User preferences in Bayesian multi-objective optimization

Paul Féliot[1], Julien Bect[2], Emmanuel Vazquez[2]

[1]Safran Aircraft Engines, (paul.feliot@safrangroup.com)

[2]Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, CNRS, Univ. Paris-Sud, Univ. Paris-Saclay, Gif-sur-Yvette, France (firstname.surname@centralesupelec.fr)

PGMO Days, EDF Lab, November the 14th, 2017

Copyright 2017 Paul Feliot, Julien Bect & Emmanuel Vazquez

You are allowed to share (copy and redistribute the material in any medium or format) this document under the terms of the

licence (https://creativecommons.org/licenses/by-nc-nd/4.0/).

You can cite this work as:

```
@misc{FBV:PGMO:2017,
  author          = {Feliot, Paul and Bect, Julien and Vazquez, Emmanuel},
  title           = {User preferences in Bayesian multi-objective optimization},
  howpublished    = {PGMO Days 2017 (slides)},
  month           = {November},
  year            = {2017}}
```

Outline of the presentation

Constrained multi-objective optimization problem

- Given
  - $\mathbb{X} \subset \mathbb{R}^d$ $\rightarrow$ search space
  - $f = (f_j)_{1 \leq j \leq p}$ $\rightarrow$ objective functions (to be minimized)
  - $c = (c_i)_{1 \leq i \leq q}$ $\rightarrow$ constraint functions

- Build an approximation of the set of feasible and non-dominated solutions:

$$\Gamma = \{x \in \mathbb{X} : c(x) \leq 0 \text{ and } \nexists\, x' \in \mathbb{X} \text{ s.t. } c(x') \leq 0 \text{ and } f(x') \prec f(x)\}$$

- Objectives and constraints computed using numerical simulations (e.g., fluid dynamics, structural analysis...): black-box optimization with expensive-to-evaluate functions.

Practical considerations

- For very expensive numerical simulations,
  - → approximating the entire Pareto front is too expensive,
  - → not all objectives are equally worth improving.
- Computational effort should focus on preferred regions of the Pareto front[1].

[1] See, e.g., Li et al. [2016]

Motivating example (1/2)

- Optimization of an aircraft mission: efficiency on cruise is more interesting than that at other engine regimes → preference for one objective

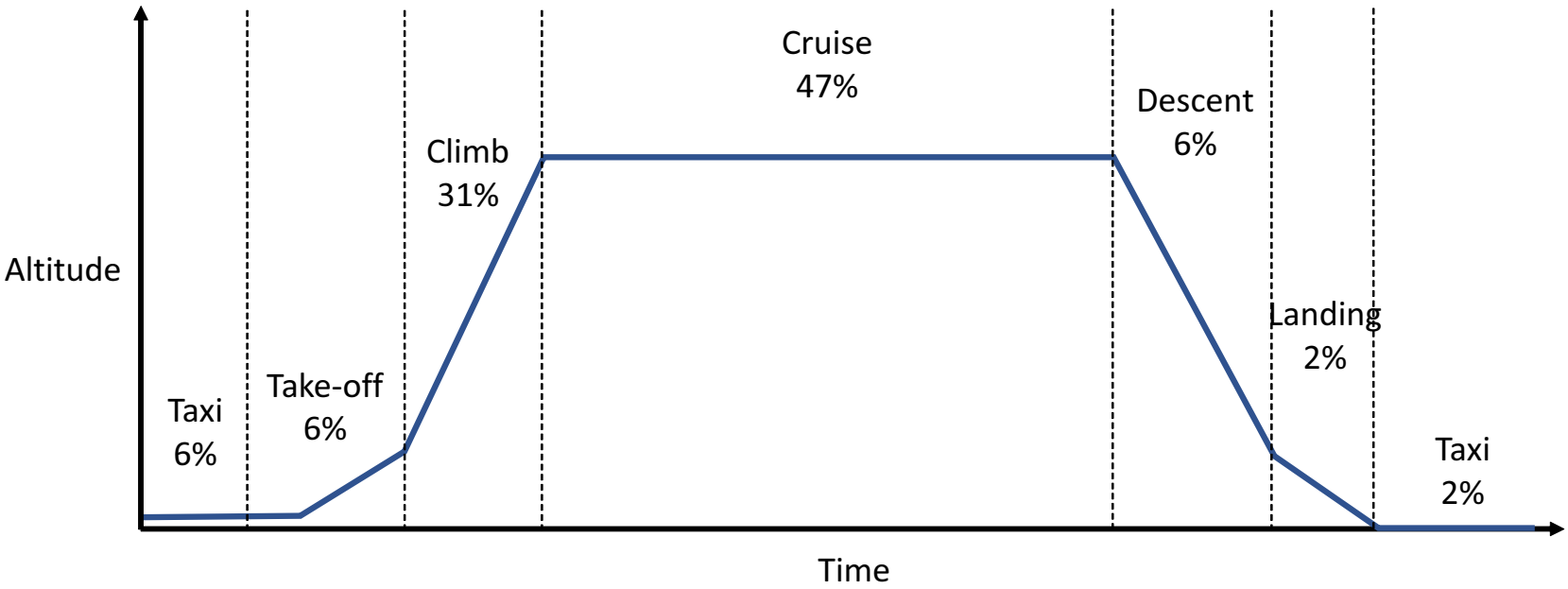


Figure: Typical aircraft mission diagram with associated CO_2 emission share.

