Stochastic and deterministic algorithms for continuous black-box optimization

Wang, H.

APPENDIX A

Gaussian Distribution

Assume the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and a measurable space $(\mathbb{R}, \mathscr{B})$, where $\mathscr{B}$ is the Borel algebra on $\mathbb{R}$. A random variable $X : \Omega \to \mathbb{R}$ is said to be normally distributed if and only if its probability distribution $\mathbb{P}_X : \mathscr{B} \to [0,1]$, defined as a push-forward measure, $\forall B \in \mathscr{B},\ \mathbb{P}_X(B) \coloneqq \mathbb{P}(X^{-1}[B])$, admits the following form:

$$\mathbb{P}_X(B) = \int_B \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \mathrm{d}\lambda, \tag{A.1}$$

where λ is the Lebesgue measure on $\mathbb{R}$ and m, σ^2 are the mean and variance of X, respectively. We typically use the notation $X \sim \mathcal{N}(m, \sigma^2)$. This distribution $\mathbb{P}_X$ is called *Gaussian measure* and the notation $\mathcal{G}_{m,\sigma^2}$ is assigned to it. The *cumulative distribution function* (c.d.f.) of X is

$$\Phi_{m,\sigma^2}(x) = \mathcal{G}_{m,\sigma^2}\left(\{X \in \mathbb{R}\ :\ X \leq x\}\right).$$

In addition, the *probability density function* (p.d.f.) of X is the Radon-Nikodym derivative of $\mathcal{G}_{m,\sigma^2}$ w.r.t. λ:

$$\phi_{m,\sigma^2}(x) = \frac{\mathrm{d}\mathcal{G}_{m,\sigma^2}}{\mathrm{d}\lambda} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \tag{A.2}$$

that is, by definition, the integrand in Eq. (A.1). In the multivariate case, consider the measurable space $(\mathbb{R}^n, \mathscr{B}^n)$ where $\mathscr{B}^n$ is the Borel algebra on $\mathbb{R}^n$. A random vector $\mathbf{x} = (X_1, X_2, \ldots, X_n)^\top : \Omega \to \mathbb{R}^n$ is said to follow the multivariate Gaussian distribution, if and only if any linear combination $\mathbf{c}^\top\mathbf{x},\ \mathbf{c} \in \mathbb{R}^n$ admits the distribution as in Eq. (A.1). In addition, the distribution of $\mathbf{x}$ is

$$\forall B \in \mathscr{B}^n,\ \mathcal{G}^n_{\mathbf{m},\mathbf{K}}(B) = \int_B (2\pi)^{-\frac{n}{2}} \det(\mathbf{K})^{\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{y}-\mathbf{m})^\top \mathbf{K}^{-1}(\mathbf{y}-\mathbf{m})\right) \mathrm{d}\lambda^n,$$

where λ^n is the n-dimensional Lebesgue measure on $(\mathbb{R}^n, \mathscr{B}^n)$ and $\mathbf{m}, \mathbf{K}$ are the mean and covariance matrix of $\mathbf{x}$. As with the univariate case, we shall take the notation $\mathbf{x} \sim \boldsymbol{\mathcal{N}}(\mathbf{m}, \mathbf{K})$ and its cumulative distribution function is,

$$\Phi^n_{\mathbf{m},\mathbf{K}}(\boldsymbol{\zeta}) = \mathscr{G}^n_{\mathbf{m},\mathbf{K}}\left(\{\mathbf{x} \in \mathbb{R}^n \ : \ \mathbf{x} \leq \boldsymbol{\zeta}\}\right).$$

Given an arbitrary partition on $\mathbf{x} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)^\top$, in which $\mathbf{x}_1$, $\mathbf{x}_2$ have n_1 and n_2 components, respectively. The distribution of $\mathbf{x}$ can be re-written as

$$\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \boldsymbol{\mathcal{N}}\left(\begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix}, \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \right),$$

where all sub-mean vectors and sub-covariance matrices are obtained by applying the same partition on $\mathbf{m}$ and $\mathbf{K}$. The *marginal distribution* of $\mathbf{x}_1$ is Gaussian:

$$\mathbf{x}_1 \sim \boldsymbol{\mathcal{N}}\left(\mathbf{m}_1, \mathbf{K}_{11}\right). \tag{A.3}$$

The result holds for $\mathbf{x}_2$ in the same manner. In addition, the *conditional distribution* of $\mathbf{x}_1$ on $\mathbf{x}_2 = \mathbf{v}$ is Gaussian (Tong, 2012):

$$\mathbf{x}_1 \mid \mathbf{x}_2 = \mathbf{v} \sim \boldsymbol{\mathcal{N}}\left(\mathbf{m}_1 + \mathbf{K}_{12}\mathbf{K}_{22}^{-1}(\mathbf{v} - \mathbf{m}_2),\ \mathbf{K}_{11} - \mathbf{K}_{12}\mathbf{K}_{22}^{-1}\mathbf{K}_{21}\right). \tag{A.4}$$

Often, the value of Gaussian random variables is restricted:

$$X \sim \mathcal{N}(m, \sigma^2), \quad X_R = \max\{0, X\}.$$

The random variable X_R is known as the *Rectified Gaussian* and its distribution shall be denoted as $\mathcal{N}_R(m, \sigma^2)$. Note that the rectification "concentrates" all the probability measure in $(-\infty, 0)$ to the rectification point 0, leading to an infinite impulse at this point. Thus, the p.d.f. of X_R is:

$$p_{X_R}(x) = \Phi_{m,\sigma^2}(0)\delta(x) + \phi_{m,\sigma^2}(x)H(x), \tag{A.5}$$

where δ is the Dirac delta (distribution)[1] and H is the step function:

$$\delta(x) = \begin{cases} \infty & x = 0, \\ 0 & x \neq 0. \end{cases}, \quad H(x) = \begin{cases} 0 & x \leq 0, \\ 1 & x > 0. \end{cases}$$

The rectification is sometimes confused with the so-called *truncated Gaussian*, which is the distribution of a Gaussian variable $X \sim \mathcal{N}(m, \sigma^2)$ conditioning on an interval $(a, b) \subset \mathbb{R}$:

$$p(x \mid a < X < b) = \frac{\phi_{m,\sigma^2}(x)}{\Phi_{m,\sigma^2}(b) - \Phi_{m,\sigma^2}(a)}.$$

[1] Formally, the Dirac delta should be defined either as a distribution or measure. We use the heuristic characterization here for the sake of simplicity.

