A Bayesian approach to constrained multi-objective optimization

Paul Feliot, Julien Bect, Emmanuel Vazquez

A Bayesian approach to constrained multi-objective optimization

FELIOT Paul, BECT Julien, VAZQUEZ Emmanuel

Context : problem formulation

$$\begin{cases}\text{Minimize} & f(x) \\ \text{Subject to} & x \in \mathbb{X} \quad \text{and} \quad c(x) \leq 0\end{cases}$$

- $f = (f_j)_{1 \leq j \leq p}$ is a **vector** of objective functions to be minimized.
- $\mathbb{X} \subset \mathbb{R}^d$ is the design space.
- $c = (c_i)_{1 \leq i \leq q}$ is the **vector** of constraint functions.

Both the objective functions f_j and the constraint functions c_i are assumed to be **smooth, nonlinear, expensive-to-evaluate black-box** functions.

Method proposed

The following presents a Bayesian approach for solving constrained multi-objective optimization problems. The functions of the problem are modeled by random processes [5,8,9] and we use an *extended* domination rule taking both constraints and objectives into account in a multi-objective framework to define a proper expected improvement (EI) criterion.

Extended domination rule

Let $y = (y_1, \dots, y_r) \in \mathbb{Y}$ and $z = (z_1, \dots, z_r) \in \mathbb{Y}$, where $\mathbb{Y} \subseteq \mathbb{R}^r$ denotes the output space. We usually say that y dominates z, which is denoted by $y \prec z$, if the following holds:

$$\begin{cases}\forall i \in [1, \mathrm{r}], & y_i \leq z_i \\ \exists j \in [1, \mathrm{r}], & y_j < z_j\end{cases}$$

In order to adapt this domination rule to the constrained setup [3], we define the following function:

$$\begin{cases}\psi : & \mathbb{F} \times \mathbb{C} \to \mathbb{R}^p \times \mathbb{R}^q \\ & (y_f, y_c) \mapsto \begin{cases}(y_f, 0) & \text{if } y_c \leq 0 \\ (+\infty, \max(y_c, 0)) & \text{otherwise}\end{cases}\end{cases}$$

where $\mathbb{F} \subset \mathbb{R}^p$ and $\mathbb{C} \subset \mathbb{R}^q$ are respectively the objective and constraint spaces. The new domination rule $\lhd$ is then defined by

$$y \lhd z \Leftrightarrow \psi(y) \prec \psi(z), \qquad y, z \in \mathbb{F} \times \mathbb{C}$$

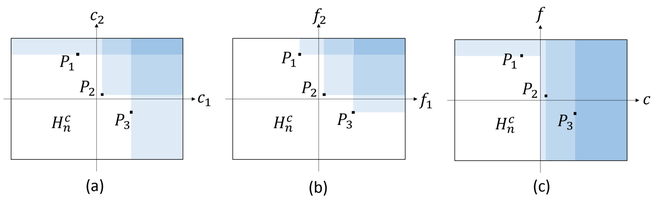


Fig. 1. Illustration of the extended domination rule in different situations. The region dominated by each point is represented by a shaded area. Darker shades of gray indicate overlapping regions. (a) Feasible solutions are compared with respect to their objective values using the usual domination rule in the objective space. (b) Non-feasible solutions are compared using the Pareto domination rule applied to the vectors of constraint violations. (c) Feasible solutions always dominate non-feasible solutions; other cases are handled as in the first two figures.

Expected Improvement

Much like [2,4,10], we then define the improvement yielded by the observation of a new point as the volume increase of the dominated region (see figure 2 bellow).

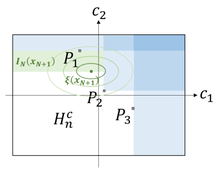


Fig. 2. Illustration of the improvement I_n yielded by the observation of a new point at location x_{n+1}

Denote by H_n the subset of $\mathbb{Y}$ dominated by the observations $X_n = \{x_1, \dots, x_n\} \in \mathbb{X}^n$ and $|H_n|_{\mathbb{Y}}$ its volume.

$$\forall x_{n+1} \in \mathbb{X}, I_n(x_{n+1}) = |H_{n+1}|_{\mathbb{Y}} - |H_n|_{\mathbb{Y}}$$

The Expected Improvement follows by taking the expectation of the above defined improvement:

$$EI_n(x) = \int_{H_n^c} \mathbb{P}_n(\xi(x) \lhd y)\, \mathrm{d}y,$$

where $\mathbb{P}_n$ denotes the probability conditioned on the observations and ξ is a vector of Gaussian process priors associated with the functions of the problem.