Motivating example (2/2)

- Improve upon a reference design → preference for a given region of the Pareto front.

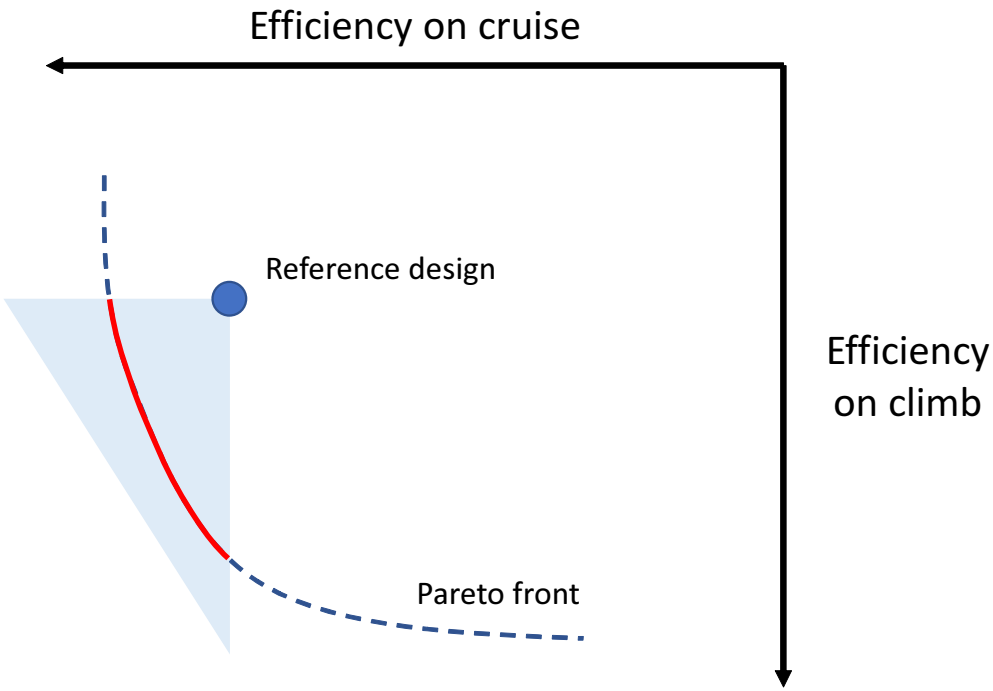


Figure: Efficiency trade-off compared to a reference design.

Statistical model

- Classical Bayesian optimization setting[2]: f modeled using a stationary Gaussian process model ξ.

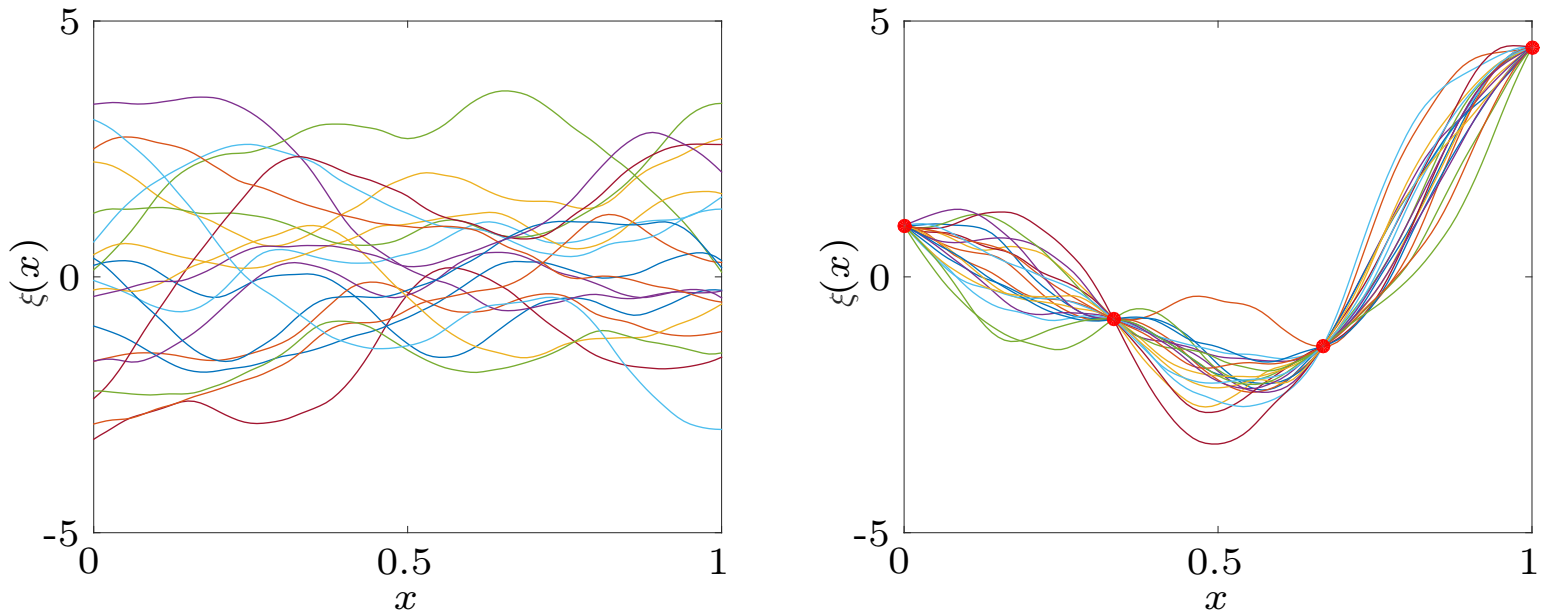


Figure: Sample paths of ξ under a Gaussian process prior distribution (left). Sample paths of ξ when four observations have been made (right)

