Wiley StatsRef:
Statistics Reference Online
DOI: 10.1002/ISBN.stat00999.pub9

Sequential Design of Computer Experiments

David Ginsbourger[*ab]

Keywords: Gaussian Processes; Numerical simulation; Evaluation strategies; Sampling criteria

Abstract: Numerical simulations are increasingly used in science and society for the study of complex systems, be it as a complement or a substitute to classical experiments. The design of computer experiments has become a research field per se, complementing the legacy of classical experimental design while accounting for novel specificities. Sequentiality is essential in many of the goal-oriented design approaches that have flourished in such framework.

1. On GP-based design of computer experiments

While multiple linear regression can be considered as a baseline statistical approach to model a response of interest as a function of covariates in the set up of classical design of experiments and response surface methodology (Steinberg and Kennet, 2014; Draper, 2014; Atkinson, 2015), in computer experiments (Morris, 2015) the approach that has arguably become the most popular in the last decades is Gaussian Process (GP) modelling (Sacks et al., 1989b; Koehler and Owen, 1996; Santner et al., 2003; Davis, 2014). One of the convenient features of GP modelling is that predictive mean and variance functions are analytically tractable. In particular, the predictive variance (a.k.a. MSE, for Mean-squared error, following the terminology of Sacks et al. (1989b)) and covariance (Samson, 2014) can be used as basic bricks to drive the choice of design points. When working with a GP conditioned on available deterministic evaluation results, the full conditional covariance structure (hence including the corresponding MSE function) solely depends on the observation points and not on the associated responses — in the case of noisy observations with a centred Gaussian noise independent of the GP, the conditional covariance depends also on the noise covariance but still not on the noisy responses. It is thus relevant to define design criteria, i.e. functions quantifying in some sense the quality of finite sets of points, in terms of the conditional covariance of the GP knowing that experiments are to be performed at these points. One natural approach from the perspective of prediction, that generalizes the concept of G-optimality from classical design of experiments, consists in minimizing the so-called MMSE criterion: the fitness of candidate point designs is assessed in terms of the maximum MSE over the domain of interest assuming that the GP is to be observed at these points. Minimizing MMSE is known to be cumbersome because the criterion itself is generally non-convex, and also costly to evaluate as it is itself the global maximum of a future MSE function. Deterministic and stochastic algorithms are used in Sacks and Schiller (1988) to approximately minimize MMSE relying on sequences of addition/deletion operations on candidate designs. Beside MMSE, these algorithms are also applied to Maximum

[a]**Uncertainty Quantification and Optimal Design group, Idiap Research Institute, Martigny, Switzerland**
[b]**Institute for Mathematical Statistics and Actuarial Science, University of Bern, Bern, Switzerland**
[*]**Email: ginsbourger@idiap.ch**

Prepared using WileySTAT.cls [Version: v1.00]

