The Kalai-Smorodinski solution for many-objective Bayesian optimization

Mickaël Binois, Victor Picheny, Abderrahmane Habbal

The Kalai-Smorodinski solution for many-objective Bayesian optimization

M. Binois
University of Chicago
mbinois@chicagobooth.edu

V. Picheny
MIAT, Université de Toulouse, INRA
victor.picheny@inra.fr

A. Habbal
Laboratoire J.A. Dieudonne, University Côte d'Azur, Inria, CNRS
habbal@unice.fr

Abstract

An ongoing scope of research in multi-objective Bayesian optimization is to extend its applicability to a large number of objectives. Recovering the set of optimal compromise solution generally requires lots of observations while being less interpretable, since this set tends to grow larger with the number of objectives. We thus propose to focus on a choice of a specific solution originating from game theory, the Kalai-Smorodinski solution, that possesses attractive properties. We further make it insensitive to a monotone transformation of the objectives by considering the objectives in the copula space. A tailored algorithm is proposed to search for the solution, which is tested on a synthetic problem.

1 Introduction

We consider here the problem of minimizing simultaneously a set of p objectives with respect to a set of input variables over a bounded domain $\mathbb{X} \subset \mathbb{R}^d$: $\min_{\mathbf{x}} \left\{ y^{(1)}(\mathbf{x}), \dots, y^{(p)}(\mathbf{x}) \right\}$. We assume that the $y^{(i)} : \mathbb{X} \to \mathbb{R}$ functions are expensive to compute, non-linear and potentially observed in noise. The usual goal of multi-objective optimization (MOO) is to uncover the Pareto set, that is, the subset $\mathbb{X}^* \subset \mathbb{X}$ containing all the Pareto non-dominated solutions.

Most of the well-established algorithms (either evolutionary [1], descent-based [2], NNCM [3] or BO [4]) perform quite well on two or three objectives problems, but poorly when $p \geq 4$, the so-called many-objective optimization (MaO). Indeed, difficulties inherent to the higher dimension of the objective space [5] arise, such as the exponential increase in the number of points necessary to approximate the –possibly singular– Pareto front hyper-surface, and the difficulties in its graphical representation. Besides, one has to deal with a more MaO intrinsic problem, which is that almost any admissible design becomes non dominated.

To circumvent these issues, some authors propose methods to reduce the number of objectives to a manageable one [6], or use the so-called decomposition based approaches [7]. Our present proposition amounts to search for a single, but remarkable in some sense, solution. To do so, we adopt a game-theoretic perspective, where the selected solution arises as an equilibrium in a non-cooperative game played by p virtual agents who own the respective p objectives [8]. In the following, we show that the so-called Kalai-Smorodinsky (KS) solution [9] is a very appealing alternative.

Intuitively, solutions at the 'center' of the Pareto front are somewhat more appealing than those at extremities. Yet, this is very arbitrary as transforming the objectives (non-linearly, e.g., with a log scale) would affect the decision. Still, most MOO methods are sensitive to a rescaling of the

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

objective, which is not desirable [10]. Our second proposition is to make the KS solution insensitive to monotone transformations, by operating in the copula space [11, 12].

Uncovering the KS solution is a non-trivial task, for which, up to our knowledge, no available algorithm is available in an expensive black-box framework. Our third contribution is a novel Gaussian-process based algorithm, based on the Stepwise Uncertainty Reduction (SUR) paradigm [13]. SUR, which is closely related to information-based approaches [14], has proven to be efficient to solve single and multi-objective optimization problems [15, 16], while enjoying strong asymptotic properties [17]. The proposed algorithm is implemented in the `R` package `GPGame`, available on CRAN at `https://cran.r-project.org/package=GPGame`.