APPENDIX B

Proof

B.1 Theorem 5.3

Proof. Let us define $\mathbf{a} \coloneqq -\nabla f_1^{(2)}$ and $\mathbf{b} \coloneqq \nabla f_2^{(1)}$, such that $\tilde{\mathbf{A}}_1 = \mathbf{b}\mathbf{a}^\top$ and

$$\nabla^2 \mathcal{H}_{\mathbf{F}}(\mathbf{X}) = \begin{pmatrix} \mathbf{D}_1 & \mathbf{b}\mathbf{a}^\top \\ \mathbf{a}\mathbf{b}^\top & \mathbf{D}_2 \end{pmatrix}.$$

For two block matrices, their column vectors are denoted as: $\mathbf{D}_1 = (\mathbf{d}_1, \ldots, \mathbf{d}_n)$ and $\mathbf{D}_2 = (\mathbf{d}'_1, \ldots, \mathbf{d}'_n)$. The hypervolume Hessian is of size $2n \times 2n$ and its determinant can be simplified using the Laplace expansion along **the first** n **rows** of the $\nabla^2 \mathcal{H}_{\mathbf{F}}(\mathbf{X})$. To achieve this, n distinct columns need to be selected out of $2n$ rows. Let S be the set of the n-element subsets of $\{1, 2, \ldots, 2n\}$:

$$S = \{\{1, 2, \ldots, n\}, \{1, 2, \ldots, n-1, n+1\}, \ldots\}$$

For every $L \in S$, we define its complement $L' \coloneqq \{1, 2, \ldots, 2n\} \setminus L$. Note that a permutation is defined on $\{1, 2, \ldots, 2n\}$, by appending L' to L: $\{L, L'\}$ and we shall use $N(L)$ to denote the number of inversions in $\{L, L'\}$. According to the Laplace expansion, such a determinant can be expressed as:

$$\det\left(\nabla^2 \mathcal{H}_{\mathbf{F}}(\mathbf{X})\right) = \sum_{L \in S} (-1)^{N(L)}\, b_L c_{L'},$$

where b_L is the cofactor of the hypervolume Hessian, which is the determinant of the minor matrix obtained by keeping the first n rows and n columns given in L. Similarly, $c_{L'}$ is the *complementary* cofactor of b_L, obtained by removing the first n rows and n columns given in L. For example, if $L = \{1, 2, \ldots, n\}$, then $b_L = \det(\mathbf{D}_1)$ and $c_{L'} = \det(\mathbf{D}_2)$. In particular, when L contains two or more elements from $\{2n+1, 2n+2, \ldots, 2n\}$, meaning that at least two columns from

$\mathbf{b}\mathbf{a}^\top$ are chosen to compute b_L, it is obvious that the cofactor b_L is zeros because all the columns from $\mathbf{b}\mathbf{a}^\top$ are linear dependent. Using this observation, the expansion can be simplified:

$$\begin{aligned}
\det\left(\nabla^2\mathcal{H}_{\mathbf{F}}(\mathbf{X})\right) = &\underbrace{\det(\mathbf{D}_1)\det(\mathbf{D}_2)}_{L=\{1,2,\ldots,n\}} \\
&+ \underbrace{(-1)^1\det((\mathbf{d}_1,\ldots,\mathbf{d}_{n-1},a_1\mathbf{b}))\det((b_n\mathbf{a},\mathbf{d}'_2\ldots,\mathbf{d}'_n))}_{L=\{1,2,\ldots,n-1,n+1\}} \\
&+ \underbrace{(-1)^2\det((\mathbf{d}_1,\ldots,\mathbf{d}_{n-1},a_2\mathbf{b}))\det((b_n\mathbf{a},\mathbf{d}'_1,\mathbf{d}'_3,\ldots,\mathbf{d}'_n))}_{L=\{1,2,\ldots,n-1,n+2\}} + \cdots \\
&+ \underbrace{(-1)^n\det((\mathbf{d}_1,\ldots,\mathbf{d}_{n-1},a_2\mathbf{b}))\det\left((b_n\mathbf{a},\mathbf{d}'_1,\mathbf{d}'_2,\ldots,\mathbf{d}'_{n-1})\right)}_{L=\{1,2,\ldots,n-1,2n\}} + \cdots
\end{aligned}$$

There are n terms shown in the equation above, resulting from choosing the first $n-1$ columns and one column from $\{2n+1, 2n+2, \ldots, 2n\}$. Those terms can also be simplified:

$$\begin{aligned}
&(-1)^1 a_1 b_n\det((\mathbf{d}_1,\ldots,\mathbf{d}_{n-1},\mathbf{b}))\det((\mathbf{a},\mathbf{d}'_2\ldots,\mathbf{d}'_n)) \\
&+(-1)^3 a_2 b_n\det((\mathbf{d}_1,\ldots,\mathbf{d}_{n-1},\mathbf{b}))\det((\mathbf{d}'_1,\mathbf{a},\mathbf{d}'_3\ldots,\mathbf{d}'_n)) + \cdots \\
&+(-1)^{2i-1} a_i b_n\det((\mathbf{d}_1,\ldots,\mathbf{d}_{n-1},\mathbf{b}))\underbrace{\det\left((\mathbf{d}'_1,\ldots,\mathbf{d}'_{i-1},\mathbf{a},\mathbf{d}'_{i+1},\ldots,\mathbf{d}'_n)\right)}_{\text{move } \mathbf{a} \text{ to the } i\text{-th column}} \\
&= -b_n\det((\mathbf{d}_1,\ldots,\mathbf{d}_{n-1},\mathbf{b}))\det(\mathbf{D}_2)\sum_{i=1}^{n} a_i\frac{\det\left((\mathbf{d}'_1,\ldots,\mathbf{d}'_{i-1},\mathbf{a},\mathbf{d}'_{i+1},\ldots,\mathbf{d}'_n)\right)}{\det((\mathbf{d}'_1,\mathbf{d}'_2,\ldots,\mathbf{d}'_n))} \qquad \text{(B.1)} \\
&= -b_n\det((\mathbf{d}_1,\ldots,\mathbf{d}_{n-1},\mathbf{b}))\det(\mathbf{D}_2)\,\mathbf{a}^\top\mathbf{D}_2^{-1}\mathbf{a}
\end{aligned}$$