Entropy Sampling in the sense of Shewry and Wynn (1987). Let us note that the entropy criterion in question boils down under standard Gaussian Process assumption to the determinant of the posterior covariance matrix associated with candidate points, so that the Maximum Entropy Sampling approach coincides in the one-point case with MSE maximization. Enriching a starting design by incorporating MSE-optimal points one at a time both in the design and in the GP model (without hyperparameter estimation) until the target number of design points is met constitutes a greedy approach to MMSE minimization, that is known to be sub-optimal but to achieve the same rates of decay than the optimal strategy (Vazquez and Bect, 2011). On a different note, as pointed out in Gramacy (2005), such approach is known to be similar to MacKay's Active Learning (often referred to as "ALM") strategy (MacKay, 1992). Coming back to the entropy criterion, an alternative to Maximum Entropy Sampling proposed in Caselton and Zidek (1984) and further studied in Krause et al. (2008) consists in choosing from a given finite set of reference points a subset so as to maximally reduce the entropy over the remaining reference points; as mentioned in the latter reference, this amounts to maximizing a *Mutual Information* criterion. Yet another criterion that is more global than MSE but less cumbersome to calculate than MMSE is the so-called IMSE criterion, the integral over prediction points (given any candidate point as input) of the future MSE (See, e.g., Sacks et al., 1989a). Recent contributions towards efficient computation and optimization of the IMSE criterion include Harari and Steinberg (2014); Gauthier and Pronzato (2014); Gorodetsky and Marzouk (2016); Gauthier and Pronzato (2017) and references therein. Under standard assumptions and unless hyperparameters are re-estimated after intermediate evaluations, the criteria mentioned so far have in common not to depend on observed responses but solely on previously visited points, so that the sequential generation of such designs can be made *off-line*. In contrast, a number of strategies used in sequential design of computer experiments are response-adaptive (Flournoy and Oron, 2014), i.e. visited points depend on previous responses. Incorporating cross-validation errors in sampling criteria is one way to make sequential design response-adaptive (See, e.g., Le Gratiet et al. (2015)); in the next section, we will focus on response-adaptive criteria for specific goals such as global optimization and level set estimation. Examples of the use of MSE/IMSE-based design abound in the methodological and applied literature. In Sacks et al. (1989a), chemical kinetics problems were considered, while Sacks et al. (1989b) tackled a circuit-simulator application. Also, a variant of IMSE integrating uncertainties on model parameters was introduced in Allen et al. (2003), with an application to die casting process design. Christen and Sansó (2011) investigated analogies between IMSE-based design and Active Learning in the sense of Cohn (1996) (referred to as ALC, following Gramacy (2005)) and applied sequential design approaches to a climate computer model. On the other hand, the use of GP-based design relying on maximum entropy sampling was illustrated in Currin et al. (1991) on a thermal energy storage system example as well as on a circuit simulation example. Coming to algorithms relying on mutual entropy maximization, while the motivating problem of Krause et al. (2008) is in sensor placement, other example applications relying both on in-field experiments and on simulations on real world sensing datasets are tackled in Singh et al. (2009), where the focus is on lake and river monitoring using multi-robot informative path planning. In Beck and S. (2016), the Mutual Information approach of Krause et al. (2008) was revisited, and an original modification of it was demonstrated to outperform competitor strategies such as ALM and ALC on test cases including the emulation of a tsunami model.

2. Focus on selected goal-oriented approaches

2.1. Global optimization

One of the reasons that have motivated the development of sequential design algorithms based on GPs is global optimization. Due to their versatility and their feature of providing both a mean predictor and an associated predictive variance in closed form, GP models constitute indeed a practical tool to address exploration/exploitation trade-offs and define sequential design strategies aimed at globally optimizing functions (or improving over current best candidate

points under tight evaluation budgets). Building upon seminal works, including notably Kushner (1964) and Mockus et al. (1978); Mockus (1989) that laid down the foundations of sequential design with criteria like the Probability of Improvement (PI) and the Expected Improvement (EI), global optimization algorithms based on GP models have known a substantial development in the engineering community and beyond following Schonlau and Welch (1996); Jones et al. (1998) and related works. The EI criterion remains as a quite popular infill sampling criterion (such criteria are also known as aquisition functions in the machine learning literature, see e.g. Snoek and Adams (2012)), and it enjoys both an analytical form combining the GP predictive mean and variance as well as theoretical results for associated design strategies (Vazquez and Bect, 2010; Bull, 2011; Yarotsky, 2013; Rhyzov, 2016). Besides, EI has been adapted to a number of situations including distributed optimization (Schonlau, 1997; Ginsbourger et al., 2010; Janusevskis et al., 2012; Chevalier et al., 2014), constrained (Parr, 2012; Gramacy et al., 2016) and multiobjective optimization (Knowles, 2005; Emmerich et al., 2006, 2011; Binois et al., 2015a; Feliot et al., 2017). Also, variations of EI exist that handle uncertain GP model parameters (Ginsbourger et al., 2008; Benassi et al., 2011; Gramacy and Polson, 2011), uncontrolled input variables (Williams et al., 2000; Janusevskis and Le Riche, 2013; Marzat et al., 2013; Ginsbourger et al., 2014), as well as noisy (Huang et al., 2006b; Picheny et al., 2013) and multifidelity (Huang et al., 2006a; Forrester et al., 2006, 2007; He et al., 2017) responses. One well-known limitation of EI is that it is a myopic criterion, in the sense that it evaluates the benefit of performing an evaluation (or a batch of them) as if it were the last one, and while optimal strategies accounting for multiple steps are considered intractable (Osborne et al., 2009; Ginsbourger and Le Riche, 2010), heuristic approaches have been proposed to alleviate the myopicity of EI while remaining computationally tractable Gonzalez et al. (2016); Lam et al. (2016). On a different note, several paradigms do exist besides EI to fuel GP-based sequential design algorithms dedicated to optimization purposes. The Knowledge Gradient approach of Frazier et al. (2008); Scott et al. (2011) is one instance of such paradigm offering much versatility, accomodating notably parallel computing (Wu and Frazier, 2016), as well as gradient evaluations (Wu et al., 2017), to cite a few recent extensions. In other respects, a stream of works mostly associated with machine learning and building upon the literature on "bandits" (Robbins., 1952; Berry and Fristedt, 1985) has been quite impactful in recent years, relying notably on the Upper Confidence Bound (UCB) framework. While UCB algorithms are traditionally analyzed in terms of cumulative regret (Auer, 2002), with the expansion of UCB and variants thereof for GP-based optimization, results in terms of simple regret have also been established (Grünewälder et al., 2010; Srinivas et al., 2010, 2012; de Freitas et al., 2012). Also, informational approaches have been used in further settings, notably to decrease the entropy of the function's global minimizer or minimum (Villemonteix et al., 2009; Hernández-Lobato et al., 2014), and a GP-based global optimization algorithm based on the notion of Mutual Information was introduced in Contal et al. (2014). Overall, current challenges in sequential design of computer experiments for global optimization include deadling with high-dimensional (Wang et al., 2013; Kandasamy et al., 2015; Binois et al., 2015b; Wang et al., 2017) and large-scale problems (Krityakierne and Ginsbourger, 2015; Wang et al., 2018), as well as pushing the limits of multi-fidelity (Perdikaris et al., 2016) and also multi-information source (Poloczek et al., 2017) approaches.