2 The Kalai-Smorodinski solution

2.1 The standard KS solution

The Kalai-Smorodinsky solution (KS) was first proposed by [9] as an alternative to the Nash bargaining solution in cooperative bargaining. The problem is as follows: starting from a *disagreement* or *status quo* point D in the objective space, the players aim at maximizing their own benefit while moving from D towards the Pareto front (i.e., the efficiency set). The KS solution is of egalitarian inspiration [18] and states that the selected efficient solution should yield equal benefit ratio to all the players. Indeed, given the utopia (or ideal, or shadow) point $U \in \mathbb{R}^p$ defined by $U_i = \min_{\mathbf{x}\in\mathbb{X}^*} \; y^{(i)}(\mathbf{x})$, selecting any compromise solution S would yield, for objective i a benefit ratio $r^{(i)}(S) = \frac{y^{(i)}(D)-y^{(i)}(S)}{y^{(i)}(D)-y^{(i)}(U)}$. Notice that the benefit from staying at D is zero, while it is maximal for the, generically unfeasible, choice $S = U$. The KS solution is the Pareto optimal choice S^* for which all the benefit ratios $r^{(i)}(S)$ are equal. It is straightforward to show that S^* is the intersection point of the Pareto front and the line (D, U).

For $p \geq 3$, the KS solution, defined as the intersection point above, fulfills some of the bargaining axioms: Pareto optimality, affine invariance and equity in benefit ratio [9]. It is particularly attractive in a many-objective context since it scales naturally to a large number of objectives and returns a single solution, avoiding the difficulty of exploring and approximating large p-dimensional Pareto fronts.

The KS solution is known to strongly depend on the choice of the disagreement point D. A standard choice is the Nadir point N given by $N_i = \max_{\mathbf{x}\in\mathbb{X}^*} \; y^{(i)}(\mathbf{x})$. Some authors introduced alternative definitions, called extended KS, to alleviate the dependence on choice of D [19].

2.2 Robust KS using copulas (CKS)

A drawback of KS is that it is not invariant by a monotone (non affine) transformation of the objectives. In a less cooperative framework, some players could be tempted to rely on such transformations to influence the choice of a point on the Pareto front. To circumvent this problem, we make use of copula theory, which has been linked to Pareto optimality in [12]. In short, the Pareto front is related to the zero level-line of the multivariate cumulative density function F_Y of $Y = (y_1(X), \ldots, y_p(X))$ with X a random vector with support in $\mathbb{X}$ (for instance uniformly distributed). A copula C is a function linking a multivariate CDF to its univariate counterparts, i.e., such that for $\mathbf{y} \in \mathbb{R}^p$, $F_Y(\mathbf{y}) = C(F_1(y_1), \ldots, F_p(y_p))$ with $F_i = \mathbf{P}(Y_i \leq y_i)$, $1 \leq i \leq p$ [11]. As shown in [12], learning the marginal CDFs F_i's and the copula C amounts to learning the Pareto front.

A remarkable property of copulas is their invariance under monotone increasing transformations of univariate CDFs [11]. Now, considering the multi-objective problem in the copula space, that is, $F_1, \ldots, F_p$, the corresponding KS (henceforth CKS) solution becomes invariant to such transformations. In addition, the utopia point is always $(0, \ldots, 0)$. While the Nadir remains unknown, the point $(1, \ldots, 1)$ may serve as an alternative disagreement point, as it corresponds to the worst solution for each objective. This removes the difficult task of learning the D-U line. Let us finally remark than when $\mathbb{X}$ is finite, then our proposed solution amounts to work on ranks: the CKS solution is simply the Pareto optimal solution with the closest ranks over all objectives.

2.3 Illustration

Figure 1 shows the KS and CKS solutions on a two-variables, bi-objective problem (P1, see [20]). The two objectives are computed over a 41×41 regular grid and the corresponding 1,681 points are reported in the objective and copula spaces. The KS solution has a central position in the Pareto front, which makes it a well-balanced compromise. The $Y_1 \rightarrow F_1$ transformation is mostly driven by the high density of points in the interval $[0, 50]$. The CKS solution seems here to favor the first objective, but would roughly be equivalent to the KS solution under a log transformation of Y_1.