Note that the last step above is according to Cramer's rule for the equation $\mathbf{D}_2\mathbf{x} = \mathbf{a}$ ($\mathbf{D}_1$ and $\mathbf{D}_2$ are assumed to be nonsingular):

$$x_i = \frac{\det\left((\mathbf{d}'_1,\ldots,\mathbf{d}'_{i-1},\mathbf{a},\mathbf{d}'_{i+1},\ldots,\mathbf{d}'_n)\right)}{\det((\mathbf{d}'_1,\mathbf{d}'_2,\ldots,\mathbf{d}'_n))}$$

In principle, the same simplification here can be applied to other terms in the hypervolume Hessian determinant:

$$
\begin{aligned}
&\det\left(\nabla^2 \mathcal{H}_{\mathbf{F}}(\mathbf{X})\right) = \det\left(\mathbf{D}_1\right)\det\left(\mathbf{D}_2\right) \\
&- \underbrace{b_n \det\left(\left(\mathbf{d}_1, \ldots, \mathbf{d}_{n-1}, \mathbf{b}\right)\right)\det\left(\mathbf{D}_2\right)\mathbf{a}^\top \mathbf{D}_2^{-1}\mathbf{a}}_{\text{drop column } n \text{ from } \mathbf{D}_1} \\
&- \underbrace{b_{n-1} \det\left(\left(\mathbf{d}_1, \ldots, \mathbf{d}_{n-2}, \mathbf{b}, \mathbf{d}_n\right)\right)\det\left(\mathbf{D}_2\right)\mathbf{a}^\top \mathbf{D}_2^{-1}\mathbf{a}}_{\text{drop column } n-1 \text{ from } \mathbf{D}_1} - \cdots \\
&- \underbrace{b_1 \det\left(\left(\mathbf{b}, \mathbf{d}_2, \ldots, \mathbf{d}_n\right)\right)\det\left(\mathbf{D}_2\right)\mathbf{a}^\top \mathbf{D}_2^{-1}\mathbf{a}}_{\text{drop column } 1 \text{ from } \mathbf{D}_1} \\
&= \det\left(\mathbf{D}_1\right)\det\left(\mathbf{D}_2\right)\left[1 - \mathbf{a}^\top \mathbf{D}_2^{-1}\mathbf{a}\sum_{i=1}^{n} b_i \frac{\det\left(\left(\mathbf{d}_1, \ldots, \mathbf{d}_{i-1}, \mathbf{b}, \mathbf{d}_{i+1}, \ldots, \mathbf{d}_n\right)\right)}{\det\left(\left(\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_n\right)\right)}\right] \\
&= \left(1 - \left(\mathbf{a}^\top \mathbf{D}_2^{-1}\mathbf{a}\right)\left(\mathbf{b}^\top \mathbf{D}_1^{-1}\mathbf{b}\right)\right)\det\left(\mathbf{D}_1\right)\det\left(\mathbf{D}_2\right)
\end{aligned}
$$

Again, in the last step above the same argument as in Eq. (B.1) is applied. Because matrices $\mathbf{D}_1$ and $\mathbf{D}_2$ are nonsingular, the hypervolume Hessian matrix is nonsingular as long as $1 - \left(\mathbf{a}^\top \mathbf{D}_2^{-1}\mathbf{a}\right)\left(\mathbf{b}^\top \mathbf{D}_1^{-1}\mathbf{b}\right)$ is not zero. $\square$

Bibliography

Ababou, R., A. C. Bagtzoglou, and E. F. Wood (1994). On the Condition Number of Covariance Matrices in Kriging, Estimation, and Simulation of Random Fields. *Mathematical Geology 26*(1), 99–133.

Agrawal, R. B., K. Deb, and R. B. Agrawal (1995). Simulated Binary Crossover for Continuous Search Space. *Complex systems 9*(2), 115–148.

Andrianakis, I. and P. G. Challenor (2012). The Effect of the Nugget on Gaussian Process Emulators of Computer Models. *Computational Statistics & Data Analysis 56*(12), 4215–4228.

Aronszajn, N. (1950). Theory of Reproducing Kernels. *Transactions of the American mathematical society 68*(3), 337–404.

Auer, P. (2002). Using Confidence Bounds for Exploitation-exploration Trade-offs. *Journal of Machine Learning Research 3*(Nov), 397–422.

Auer, P., N. Cesa-Bianchi, and P. Fischer (2002). Finite-time Analysis of the Multiarmed Bandit Problem. *Machine learning 47*(2-3), 235–256.

Auger, A., D. Brockhoff, and N. Hansen (2010). Mirrored Variants of the $(1, 2)$-CMA-ES Compared on the Noisy BBOB-2010 Testbed. In *Proceedings of the 12th annual conference companion on Genetic and evolutionary computation*, GECCO '10, New York, NY, USA, pp. 1575–1582. ACM.

Auger, A., D. Brockhoff, and N. Hansen (2011a). Analyzing the Impact of Mirrored Sampling and Sequential Selection in Elitist Evolution Strategies. In H.-G. Beyer and W. B. Langdon (Eds.), *Proceedings of the 11th workshop proceedings on Foundations of genetic algorithms*, FOGA '11, ACM, New York, pp. 127–138.

Auger, A., D. Brockhoff, and N. Hansen (2011b). Mirrored Sampling in Evolution Strategies with Weighted Recombination. In *Proceedings of the 13th Annual*

Conference on Genetic and Evolutionary Computation, GECCO ’11, New York, NY, USA, pp. 861–868. ACM.