[2] see, e.g., Santner et al. [2003], Williams and Rasmussen [2006]

Bayesian optimization strategy

- Define

  - $\underline{X} : (X_1, X_2, \ldots) \quad \rightarrow \quad$ optimization strategy
  - $\varepsilon_n \quad \rightarrow \quad$ loss at time $n \geq 1$

- One-step look-ahead optimal choice at time n:

$$X_{n+1} = \underset{x \in \mathbb{X}}{\operatorname{argmax}}\, \rho_n(x),$$

where

$$\rho_n(x) = \mathbb{E}\big(\varepsilon_n - \varepsilon_{n+1} | X_{n+1} = x\big), \quad x \in \mathbb{X}$$

- ρ_n is called the EI (expected improvement) sampling criterion[3].

[3] See, e.g., Mockus et al. [1978], Jones et al. [1998]

Illustration: unconstrained single-objective

- Loss function:

$$\varepsilon_n = m_n - m$$

where

$$\begin{cases} m & = & \min\limits_{\mathbb{X}} f, \\ m_n & = & \min\left(f(X_1), \ldots, f(X_n)\right), \quad n \geq 0. \end{cases}$$

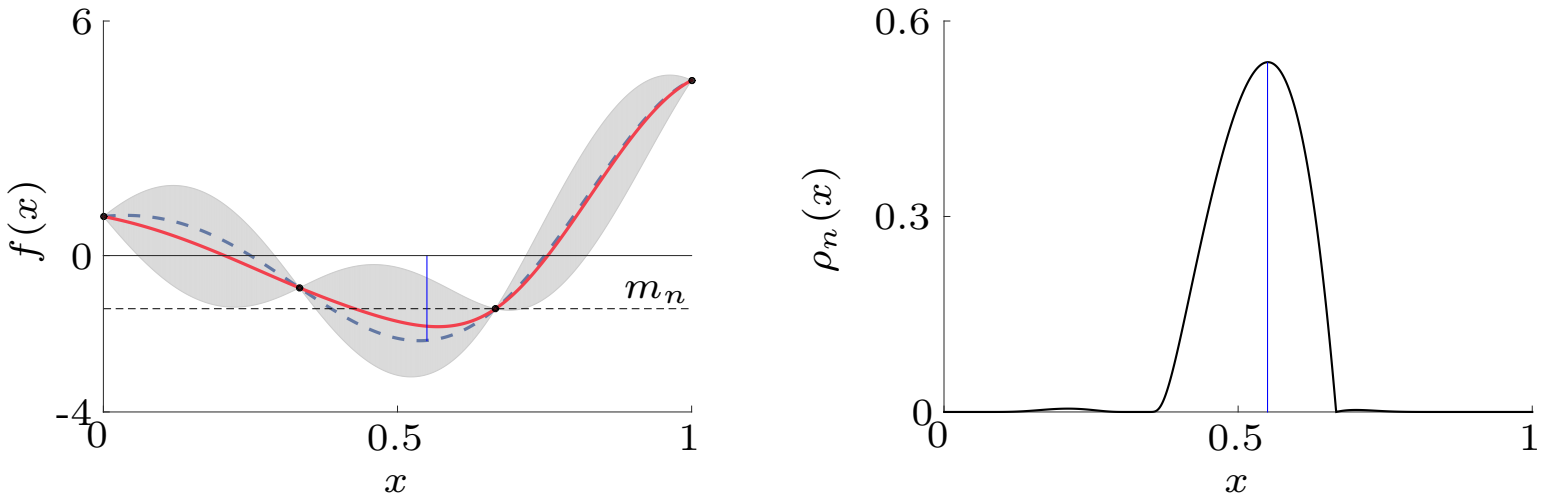


Figure: Bayesian optimization with the EI criterion

Illustration: unconstrained single-objective

- Loss function:

$$\varepsilon_n = m_n - m$$

where

$$\begin{cases} m & = & \min\limits_{\mathbb{X}} f, \\ m_n & = & \min\left(f(X_1), \ldots, f(X_n)\right), \quad n \geq 0. \end{cases}$$

