

Comment: some enhancements over the augmented Lagrangian approach

Victor Picheny, David Ginsbourger, Tupaluck Krityakierne

► **To cite this version:**

Victor Picheny, David Ginsbourger, Tupaluck Krityakierne. Comment: some enhancements over the augmented Lagrangian approach. *Technometrics*, Taylor & Francis, 2016, 58 (1), pp.17-21. 10.1080/00401706.2015.1079246 . hal-02640140

HAL Id: hal-02640140

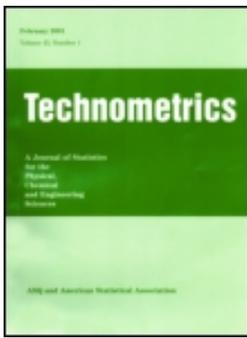
<https://hal.inrae.fr/hal-02640140>

Submitted on 28 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright



Comment: Some Enhancements Over the Augmented Lagrangian Approach

Victor Picheny, David Ginsbourger & Tupaluck Krityakierne

To cite this article: Victor Picheny, David Ginsbourger & Tupaluck Krityakierne (2016) Comment: Some Enhancements Over the Augmented Lagrangian Approach, *Technometrics*, 58:1, 17-21, DOI: [10.1080/00401706.2015.1079246](https://doi.org/10.1080/00401706.2015.1079246)

To link to this article: <http://dx.doi.org/10.1080/00401706.2015.1079246>



Published online: 22 Feb 2016.



Submit your article to this journal [↗](#)



Article views: 235



View related articles [↗](#)



View Crossmark data [↗](#)

Liang, F., Liu, C., and Carroll, R. J. (2007), “Stochastic Approximation in Monte Carlo Computation,” *Journal of the American Statistical Association*, 102, 305–320. [16]

Song, Q., Wu, M., and Liang, F. (2014), “Weak Convergence Rates of Population Versus Single-Chain Stochastic Approximation MCMC Algorithms,” *Advances in Applied Probability*, 46, 1059–1083. [16]

Comment: Some Enhancements Over the Augmented Lagrangian Approach

Victor PICHENY

MIAT, INRA
Castanet-Tolosan 31326, France
(victor.picheny@toulouse.inra.fr)

**David GINSBOURGER and
Típaluck KRITYAKIERNE**

Department of Mathematics and Statistics
University of Bern
Bern 3012, Switzerland
(david.ginsbourger@stat.unibe.ch;
tk338@cornell.edu)

We would like to begin by congratulating Gramacy, Gray, Le Digabel, Lee, Ranjan, Wells, and Wild (henceforth Gramacy et al.) for their important and timely contribution to the Bayesian optimization field. Combining augmented Lagrangian (AL) and Gaussian process modeling is an elegant solution to address constrained optimization of computer experiments. Their algorithm may open new doors to solve such types of problem, for which practitioners are yet waiting for efficient and user-friendly tools. Supplementary materials for this article are available online.

We first seize the opportunity of this discussion to review the main situations one may encounter in constrained optimization of computer experiments, some of the corresponding existing tools and areas open for innovation (Section 1). In Section 2, we deliver analytical formulas of the main criterion of Gramacy et al. for the single constraint problem, as well as directions for more complex cases. We believe that this could improve substantially the efficiency of the proposed workflow, in particular as it allows the use of standard optimizers for infill sampling criteria. In Section 3, we discuss the general behavior of the AL strategy, and propose some modifications to improve its efficiency. We close this comment by pointing out potential extensions for this approach.

1. CONSTRAINED OPTIMIZATION IN THE COMPUTER EXPERIMENTS COMMUNITY

Since the seminal article of Jones, Schonlau, and Welch (1998) and its EGO algorithm, heuristics have flourished to tackle many variations of the box-constrained, global deterministic optimization problem (including gradient enhanced, noisy, multi-objective, multi-fidelity, high-dimensional, etc.). Yet, we must agree with Gramacy et al. that constraints have been relatively neglected compared to other aspects. We wish here to build on the literature review of Gramacy et al. without pretend-

ing to be exhaustive but to create a roadmap to practitioners and reveal the needs for new methods.