2.2. Preferential exploration schemes for further goals

GP models have also be found quite convenient for the sequential design of experiments when the goal is to learn further quantities or sets involving expensive-to-evaluate functions. In a number of practical situations, it is of interest to estimate regions of the input domain where a real-valued response takes critical values (be it, e.g., excursion sets above some prescribed threshold or contour lines), or less ambitiously to estimate the measure of such set (with respect to some given measure, often a probability measure in reliability applications) or quantiles of the image probability measure on the set of responses —in cases when inputs are treated as random. To give an excerpt of contributions around these topics, let us cite Ranjan et al. (2008) and Picheny et al. (2010) where variations of EI and IMSE criteria are considered, respectively, with a focus on contour line estimation. An overview of neighbouring criteria

and more specifically approaches for sequentially reducing uncertainty on probabilities of excursion were presented in Bect et al. (2012). Sequential Uncertainty Reduction (SUR) principles such as presented in Bect et al. (2012) were already tackled in Vazquez and Bect (2009), and have since then been further developed in probability/set of excursion setting with fast implementations (Chevalier et al., 2014) supported by Kriging update formulae for batch-sequential data assimilation (Chevalier et al., 2013). Also, SUR ideas were revisited for constrained (Picheny, 2014) and multi-objective optimization (Picheny, 2015; Binois, 2015), as well as within further developments in sequential design of computer experiments for set estimation (Chevalier, 2013; Azzimonti, 2016). From the perspective of percentile estimation (Oakley, 2004), a sequential design of computer experiments for the assessment of fetal exposure to electromagnetic fields has been considered in Jala et al. (2016), and also a SUR approach was proposed in Labopin-Richard and Picheny (2017). Inspired by reliability issues, preferential exploration schemes based on GP have also been used for refining approximations of the optimal instrumental distribution in importance sampling (See, e.g., Balesdent et al. (2013); Dubourg et al. (2013)) and an algorithm, coined Bayesian subset simulation algorithm, that combines subset simulation (Au and Beck, 2001) with SUR principles was proposed in Bect et al. (2017) for the estimation of small volumes of excursion under very limited numbers of evaluations. Also, a Bayesian experimental design approach relying on the relative entropy between prior and posterior densities of interest under a GP model was used in Wang et al. (2016) in a failure detection context. In the spirit of UCB methods, level set estimation was addressed in Gotovos et al. (2013), and recent works adressing preferential exploration for hybrid goals include Berkenkamp et al. (2016), where safe optimization is sought, and an algorithm called truncated variance reduction was proposed in Bogunovic et al. (2016) that addresses Bayesian optimization and level set estimation in a unified framework.