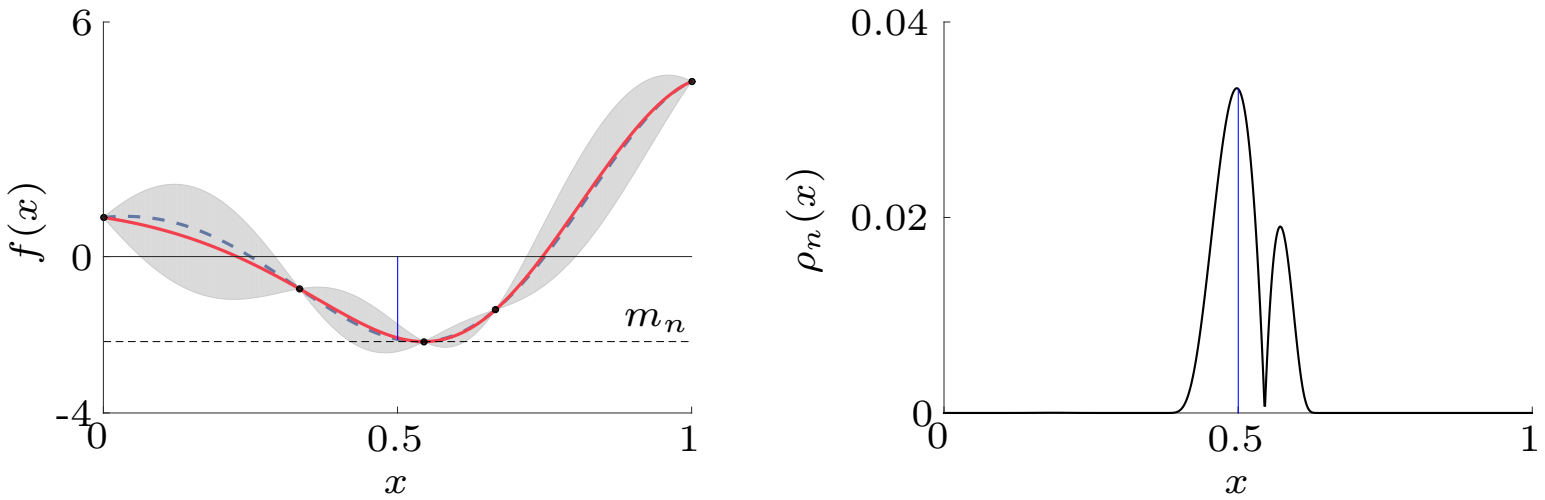


Figure: Bayesian optimization with the EI criterion

Illustration: unconstrained single-objective

- Loss function:

$$\varepsilon_n = m_n - m$$

where

$$\begin{cases} m & = & \min\limits_{\mathbb{X}} f, \\ m_n & = & \min\left(f(X_1), \ldots, f(X_n)\right), \quad n \geq 0. \end{cases}$$

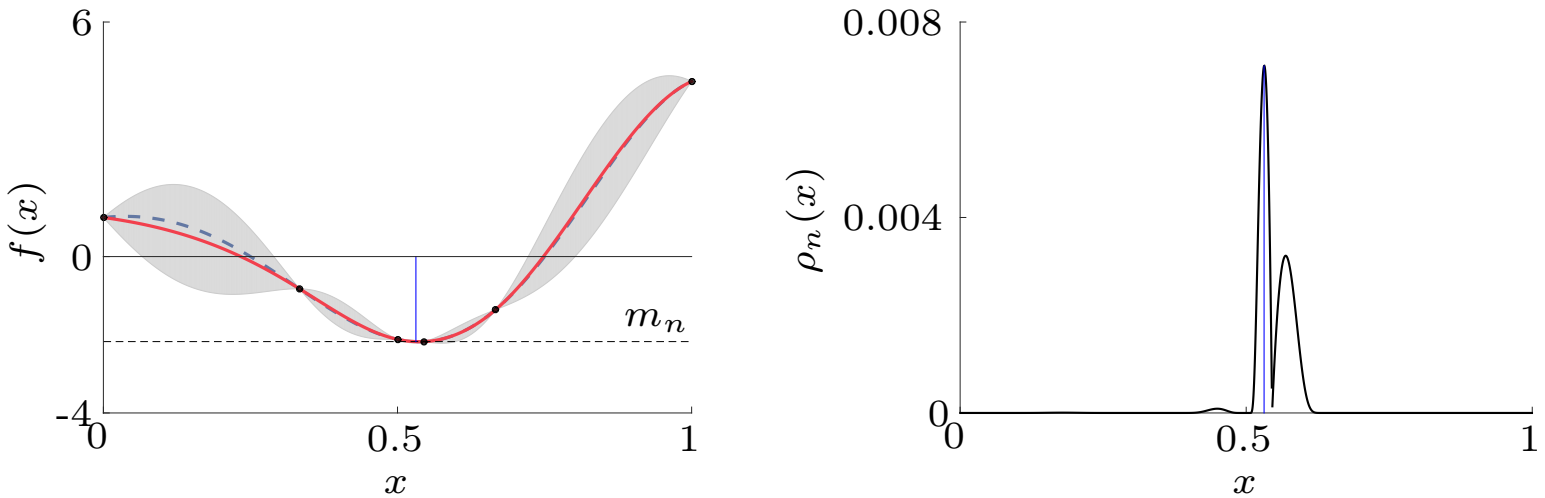


Figure: Bayesian optimization with the EI criterion

Unconstrained multi-objective setting

- Define:

$$\begin{cases} \mathbb{B} & = & \{y \in \mathbb{R}^p;\ y \preceq y^{\text{upp}}\}, y^{\text{upp}} \in \mathbb{R}^p \\ H & = & \{y \in \mathbb{B}; \exists x \in \mathbb{X}, f(x) \prec y\} \\ H_n & = & \{y \in \mathbb{B}; \exists i \leq n,\ f(X_i) \prec y\} \end{cases}$$