Starting with the classical equation $\min_x \{f(x) : c(x) \leq 0\}$, we see four main cases (plus all the hybrid situations):

1. f is expensive to compute (timewise) and c is cheap, typically when some combinations of the input parameters are known in advance to be infeasible;
2. both f and c are expensive, when both f and c are different outputs of the same computer model;
3. f is cheap and c is expensive, when objectives are weights, costs, etc., that are easy to compute while constraints cover complex processes, such as mechanical stress, sustainability, etc.;
4. f is expensive and c represents a crash constraint (the computer model does not return values at the end of the run for some unknown parameter combinations).

The first case is relatively straightforward to address as c can be embedded in the expected improvement (EI) maximization loop. The original branch-and-bound strategy of Jones et al. cannot be used however, but common practice nowadays mostly rely on meta-heuristics, which usually account for constraints without difficulty.

The second case has received the largest attention (for instance Sasena, Papalambros, and Goovaerts 2002; Parr et al. 2010, 2012; Picheny 2014, or the Gramacy et al. article), perhaps since the synchronicity of f and c evaluations makes the problem closer to the original EGO framework. The overall difficulty is to balance efficiently the search for the best objective values and learning the constraints. Parr et al. (2012) showed

© 2016 American Statistical Association and
the American Society for Quality
TECHNOMETRICS, FEBRUARY 2016, VOL. 58, NO. 1
DOI: 10.1080/00401706.2015.1079246

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/r/tech.

that those aims are difficult to reconcile within an EI criterion (as the *expected feasible improvement* of Sasena, Papalambros, and Goovaerts 2002) and proposed a multi-objective treatment of the problem. In the same spirit, another solution would be to build on the work on contour line estimation (see Chevalier, Picheny, and Ginsbourger 2014, for a review), as problem can be seen as learning a contour line $c = 0$. Of course, an efficient heuristic would have to alter existing algorithms to bias the search toward low values of f .

The third case is the main target of Gramacy et al. While such a case may be fairly standard, the approach proposed by the authors seems to be the first one dedicated to this problem. The critical element is to make the best use of the “complete” information regarding the objective. Note that an alternative would be to use the methods developed for the second case, replacing the Gaussian process quantities of the objective by the true objective function (and a zero prediction variance), as proposed here by Gramacy et al. but the effect on the fragile trade-off between exploration and intensification would require a careful attention. Besides this, variants of contour line estimation approaches integrating an incentive to explore regions with low response values could also prove to be an efficient alternative here.

The final case occurs when the computer model “crashes” for some combination of input parameters and hence does not return any objective value. This is the most critical case from the user’s perspective as the information provided by the expensive numerical experiments reduces to a Boolean. Gramacy and Lee (2011) and Lindberg and Lee (2015) proposed to treat the constraint using Gaussian process classification, which might also be feasible for most of the methods cited above.

It is also noteworthy that the case of equality constraints $c = 0$ has hardly been addressed by the statistical community. Nevertheless, it may be particularly convenient; for instance, when the optimum is expected to be on the constraint boundary, the search space would reduce considerably. As suggested in their discussion by Gramacy et al., the AL framework may prove particularly useful here (see also Sections 2.3 and 3).

In summary, as justly observed by the authors, there does not seem to be an efficient, unique method to cover the different situations. To our knowledge, no extensive benchmarking has already been performed that allows us to advocate a particular strategy. Using the augmented lagrangian is indeed a welcome step in this direction. Entropy-based approaches (Villemonteix et al. 2009; Hennig and Schuler 2012) might offer a more comfortable framework (statistically) for unified formulations, by pursuing the efforts of Gramacy and Lee (2011), Picheny (2014), and Hernández-Lobato et al. (2015) on constrained cases, but these often come at the price of numerical complexity.

2. ON THE EXPECTED AL IMPROVEMENT

2.1 Analytical Formulation for a Known Objective, Single Constraint

We first recall the general form of the random process standing for the augmented Lagrangian, following the notations in

Gramacy et al.:

$$Y(x) = Y_f(x) + \lambda^T Y_c(x) + \frac{1}{2\rho} \sum_{j=1}^m \max(0, Y_{c_j}(x))^2, \quad (1)$$

from which we can define the corresponding improvement function:

$$I_Y(x) = (0, y_{\min}^n - Y(\mathbf{x}))^+, \quad (2)$$

with $(\cdot)^+ = \max(0, \cdot)$, the expected improvement function being $\mathbb{E}[I_Y(x)]$. As in Schonlau (1997), the expected improvement is to be used as an infill criterion, providing a trade-off between exploration (sampling in regions with high $Y(x)$ variance) and intensification (sampling in regions with low $Y(x)$ mean).