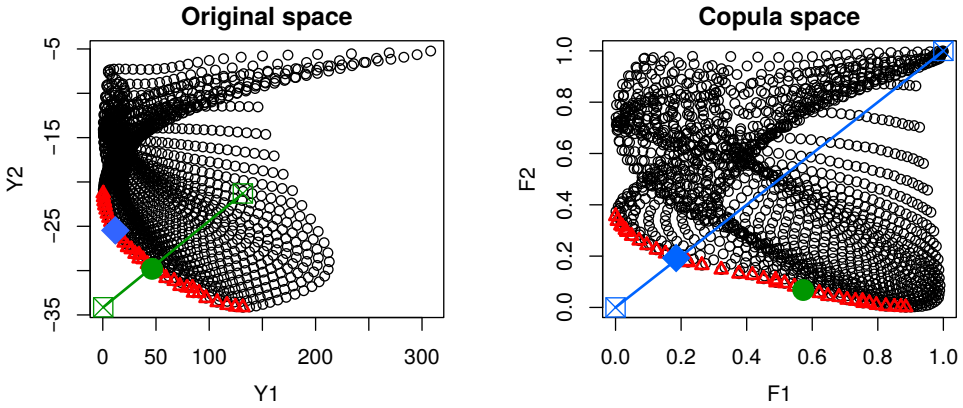

Figure 1: KS (green disk) and CKS (blue diamond) solutions for a bi-objective problem. The black circles show all the dominated values of the grid and the red triangles the Pareto-optimal ones. The D-U lines are shown with matching colors.

3 A Bayesian optimization algorithm to find Kalai-Smorodinski solutions

Computing the KS and CKS solutions is a very challenging problem, as it requires for KS learning the ideal and disagreement points U and D (which are challenging problems on their own, see for instance [21]) and for CKS the marginals and copula, as well as the part of the Pareto front that intersects the D-U line.

For our algorithm, we consider a classical BO framework, where independent Gaussian processes (GPs) priors are put on the objectives: $Y_i(.) \sim \mathcal{GP}\left(\mu_i(.), \Sigma_i(.,.)\right), \forall i \in 1, \ldots, p$, where the mean μ_i and covariance Σ_i have pre-determined parametric forms whose parameters are estimated by maximum likelihood [22]. Conditioning on a set of observations, GPs provide flexible response fits associated with uncertainty estimates. They enable operating sequential design decisions via an *acquisition function* $J(\mathbf{x})$, which balances between exploration and exploitation in seeking global optima. Hence, the design consists of a first set of n_0 observations generated using a space-filling design to obtain a first predictive distribution of $\mathbf{Y}(.)$, and a second set of sequential observations chosen as $\mathbf{x}_{n+1} \in \arg\max_{\mathbf{x} \in \mathbb{X}} J(\mathbf{x})$ $(n \geq n_0)$.

To tailor $J(\mathbf{x})$ to our problem, we follow a SUR approach, whose principle is to reduce as quickly as possible an uncertainty measure related to the quantity of interest (in the present case: the equilibrium). The approach detailed here is similar to the one proposed in [23] to solve Nash equilibria problems. Let us first denote by $\Psi(\mathbf{y})$ the application that associates a KS solution with a multivariate function. If we consider the random process $\mathbf{Y}$ (or $\mathbf{F}$ for CKS) in lieu of the deterministic objective $\mathbf{y}$, the equilibrium $\Psi(\mathbf{Y})$ (resp. $\Psi(\mathbf{F})$) is a random vector of $\mathbb{R}^p$ with unknown distribution. We represent the uncertainty regarding our knowledge of the equilibrium as

$$\Gamma(\mathbf{Y}) = \det\left[\mathrm{cov}\left(\Psi(\mathbf{Y})\right)\right]. \tag{1}$$

Intuitively, $\Gamma(\mathbf{Y})$ (resp. $\Gamma(\mathbf{F})$) tends to zero when all the components of $\Psi(\mathbf{Y})$ become known accurately. The SUR strategy aims at reducing Γ by adding sequentially observations $\mathbf{y}(\mathbf{x})$ on which $\mathbf{Y}$ is conditioned. To do so, we propose as an acquisition function to minimize:

$$J(\mathbf{x}_{n+1}) = \mathbb{E}\left(\Gamma\left[\mathbf{Y} | \mathbf{Y}(\mathbf{x}_{n+1})\right]\right), \tag{2}$$

where $\mathbf{Y}$ is the GP conditioned on the n current observations and a new random observation event $\mathbf{Y}(\mathbf{x}_{n+1})$ that follows the posterior distribution.