Auger, A. and N. Hansen (2011). Theory of Evolution Strategies: A New Perspective. *Theory of Randomized Search Heuristics: Foundations and Recent Developments 1*, 289–325.

Bache, K. and M. Lichman (2013). UCI Machine Learning Repository. `http://archive.ics.uci.edu/ml`.

Bäck, T. (1995). Order Statistics for Convergence Velocity Analysis of Simplified Evolutionary Algorithms. Volume 3 of *Foundations of Genetic Algorithms*, pp. 91 – 102. Elsevier.

Bäck, T. (1996). *Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms.* Oxford, UK: Oxford University Press.

Bäck, T., C. Foussette, and P. Krause (2013). *Contemporary Evolution Strategies* (1 ed.). Natural Computing Series. Springer-Verlag Berlin Heidelberg.

Bäck, T. and H.-P. Schwefel (1993, March). An Overview of Evolutionary Algorithms for Parameter Optimization. *Evol. Comput. 1*(1), 1–23.

Bartz-Beielstein, T., C. W. G. Lasarczyk, and M. Preuss (2005, Sept). Sequential Parameter Optimization. In *2005 IEEE Congress on Evolutionary Computation*, Volume 1, pp. 773–780 Vol.1.

Beume, N., B. Naujoks, and M. Emmerich (2007). SMS-EMOA Multiobjective Selection Based on Dominated Hypervolume. *European Journal of Operational Research 181*(3), 1653–1669.

Beyer, H.-G. (1993, June). Toward a Theory of ‘Evolution Strategies’: Some Asymptotical Results from the $(1, ^+ \lambda)$-Theory. *Evolution Computation 1*(2), 165–188.

Beyer, H.-G. (2013). *The Theory of Evolution Strategies.* Springer Science & Business Media.

Bischl, B., S. Wessing, N. Bauer, K. Friedrichs, and C. Weihs (2014). MOI-MBO: Multiobjective Infill for Parallel Model-Based Optimization. In *International Conference on Learning and Intelligent Optimization*, pp. 173–186. Springer.

Björck, A. (1994, Jan). Numerics of Gram-Schmidt Orthogonalization. *Linear Algebra and its Applications 197-198*, 297–316.

Bonabeau, E., M. Dorigo, and G. Theraulaz (1999). *Swarm Intelligence: from Natural to Artificial Systems.* Number 1. Oxford university press.

Bonnans, J.-F., J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal (2006). *Numerical Optimization: Theoretical and Practical Aspects.* Springer Science & Business Media.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization.* New York, NY, USA: Cambridge University Press.

Breiman, L. (1996). Bagging Predictors. *Machine Learning 24*(2), 123–140.

Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984). *Classification and Regression Trees.* CRC press.

Brockhoff, D., A. Auger, N. Hansen, D. V. Arnold, and T. Hohm (2010). Mirrored Sampling and Sequential Selection for Evolution Strategies. In R. Schaefer, C. Cotta, J. Kolodziej, and G. Rudolph (Eds.), *Proceedings of the 11th international conference on Parallel problem solving from nature: Part I*, PPSN'10, Springer-Verlag, Berlin, pp. 11–21.

Buhmann, M. D. (2003). *Radial Basis Functions: Theory and Implementations*, Volume 12. Cambridge university press.

Chalupka, K., C. K. I. Williams, and I. Murray (2013, February). A Framework for Evaluating Approximation Methods for Gaussian Process Regression. *J. Mach. Learn. Res. 14*(1), 333–350.

Chen, T. and J. Ren (2009, March). Bagging for Gaussian Process Regression. *Neurocomput. 72*(7-9), 1605–1610.

Chevalier, C. and D. Ginsbourger (2013). Fast Computation of the Multi-points Expected Improvement with Applications in Batch Selection. In G. Nicosia and P. Pardalos (Eds.), *Learning and Intelligent Optimization*, Berlin, Heidelberg, pp. 59–69. Springer Berlin Heidelberg.

Cressie, N. (1990). The Origins of Kriging. *Mathematical geology 22*(3), 239–252.

Cressie, N. (2015). *Statistics for spatial data.* John Wiley & Sons.

Csató, L. and M. Opper (2002). Sparse On-line Gaussian Processes. *Neural computation 14*(3), 641–668.

De Jong, K. A. (1975). *An Analysis of the Behavior of a Class of Genetic Adaptive Systems.* Ph. D. thesis, Ann Arbor, MI, USA. AAI7609381.

Deb, K., S. Agrawal, A. Pratap, and T. Meyarivan (2000). A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. In *Parallel Problem Solving from Nature, 2000 Conference on*, pp. 849–858. Berlin, Heidelberg: Springer.

den Hertog, D., J. P. C. Kleijnen, and a. Y. D. Siem (2006). The Correct Kriging Variance Estimated by Bootstrapping. *Journal of the Operational Research Society 57*(4), 400–409.

Dick, J. and F. Pillichshammer (2010). *Digital Nets and Sequences: Discrepancy Theory and Quasi-Monte Carlo Integration.* New York, NY, USA: Cambridge University Press.

Dunn, J. C. (1973). A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Journal of Cybernetics 3*(3), 32–57.

Eaton, M. L. (1983). *Multivariate Statistics: A Vector Space Approach.* Wiley, New York.

Ehrgott, M. (2006). *Multicriteria Optimization.* Springer Science & Business Media.

Emmerich, M. (2005). *Single-and Multi-Objective Evolutionary Design Optimization assisted by Gaussian Random Field Metamodels.* Ph. D. thesis, FB Informatik, TU Dortmund.

Emmerich, M. and A. Deutz (2014). Time Complexity and Zeros of the Hypervolume Indicator Gradient Field. In A. Tantar et al. (Eds.), *EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation III*, Volume 500 of *Studies in Computational Intelligence*, pp. 169–193. Springer.

Emmerich, M., A. Deutz, and N. Beume (2007). Gradient-Based/Evolutionary Relay Hybrid for Computing Pareto Front Approximations Maximizing the $\mathcal{S}$-metric. In *Proceedings of the 4th International Conference on Hybrid Metaheuristics*, HM'07, pp. 140–156. Springer.

