Stochastic and deterministic algorithms for continuous black-box optimization

Wang, H.

Stochastic and Deterministic Algorithms for Continuous Black-Box Optimization

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 1 november 2018
klokke 10.00 uur

door

Hao Wang

geboren te Baoji, China
in 1989

Promotiecommissie

Promotor:	Prof. Dr. T.H.W. Bäck	
Co-promotor:	Dr. M.T.M. Emmerich	
Overige leden:	Prof. Dr. A. Plaat	(voorzitter)
	Prof. Dr. H. Trautmann	(WWU Münster and LIACS)
	Dr. C. Dörr	(CNRS and Sorbonne University, FR)
	Prof. Dr. S. Manegold	(secretaris, CWI and LIACS, NL)
	Dr. W.A. Kosters	
	Prof. X. Liu	(Brunel University, UK)

This research is financially supported by the Dutch funding agency NWO, under project number 650.002.001 (the PROMIMOOC project), in collaboration with Tata Steel IJmuiden, BMW Group Regensburg, Centrum voor Wiskunde en Informatica (CWI) and MonetDB.

Figures and diagrams are generated using GGPLOT2, PGF/TIKZ and MATPLOTLIB.

Abstract

Continuous optimization is never easy: the exact solution is always a luxury demand and the theory of it is not always analytical and elegant. Continuous optimization, in practice, is essentially about the efficiency: how to obtain the solution with same quality using as minimal resources (e.g., CPU time or memory usage) as possible? In this thesis, the number of function evaluations is considered as the most important resource to save. To achieve this goal, various efforts have been implemented and applied successfully. One research stream focuses on the so-called stochastic variation (mutation) operator, which conducts an (local) exploration of the search space. The efficiency of those operator has been investigated closely, which shows a good stochastic variation should be able to generate a good coverage of the local neighbourhood around the current search solution. The first part (Chapter 2) of this thesis contributes on this issue by formulating a novel stochastic variation that yields good space coverage.

Alternative research stream approaches the efficiency issue differently: we should keep record of the evaluated solutions and re-use them as they carry partially information about the objective function. This leads to studies on the so-called surrogate modeling. Here, the second part (Chapter 3) of this thesis dives into the one specific surrogate modeling technique, called Kriging/Gaussian Process Regression (GPR). In addition, we try to keep a precise and theoretical treatment on Kriging/GPR and several improvements over it. Closely related to the surrogate modeling, it is crucial to exploit the surrogate model properly. A common approach is to take the so-called infill criteria/acquisition function, which measures the potential gain we could obtain by evaluate one candidate solution on the model. Lastly, the efficiency issues can also be tackled by generalizing a well-performing algorithm from a specific domain to a broader class of problems. One prominent example is to extend the gradient-based optimization algorithms (that are devised

for single objective problems) to the multi-objective scenario. The last part of this thesis (Chapter 5) is on this topic, where both the first- and second-order methods are generalized for multi-objective optimization problems.

Contents

Abstract **i**

List of Symbols

1 Introduction **1**

1.1 Stochastic Optimization 2
1.2 Multi-objective Optimization 6
1.3 Matrix Calculus 8
1.4 Outline of the Dissertation 10

2 Stochastic Variation **15**

2.1 Quasi-Random Sampling 16
2.2 Mirroring and Orthogonalization 18
2.2.1 Deterministic Orthogonal Sampling 19
2.2.2 Mirrored Orthogonal Sampling 21
2.2.3 Implementation of Random Orthogonal Sampling 24
2.3 Convergence Analysis of Mirroring and Orthogonalization 26
2.3.1 Mirrored Sampling 27
2.3.2 Mirrored Orthogonal Sampling 31
2.4 Empirical Results on Mirroring and Orthogonalization 34
2.4.1 Experiments on BBOB 35
2.5 Efficient Global Optimization 37
2.6 Summary 41

3 Kriging/Gaussian Process Regression **43**

3.1 General Discussion 44
3.1.1 Best Linear Unbiased Predictor 47
3.1.2 Reproducing Kernel Hilbert Space 53

3.1.3 Bayesian Inference 58
3.1.4 Differentiation 61
3.2 Cluster Kriging 63
3.2.1 Clustering 66
3.2.2 Modeling 69
3.2.3 Cluster Kriging Predictor 69
3.2.4 Experiments 74
3.3 Cluster Kriging and EGO 77
3.3.1 The algorithm 78
3.3.2 Experiments 82
3.4 Summary 83

