$

$$mEI(\mathbf{x}; \mathbf{z}) = \prod_{i=1}^m EI_i(\mathbf{x}; z_i)$$

where $EI_i(\cdot, z)$ stands for the Expected Improvement in objective i , considering z as the current minimum (y_{min}) in (1).

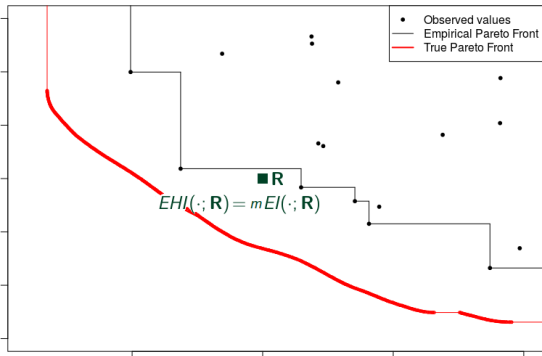


D. R. Jones, M. Schonlau, and W. J. Welch,

Efficient global optimization of expensive black-box functions
Journal of Global Optimization, 13, 455–492 (1998).

Expected Improvement

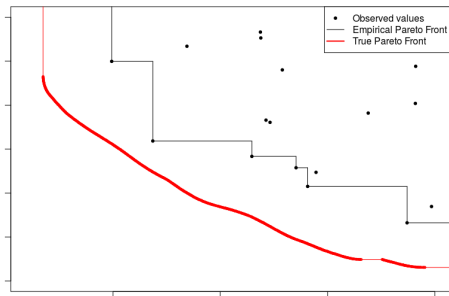
- Proposition : let $Y_i(\cdot)$ be independent GPs and $\mathbf{R} \in Y$ be a non-dominated reference point for EHI, that is $\widehat{\mathcal{P}}_Y \not\subseteq \mathbf{R}$. Then $EHI(\mathbf{x}; \mathbf{R}) = mEI(\mathbf{x}; \mathbf{R}) \forall \mathbf{x} \in X$



- Cheaper, parallelizable, analytic expression

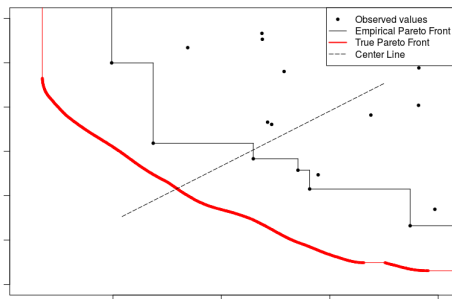
Targeting via Tailored Infill Criteria

- These three infill criteria can target regions of the objective space, by controlling a reference point
- Target the center



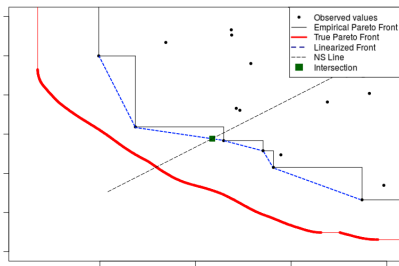
Targeting via Tailored Infill Criteria

- These three infill criteria can target regions of the objective space, by controlling a reference point
- Target the center \Rightarrow estimated Nadir-Shadow line $\hat{\mathcal{L}}$



Targeting via Tailored Infill Criteria

- Need a reference point \Rightarrow intersection between meshed empirical Pareto Front $\widehat{\mathcal{P}}_y$ and $\widehat{\mathcal{L}}$ (if possible)
- Slightly optimistic reference point : challenge the algorithm, avoid getting stuck, *EHI-mEI* equivalence



- Optimization directed towards the (estimated) center of the Pareto Front

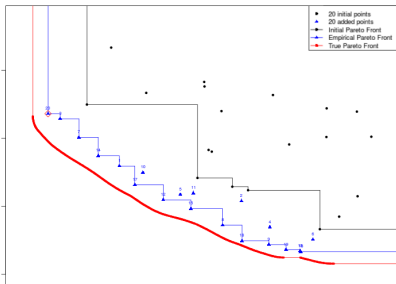
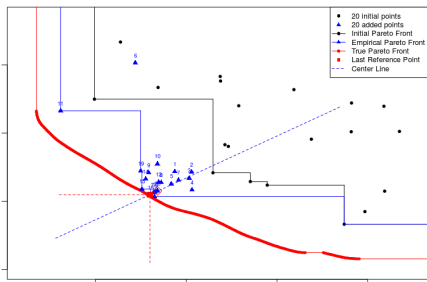