Here, we present a modest contribution consisting in analytically calculating the genuine criterion $\mathbb{E}[I_Y(x)]$ in the case of a single constraint and known objective.

With one constraint, the criterion can be written in the simple following form:

$$\mathbb{E}[I_Y(x)] = \frac{1}{2\rho} \mathbb{E}[(T - \alpha Z - \max(0, Z^2))^+], \quad (3)$$

where $T = 2\rho(y_{\min}^n - f(x))$, $\alpha = 2\rho\lambda$, and $Z = Y_c(x)$. Here, $Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)$ with $\mu_Z = \mu_c^n(x)$ and $\sigma_Z = \sigma_c^n(x)$.

We first decompose Equation (3) into two parts:

$$\begin{aligned} \mathbb{E}[(T - \alpha Z - \max(0, Z^2))^+] &= \mathbb{E}[(T - \alpha Z)^+ \mathbf{1}_{Z < 0}] \\ &+ \mathbb{E}[(T - \alpha Z - Z^2)^+ \mathbf{1}_{Z \geq 0}]. \end{aligned} \quad (4)$$

Letting $u := \frac{\min(0, \frac{T}{\alpha}) - \mu_Z}{\sigma_Z}$, Equation (A.1) in Appendix gives

$$\begin{aligned} \mathbb{E}[(T - \alpha Z)^+ \mathbf{1}_{Z < 0}] &= \mathbb{E}[(T - \alpha Z) \mathbf{1}_{\{Z < 0\} \cap \{Z < \frac{T}{\alpha}\}}] \\ &= T \mathbb{P}\left(Z < \min\left(0, \frac{T}{\alpha}\right)\right) \\ &\quad - \alpha \mathbb{E}\left[Z \mathbf{1}_{Z < \min(0, \frac{T}{\alpha})}\right] \\ &= T \Phi(u) - \alpha (\mu_Z \Phi(u) - \sigma_Z \phi(u)) \\ &:= w. \end{aligned}$$

Assume now that $T \geq 0$ and denote $T' = T + \frac{\alpha^2}{4}$ and $Z' = Z + \frac{\alpha}{2}$:

$$\begin{aligned} \mathbb{E}[(T - \alpha Z - Z^2)^+ \mathbf{1}_{Z \geq 0}] &= \mathbb{E}[(T' - Z'^2)^+ \mathbf{1}_{\frac{\alpha}{2} \leq Z'}] \\ &= \mathbb{E}[(T' - Z'^2) \mathbf{1}_{\frac{\alpha}{2} \leq Z' \leq \sqrt{T'}}] \\ &= T' \mathbb{P}\left(\frac{\alpha}{2} \leq Z' \leq \sqrt{T'}\right) \\ &\quad - \mathbb{E}\left[Z'^2 \mathbf{1}_{\frac{\alpha}{2} \leq Z' \leq \sqrt{T'}}\right]. \end{aligned}$$

Setting $\tilde{a} := \frac{-\mu_Z}{\sigma_Z}$, and $\tilde{b} := \frac{\sqrt{T'} - \frac{\alpha}{2} - \mu_Z}{\sigma_Z}$, Equation (A.2) in Appendix gives us:

$$\begin{aligned} \mathbb{E}[Z'^2 1_{\frac{\alpha}{2} \leq Z' \leq \sqrt{T'}}] &= \left(\frac{\alpha}{2} + \mu_Z\right)^2 (\Phi(\tilde{b}) - \Phi(\tilde{a})) \\ &\quad + (\alpha + 2\mu_Z)\sigma_Z (\phi(\tilde{a}) - \phi(\tilde{b})) \\ &\quad + \sigma_Z^2 (\tilde{a}\phi(\tilde{a}) - \tilde{b}\phi(\tilde{b})) \\ &\quad + (\Phi(\tilde{b}) - \Phi(\tilde{a})). \end{aligned}$$

Putting all of the pieces together, we have:

$$2\rho\mathbb{E}[I_Y(x)] = \begin{cases} w & \text{if } T < 0 \\ w + (T' - (\frac{\alpha}{2} + \mu_Z)^2 - \sigma_Z^2) (\Phi(\tilde{b}) - \Phi(\tilde{a})) \\ \quad + (\alpha + 2\mu_Z)\sigma_Z (\phi(\tilde{b}) - \phi(\tilde{a})) + \sigma_Z^2 (\tilde{b}\phi(\tilde{b}) - \tilde{a}\phi(\tilde{a})) & \text{otherwise.} \end{cases} \quad (5)$$

2.2 Extension to More Complex Cases

We have also considered two more complex cases: two constraints, and also costly objective. On both cases, this amounts to consider T in Equation (3) as a Gaussian random variable. Unfortunately, calculations seem intractable.