3. Concluding remarks

Thanks to the tractability of posterior distributions assuming future evaluations at candidate points, Gaussian Process models are especially convenient for the sequential design of experiments, and have lead to a number of algorithms and applications to computer experiments as touched upon throughout this short article. Of course, the practical relevance of such algorithms relies on the adequacy between the hypotheses made through the considered family of GP models and the actual function of interest. While the art of designing so-called "initial" experiments was quietly skipped here, it is a field of research per se (Fang, 2006; Pronzato and Müller, 2012; Gamblin et al., 2013; Dupuy et al., 2015). Indeed, even if MSE/IMSE-based designs give a convenient way to explore the input domain, they assume a known model; yet, choosing a GP model and estimating its parameters require to perform experiments in the first place, a circularity problem that is well-known in spatial statistics (See, e.g., Stein, 1999; Müller, 2007). The mitigation of parameter misestimation in sequential optimization using GP models has attracted attention, e.g. in den Hertog et al. (2006); Benassi et al. (2011). Besides, Student t-processes were proposed as an alternative to GPs in Shah et al. (2014) where Bayesian optimization was applied with an EI formula for the t-process model, EI being itself averaged over posterior samples of model parameters. Accounting for and learning non-stationarities in Gaussian and related frameworks constitute further directions of interest, that have been tackled notably in Gramacy and Taddy (2010); Damianou and Lawrence (2013); Marmin (2017) and references therein. Of course, several other classes of surrogate models have been used beyond GP for the sequential design of computer experiments. For instance, Snoek et al. (2015) tackles scalable Bayesian optimization thanks to a neural network surrogate model. On the other hand, Burnaev et al. (2016) focuses on sequential design for global sensitivity analysis (Fassò, 2015) using Polynomial Chaos Expansions (PCE) and Pronzato furthermore considers generalized settings where PCEs are combined with GPs.

Related Articles

A.C. Atkinson. Optimal Design. Wiley StatsRef: Statistics Reference Online, 2015.

R. A. Davis. Gaussian Process: Theory. Wiley StatsRef: Statistics Reference Online, 2014.

N. R. Draper. Response Surface Designs. Wiley StatsRef: Statistics Reference Online, 2014.

A. Fassò. Sensitivity Analysis of Computer Models. Wiley StatsRef: Statistics Reference Online, 2015.

N. Flournoy, and A. P. Oron. Adaptive Designs. Wiley StatsRef: Statistics Reference Online, 2014.

M. D. Morris. Computer Experiments. Wiley StatsRef: Statistics Reference Online, 2015.

P. D. Sampson. Spatial Covariance. Wiley StatsRef: Statistics Reference Online, 2014.

D. M. Steinberg and R. S. Kenett. Response Surface Methodology. Wiley StatsRef: Statistics Reference Online, 2014.

References

T.T. Allen, L. Yu, and J. Schmitz. An experimental design criterion for minimizing meta-model prediction errors applied to die casting process design. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52:103–117, 2003.

S. K. Au and J. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probab. Engrg. Mechan.*, 16(4):263–277, 2001.

P. Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3: 397–422, 2002.

D. Azzimonti. *Contributions to Bayesian set estimation relying on random field priors*. PhD thesis, University of Bern, 2016.

M . Balesdent, J. Morio, and J. Marzat. Kriging-based adaptive Importance Sampling algorithms for rare event estimation. *Structural Safety 44, 1-10*, 44:1–10, 2013.

J. Beck and S. Guillas. Sequential design with mutual information for computer experiments (MICE): emulation of a tsunami model. *SIAM/ASA J. Uncertainty Quantification*, 4:739–766, 2016.

J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22 (3):773–793, 2012.

J. Bect, L. Li, and E. Vazquez. Bayesian subset simulation. *SIAM/ASA J. Uncertainty Quantification*, 5(1):762–786, 2017.

R. Benassi, J. Bect, and E. Vazquez. *Learning and Intelligent Optimization*, chapter Robust Gaussian Process-based global optimization using a fully Bayesian expected improvement criterion, pages 176–190. Springer, 2011.

F. Berkenkamp, A.P. Schoellig, and A. Krause. Safe controller optimization for quadrotors with Gaussian processes. In *Proc. of the IEEE International Conference on Robotics and Automation*, pages 493–496, 2016.

D. A. Berry and B. Fristedt. *Bandit problems: sequential allocation of experiments*. Chapman & Hall, 1985.

M. Binois. *Uncertainty quantification on Pareto fronts and high-dimensional strategies in Bayesian optimization, with applications in multi-objective automotive design*. PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2015.

M. Binois, D. Ginsbourger, and O. Roustant. Uncertainty quantification on Pareto fronts using Gaussian Process conditional simulations. *European Journal of Operational Research*, 243 (2):386–394, 2015a.