Hartikainen, M., Miettinen, K., Wiecek, M. M.,

PAINT: Pareto front interpolation for nonlinear multiobjective optimization
Computational optimization and applications, 52(3), 845-867 (2012).

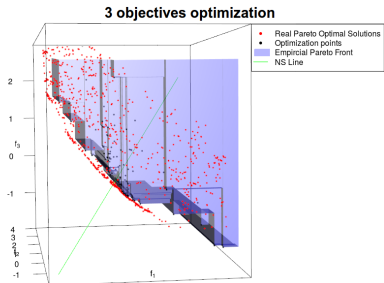
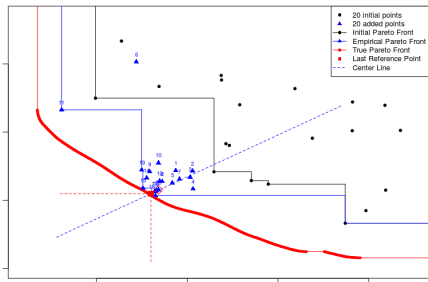
Example : Center-Targeting MOO with mEI

- Budget of 20+20 calls to the expensive function :
Center-Targeting (left) and Classical Approach (right)



Example : Center-Targeting MOO with mEI

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



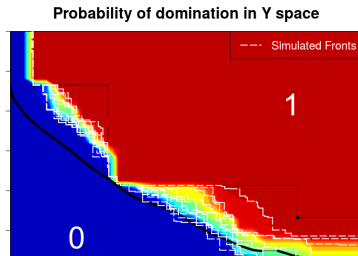
- Local convergence towards the Pareto Front achieved
- When to stop? Remaining budget \Rightarrow what to do next?

Local convergence towards the Pareto Front

- $\hat{\mathbb{P}}_{dom} : Y \rightarrow [0, 1]$ probability that vector \mathbf{y} is dominated
- $\hat{\mathbb{P}}_{dom}(\mathbf{y})$ estimated using GP simulations

$$\hat{\mathbb{P}}_{dom}(\mathbf{y}) \approx \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\hat{\mathcal{P}}_i \preceq \mathbf{y}} =: p$$

with $\hat{\mathcal{P}}_i$ Pareto Front resulting from i -th simulation



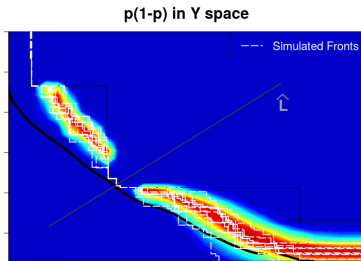
- Convergence in areas where p goes quickly from 0 to 1

Local convergence towards the Pareto Front

- $\hat{\mathbb{P}}_{dom} : Y \rightarrow [0, 1]$ probability that vector \mathbf{y} is dominated
- $\hat{\mathbb{P}}_{dom}(\mathbf{y})$ estimated using GP simulations

$$\hat{\mathbb{P}}_{dom}(\mathbf{y}) \approx \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\hat{\mathcal{P}}_i \preceq \mathbf{y}} =: p$$