- Loss function:

$$\varepsilon_n = |H \setminus H_n| = |H| - |H_n|$$

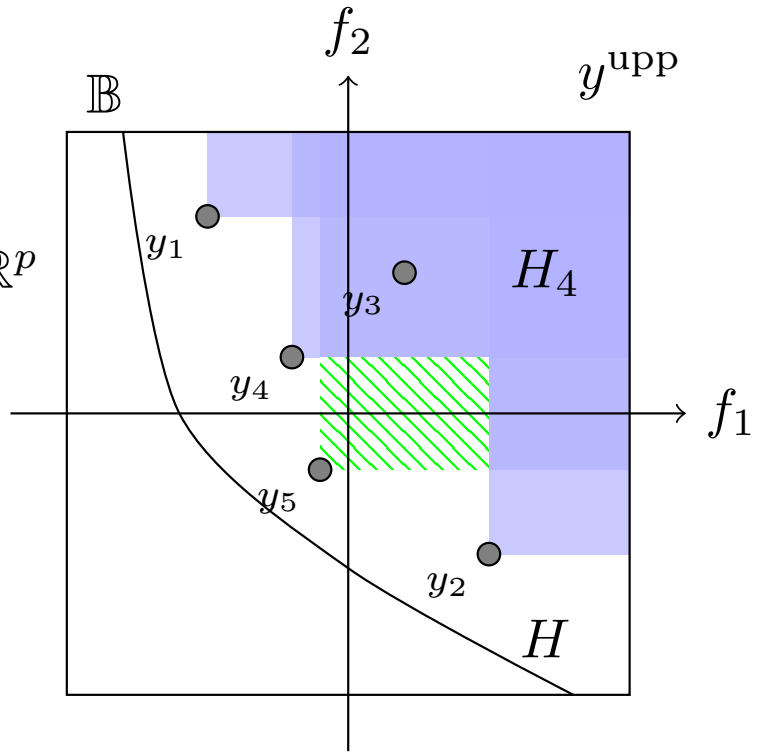


Figure: Hypervolume improvement yielded by the observation of y_5

EHVI (expected hypervolume improvement)

- The EHVI sampling criterion[4] takes the form of an integral over the non-dominated subset of the objectives space:

$$\rho_n(x) = \int_{\mathbb{B}\setminus H_n} \mathbb{P}_n\left(\xi(x) \prec y\right) \mathrm{d}y\,, \quad x \in \mathbb{X}.$$

[4] See Emmerich [2005], Emmerich et al. [2006], Emmerich and Klinkenberg [2008], Bader and Zitzler [2011], Hupkens et al. [2014], Couckuyt et al. [2014]

Weighted hypervolume[5] loss function

- The EHVI corresponds to using the Lebesgue measure of the non-dominated subset
- Our proposal:

$$\varepsilon_n = \mu(H) - \mu(H_n),$$

where μ is a positive measure defined by

$$\mathrm{d}\mu = \omega \mathrm{d}y,$$

for some weight function

$$\omega : \mathbb{R}^p \mapsto \mathbb{R}^+.$$

[5] Zitzler et al. [2007], Friedrich et al. [2013], Emmerich et al. [2014]

EWHI (Expected weighted hypervolume improvement)

- The new criterion can be written as

$$\begin{aligned}
\rho_n(x) &= \mathbb{E}\big(\varepsilon_n - \varepsilon_{n+1} | X_{n+1} = x\big)\,, \\
&= \mathbb{E}_n\big(\mu(H_n \cup \{y \in \mathbb{R}^p; \xi(x) \prec y\}) - \mu(H_n)\big)\,, \\
&= \mathbb{E}_n\left(\int_{H_n^c} \omega(y) \cdot \mathbb{1}_{\xi(x) \prec y}\, \mathrm{d}y\right)\,, \\
&= \int_{H_n^c} \omega(y) \cdot \mathbb{P}_n\left(\xi(x) \prec y\right)\, \mathrm{d}y\,, \quad x \in \mathbb{X}.
\end{aligned}$$

- EHVI is recovered by taking $\omega = \mathbb{1}_{\mathbb{B}}$.

Computation of the criterion

$$\rho_n(x) = \int_{H_n^c} \omega(y) \cdot \mathbb{P}_n \left(\xi(x) \prec y\right) \mathrm{d}y\,, \quad x \in \mathbb{X}.$$

- No closed form expression of the integral is known in the general case
- Our proposal: sequential Monte-Carlo[6] approximation

[6] See,e.g., Del Moral et al. [2006], Liu [2001]

Bi-objective test problem from Chafekar et al. [2003]

$$\begin{cases} (x_1, x_2) & \in & [0,5] \times [0,3]\,, \\ f_1(x_1, x_2) & = & 4x_1^2 + 4x_2^2\,, \\ f_2(x_1, x_2) & = & (x_1 - 5)^2 + (x_2 - 5)^2\,, \\ c_1(x_1, x_2) & = & (x_1 - 5)^2 + x_2^2 - 25\,, \\ c_1(x_1, x_2) & = & 7.7 - (x_1 - 8)^2 - (x_2 + 3)^2\,, \end{cases}$$