M. Binois, D. Ginsbourger, and O. Roustant. A warped kernel improving robustness in Bayesian optimization via random embeddings. In C. Dhaenens, L. Jourdan, and M.-E. Marmion, editors, *Learning and Intelligent Optimization*, volume 8994 of *Lecture Notes in Computer Science*, pages 281–286. Springer, 2015b.

I. Bogunovic, J. Scarlett, V. Cevher, and A. Krause. Truncated variance reduction: a unified approach to Bayesian optimization and level-set estimation. In *Advances in Neural Information Processing Systems 29*, 2016.

A. Bull. Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, 12: 2879–2904, 2011.

E. Burnaev, I. Panin, and B. Sudret. Effective design for Sobol indices estimation based on polynomial chaos expansions. In *Symposium on Conformal and Probabilistic Prediction with Applications*, pages 165–184. Springer, 2016.

W.F. Caselton and J.V. Zidek. Optimal monitoring network designs. *Statistics & Probability Letters*, 2(4):223–227, 1984.

C. Chevalier. *Fast uncertainty reduction strategies relying on Gaussian process models*. PhD thesis, University of Bern, 2013.

C. Chevalier, D. Ginsbourger, and X. Emery. Corrected Kriging update formulae for batch-sequential data assimilation. In: Pardo-Igúzquiza E., Guardiola-Albert C., Heredia J., Moreno-Merino L., Durán J., Vargas-Guzmán J. (eds) Mathematics of Planet Earth. Lecture Notes in Earth System Sciences. Springer, Berlin, Heidelberg 2014.

C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, and Y. Richet. Fast Kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56 (4):455–465, 2014.

J.A. Christen and B. Sansó. Advances in the sequential design of computer experiments based on active learning. *Communications in Statistics - Theory and Methods*, 40(24):4467–4483, 2011.

D. A. Cohn. Neural network exploration using optimal experiment design. *Neur. Netw.*, 9(6):1071-1083, 1996.

E. Contal, V. Perchet, and N. Vayatis. Gaussian process optimization with mutual information. In *Proceedings of the 31st International Conference on Machine Learning*, 2014.

C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *J. Amer. Statist. Assoc.*, 86(416):953–963, 1991.

A.C. Damianou and N. D. Lawrence. Deep Gaussian processes. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*, 2013.

N. de Freitas, A.J. Smola, and M. Zogui. Exponential regret bounds for Gaussian process bandits with deterministic observations. In *Proceedings of the 29th International Conference on Machine Learning*, 2012.

D. den Hertog, J. P. C. Kleijnen, and A. Y. D. Siem. The correct Kriging variance estimated by bootstrapping. *The Journal of the Operational Research Society*, 57 (4):400–409, 2006.

V. Dubourg, Sudret. B., and F. Deheeger. Metamodel-based importance sampling for structural reliability analysis. *Probabilistic Engineering Mechanics*, 33:47–57, 2013.

D. Dupuy, C. Helbert, and J. Franco. Dicedesign and Diceeval: two R packages for design and analysis of computer experiments. *Journal of Statistical Software*, 65(11), 2015.

M. Emmerich, K. Giannakoglou, and B. Naujoks. Single-and multiobjective optimization assisted by Gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10 (4), 2006.

M.T.M. Emmerich, A.H. Deutz, and J.W. Klinkenberg. Hypervolume-based expected improvement: monotonicity properties and exact computation. In *Evolutionary Computation (CEC), 2011 IEEE Congress on*, pages 2147–2154. IEEE, 2011.

K. Fang. *Design and modeling for computer experiments*, volume 6. CRC Press, 2006.

P. Feliot, J. Bect, and E. Vazquez. A Bayesian approach to constrained single- and multi-objective optimization. *Journal of Global Optimization*, 67(1-2):97–133, 2017.

A. I. J. Forrester, N. W. Bressloff, and A. J Keane. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proc. R. Soc. A*, 462:2177–2204, 2006.

A. I. J. Forrester, A. Sóbester, and A. J Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 463:3251–3269, 2007.

P. I. Frazier, W. B. Powell, and S. Dayanik. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439, 2008.

G. Gamblin, M. Couplet, and B. looss. Numerical studies of space-filling designs: optimization of latin hypercube samples and subprojection properties. *Journal of Simulation*, 7(4):276–289, 2013.

B. Gauthier and L. Pronzato. Spectral approximation of the imse criterion for optimal designs in kernel-based interpolation meta-models. *SIAM/ASA J. Uncertainty Quantification*, 2(1):805–825, 2014.