Emmerich, M., O. M. Shir, and H. Wang (2018). *Evolution Strategies*, pp. 1–31. Cham: Springer International Publishing.

Emmerich, M., K. Yang, A. Deutz, H. Wang, and C. M. Fonseca (2016). *A Multicriteria Generalization of Bayesian Global Optimization*, pp. 229–242. Cham: Springer International Publishing.

Fletcher, R. (2013). *Practical Methods of Optimization.* John Wiley & Sons.

Fliege, J. and B. F. Svaiter (2000). Steepest Descent Methods for Multicriteria Optimization. *Mathematical Methods of Operations Research 51*(3), 479–494.

Forrester, A., A. Keane, et al. (2008). *Engineering Design Via Surrogate Modelling: A Practical Guide.* John Wiley & Sons.

Fortin, F., F. Michel, M. A. Gardner, M. Parizeau, and C. Gagné (2012, jul). DEAP: Evolutionary Algorithms Made Easy. *Journal of Machine Learning Research 13*, 2171–2175.

Gabillon, V., M. Ghavamzadeh, and A. Lazaric (2012). Best Arm Identification: A Unified Approach to Fixed Budget and Fixed Confidence. In *Advances in Neural Information Processing Systems*, pp. 3212–3220.

Ginsbourger, D., R. Le Riche, and L. Carraro (2010). *Kriging Is Well-Suited to Parallelize Optimization*, pp. 131–162. Berlin, Heidelberg: Springer Berlin Heidelberg.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning* (1st ed.). Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Goldberg, D. E. and J. Richardson (1987). Genetic Algorithms with Sharing for Multimodal Function Optimization. In *Genetic algorithms and their applications: Proceedings of the Second International Conference on Genetic Algorithms*, pp. 41–49. Hillsdale, NJ: Lawrence Erlbaum.

Hansen, N. (2006). *The CMA Evolution Strategy: A Comparing Review*, pp. 75–102. Berlin, Heidelberg: Springer Berlin Heidelberg.

Hansen, N. (2008, September). Adaptive Encoding: How to Render Search Coordinate System Invariant. In *Parallel Problem Solving from Nature - PPSN X*, Dortmund, Germany, pp. 205–214.

Hansen, N., A. Auger, S. Finck, and R. Ros (2010, March). Real-Parameter Black-Box Optimization Benchmarking 2010: Experimental Setup. Research Report RR-7215, INRIA.

Hansen, N., S. Finck, R. Ros, and A. Auger (2009). Real-Parameter Black-Box Optimization Benchmarking 2009: Noiseless Functions Definitions. Research Report RR-6829, INRIA.

Hansen, N., S. D. Müller, and P. Koumoutsakos (2003, March). Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). *Evol. Comput. 11*(1), 1–18.

Hansen, N. and A. Ostermeier (2001, June). Completely Derandomized Self-Adaptation in Evolution Strategies. *Evol. Comput. 9*(2), 159–195.

Hartman, L. and O. Hössjer (2008). Fast Kriging of Large Data Sets with Gaussian Markov Random Fields. *Computational Statistics & Data Analysis 52*(5), 2331–2349.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning* (2 ed.). Springer Series in Statistics. New York, NY, USA: Springer-Verlag New York.

Hillermeier, C. (2001). Generalized Homotopy Approach to Multiobjective Optimization. *Journal of Optimization Theory and Applications 110*(3), 557–583.

Hoffman, M., E. Brochu, and N. D. Freitas (2011). Portfolio Allocation for Bayesian Optimization. *Conference on Uncertainty in Artificial Intelligence*, 327–336.

Hoffman, M. W., B. Shahriari, and N. de Freitas (2014). On Correlation and Budget Constraints in Model-Based Bandit Optimization with Application to Automatic Machine Learning. *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics 33*, 365–374.

Hutter, F., H. H. Hoos, and K. Leyton-Brown (2011). *Sequential Model-Based Optimization for General Algorithm Configuration*, pp. 507–523. Berlin, Heidelberg: Springer Berlin Heidelberg.

Hutter, F., H. H. Hoos, and K. Leyton-Brown (2012). Parallel Algorithm Configuration. *LION 6*, 55–70.

Hutter, F., H. H. Hoos, K. Leyton-Brown, and K. Murphy (2010). Time-Bounded Sequential Parameter Optimization. In *International Conference on Learning and Intelligent Optimization*, pp. 281–298. Springer.

Jin, Y. (2011). Surrogate-assisted Evolutionary Computation: Recent Advances and Future Challenges. *Swarm and Evolutionary Computation 1*(2), 61–70.

Jones, D. R. (2001). A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of global optimization 21*(4), 345–383.

Jones, D. R., M. Schonlau, and W. J. Welch (1998). Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global optimization 13*(4), 455–492.

Journel, A. G. and M. Rossi (1989). When do we need a trend model in Kriging? *Mathematical Geology 21*(7), 715–739.

Kaya, H., P. Tüfekci, and S. F. Gürgen (2012). Local and Global Learning Methods for Predicting Power of a Combined Gas & Steam Turbine. *International Conference on Emerging Trends in Computer and Electronics Engineering (ICETCEE 2012)*, 13–18.

Kennedy, J. and R. Eberhart (1995, Nov). Particle Swarm Optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, Volume 4, pp. 1942–1948 vol.4.

Kerschke, P., M. Preuss, S. Wessing, and H. Trautmann (2016). Low-Budget Exploratory Landscape Analysis on Multiple Peaks Models. In *Proceedings of the 18th Annual Conference on Genetic and Evolutionary Computation*, GECCO '16. ACM.

Kerschke, P., H. Wang, M. Preuss, C. Grimme, A. H. Deutz, H. Trautmann, and M. Emmerich (2016). Towards Analyzing Multimodality of Continuous Multiobjective Landscapes. In J. Handl, E. Hart, P. R. Lewis, M. López-Ibáñez, G. Ochoa, and B. Paechter (Eds.), *Proceedings of the 14th International Conference on Parallel Problem Solving from Nature (PPSN XIV)*, Volume 9921 of *Lecture Notes in Computer Science*, pp. 962–972. Springer.