In practice, computing $J(\mathbf{x})$ is a complex task, as no closed-form expression is available, and one must rely on Monte-Carlo approaches. We refer to [23], Section 3.3 for the technical details, since the same algorithm is used here. In short, computing $J(\mathbf{x})$ involves generating posterior samples of the GPs indexed on discretizations of $\mathbb{X}$, and conditioning further such samples on virtual observations that are drawings of $\mathbf{Y}(\mathbf{x}_{n+1})$. The optimization of $J(\mathbf{x})$ is done over the discretized $\mathbb{X}$. On both cases, the computational burden is contained by restricting $\mathbb{X}$ to "useful" elements.

4 Results

As a proof of concept, we consider the DTLZ2 function [1], with 5 variables and 4 objectives, that has a concave dome-shaped Pareto front. We consider a finite candidate set $\mathbb{X}$ with 10^5 elements (uniformly distributed in $[0,1]^5$), which allows the computation of reference KS and CKS solutions.

For KS, we use $n_0 = 50$ and 50 infill points, and for CKS, $n_0 = 80$ and 20 infills, since previous tests (not shown) indicated that on this problem KS is more efficient with a larger sequential budget while CKS is more stable with a large initial design. For $J(\mathbf{x})$ computation and optimization, the space $\mathbb{X}$ is discretized over a 1000-point LHS, renewed at each iteration, which was found as a satisfactory trade-off between speed and accuracy. The other parameters are the default ones of `GPGame`, see [24] for details.

Results on one run are given in Figure 2 in the form of projections on the marginal 2D spaces. Notice the central location of the KS point on this problem (Figure 2, first row), while the CKS point leans toward areas of larger densities (for instance, second line, fifth plot, CKS is close to the upper left corner). For KS, new points are added close to the actual solution but some are also more exploratory, in particular near the individual minimas to reduce uncertainty on the D-U line. For CKS, the behavior is more local, with points mostly added around the actual solution. This can also be explained by a larger n_0.

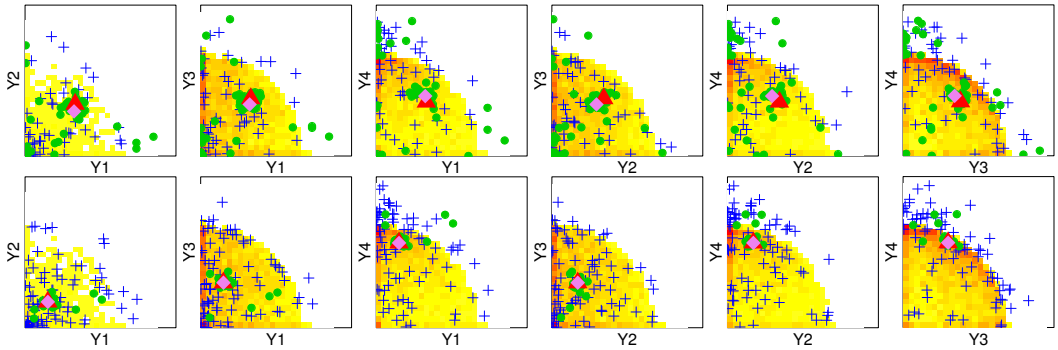

Figure 2: Results of one run for KS (top) and CKS (bottom), represented using marginal 2D projections. The blue crosses are the initial design points, the green points the added designs, the red triangles the target true equilibria and pink diamonds the predicted ones. The heatmap represents the density of Pareto front points on the projected spaces.

Convergence results are provided in Figure 3 in terms of Euclidean distance to the actual equilibrium. The optimum is found in exactly in 2 out of 10 tests for KS and 7 out of 10 for CKS, and approximately on all runs, despite the very restricted budgets.

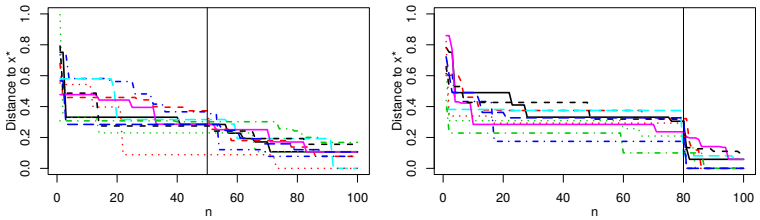

Figure 3: Results on 10 runs of KS and CKS, Euclidean distance to the best design as a function of number of evaluations. Thin vertical lines mark the start of the sequential procedure.