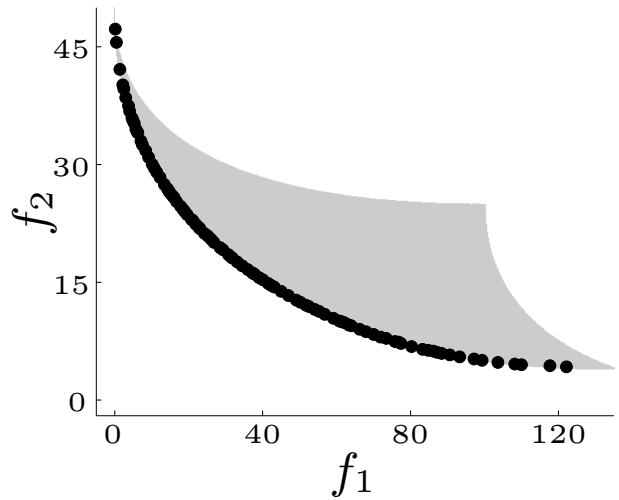


Figure: Pareto front for the BNH problem. In gray the feasible region.

Examples of weight functions

- Define weight functions[7] w_1, w_2 on $\mathbb{R}^2$:

$$\begin{cases} \omega_1(y) &= \frac{1}{15} e^{-\frac{y_1}{15}} \cdot \frac{\mathbb{1}_{[0,150]}(y_1)}{150} \cdot \frac{\mathbb{1}_{[0,60]}(y_2)}{60}, \\ \omega_2(y) &= \frac{1}{2}\big(\varphi\left(y, \mu_1, C\right) + \varphi\left(y, \mu_2, C\right)\big), \end{cases}$$

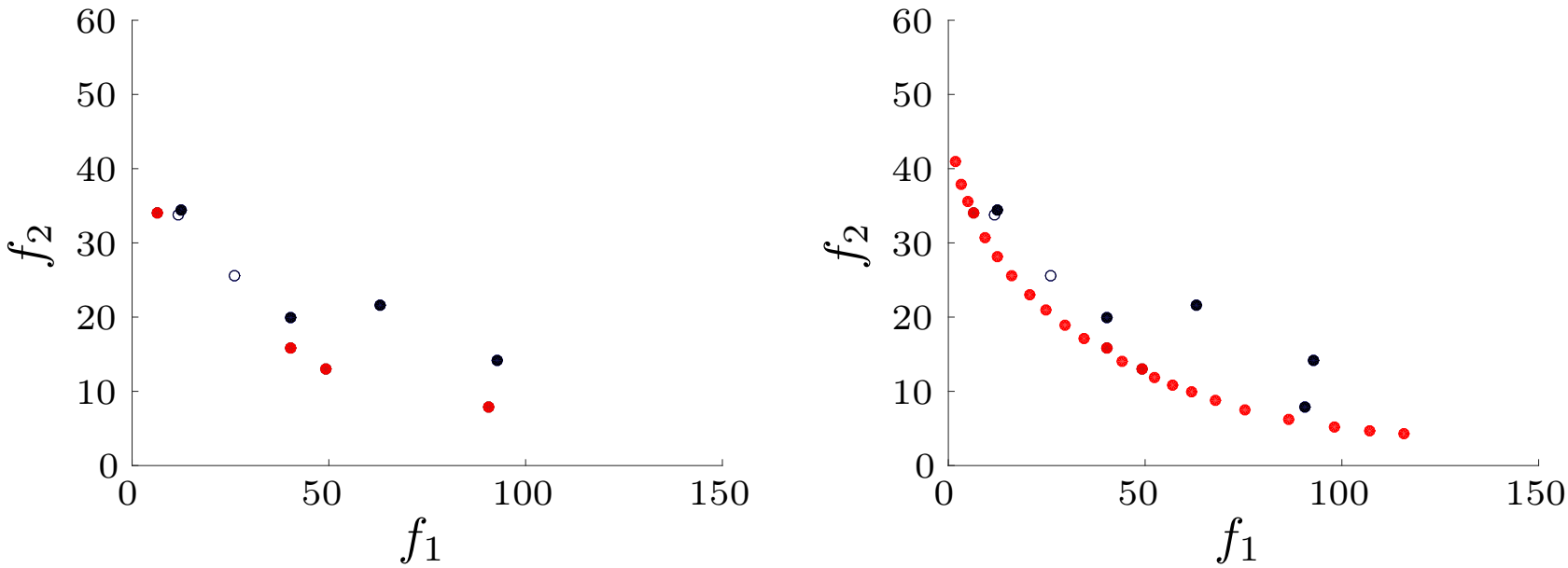
  where $\varphi(y, \mu, C)$ denotes the Gaussian pdf with mean μ and covariance matrix C, evaluated at y.

- Weight function ω_1 encodes preference for the first objective.