Kimura, S. and K. Matsumura (2005). Genetic Algorithms Using Low-discrepancy Sequences. In *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*, GECCO '05, New York, NY, USA, pp. 1341–1346. ACM.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi (1983). Optimization by Simulated Annealing. *science 220*(4598), 671–680.

Kleijnen, J. P. C., W. van Beers, and I. van Nieuwenhuyse (2012). Expected Improvement in Efficient Global Optimization through Bootstrapped Kriging. *Journal of Global Optimization 54*(1), 59–73.

Kollo, T. o. and D. von Rosen (2005). *Advanced Multivariate Statistics with Matrices* (1 ed.), Volume 579 of *Mathematics and Its Applications*. Springer Netherlands.

Krige, D. G. (1951, December). A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa 52*(6), 119–139.

Lawrence, N. D. (2004). Gaussian Process Latent Variable Models for Visualisation of High Dimensional Data. *Advances in neural information processing systems 16*(3), 329–336.

López, A. L., C. A. C. Coello, and O. Schütze (2012, May). *Using Gradient Based Information to Build Hybrid Multi-objective Evolutionary Algorithms*. Ph. D. thesis, CINVESTAV-IPN, Mexico City.

Lophaven, S. N., H. B. Nielsen, and J. Søndergaard (2002). *DACE: a Matlab Kriging Toolbox*, Volume 2. Citeseer.

Loshchilov, I., M. Schoenauer, and M. Sebag (2011). Adaptive Coordinate Descent. In *Proceedings of the 13th annual conference on Genetic and evolutionary computation*, pp. 885–892. ACM.

Luke, S. (2009). *Essentials of Metaheuristics*, Volume 113. Lulu Raleigh.

MacQueen, J. et al. (1967). Some Methods for Classification and Analysis of Multivariate Observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Volume 1, pp. 281–297. Oakland, CA, USA.

Martinez-Cantin, R. (2014). Bayesopt: A Bayesian Optimization Library for Nonlinear Optimization, Experimental Design and Bandits. *The Journal of Machine Learning Research 15*(1), 3735–3739.

Miller, B. L. and M. J. Shaw (1996). Genetic Algorithms with Dynamic Niche Sharing for Multimodal Function Optimization. In *Evolutionary Computation, 1996., Proceedings of IEEE International Conference on*, pp. 786–791. IEEE.

Močkus, J. (1975). On Bayesian Methods for Seeking the Extremum. In *Optimization Techniques IFIP Technical Conference*, pp. 400–404. Springer.

Močkus, J. (2012). *Bayesian Approach to Global Optimization: Theory and Applications*, Volume 37. Springer Science & Business Media.

Moscato, P. et al. (1989). On Evolution, Search, Optimization, Genetic Algorithms and Martial Arts: Towards Memetic Algorithms. *Caltech concurrent computation program, C3P Report 826*, 1989.

Naish-Guzman, A. and S. Holden (2007). The Generalized FITC Approximation. In *Advances in Neural Information Processing Systems*, pp. 1057–1064.

Nguyen-Tuong, D., M. Seeger, and J. Peters (2009). Model Learning with Local Gaussian Process Regression. *Advanced Robotics 23*(15), 2015–2034.

Niederreiter, H. (1992). *Random Number Generation and quasi-Monte Carlo Methods.* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.

Nocedal, J. and S. Wright (2000). *Numerical Optimization* (2 ed.). Springer Series in Operations Research and Financial Engineering. Springer-Verlag New York.

Nourani, Y. and B. Andresen (1998). A Comparison of Simulated Annealing Cooling Strategies. *Journal of Physics A: Mathematical and General 31*(41), 8373.

O'Hagan, A. and J. Kingman (1978). Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–42.

Øksendal, B. (2003). *Stochastic Differential Equations.* Universitext. Springer-Verlag Berlin Heidelberg.

Omre, H. (1987). Bayesian Kriging — Merging Observations and Qualified Guesses in Kriging. *Mathematical Geology 19*(1), 25–39.

Ostermeier, A., A. Gawelczyk, and N. Hansen (1994). Step-Size Adaption Based on Non-Local Use of Selection Information. In *Proceedings of the International Conference on Evolutionary Computation. The Third Conference on Parallel*

Problem Solving from Nature: Parallel Problem Solving from Nature, PPSN III, London, UK, UK, pp. 189–198. Springer-Verlag.

Ponweiser, W., T. Wagner, D. Biermann, and M. Vincze (2008). Multiobjective Optimization on a Limited Budget of Evaluations Using Model-Assisted $\mathcal{S}$-Metric Selection. In G. Rudolph, T. Jansen, N. Beume, S. Lucas, and C. Poloni (Eds.), *Parallel Problem Solving from Nature – PPSN X*, Berlin, Heidelberg, pp. 784–794. Springer Berlin Heidelberg.

Ponweiser, W., T. Wagner, and M. Vincze (2008, June). Clustered Multiple Generalized Expected Improvement: A Novel Infill Sampling Criterion for Surrogate Models. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*, pp. 3515–3522.

Quiñonero-Candela, J. and C. E. Rasmussen (2005). A Unifying View of Sparse Approximate Gaussian Process Regression. *The Journal of Machine Learning Research 6*(1), 1939–1959.

Rao, C. R., H. Toutenburg, Shalabh, and C. Heumann. *Linear Models and Generalizations* (3rd ed.). Springer Series in Statistics. Springer-Verlag Berlin Heidelberg.

Rasmussen, C. and C. Williams (2006). *Gaussian Processes for Machine Learning*. Adaptative computation and machine learning series. University Press Group Limited.

Ren, Y., A. Deutz, and M. Emmerich (2015). On Steering Dominated Points in Hypervolume Gradient Ascent for Bicriteria Continuous Optimization (extended abstract). In *Numerical and Evolutionary Optimization, NEO 2015, Tijuana, Mexico (Book of abstracts)*.

Reynolds, D. (2009). Gaussian Mixture Models. In *Encyclopedia of Biometrics*, pp. 659–663. Springer.

