

CONTENTS

PREFACE	ix
NOTATION	xiii
1 INTRODUCTION	1
1.1 Formalization of Optimization	2
1.2 The Bayesian Approach	5
2 GAUSSIAN PROCESSES	15
2.1 Definition and Basic Properties	16
2.2 Inference with Exact and Noisy Observations	18
2.3 Overview of Remainder of Chapter	26
2.4 Joint Gaussian Processes	26
2.5 Continuity	28
2.6 Differentiability	30
2.7 Existence and Uniqueness of Global Maxima	33
2.8 Inference with Non-Gaussian Observations and Constraints	35
2.9 Summary of Major Ideas	41
3 MODELING WITH GAUSSIAN PROCESSES	45
3.1 The Prior Mean Function	46
3.2 The Prior Covariance Function	49
3.3 Notable Covariance Functions	51
3.4 Modifying and Combining Covariance Functions	54
3.5 Modeling Functions on High-Dimensional Domains	61
3.6 Summary of Major Ideas	64
4 MODEL ASSESSMENT, SELECTION, AND AVERAGING	67
4.1 Models and Model Structures	68
4.2 Bayesian Inference over Parametric Model Spaces	70
4.3 Model Selection via Posterior Maximization	73
4.4 Model Averaging	74
4.5 Multiple Model Structures	78
4.6 Automating Model Structure Search	81
4.7 Summary of Major Ideas	84
5 DECISION THEORY FOR OPTIMIZATION	87
5.1 Introduction to Bayesian Decision Theory	89
5.2 Sequential Decisions with a Fixed Budget	91
5.3 Cost and Approximation of the Optimal Policy	99
5.4 Cost-Aware Optimization and Termination as a Decision	103
5.5 Summary of Major Ideas	106
6 UTILITY FUNCTIONS FOR OPTIMIZATION	109
6.1 Expected Utility of Terminal Recommendation	109
6.2 Cumulative Reward	114

6.3	Information Gain	115
6.4	Dependence on Model of Objective Function	116
6.5	Comparison of Utility Functions	117
6.6	Summary of Major Ideas	119
7	COMMON BAYESIAN OPTIMIZATION POLICIES	123
7.1	Example Optimization Scenario	124
7.2	Decision-Theoretic Policies	124
7.3	Expected Improvement	127
7.4	Knowledge Gradient	129
7.5	Probability of Improvement	131
7.6	Mutual Information and Entropy Search	135
7.7	Multi-Armed Bandits and Optimization	141
7.8	Maximizing a Statistical Upper Bound	145
7.9	Thompson Sampling	148
7.10	Other Ideas in Policy Construction	150
7.11	Summary of Major Ideas	156
8	COMPUTING POLICIES WITH GAUSSIAN PROCESSES	157
8.1	Notation for Objective Function Model	157
8.2	Expected Improvement	158
8.3	Probability of Improvement	167
8.4	Upper Confidence Bound	170
8.5	Approximate Computation for One-Step Lookahead	171
8.6	Knowledge Gradient	172
8.7	Thompson Sampling	176
8.8	Mutual Information with x^*	180
8.9	Mutual Information with f^*	187
8.10	Averaging over a Space of Gaussian Processes	192
8.11	Alternative Models: Bayesian Neural Networks, etc.	196
8.12	Summary of Major Ideas	200
9	IMPLEMENTATION	201
9.1	Gaussian Process Inference, Scaling, and Approximation	201
9.2	Optimizing Acquisition Functions	207
9.3	Starting and Stopping Optimization	210
9.4	Summary of Major Ideas	212
10	THEORETICAL ANALYSIS	213
10.1	Regret	213
10.2	Useful Function Spaces for Studying Convergence	215
10.3	Relevant Properties of Covariance Functions	220
10.4	Bayesian Regret with Observation Noise	224
10.5	Worst-Case Regret with Observation Noise	232
10.6	The Exact Observation Case	237
10.7	The Effect of Unknown Hyperparameters	241
10.8	Summary of Major Ideas	243

11	EXTENSIONS AND RELATED SETTINGS	245
11.1	Unknown Observation Costs	245
11.2	Constrained Optimization and Unknown Constraints	249
11.3	Synchronous Batch Observations	252
11.4	Asynchronous Observation with Pending Experiments	262
11.5	Multifidelity Optimization	263
11.6	Multitask Optimization	266
11.7	Multiobjective Optimization	269
11.8	Gradient Observations	276
11.9	Stochastic and Robust Optimization	277
11.10	Incremental Optimization of Sequential Procedures	281
11.11	Non-Gaussian Observation Models and Active Search	282
11.12	Local Optimization	285
12	A BRIEF HISTORY OF BAYESIAN OPTIMIZATION	287
12.1	Historical Precursors and Optimal Design	287
12.2	Sequential Analysis and Bayesian Experimental Design	287
12.3	The Rise of Bayesian Optimization	289
12.4	Later Rediscovery and Development	290
12.5	Multi-Armed Bandits to Infinite-Armed Bandits	292
12.6	What's Next?	294
A	THE GAUSSIAN DISTRIBUTION	295
B	METHODS FOR APPROXIMATE BAYESIAN INFERENCE	301
C	GRADIENTS	307
D	ANNOTATED BIBLIOGRAPHY OF APPLICATIONS	313
	REFERENCES	331
	INDEX	353