However, it is actually possible to apply Equation (5) to improve over crude Monte Carlo using the following approach. Depending on the case, let us replace T by \tilde{T} in Equation (3), with either

$$\tilde{T} := \begin{cases} T - \alpha_2 Z_2 - \max(0, Z_2)^2, \\ \quad Z_2 \text{ standing for the second constraint, or} \\ 2\rho (y_{\min}^n - Y_f(x)), \\ \quad Y_f(x) \text{ standing for the GP model of } f \text{ at } x. \end{cases} \quad (6)$$

Then, with N replicates $\tilde{T}_1, \dots, \tilde{T}_N$ of \tilde{T} , we can approximate the criterion by

$$\mathbb{E}[I_Y(x)] \approx \frac{1}{N} \sum_{i=1}^N \mathbb{E}[I_{Y; \tilde{T}_i}(x)], \quad (7)$$

where $\mathbb{E}[I_{Y; \tilde{T}_i}(x)]$ stands for the criterion of Equation (5) with T replaced by \tilde{T}_i (following Equation (6)) and can be analytically calculated for every replicate by appealing to Equation (5). Using the law of total variance, it is easy to see that for a fixed sample size, the improved estimator has a smaller variance than the crude estimator.

2.3 AL With Equality Constraints

With a single equality constraint, the AL is defined as

$$Y(x) = Y_f(x) + \lambda^T Y_c(x) + \frac{1}{2\rho} Y_c(x)^2, \quad (8)$$

and the expected improvement is of the form $\mathbb{E}[(T - \alpha Z - Z^2)^+]$. It follows that

$$2\rho\mathbb{E}[I_Y(x)]$$

$$= \begin{cases} 0 & \text{if } T' < 0 \\ (T' - (\frac{\alpha}{2} + \mu_Z)^2 - \sigma_Z^2) (\Phi(\tilde{b}) - \Phi(\tilde{a})) \\ \quad + (\alpha + 2\mu_Z)\sigma_Z (\phi(\tilde{b}) - \phi(\tilde{a})) + \sigma_Z^2 (\tilde{b}\phi(\tilde{b}) - \tilde{a}\phi(\tilde{a})) & \text{otherwise.} \end{cases} \quad (9)$$

With $\tilde{a} := \frac{-\sqrt{T'} - \frac{\alpha}{2} - \mu_Z}{\sigma_Z}$ and T', T, α , and \tilde{b} defined as above, Equation (9) yields the same formula as Equation (11) of the Gramacy et al. article.

2.4 Multiple Infill Points

Another extension of the Gramacy et al. expected improvement is to consider multiple infill points, which can be particularly useful if several runs of the simulator can be run in parallel. Noting $X^* = \{x_1^*, \dots, x_p^*\}$ a batch of candidate points, a natural candidate for a multipoint AL expected improvement would be

$$\begin{aligned} \mathbb{E}[I_Y(X^*)] &= \mathbb{E} \left[\max \left(0, y_{\min}^n - \min_{1 \leq i \leq p} \left\{ Y_f(x_i^*) + \lambda^T Y_c(x_i^*) \right. \right. \right. \\ &\quad \left. \left. \left. + \frac{1}{2\rho} \sum_{j=1}^m \max(0, Y_{c_j}(x_i^*))^2 \right\} \right) \right]. \end{aligned} \quad (10)$$

This form is readily usable with Monte Carlo estimations as in Equation (10) of Gramacy et al. provided that the drawings of the vectors $[Y_{c_j}(x_1), \dots, Y_{c_j}(x_p)]$ ($1 \leq j \leq m$) are generated using their conditional covariance matrix; see Schonlau (1997) and Ginsbourger, Le Riche, and Carraro (2010). On the other hand, analytical expressions seem to be out of reach.

3. BEHAVIOR OF THE AL-BASED STRATEGY

In this section, we wish to examine empirically the behavior of the Gramacy et al. strategy on a restricted scenario (known objective and single constraint). Compared to the original EGO, the AL-based strategy (henceforth AL-EGO) is substantially more complex, as:

1. it contains an inner and an outer loop;
2. only candidates improving the objective are considered;
3. if EI is too small, EI maximization is switched to best predictor minimization.