B. Gauthier and L. Pronzato. Convex relaxation for imse optimal design in random-field models. *Computational Statistics and Data Analysis*, 113:375–394, 2017.

D. Ginsbourger and R. Le Riche. Towards Gaussian process-based optimization with finite time horizon. In Alessandra Giovagnoli, Anthony C. Atkinson, Bernard Torsney, and Caterina May, editors, *mODa 9 Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, pages 89–96. Physica-Verlag HD, 2010.

D. Ginsbourger, C. Helbert, and L. Carraro. Discrete mixtures of kernels for Kriging-based optimization. *Quality and Reliability Eng. Int.*, 24(6):681–691, August 2008.

D. Ginsbourger, R. Le Riche, and L. Carraro. Kriging is well-suited to parallelize optimization. In Lim Meng Hiot, Yew Soon Ong, Yoel Tenne, and Chi-Keong Goh, editors, *Computational Intelligence in Expensive Optimization Problems*, Adaptation Learning and Optimization, pages 131–162. Springer Berlin Heidelberg, 2010.

D. Ginsbourger, J. Baccou, C. Chevalier, N. Garland, F. Perales, and Y. Monerie. Bayesian adaptive reconstruction of profile optima and optimizers. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):490–510, 2014.

J. Gonzalez, M. Osborne, and N. Lawrence. GLASSES: relieving the myopia of Bayesian optimisation. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, 2016.

A. Gorodetsky and Y. Marzouk. Mercer kernels and integrated variance experimental design: connections between gaussian process regression and polynomial approximation. *SIAM/ASA J. Uncertainty Quantification*, 4(1):796–828, 2016.

A. Gotovos, N. Casati, G. Hitz, and A. Krause. Active learning for level set estimation. In *IJCAI*, pages 1344–1350, 2013.

R. Gramacy. *Bayesian Treed Gaussian Process Models*. PhD thesis, University of California at Santa Cruz, 2005.

R. B. Gramacy and N.G. Polson. Particle learning of Gaussian process models for sequential design and optimization. *Journal of Computational and Graphical Statistics*, 20(1):102–118, 2011.

R. B. Gramacy and M. A. Taddy. Categorical inputs, sensitivity analysis, optimization and importance tempering with tgp version 2, an r package for treed Gaussian process models. *Journal of Statistical Software*, 33(6), 2010.

R.B. Gramacy, G. Gray, S. Le Digabel, H. Lee, P. Ranjan, G. Wells, and S. Wild. Modeling an augmented lagrangian for blackbox constrained optimization. *Technometrics*, 58(1):1–11, 2016.

S. Grünewälder, J.-Y. Audibert, M. Opper, and J. Shawe-Taylor. Regret bounds for Gaussian process bandit problems. In *International Conference on Artificial Intelligence and Statistics*, 2010.

O. Harari and D. Steinberg. Optimal designs for Gaussian process models via spectral decomposition. *Journal of Statistical Planning and Inference*, 154,:87–101, November 2014.

X. He, R. Tuo, and C. F. J. Wu. Optimization of multi-fidelity computer experiments via the EQIE criterion. *Technometrics*, 59(1), 2017.

J. M. Hernández-Lobato, M. W. Hoffman, and Z. Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In *Neural Information Processing Systems*, 2014.

D. Huang, T.T. Allen, W. Notz, and R.A. Miller. Sequential Kriging optimization using multiple fidelity evaluations. *Sructural and Multidisciplinary Optimization*, 32:pp. 369–382 (14), 2006.

D. Huang, T.T. Allen, W. Notz, and N. Zheng. Global optimization of stochastic black-box systems via sequential Kriging meta-models. *Journal of Global Optimization*, 34:441–466, 2006.

M. Jala, C. Levy-Leduc, E. Moulines, E. Conil, and J. Wiart. Sequential design of computer experiments for the assessment of fetus exposure to electromagnetic fields. *Technometrics*, 58:30–42, 2016.

J. Janusevskis and R. Le Riche. Simultaneous Kriging-based estimation and optimization of mean response. *Journal of Global Optimization*, 55(2):313-336, 2013.

J. Janusevskis, R. Le Riche, and D. Ginsbourger. Expected improvements for the asynchronous parallel global optimization of expensive functions : potentials and challenges. In *Learning and Intelligent Optimization*, Lecture Notes in Computer Science, pages 413–418, 2012.