Rosenbrock, H. H. (1960, January). An Automatic Method for Finding the Greatest or Least Value of a Function. *The Computer Journal 3*(3), 175–184.

Roustant, O., D. Ginsbourger, and Y. Deville (2012, 10). DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. *Journal of Statistical Software 51*(1), 1–55.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989). Design and Analysis of Computer Experiments. *Statistical Science 4*(4), 409–423.

Santner, T., B. Williams, and W. Notz (2003). *The Design and Analysis of Computer Experiments.* Springer Series in Statistics. Springer.

Sasena, M. J., P. Papalambros, and P. Goovaerts (2002). Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization. *Engineering optimization 34*(3), 263–278.

Schölkopf, B., R. Herbrich, and A. J. Smola (2001). A Generalized Representer Theorem. In *International conference on computational learning theory*, pp. 416–426. Springer.

Schonlau, M. (1998). *Computer Experiments and Global Optimization.* University of Waterloo.

Schonlau, M., W. J. Welch, and D. R. Jones (1998). Global versus Local Search in Constrained Optimization of Computer Models. *Lecture Notes-Monograph Series*, 11–25.

Schütze, O., C. Domínguez-Medina, N. Cruz-Cortés, L. Gerardo de la Fraga, J.-Q. Sun, G. Toscano, and R. Landa (2016). A Scalar Optimization Approach for Averaged Hausdorff Approximations of the Pareto front. *Engineering Optimization 48*(9), 1593–1617.

Schütze, O., X. Esquivel, A. Lara, and C. A. C. Coello (2012, Aug). Using the Averaged Hausdorff Distance as a Performance Measure in Evolutionary Multiobjective Optimization. *IEEE Transactions on Evolutionary Computation 16*(4), 504–522.

Schütze, O., A. Lara, and C. Coello Coello (2011). The Directed Search Method for Unconstrained Multi-objective Optimization Problems. *Proceedings of the EVOLVE–A Bridge Between Probability, Set Oriented Numerics, and Evolutionary Computation.*

Schwefel, H.-P. (1993). *Evolution and Optimum Seeking: The Sixth Generation.* New York, NY, USA: John Wiley & Sons, Inc.

Shir, O. M. et al. (2008). *Niching in Derandomized Evolution Strategies and its Applications in Quantum Control.* Natural Computing Group, LIACS, Faculty of Science, Leiden University.

Shir, O. M. and T. Bäck (2005a). Dynamic Niching in Evolution Strategies with Covariance Matrix Adaptation. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, Volume 3, pp. 2584–2591. IEEE.

Shir, O. M. and T. Bäck (2005b). Niching in Evolution Strategies. In *Proceedings of the 7th annual conference on Genetic and evolutionary computation*, pp. 915–916. ACM.

Shir, O. M., M. Emmerich, and T. Bäck (2007). Self-Adaptive Niching CMA-ES with Mahalanobis Metric. In *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, pp. 820–827. IEEE.

Silver, D., A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al. (2016). Mastering the Game of Go with Deep Neural Networks and Tree Search. *nature 529*(7587), 484–489.

Silverman, B. W. (1985). Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting. *Journal of the Royal Statistical Society. Series B (Methodological) 47*(1), 1–52.

Slater, M. (2014). Lagrange Multipliers Revisited. In *Traces and Emergence of Nonlinear Programming*, pp. 293–306. Springer.

Snelson, E. and Z. Ghahramani (2005). Sparse Gaussian Processes using Pseudo-inputs. In *Advances in neural information processing systems*, pp. 1257–1264.

Snoek, J., H. Larochelle, and R. P. Adams (2012). Practical Bayesian Optimization of Machine Learning Algorithms. In *Advances in neural information processing systems*, pp. 2951–2959.

Sóbester, A., S. J. Leary, and A. J. Keane (2005). On the Design of Optimization Strategies Based on Global Response Surface Approximation Models. *Journal of Global Optimization 33*(1), 31–59.

Sosa Hernández, V. A., O. Schütze, and M. Emmerich (2014). *EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V*, Chapter Hypervolume Maximization via Set Based Newton's Method, pp. 15–28. Cham, Switzerland: Springer International Publishing.

Srinivas, N. and K. Deb (1994). Muiltiobjective Optimization using Nondominated Sorting in Genetic Algorithms. *Evolutionary computation 2*(3), 221–248.

Srinivas, N., A. Krause, S. Kakade, and M. Seeger (2010). Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, USA, pp. 1015–1022. Omnipress.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging* (1 ed.). Springer Series in Statistics. Springer-Verlag New York.

Storn, R. and K. Price (1997, Dec). Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of global optimization 11*(4), 341–359.

Sundberg, R. (1974). Maximum Likelihood Theory for Incomplete Data from an Exponential Family. *Scandinavian Journal of Statistics*, 49–58.

Talbi, E.-G. (2009). *Metaheuristics: from Design to Implementation*, Volume 74. John Wiley & Sons.

Teytaud, O. and S. Gelly (2007). DCMA: yet another derandomization in Covariance-Matrix-Adaptation. In *Proceedings of the 9th annual conference on Genetic and evolutionary computation*, pp. 955–963. ACM.

Tong, Y. L. (2012). *The Multivariate Normal Distribution.* Springer Science & Business Media.

Torn, A. and A. Zilinskas (1989). Global Optimization. *Lecture Notes in Computer Science 350.*

Tresp, V. (2000). A Bayesian Committee Machine. *Neural computation 12*(11), 2719–2741.

Ursem, R. K. (2014). From Expected Improvement to Investment Portfolio Improvement: Spreading the Risk in Kriging-Based Optimization. In *International Conference on Parallel Problem Solving from Nature*, pp. 362–372. Springer.

van der Vaart, A. W. and J. H. van Zanten (2008). Rates of Contraction of Posterior Distributions Based on Gaussian Process Priors. *The Annals of Statistics*, 1435–1463.

van Stein, B., H. Wang, W. Kowalczyk, T. Bäck, and M. Emmerich (2015). Optimally Weighted Cluster Kriging for Big Data Regression. In E. Fromont, T. De Bie, and M. van Leeuwen (Eds.), *Advances in Intelligent Data Analysis XIV*, pp. 310–321. Cham: Springer International Publishing.

van Stein, B., H. Wang, W. Kowalczyk, M. Emmerich, and T. Bäck (2016). Fuzzy clustering for Optimally Weighted Cluster Kriging. In *Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on*, pp. 939–945. IEEE.