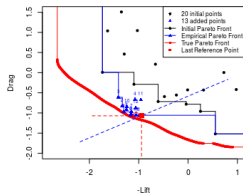
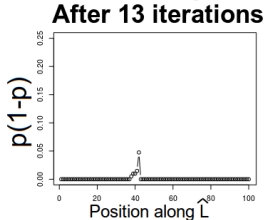
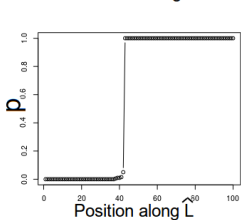
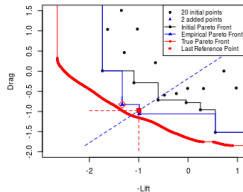
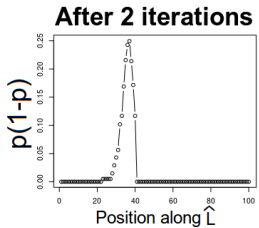
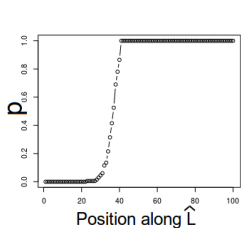
with $\hat{\mathcal{P}}_i$ Pareto Front resulting from i -th simulation



- Convergence in areas where $p(1 - p)$ is close to 0

Local convergence towards the Pareto Front

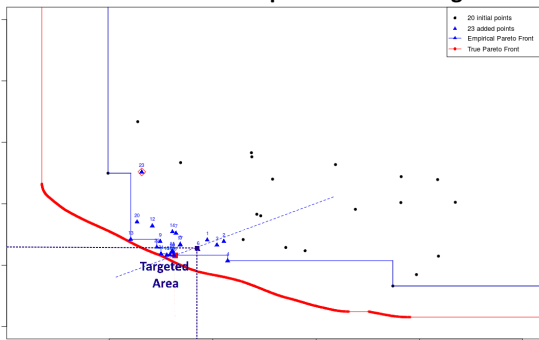
- Assume (local) convergence towards the central region of the Pareto Front if $\int_{\hat{L}} p(1-p)dy \leq \varepsilon$



Ongoing work : expanding the PF approximation

- Expansion of the \mathbf{y} 's considered for improvement around converged area \Rightarrow wider range of solutions
- Well-balanced solutions : use $EHI(\cdot, \mathbf{R})$ with \mathbf{R} along $\hat{\mathcal{L}}$

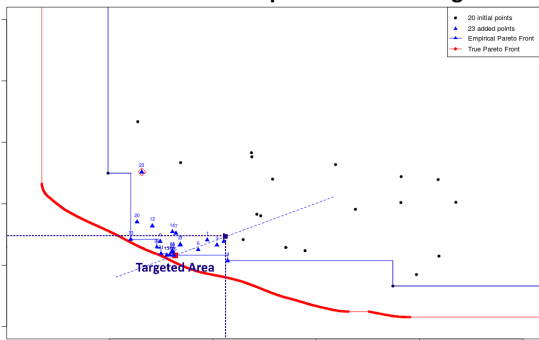
Possible reference point after convergence



Ongoing work : expanding the PF approximation

- Expansion of the \mathbf{y} 's considered for improvement around converged area \Rightarrow wider range of solutions
- Well-balanced solutions : use $EHI(\cdot, \mathbf{R})$ with \mathbf{R} along $\hat{\mathcal{L}}$

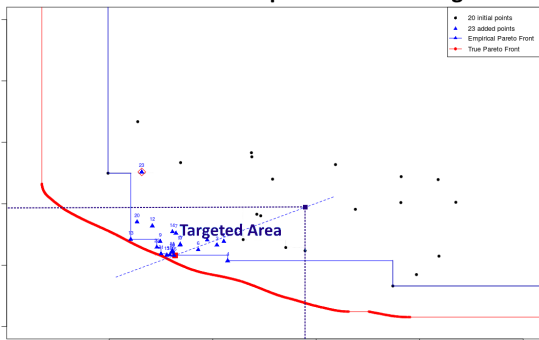
Possible reference point after convergence



Ongoing work : expanding the PF approximation

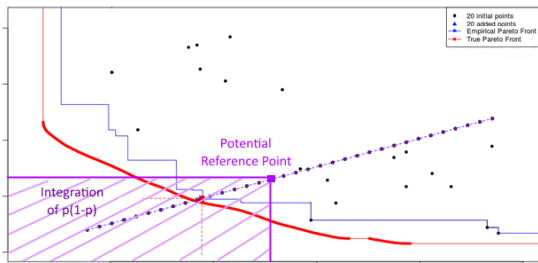
- Expansion of the \mathbf{y} 's considered for improvement around converged area \Rightarrow wider range of solutions
- Well-balanced solutions : use $EHI(\cdot, \mathbf{R})$ with \mathbf{R} along $\hat{\mathcal{L}}$