The outer loop controls the update of the values of the AL parameters. Although being at the core of the AL framework, one may wonder if this is the best solution for expensive experiments. If the stopping criterion of the inner loop is not well tuned, a substantial number of experiments are likely to be spent on finding minimizers of AL functions that are far away from the minimizer(s) of the actual constrained problem.

Using only improving candidates is natural, but it does not allow the method to find the true minimizers of the ALs. Hence, there is a risk that the efficiency of the method will deteriorate rather than improve.

Besides, inequality constraints are considered. In the examples considered here the solution must be in the boundary (since the objective is linear); hence, equality constraints could be used.

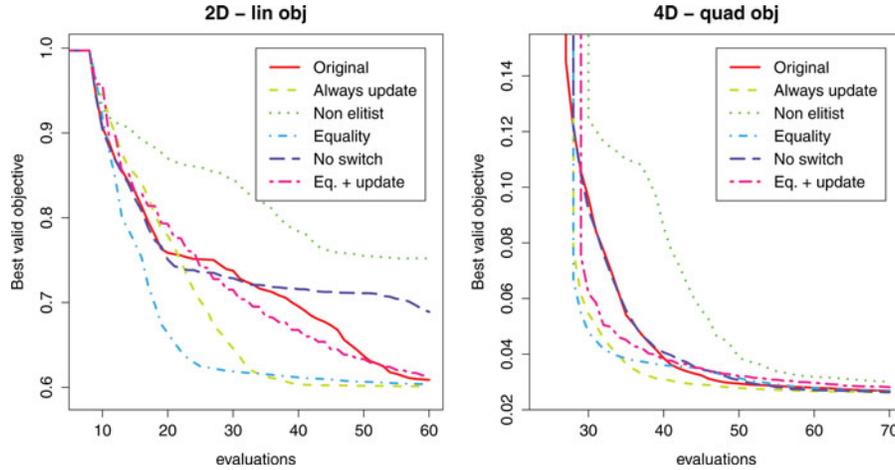


Figure 1. Average performance of six variants of AL-EGO on two toy problems.

To assess the criticality of each of these points, we perform the following numerical experiments. Two test problems are considered. For each, we compare six variants of AL-EGO:

- the original version, as proposed by Gramacy et al.: the infill criterion is $\mathbb{E}[I_Y(x)] \mathbb{1}_{f(x) \leq y_{\min}^n}$, computed using Equation (5). $\mathbb{1}_{f(x) \leq y_{\min}^n}$ accounts for the fact that only improving candidates are chosen. The criterion switches to $-\mathbb{E}[Y(x)] \mathbb{1}_{f(x) \leq y_{\min}^n}$ if too few nonzero AL-EI values are found;
- an “always-update” AL-EGO, whose AL parameters are updated every iteration;
- a “nonelitist” AL-EGO, for which $\mathbb{1}_{f(x) \leq y_{\min}^n}$ is removed;
- a “nonswitching” AL-EGO, where $\mathbb{E}[I_Y(x)] \mathbb{1}_{f(x) \leq y_{\min}^n}$ is always used;
- an “equality-based” AL-EGO that is based on Equation (8) rather than Equation (3);
- an equality and always-update AL-EGO, where the two variants “equality-based” and “always-update” are considered simultaneously.

Note that this last version is equivalent to using the “EI-nomax” of Gramacy et al. with λ that can take negative values. We use as a secondary criterion $\mathbb{E}[Y(x)] = f(x) + \lambda \mu_c^n(x) + \frac{1}{2\rho} (\sigma_c^{2n}(x) + \mu_c^{2n}(x))$, which we suppose is the same as “EY-nomax.”

Each experiment is repeated 100 times with different initial designs (generated using improved Latin hypercube sampling). For each repetition, all AL-EGO variants start with the same initial design and GP model. The GP models are built using the DiceKriging R package (Roustant, Ginsbourger, and Deville 2012). We found that our program behaves similarly to 1aGP.

The first test problem is the one of Gramacy et al. The second one uses the 4D *Hartman* function minus two (Dixon and Szegö 1978) as a constraint, and the objective is quadratic $(-\sum_{i=1}^4 (x_i - 1/2)^2)$. The results are shown in Figure 1 in the form of average performance (best valid objective) over the 100 repetitions.