- Weight function ω_2 encodes preference for regions centered on μ_1 and μ_2

[7] Zitzler et al. [2007], Auger et al. [2009a]

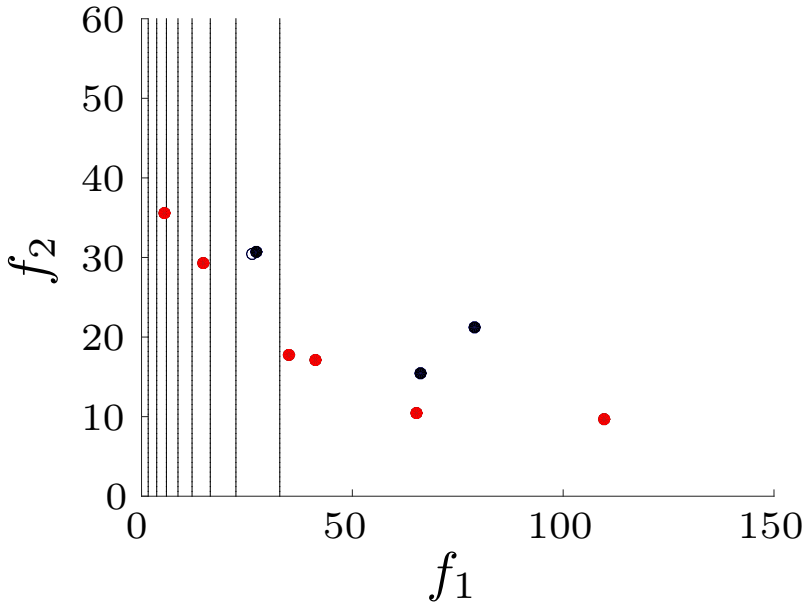
Results obtained using the EHVI criterion



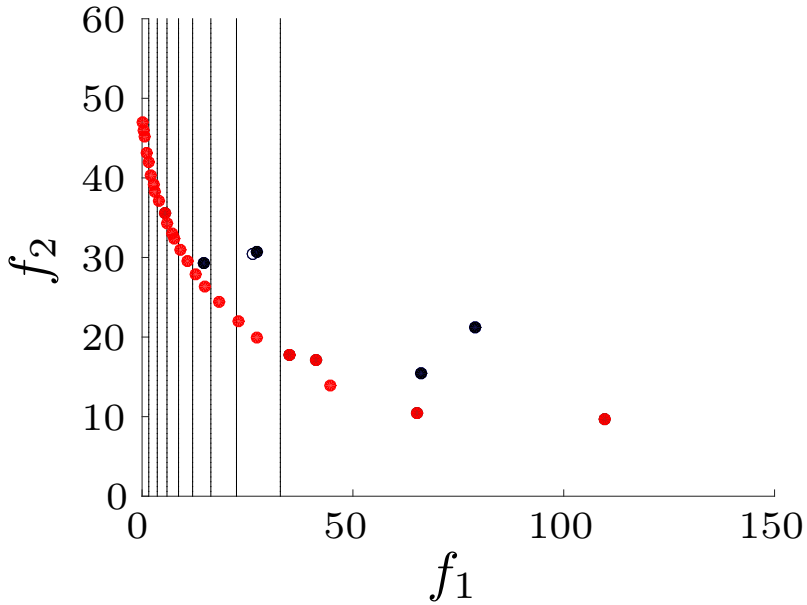
(a) $N = 10$ (b) $N = 30$

Figure: Results obtained using the EHVI criterion

Preference for one objective



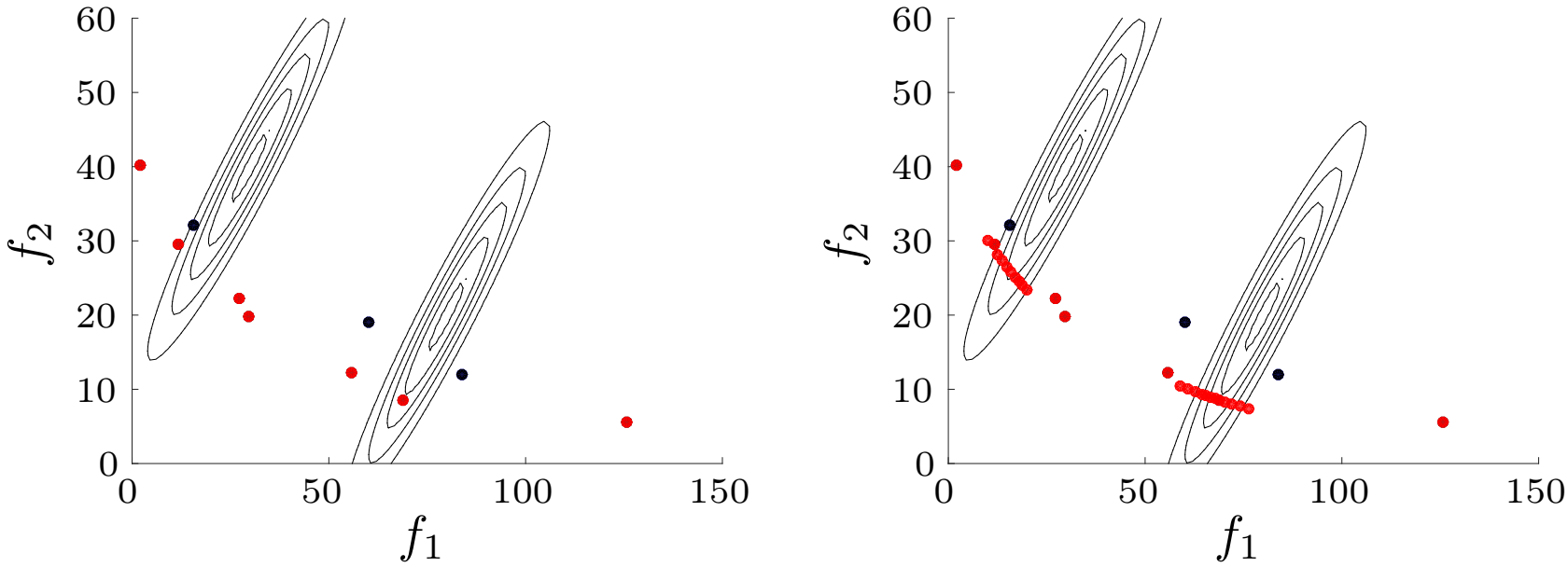
(a) $N = 10$



(b) $N = 30$

Figure: Results obtained using the EWHI criterion with an exponential weight function