Possible reference point after convergence



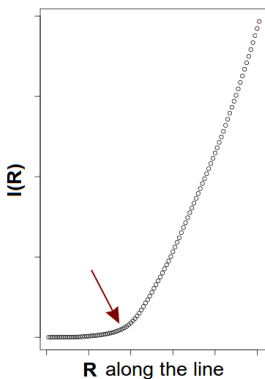
Ongoing work : expanding the PF approximation

- Expansion of the \mathbf{y} 's considered for improvement around converged area \Rightarrow wider range of solutions
- Well-balanced solutions : use $EHI(\cdot, \mathbf{R})$ with \mathbf{R} along $\hat{\mathcal{L}}$
- How to choose \mathbf{R} ? Integration of $p(1-p)$ in the area dominating \mathbf{R} : $I(\mathbf{R}) = \int_{\mathbf{y} \preceq \mathbf{R}} p(1-p) d\mathbf{y}$ for varying \mathbf{R} 's.



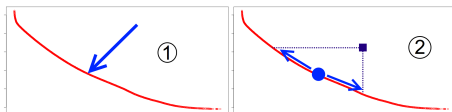
Ongoing work : expanding the PF approximation

- Elbow indicates increase of uncertainty \Rightarrow attractive choice for \mathbf{R}



Summary : A two-step algorithm for targeting well-balanced solutions

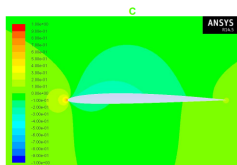
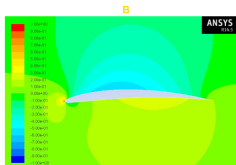
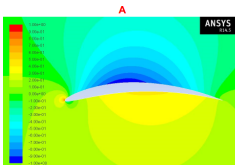
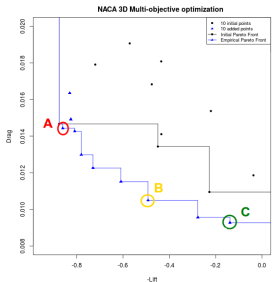
- Compute the estimated NS line $\hat{\mathcal{L}}$: the center of the Pareto Front is expected to be located on this line (1)
- Define a reference point \mathbf{R} combining $\hat{\mathcal{L}}$ and the current approximation front $\widehat{\mathcal{P}}_y$
- Target the estimated central part of \mathcal{P}_y using \mathbf{R} and a targeting infill criterion (e.g. mEI)
- When convergence is detected, widen $\widehat{\mathcal{P}}_y$ defining \mathbf{R} in an adequate way (2)
- \Rightarrow Convergence towards the Pareto Front in the region of well-balanced solutions, in spite of the restricted budget



Next steps

- Finalize the second phase
- Apply the whole algorithm to test cases with $m = 2$ to 4 objectives
- Parallelize the algorithm

Attractive solutions : central solutions



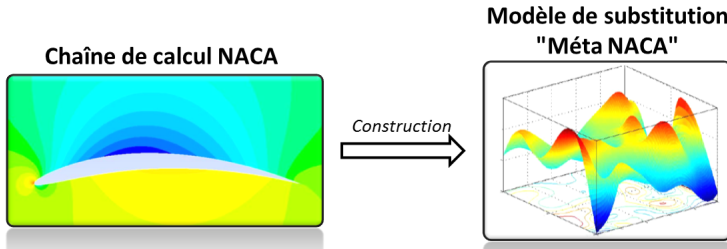
← +

Drag (minimize)

Lift (maximize)

Benchmark : "Meta NACA"

- For each case (dimension $d = 3, 8$ or 22), creation of a surrogate of 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



Nadir-Shadow line for an empirical Pareto Front

- Center of a Pareto Front \Rightarrow Nadir-Shadow line
- Not robust when applied to an approximation Front $\widehat{\mathcal{P}}_y$