The *no-switch* curves illustrate clearly the soundness of switching to the best predictor as an infill criterion: on the left graph, the curve is identical to the original strategy up to 20 evaluations (which, in 2D, is quite sufficient to get an accurate approximation of the constraint) before becoming almost flat as it fails to progress further. The effect is not visible on the 4D problem, probably because 70 observations are not sufficient to obtain an accurate approximation of the function.

The *nonelitist* curves also show that choosing improving candidates only greatly accelerates the progress of the strategy. This seems to indicate that departing from the augmented lagrangian framework is not necessarily a bad thing here.

The *equality* curves show contrasting results: it is on both cases the best method for the first iterations, in particular in the 2D setting. However, it seems rather slow to progress locally toward the optimum.

Always updating λ and ρ seems to improve substantially the efficiency of AL-EGO, both during the exploratory steps and local convergence. A close look at a single run (Figure 2) confirms our intuition: unfavorable values of the parameters tend to enhance exploration in noncritical regions (“+” and “x” symbols for instance), resulting in a “waste” of several consecutive iterations.

Finally, we used as a last variation of AL-EGO the equality setting with systematic update of the parameters. Despite being based on two positive modifications of the original algorithm, this variation does not perform better than the original, which is somehow puzzling.

To conclude this section, and without generalizing too much from our limited experimental setup, it seems that although AL offers a nice framework for GP-based optimization, departing from it (through “elitist” sampling and systematic update) clearly seems beneficial here. How to combine optimally (both in terms of efficiency and robustness) those different aspects appears to us yet as an open question, whose answer might even differ according to the optimization problem considered (known/unknown objective, number of constraints, etc.).

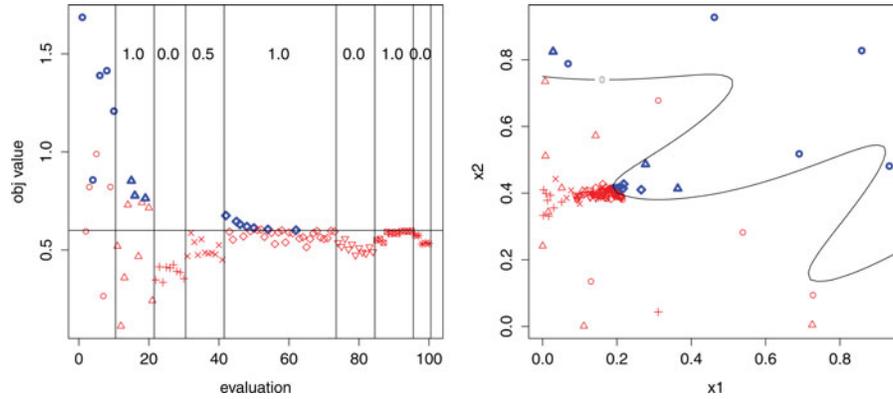


Figure 2. A single run of AL-EGO. Left: objective function values over iterations. The vertical bars show when the parameters are updated, and the numbers in each column correspond to the λ values. Right: experiments and boundary of the feasible space. The feasible points are shown in bold symbols on both graphs.

4. CLOSING REMARKS

The results and concepts presented in the article by Gramacy et al. and discussed here open a number of perspectives beyond the use of the augmented Lagrangian with deterministic settings and independent responses and constraints. Indeed, what is proposed for AL consists in extending the EI framework to an instrumental objective function obtained by combining nonlinearly several responses or constraints, modeled here as independent Gaussian processes. Similar ideas may appear beneficial for other kinds of instrumental functions, be it for constrained or multi-objective optimization, or for other purposes. Besides this, it might be useful in some cases to depart from the assumption of independence between the various processes, and to appeal for instance to co-kriging. Last but not least, for the case of noisy responses and/or noisy constraint evaluations, Gelbart, Snoek, and Adams (2014) seemed to pose important challenges for the extension of the presented approaches.