Vijayakumar, S., A. D'souza, and S. Schaal (2005). Incremental Online Learning in High Dimensions. *Neural computation 17*(12), 2602–2634.

Wang, H., T. Bäck, and M. T. M. Emmerich (2018). Multi-point Efficient Global Optimization Using Niching Evolution Strategy. In A.-A. Tantar, E. Tantar, M. Emmerich, P. Legrand, L. Alboaie, and H. Luchian (Eds.), *EVOLVE - A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation VI*, Cham, pp. 146–162. Springer International Publishing.

Wang, H., A. Deutz, T. Bäck, and M. Emmerich (2017). Hypervolume Indicator Gradient Ascent Multi-objective Optimization. In H. Trautmann, G. Rudolph, K. Klamroth, O. Schütze, M. Wiecek, Y. Jin, and C. Grimme (Eds.), *Evolutionary Multi-Criterion Optimization*, Cham, pp. 654–669. Springer International Publishing.

Wang, H., M. Emmerich, and T. Bäck (2014). Mirrored Orthogonal Sampling with Pairwise Selection in Evolution Strategies. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, SAC '14, New York, NY, USA, pp. 154–156. ACM.

Wang, H., M. Emmerich, and T. Bäck (2016). Balancing Risk and Expected Gain in Kriging-Based Global Optimization. In *Evolutionary Computation (CEC), 2016 IEEE Congress on*, pp. 719–727. IEEE.

Wang, H., M. Emmerich, and T. Bäck (2018). Cooling Strategies for the Moment-Generating Function in Bayesian Global Optimization. In *Evolutionary Computation (CEC), 2018 IEEE Congress on*, pp. to appear. IEEE.

Wang, H., Y. Ren, A. Deutz, and M. Emmerich (2017). *On Steering Dominated Points in Hypervolume Indicator Gradient Ascent for Bi-Objective Optimization*, pp. 175–203. Cham: Springer International Publishing.

Wang, H., B. van Stein, M. Emmerich, and T. Bäck (2017). A New Acquisition Function for Bayesian Optimization Based on the Moment-Generating Function. In *Systems, Man, and Cybernetics (SMC), 2017 IEEE International Conference on*, pp. 507–512. IEEE.

Wang, H., B. van Stein, M. Emmerich, and T. Bäck (2017). Time Complexity Reduction in Efficient Global Optimization using Cluster Kriging. In *Proceedings of the Genetic and Evolutionary Computation Conference*, GECCO '17, New York, NY, USA, pp. 889–896. ACM.

Webster, R. and M. A. Oliver (2007). *Geostatistics for Environmental Scientists.* John Wiley & Sons.

Wessing, S. (2015). *Two-Stage Methods for Multimodal Optimization.* Ph. D. thesis, Technische Universität Dortmund.

Wessing, S. (2016). *optproblems: Infrastructure to Define Optimization Problems and Some Test Problems for Black-Box Optimization.* python package version 0.9.

Wright, S. and J. Nocedal (1999). Numerical Optimization. *Springer Science 35*(67-68), 7.

Yeh, I.-C. (1998). Modeling of Strength of High-performance Concrete using Artificial Neural Networks. *Cement and Concrete research 28*(12), 1797–1808.

Zhang, Q. and H. Li (2007). MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition. *IEEE Transactions on evolutionary computation 11*(6), 712–731.

Žilinskas, A. (1992). A Review of Statistical Models for Global Optimization. *Journal of Global Optimization 2*(2), 145–153.

Zimmerman, D., C. Pavlik, A. Ruggles, and M. P. Armstrong (1999). An Experimental Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting. *Mathematical Geology 31*(4), 375–390.

Zitzler, E., K. Deb, and L. Thiele (2000, June). Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation* (2), 173 – 195.

Zitzler, E., M. Laumanns, L. Thiele, et al. (2001). SPEA2: Improving the Strength Pareto Evolutionary Algorithm. In *Eurogen*, Volume 3242, pp. 95–100.

Zitzler, E. and L. Thiele (1998). Multiobjective Optimization Using Evolutionary Algorithms — A Comparative Case Study. In A. E. Eiben, T. Bäck, M. Schoenauer, and H.-P. Schwefel (Eds.), *Parallel Problem Solving from Nature — PPSN V*, Berlin, Heidelberg, pp. 292–301. Springer Berlin Heidelberg.

Zitzler, E., L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca (2003, April). Performance Assessment of Multiobjective Optimizers: An Analysis and Review. *IEEE Transactions on Evolutionary Computation 7*(2), 117–132.

Acronyms

BLP Best Linear Estimator. 50

BLUE Best Linear Unbiased Estimator. 48

BLUP Best Linear Unbiased Predictor. 48

ECDF Empirical Cumulative Distribution Function. 36, 101

EGO Efficient Global Optimization. 37, 90, 113, 114

EI Expected Improvement. 78, 90–93, 96, 103, 108, 114, 154

GEI Generalized Expected Improvement. 91, 92, 98, 100

GLS Generalized Least Squares. 48

GPR Gaussian Process Regression. 39, 43, 44, 56, 59, 63, 69, 87, 88, 95, 100, 152

KKT Karush-Kuhn-Tucker conditions. 57, 95, 125

LHS Latin Hypercube Sampling. 57, 101

LUP Linear Unbiased Predictor. 47

MAP Maximum a Posterior. 59

MGF Moment-Generating Function. 96

MGFI Moment-Generating Function of Improvement. 98, 100, 101, 103, 107, 154, 155

MSE Mean Squared Error. 37, 39, 49, 50, 69, 71, 73, 80, 82, 94, 108

Acronyms

RBF Radial Basis Functions. 46

RKHS Reproducing Kernel Hilbert Space. 54, 56, 57, 155

Index