Acknowledgments

The authors acknowledge inspiration from Lorentz Center Workshop "SAMCO - Surrogate Model Assisted Multicriteria Optimization", at Leiden University Feb 29 - March 4, 2016. M. B. is grateful for support from National Science Foundation grant DMS-1521702.

References

[1] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

[2] Indraneel Das and John E Dennis. Normal-boundary intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. *SIAM Journal on Optimization*, 8(3):631–657, 1998.

[3] Achille Messac, Amir Ismail-Yahaya, and Christopher A Mattson. The normalized normal constraint method for generating the Pareto frontier. *Structural and multidisciplinary optimization*, 25(2):86–98, 2003.

[4] 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.

[5] Hisao Ishibuchi, Noritaka Tsukamoto, and Yusuke Nojima. Evolutionary many-objective optimization: A short review. In *Evolutionary Computation, 2008. CEC 2008.(IEEE World Congress on Computational Intelligence). IEEE Congress on*, pages 2419–2426. IEEE, 2008.

[6] Hemant Kumar Singh, Amitay Isaacs, and Tapabrata Ray. A Pareto corner search evolutionary algorithm and dimensionality reduction in many-objective optimization problems. *IEEE Transactions on Evolutionary Computation*, 15(4):539–556, 2011.

[7] Md Asafuddoula, Tapabrata Ray, and Ruhul Sarker. A decomposition-based evolutionary algorithm for many objective optimization. *IEEE Transactions on Evolutionary Computation*, 19(3):445–460, 2015.

[8] J.-A. Désidéri, R. Duvigneau, and A. Habbal. *Computational Intelligence in Aerospace Sciences, V. M. Becerra and M. Vassile Eds.*, volume 244 of *Progress in Astronautics and Aeronautics*, chapter Multi-Objective Design Optimization Using Nash Games. AIAA, 2014.

[9] E. Kalai and M. Smorodinsky. Other solutions to Nash's bargaining problem. *Econometrica*, 43:513–518, 1975.

[10] Joshua D. Svenson. *Computer Experiments: Multiobjective Optimization and Sensitivity Analysis*. PhD thesis, The Ohio State University, 2011.

[11] Roger B Nelsen. *An introduction to copulas*. Springer, 2006.

[12] Mickaël Binois, Didier Rullière, and Olivier Roustant. On the estimation of Pareto fronts from the point of view of copula theory. *Information Sciences*, 324:270 – 285, 2015.

[13] Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, and Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793, 2012.

[14] José Miguel Hernández-Lobato, Michael A Gelbart, Ryan P Adams, Matthew W Hoffman, and Zoubin Ghahramani. A general framework for constrained Bayesian optimization using information-based search. *Journal of Machine Learning Research*, 17(160):1–53, 2016.

[15] Julien Villemonteix, Emmanuel Vazquez, and Eric Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534, 2009.

[16] Victor Picheny. Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction. *Statistics and Computing*, pages 1–16, 2013.

[17] Julien Bect, François Bachoc, and David Ginsbourger. A supermartingale approach to Gaussian process based sequential design of experiments. *arXiv preprint arXiv:1608.01118*, 2016.

[18] John P Conley and Simon Wilkie. The bargaining problem without convexity: extending the egalitarian and Kalai-Smorodinsky solutions. *Economics Letters*, 36(4):365–369, 1991.

[19] Irem Bozbay, Franz Dietrich, and Hans Peters. Bargaining with endogenous disagreement: The extended Kalai–Smorodinsky solution. *Games and Economic Behavior*, 74(1):407–417, 2012.

[20] James Parr. *Improvement criteria for constraint handling and multiobjective optimization*. PhD thesis, University of Southampton, 2013.

[21] Slim Bechikh, Lamjed Ben Said, and Khaled Ghedira. Estimating nadir point in multi-objective optimization using mobile reference points. In *Evolutionary computation (CEC), 2010 IEEE congress on*, pages 1–9. IEEE, 2010.

[22] Carl E. Rasmussen and Christopher Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[23] Victor Picheny, Mickael Binois, and Abderrahmane Habbal. A Bayesian optimization approach to find Nash equilibria. *arXiv preprint arXiv:1611.02440*, 2016.

[24] Victor Picheny and Mickael Binois. *GPGame: Solving Complex Game problems using Gaussian processes*, 2017. R package version 1.0.0.