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.

K. Kandasamy, J. Schneider, and B. Poczos. High dimensional Bayesian optimisation and bandits via additive models. In *Proceedings of The 32nd International Conference on Machine Learning, pp. 295-304*, 2015.

J. Knowles. Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. *IEEE transactions on evolutionnary computation*, 2005.

J.R. Koehler and A.B. Owen. *Handbook of Statistics*, chapter Computer Experiments, pages 261–308. 1996.

A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.

T. Krityakierne and D. Ginsbourger. Global Optimization with Sparse and Local Gaussian Process Models. In *International Workshop on Machine Learning, Optimization and big Data*, 2015.

H. J. Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *J. Basic Engineering*, 86:97–106, 1964.

T. Labopin-Richard and V. Picheny. Sequential design of experiments for estimating percentiles of black-box functions. *Statistica Sinica*, 2017.

R. Lam, K. Wilcox, and D.H. Wolpert. Bayesian optimization with a finite budget: An approximate dynamic programming approach. In *Advances In Neural Information Processing Systems 29 (NIPS 2016)*, 2016.

L. Le Gratiet, C. Cannamela, and B. looss. Cokriging-based sequential design strategies using fast cross-validation for multi-fidelity computer codes. *Technometrics*, 57:418–427, 2015.

D. J. C. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):589-603, 1992.

S. Marmin. *Warping and sampling approaches to non-stationary Gaussian process modelling*. PhD thesis, Ecole Centrale de Marseille and University of Bern, 2017.

J. Marzat, E. Walter, and H. Piet-Lahanier. Worst-case global optimization of black-box functions through Kriging and relaxation. *Journal of Global Optimization*, 55(4):707–727, 2013.

J. Mockus. *Bayesian Approach to Global Optimization. Theory and Applications*. Kluwer Academic Publisher, Dordrecht, 1989.

J. Mockus, V. Tiesis, and A. Zilinskas. The application of Bayesian methods for seeking the extremum. In L. Dixon and Eds G. Szego, editors, *Towards Global Optimization*, volume 2, pages 117–129. Elsevier, 1978.

W. G. Müller. *Collecting Spatial Data: Optimum Design of Experiments for Random Fields (Third revised and extended edition)*. Springer, 2007.

J. Oakley. Estimating percentiles of uncertain computer code outputs. *J. Roy. Statist. Soc. Ser. C*, 53(1):83–93, 2004.

M. A. Osborne, R. Garnett, and S. J. Roberts. Gaussian processes for global optimization. In *3rd International Conference on Learning and Intelligent Optimization (LION3)*, 2009.

J.M. Parr. *Improvement Criteria for Constraint Handling and Multiobjective Optimization*. PhD thesis, University of Southampton, 2012.

P. Perdikaris, D. Venturi, and G.E. Karniadakis. Multifidelity information fusion algorithms for high-dimensional systems and massive data sets. *SIAM J. Sci. Comput.*, 38.4:521–538, 2016.

V. Picheny. A stepwise uncertainty reduction approach to constrained global optimization. In *Artificial Intelligence and Statistics*, pages 787–795, 2014.

V. Picheny. Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction. *Statistics and Computing*, 25(6):1265–1280, 2015.

V. Picheny, D. Ginsbourger, O. Roustant, R.T. Haftka, and N.-H. Kim. Adaptive designs of experiments for accurate approximation of target regions. *Journal of Mechanical Design*, 132(7), July 2010.

V. Picheny, T. Wagner, and D. Ginsbourger. Benchmark of kriging-based infill criteria for noisy optimization. *Structural and Multidisciplinary Optimization*, 48(3):607–626, 2013. doi: 10.1007/s00158-013-0919-4.

M.U. Poloczek, J. Wang, and P.I. Frazier. Multi-information source optimization. In *Neural Information Processing Systems (NIPS)*, 2017.

L. Pronzato. Sensitivity analysis via Karhunen-Loève expansion of a random field model: estimation of Sobol' indices and experimental design. Reliability Engineering & System Safety, 2018.

L. Pronzato and W. Müller. Design of computer experiments: space filling and beyond. Statistics and Computing, 3: 681-701, 2012.

P. Ranjan, D. Bingham, and G. Michailidis. Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4):527-541, November 2008.

I.O. Rhyzov. On the convergence rates of expected improvement methods. *Operations Research*, pages 1515 – 1528, 2016.