APPENDIX: TWO USEFUL TRUNCATED GAUSSIAN IDENTITIES

Suppose $Y \sim \mathcal{N}(\mu, \sigma^2)$, a Gaussian random variable, and $a, b \in \mathbb{R}$, $\tilde{a} = \frac{a-\mu}{\sigma}$, $\tilde{b} = \frac{b-\mu}{\sigma}$. The following is a list of useful identities:

$$\mathbb{E}[Y \mathbf{1}_{a \leq Y \leq b}] = \mu (\Phi(\tilde{b}) - \Phi(\tilde{a})) + \sigma (\phi(\tilde{a}) - \phi(\tilde{b})) \quad (\text{A.1})$$

$$\mathbb{E}[Y^2 \mathbf{1}_{a \leq Y \leq b}] = \mu^2 (\Phi(\tilde{b}) - \Phi(\tilde{a})) + 2\mu\sigma (\phi(\tilde{a}) - \phi(\tilde{b})) + \sigma^2 (\tilde{a}\phi(\tilde{a}) - \tilde{b}\phi(\tilde{b}) + \Phi(\tilde{b}) - \Phi(\tilde{a})). \quad (\text{A.2})$$

SUPPLEMENTARY MATERIALS

R scripts used to generate the results presented in this comment. (Zipped tar file)

REFERENCES

Chevalier, C., Picheny, V., and Ginsbourger, D. (2014), “Kriginv: An Efficient and User-Friendly Implementation of Batch-Sequential Inversion Strate-

- gies Based on Kriging,” *Computational Statistics & Data Analysis*, 71, 1021–1034. [17]
- Dixon, L. C. W., and Szegő, G. P. (1978), *Towards Global Optimisation 2*, Amsterdam: North-Holland. [20]
- Gelbart, M., Snoek, J., and Adams, R. (2014), “Bayesian Optimization With Unknown Constraints,” in *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence (UAI2014)*, Quebec City, Canada, July 23–27, 2014. [21]
- Ginsbourger, D., Le Riche, R., and Carraro, L. (2010), “Kriging is Well-Suited to Parallelize Optimization,” in *Computational Intelligence in Expensive Optimization Problems (Adaptation Learning and Optimization)*, eds. L. M. Hiot, Y. S. Ong, Y. Tenne, and C.-K. Goh, Berlin, Heidelberg: Springer, pp. 131–162. [19]
- Gramacy, L., and Lee, H. (2011), “Optimization Under Unknown Constraints,” Valencia discussion paper, in *Bayesian Statistics 9*, Oxford: Oxford University Press, 9, 229. [18]
- Hennig, P., and Schuler, C. (2012), “Entropy Search for Information-Efficient Global Optimization,” *Journal of Machine Learning Research*, 13, 1809–1837. [18]
- Hernández-Lobato, J., Gelbart, M., Hoffman, M., Adams, R., and Gharamani, Z. (2015), “Predictive Entropy Search for Bayesian Optimization With Unknown Constraints,” in *Proceedings of the 32nd International Conference on Machine Learning (ICML)*. [18]
- Jones, D. R., Schonlau, M., and Welch, W. J. (1998), “Efficient Global Optimization of Expensive Black-Box Functions,” *Journal of Global optimization*, 13, 455–492. [17]
- Lindberg, D. V., and Lee, H. K. (2015), “Optimization Under Constraints by Applying an Asymmetric Entropy Measure,” *Journal of Computational and Graphical Statistics*, 24, 379–393. [18]
- Parr, J., Holden, C. M., Forrester, A. I., and Keane, A. J. (2010), “Review of Efficient Surrogate Infill Sampling Criteria With Constraint Handling,” in *2nd International Conference on Engineering Optimization*, pp. 1–10. [17]
- Parr, J., Keane, A., Forrester, A., and Holden, C. (2012), “Infill Sampling Criteria for Surrogate-Based Optimization With Constraint Handling,” *Engineering Optimization*, 44, 1147–1166. [17]
- Picheny, V. (2014), “A Stepwise Uncertainty Reduction Approach to Constrained Global Optimization,” *Journal of Machine Learning Research W&CP*, 33, 787–795. [17,18]
- Roustant, O., Ginsbourger, D., and Deville, Y. (2012), “DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization,” *Journal of Statistical Software*, 51, 1–55. [20]
- Sasena, M. J., Papalambros, P., and Goovaerts, P. (2002), “Exploration of Meta-modeling Sampling Criteria for Constrained Global Optimization,” *Engineering Optimization*, 34, 263–278. [17]
- Schonlau, M. (1997), “Computer Experiments and Global Optimization,” Ph.D. dissertation, Department of Statistics and Actuarial Science, University of Waterloo. [18,19]
- Villemonaix, J., Vazquez, E., and Walter, E. (2009), “An Informational Approach to the Global Optimization of Expensive-to-Evaluate Functions,” *Journal of Global Optimization*, 44, 509–534. [18]