Criterion calculation

The integrand of the EI formula can be calculated **in closed form** for any $x \in \mathbb{X}$ and any $y = (y_o, y_c) \in \mathbb{Y}$

$$\mathbb{P}_N(\xi(x) \lhd y) = \begin{cases}\left(\prod_{i=1}^{p} \phi\left(\frac{y_{o,i} - \hat{\xi}_{o,i,n}(x)}{\sigma_{o,i,n}(x)}\right)\right)\left(\prod_{j=1}^{q} \phi\left(-\frac{\hat{\xi}_{o,j,n}(x)}{\sigma_{o,j,n}(x)}\right)\right) & \text{if } y_c \leq 0 \\ \prod_{j=1}^{q} \phi\left(\frac{\max(y_{c,j}, 0) - \hat{\xi}_{o,j,n}(x)}{\sigma_{o,j,n}(x)}\right) & \text{otherwise}\end{cases}$$

Its integration over $\mathbb{Y}\backslash\mathrm{H}_n$ however is not trivial. In particular, we cannot use decomposition methods due to the dimension of $\mathbb{Y}$ [12]. To address this difficulty, we propose to use a **Monte Carlo approximation** of the integral:

$$EI_n(x) \approx \frac{1}{m} \sum_{k=1}^{m} \mathbb{P}_n(\xi(x) \lhd y_k),$$

where $\mathcal{Y}_n = (y_{n,k})_{1 \leq k \leq m}$ is a set of particles targeting the uniform density over $\mathbb{Y}\backslash\mathrm{H}_n$.

In principle, sampling uniformly over $\mathbb{Y}\backslash\mathrm{H}_n$ could be achieved using an accept-reject method [13]. However, when the dimension of $\mathbb{Y}$ is high, $\mathbb{Y}\backslash\mathrm{H}_n$ will probably have a small volume with respect to that of $\mathbb{Y}$. In this case, the acceptance rate of an accept–reject method becomes small and the cost of generating a uniform sample on $\mathbb{Y}\backslash\mathrm{H}_n$ becomes prohibitive.

In this work, we use a variant of a technique called **subset simulation**[11,14]. Given a set of particles $\mathcal{Y}_n = (y_{n,k})_{1 \leq k \leq m}$ uniformly distributed on $\mathbb{Y}\backslash\mathrm{H}_n$, one obtains a sample $\mathcal{Y}_{n+1}$ uniformly distributed over $\mathbb{Y}\backslash\mathrm{H}_{n+1}$ using the following Remove-Resample-Move procedure:

1) Remove particles $y_{n,k} \in \mathcal{Y}_n$ that are not in $\mathbb{Y}\backslash\mathrm{H}_{n+1}$
2) Replicate randomly the surviving particles until the population size is m again
3) Move particles using a Metropolis-Hastings algorithm with invariant density $\mathbb{1}_{\mathbb{Y}\backslash\mathrm{H}_{n+1}}$

This procedure is illustrated in the illustration panel.

Criterion optimization

The optimization of the criterion is also a difficult problem because the EI criterion is known to be **highly multi-modal** and hard to optimize. Our proposal is to conduct the optimization on a restrained set of good candidates provided at each iteration by a **Sequential Monte Carlo approach** in the spirit of [1,6]. This enables us to use results from past iterations to save computational time and converge hopefully to higher precisions.

Let $\mathcal{X}_n = (x_{n,k}, w_{n,k})_{1 \leq k \leq m}$ be a cloud of weighted particles distributed according to $\pi_n(x) \propto \mathbb{P}_n(\xi(x) \in \mathbb{Y}\backslash\mathrm{H}_n)$ (this is actually the probability of improvement). When a new sample is observed, the incremental weights of the particles can be updated to fit the new density:

$$\widetilde{w}_{k,n+1} = \frac{\pi_{n+1}(x_{n,k})}{\pi_n(x_{n,k})} \left(\sum_{k=1}^{m} \frac{\pi_{n+1}(x_{n,k})}{\pi_n(x_{n,k})}\right)^{-1}$$

The incremental weights can be used to resample the particles in the regions of high density for π_{n+1}. Then a Metropolis-Hastings algorithm with invariant density π_{n+1} is used to mix the particles. Below is a summary of the proposed method [7]:

1) Reweight particles using the incremental weights
2) Resample with a residual resampling scheme
3) Move the new particles with a metropolis Hastings algorithm