4 Infill Criteria 87
4.1 Improvement-based Infill Criteria 89
4.2 Balancing Risk and Gain 92
4.3 Moment-Generating Function of Improvement 96
4.4 Cooling Strategies for MGFI 100
4.4.1 Impact of Temperature Configurations 101
4.4.2 Benchmarking the Cooling Strategies 103
4.5 Parallelization 105
4.5.1 Multi-point Infill Criteria 106
4.5.2 Multi-instance of Infill Criteria 106
4.5.3 Multi-objective Infill Criteria 107
4.5.4 Niching-based Infill Criteria Maximization 107
4.6 Experimental Comparison 113
4.7 Summary 117

5 Numerical Multi-objective Optimization 119
5.1 Mixed-Peak Test Problem 122
5.1.1 Mixed-Peak Functions 122
5.1.2 Mixed-Peak Bi-objective Problem 124
5.2 Hypervolume Indicator Gradient 127
5.2.1 Steering Dominated Points 129
5.2.2 Step-size adaptation 132
5.2.3 Hypervolume Indicator Gradient Ascent Algorithm 134
5.2.4 Experiments 136
5.3 Hypervolume Indicator Hessian 139

5.3.1 The Bi-objective Case 142
5.3.2 Hypervolume Indicator Newton Method 145
5.4 Summary 147

6 Conclusion **151**

Appendix A Gaussian Distribution **157**

Appendix B Proof **159**
B.1 Theorem 5.3 159

Bibliography 163

Index 181

Summary 185

Samenvatting 189

About the Author 193

List of Symbols

$\mathbf{1}$	Vector of ones, whose dimension is implied in context
$\mathbf{0}$	Vector of zeros, whose dimension is implied in context
$\mathbf{I}$	Identity matrix, whose shape is implied in context
$\mathbb{N}_{>0}$	Positive natural numbers
$\mathbb{R}^d$	d-dimensional Euclidean space
S	Search space/decision space/domain of the objective function
$\mathcal{X}$	Pareto efficient set
$P_{\mathcal{X}}$	Pareto front
$\mathcal{H}$	Hilbert space of functions $f \colon \mathrm{S} \to \mathbb{R}$
$L^2(\mathrm{S})$	Space of square-integrable functions on S
$\lVert\cdot\rVert$	Euclidean norm
$\lVert\cdot\rVert_{\boldsymbol{\Sigma}}$	Mahalanobis norm with respect to a covariance matrix $\boldsymbol{\Sigma}$
$\lVert\cdot\rVert_{\mathcal{H}}$	Norm in Hilbert space $\mathcal{H}$
$\lVert\cdot\rVert_{\infty}$	supremum norm
$\langle\cdot,\cdot\rangle$	Dot product in Euclidean spaces.
$\langle\cdot,\cdot\rangle_{\mathcal{H}}$	Inner product in Hilbert space $\mathcal{H}$
$\mathcal{U}(0,1)$	Uniform distribution over $[0,1]$
$\mathcal{N}(m,\sigma^2)$	Gaussian random variable with mean m and variance σ^2
$\boldsymbol{\mathcal{N}}(\mathbf{m},\mathbf{K})$	Gaussian random vector with mean $\mathbf{m}$ and covariance matrix $\mathbf{K}$

$\mathcal{GP}(m(\cdot), k(\cdot,\cdot))$	Gaussian process with mean function $m(\cdot)$ and kernel $k(\cdot,\cdot)$
$\Pr(\cdot)$	Probability of an event
$p(\cdot)$	Probability density function (p.d.f.)
$P(\cdot)$	Cumulative distribution function (c.d.f.)
$\mathbb{P}$	Probability measure
ϕ	Probability density function of $\mathcal{N}(0,1)$
ϕ_{m,σ^2}	Probability density function of $\mathcal{N}(m,\sigma^2)$
Φ	Cumulative distribution function of $\mathcal{N}(0,1)$
Φ_{m,σ^2}	Cumulative distribution function of $\mathcal{N}(m,\sigma^2)$
$\Phi^n_{\mathbf{m},\mathbf{K}}$	Cumulative distribution function of a multivariate Gaussian $\boldsymbol{\mathcal{N}}(\mathbf{m},\mathbf{K})$
$\mathbb{E}$	Expectation
Var	Variance
Cov	Covariance
$\perp\!\!\!\perp$	Statistical independence
σ_n^2	Variance of the white noise process
$\mathscr{A}$	Infill criteria/acquisition function
det	Determinant of square matrices
κ	Condition number of matrices