Preference for given regions



(a) $N = 10$ (b) $N = 30$

Figure: Results obtained using the EWHI criterion with a Gaussian weight function

Summary of contributions & perspectives

- New EI sampling criterion called EWHI criterion.
- A generalization of the EHVI criterion.
- Makes it possible to encode preference on objectives or regions of the Pareto front[8].
- Crafting of weight functions should be further investigated:
  - → Approaches based on desirability functions[9] are promising.
  - → Theoretical results from Auger et al. [2009b] provide insights as well.

[8] See also the truncated EHVI criterion of Yang et al. [2016a,b]
[9] Wagner and Trautmann [2010], Emmerich et al. [2014]

References I

A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Articulating user preferences in many-objective problems by sampling the weighted hypervolume. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 555–562. ACM, 2009a.

A. Auger, J. Bader, D. Brockhoff, and E. Zitzler. Investigating and exploiting the bias of the weighted hypervolume to articulate user preferences. In *Proceedings of the 11th Annual conference on Genetic and evolutionary computation*, pages 563–570. ACM, 2009b.

J. Bader and E. Zitzler. Hype: An algorithm for fast hypervolume-based many-objective optimization. *Evolutionary Computation*, 19(1):45–76, 2011.

D. Chafekar, J. Xuan, and K. Rasheed. Constrained multi-objective optimization using steady state genetic algorithms. In *Genetic and Evolutionary Computation-GECCO 2003*, pages 813–824. Springer, 2003.

References II

I. Couckuyt, D. Deschrijver, and T. Dhaene. Fast calculation of multiobjective probability of improvement and expected improvement criteria for pareto optimization. *Journal of Global Optimization*, 60(3):575–594, 2014.

P. Del Moral, A. Doucet, and A. Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

M. Emmerich. *Single- and multiobjective evolutionary design optimization assisted by Gaussian random field metamodels.* PhD thesis, Technical University Dortmund, 2005.

M. Emmerich and J. W. Klinkenberg. The computation of the expected improvement in dominated hypervolume of Pareto front approximations. *Technical report, Leiden University*, 2008.

References III

M. Emmerich, K. C. Giannakoglou, and B. Naujoks. Single- and multi-objective evolutionary optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10(4):421–439, 2006.

M. Emmerich, A. H. Deutz, and I. Yevseyeva. On reference point free weighted hypervolume indicators based on desirability functions and their probabilistic interpretation. *Procedia Technology*, 16: 532–541, 2014.

T. Friedrich, T. Kroeger, and F. Neumann. Weighted preferences in evolutionary multi-objective optimization. *International Journal of Machine Learning and Cybernetics*, 4(2):139–148, 2013.

I. Hupkens, M. Emmerich, and A. Deutz. Faster computation of expected hypervolume improvement. *arXiv preprint arXiv:1408.7114*, 2014.

References IV

D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.

L. Li, I. Yevseyeva, V. Basto-Fernandes, H. Trautmann, N. Jing, and M. Emmerich. An ontology of preference-based multiobjective evolutionary algorithms. *arXiv preprint arXiv:1609.08082*, 2016.

J. S. Liu. *Monte Carlo strategies in scientific computing*. Springer, 2001.

J. Mockus, V. Tiesis, and A. Žilinskas. The application of Bayesian methods for seeking the extremum. In L. C. W. Dixon and G. P. Szegö, editors, *Towards Global Optimization*, volume 2, pages 117–129, North Holland, New York, 1978.

T. J. Santner, B. J. Williams, and W. Notz. *The design and analysis of computer experiments*. Springer, 2003.

References V

T. Wagner and H. Trautmann. Integration of preferences in hypervolume-based multiobjective evolutionary algorithms by means of desirability functions. *IEEE Transactions on Evolutionary Computation*, 14(5):688–701, 2010.

C. K. I. Williams and C. Rasmussen. Gaussian processes for machine learning. *the MIT Press*, 2(3):4, 2006.

K. Yang, A. Deutz, Z. Yang, T. Bäck, and M. Emmerich. Truncated expected hypervolume improvement: Exact computation and application. In *Evolutionary Computation (CEC), 2016 IEEE Congress on, Submitted*, 2016a.

K. Yang, L. Li, A. Deutz, T. Back, and M. Emmerich. Preference-based multiobjective optimization using truncated expected hypervolume improvement. In *Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2016 12th International Conference on*, pages 276–281. IEEE, 2016b.

References VI

E. Zitzler, D. Brockhoff, and L. Thiele. The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. In *Evolutionary multi-criterion optimization*, pages 862–876. Springer, 2007.