H. Robbins. Some aspects of the sequential design of experiments. *Bulletin American Mathematical Society*, 55: 527–535, 1952.

J. Sacks and S. Schiller. *Statistical Decision Theory and Related Topics IV (Vol. 2)*, chapter Spatial Designs, pages 385–399. Springer Verlag, 1988.

J. Sacks, S.B. Schiller, and W.J. Welch. Design of computer experiments. *Technometrics*, 31(1):41–47, 1989a.

J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989b.

T.J. Santner, B.J. Williams, and W. Notz. *The Design and Analysis of Computer Experiments*. Springer, New York, 2003.

M. Schonlau. *Computer Experiments and Global Optimization*. PhD thesis, University of Waterloo, 1997.

M. Schonlau and W.J. Welch. Global optimization with nonparametric function fitting. In *Proceedings of the ASA, Section on Physical and Engineering Sciences. Amer. Statist. Assoc.*, 1996.

W. Scott, P.I. Frazier, and W.B. Powell. The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression. *SIAM Journal on Optimization*, 2011.

A. Shah, A.G. Wilson, and Z. Ghahramani. Student-t processes as alternatives to Gaussian processes. In *International Conference on Artificial Intelligence and Statistics*, 2014.

M. C. Shewry and H.P. Wynn. Maximum entropy sampling. *Journal of Applied Statistics*, 14(2), 1987.

A. Singh, A. Krause, C. Guestrin, and W.J. Kaiser. Efficient informative sensing using multiple robots. *Journal of Artificial Intelligence Research*, 34:707–755, 2009.

H. Snoek, J.and Larochelle and R. P. Adams. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems 25*, 2012.

J. Snoek, O. Rippel, K. Swersky, S. Kiros, N. Satish, N. Sundaram, M.M.A. Patwary, Prabhat, and R.P. Adams. Scalable Bayesian optimization using deep neural networks. In *Proceedings of the 32 nd International Conference on Machine Learning*, 2015.

N. Srinivas, A. Krause, S. Kakade, and M. Seeger. Gaussian process optimization in the bandit setting: No regret and experimental design. In *Proceedings of the 27 th International Conference on Machine Learning*, 2010.

N. Srinivas, Krause. A., S. Kakade, and M. Seeger. Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58:3250–3265, 2012.

M.L. Stein. *Interpolation of spatial data, some theory for Kriging*. Springer, 1999.

E. Vazquez and J. Bect. A sequential Bayesian algorithm to estimate a probability of failure. In *Proceedings of the 15th IFAC Symposium on System Identification, SYSID 2009 15th IFAC Symposium on System Identification, SYSID 2009*, Saint-Malo France, 2009.

E. Vazquez and J. Bect. Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *Journal of Statistical Planning and Inference*, 140:11:3088–3095, 2010.

E. Vazquez and J. Bect. Sequential search based on Kriging: convergence analysis of some algorithms. In *Proc. 58th ISI World Statistics Congress.*, 2011.

J. Villemonteix, E. Vazquez, and E. Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4):509–534, 2009.

H. Wang, G. Lin, and J. Li. Gaussian process surrogates for failure detection: a Bayesian experimental design approach. *Journal of Computational Physics*, 313:247–259, 2016.

Z. Wang, M. Zoghi, F. Hutter, D. Matheson, and N. de Freitas. Bayesian optimization in high dimensions via random embeddings. In *International Joint Conferences on Artificial Intelligence*, 2013.

Z. Wang, C. Li, S. Jegelka, and P. Kohli. Batched high-dimensional Bayesian optimization via structural kernel learning. In *International Conference on Machine Learning (ICML)*, 2017.

Z. Wang, C. Gehring, P. Kohli, and S. Jegelka. Batched large-scale Bayesian optimization in high-dimensional spaces. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2018.

B. J. Williams, T. J. Santner, and W. I. Notz. Sequential design of computer experiments to minimize integrated response functions. *Statistica Sinica*, 10:1133–1152, 2000.

J. Wu and P.I. Frazier. The parallel knowledge gradient method for batch Bayesian optimization. In *Neural Information Processing Systems (NIPS)*, 2016.

J. Wu, M.U. Poloczek, A.G. Wilson, and P.I. Frazier. Bayesian optimization with gradients. In *Neural Information Processing Systems (NIPS)*, 2017.

D. Yarotsky. Examples of inconsistency in optimization by expected improvement. *Journal of Global Optimization*, 56(4):1773–1790, 2013.