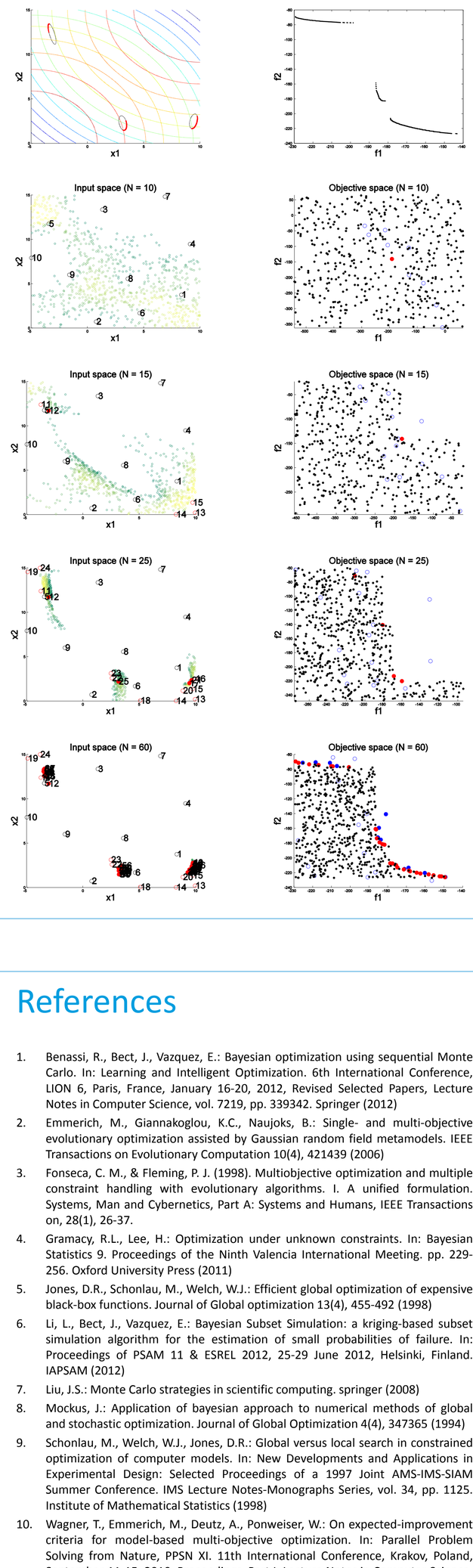
In practice, the probability of improvement can't be calculated analytically in the general case. However, the posterior density of $\xi(x)$ is Gaussian and it can be simulated empirically. If $(z_k)_{1 \leq k \leq N}$ are N simulations of $\xi(x)$, an estimator of $\mathbb{P}_n(\xi(x) \in \mathbb{Y}\backslash\mathrm{H}_n)$ is the following counting measure:

$$\mathbb{P}_n(\xi(x) \in \mathbb{Y}\backslash\mathrm{H}_n) \approx \sum_{k=1}^{N} \mathbb{1}_{\mathbb{Y}\backslash\mathrm{H}_n}(z_k)$$

The optimization method is illustrated in the illustration panel.

Illustration

The method is illustrated on a 2D toy problem with two objective functions and one constraint function. The feasible subset consists in three small regions (see figure below) and the Pareto front is composed of three corresponding disconnected fronts. In particular, note that no feasible point is given to begin with.



References

1. Benassi, R., Bect, J., Vazquez, E.: Bayesian optimization using sequential Monte Carlo. In: Learning and Intelligent Optimization. 6th International Conference, LION 6, Paris, France, January 16-20, 2012, Revised Selected Papers, Lecture Notes in Computer Science, vol. 7219, pp. 339342. Springer (2012)
2. Emmerich, M., Giannakoglou, K.C., Naujoks, B.: Single- and multi-objective evolutionary optimization assisted by Gaussian random field metamodels. IEEE Transactions on Evolutionary Computation 10(4), 421439 (2006)
3. Fonseca, C. M., & Fleming, P. J. (1998). Multiobjective optimization and multiple constraint handling with evolutionary algorithms. I. A unified formulation. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 28(1), 26-37.
4. Gramacy, R.L., Lee, H.: Optimization under unknown constraints. In: Bayesian Statistics 9. Proceedings of the Ninth Valencia International Meeting. pp. 229-256. Oxford University Press (2011)
5. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. Journal of Global optimization 13(4), 455-492 (1998)
6. Li, L., Bect, J., Vazquez, E.: Bayesian Subset Simulation: a kriging-based subset simulation algorithm for the estimation of small probabilities of failure. In: Proceedings of PSAM 11 & ESREL 2012, 25-29 June 2012, Helsinki, Finland. IAPSAM (2012)
7. Liu, J.S.: Monte Carlo strategies in scientific computing. springer (2008)
8. Mockus, J.: Application of bayesian approach to numerical methods of global and stochastic optimization. Journal of Global Optimization 4(4), 347365 (1994)
9. Schonlau, M., Welch, W.J., Jones, D.R.: Global versus local search in constrained optimization of computer models. In: New Developments and Applications in Experimental Design: Selected Proceedings of a 1997 Joint AMS-IMS-SIAM Summer Conference. IMS Lecture Notes-Monographs Series, vol. 34, pp. 1125. Institute of Mathematical Statistics (1998)
10. Wagner, T., Emmerich, M., Deutz, A., Ponweiser, W.: On expected-improvement criteria for model-based multi-objective optimization. In: Parallel Problem Solving from Nature, PPSN XI. 11th International Conference, Krakov, Poland, September 11-15, 2010, Proceedings, Part I. Lecture Notes in Computer Science, vol. 6238, pp. 718727. Springer (2010)
11. Au, S.-K., Beck, J.L.: Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic Engineering Mechanics, 16(4):263277, 2001.
12. Hupkens, I., Emmerich, M., Deutz, A.: Faster computation of expected hypervolume improvement. arXiv preprint arXiv:1408.7114, 2014.
13. Robert, C., Casella, G.: Monte Carlo statistical methods. Springer, 2013.
14. Cérou, F., Del Moral, P., Furon, T., Guyader, A.: Sequential monte carlo for rare event estimation. Statistics and Computing, 22(3):795808, 2012.