

TRUST-REGION METHODS

Andrew R. Conn
IBM-Thomas J. Watson Research Center
Yorktown Heights, New York, United States

Nicholas I. M. Gould
Rutherford Appleton Laboratory
Chilton, Oxfordshire, England

Philippe L. Toint
University of Namur
Namur, Belgium



Society for Industrial and Applied Mathematics
Philadelphia



Mathematical Programming Society
Philadelphia

Contents

Preface	xv
1 Introduction	1
1.1 What Is a Trust-Region Algorithm?	1
1.2 Historical Development of Trust-Region Methods	8
1.3 Some Applications of Trust-Region Algorithms	9
I Preliminaries	13
2 Basic Concepts	15
2.1 Basic Notation	15
2.2 Eigenvalues and Eigenvectors	16
2.3 Vector and Matrix Norms	20
2.3.1 Vector and Functional Norms	21
2.3.2 Matrix Norms	22
3 Basic Analysis and Optimality Conditions	25
3.1 Basic Real Analysis	25
3.1.1 Basic Topology	25
3.1.2 Derivatives and Taylor's Theorem	27
3.1.3 Convexity	30
3.1.4 Nonsmooth Functions	33
3.1.5 Geometry	34
3.2 Optimality Conditions	37
3.2.1 Differentiable Unconstrained Problems	38
3.2.2 Differentiable Constrained Problems	39
3.2.3 Convex Programming	44
3.2.4 Nonsmooth Problems	47
3.3 Sequences and Basic Iterations	50
3.3.1 Sequences	50
3.3.2 Newton's Method	51
3.3.3 Forcing Functions	53

4	Basic Linear Algebra	55
4.1	Linear Equations in Optimization	55
4.1.1	Structure and Sparsity	56
4.1.2	Structured Systems	57
4.2	Special Matrices	57
4.2.1	Diagonal Matrices	58
4.2.2	Triangular Matrices	58
4.2.3	Orthonormal Matrices	58
4.2.4	Tridiagonal and Band Matrices	59
4.3	Direct Methods for Solving Linear Systems	60
4.3.1	Stability in the Face of Rounding Errors	60
4.3.2	Sparse Systems	61
4.3.3	Symmetric Positive Definite Matrices	63
4.3.4	Symmetric Indefinite Matrices	65
4.3.5	Frontal and Multifrontal Methods	66
4.3.6	Matrix Modification	67
4.4	Least-Squares Problems and Projections	68
4.4.1	Least-Squares Problems	68
4.4.2	Projections, Range-, and Null-Space Bases	71
5	Krylov Subspace Methods	75
5.1	The Conjugate Gradient Method	75
5.1.1	Minimization in a Subspace	76
5.1.2	Conjugate Directions	77
5.1.3	Generating Conjugate Directions	78
5.1.4	Conjugate Gradients	82
5.1.5	Convergence of the Conjugate Gradient Method	83
5.1.6	Preconditioned Conjugate Gradients	86
5.2	The Lanczos Method	91
5.2.1	Computing an Orthonormal Basis for the Krylov Space	92
5.2.2	Relationship with the Conjugate Direction Method	95
5.2.3	Finding Conjugate Directions from an Orthonormal Krylov Basis	96
5.2.4	Approximation of Critical Points within the Subspace	99
5.2.5	Rayleigh–Ritz Approximations to Eigenpairs	101
5.2.6	Preconditioned Lanczos	102
5.3	Linear Least-Squares Problems	106
5.4	Problems with Constraints	108
5.4.1	Projected Preconditioned Conjugate Gradients	109

II	Trust-Region Methods for Unconstrained Optimization	113
6	Global Convergence of the Basic Algorithm	115
6.1	The Basic Trust-Region Algorithm	115
6.2	Assumptions	121
6.2.1	Assumptions on the Problem	121
6.2.2	Assumptions on the Algorithm	122
6.3	The Cauchy Point and the Model Decrease	123
6.3.1	The Cauchy Arc	123
6.3.2	The Cauchy Point for Quadratic Models	124
6.3.3	The Approximate Cauchy Point	128
6.3.4	The Final Condition on the Model Decrease	130
6.4	Convergence to First-Order Critical Points	133
6.5	Second-Order Convex Models	139
6.6	The Eigenpoint and Second-Order Nonconvex Models	147
6.6.1	Exploitation of Negative Curvature by Minimization	148
6.6.2	Exploitation of Negative Curvature by a Linesearch	150
6.6.3	Convergence Theorems	153
6.7	Trust-Region Scaling	162
6.7.1	Geometry and Scaling	162
6.7.2	Uniformly Equivalent Norms Again	166
7	The Trust-Region Subproblem	169
7.1	The Solution of Trust-Region Subproblems	169
7.2	Characterization of the ℓ_2 -Norm Model Minimizer	171
7.3	Finding the ℓ_2 -Norm Model Minimizer	176
7.3.1	Finding the ℓ_2 -Norm Model Minimizer	176
7.3.2	Finding the Root of $\ s(\lambda)\ _2 - \Delta = 0$	181
7.3.3	Newton's Method and the Secular Equation	182
7.3.4	Safeguarding Newton's Method	185
7.3.5	Updating the Intervals of Uncertainty	189
7.3.6	Finding λ in the Interval of Uncertainty	189
7.3.7	Finding Good Lower Bounds on $-\lambda_1$	190
7.3.8	Initial Values	192
7.3.9	The Complete Algorithm	193
7.3.10	Termination	194
7.3.11	Enhancements	197
7.4	The Scaled ℓ_2 -Norm Problem	200
7.5	Approximating the Model Minimizer	201
7.5.1	The Truncated Conjugate Gradient Method	202
7.5.2	How Good Is the Steihaug–Toint Point?	208
7.5.3	Dogleg and Double-Dogleg Paths	218
7.5.4	The Truncated Lanczos Approach	221

7.5.5	Computing the Eigenpoint	230
7.5.6	Eigenvalue-Based Approaches	231
7.6	Projection-Based Approaches	235
7.7	Norms that Reflect the Underlying Geometry	236
7.7.1	The Ideal Trust Region	236
7.7.2	The Absolute-Value Trust Region	237
7.7.3	Solving Diagonal and Block Diagonal Trust-Region Subproblems	238
7.7.4	Coping with Singularity	239
7.8	The ℓ_∞ -Norm Problem	243
8	Further Convergence Theory Issues	249
8.1	Convergence of General Measures	249
8.1.1	First-Order and Criticality Measures	249
8.1.2	First-Order Convergence Theory	251
8.1.3	Second-Order Convergence Theory	255
8.1.4	Convergence in Dual Norms	259
8.1.5	A More General Cauchy Point	265
8.1.6	The Scaled Cauchy Point	269
8.2	Weakly Equivalent Norms	272
8.3	Minimization in Infinite-Dimensional Spaces	274
8.3.1	Hilbert Spaces	275
8.3.2	Banach Spaces	277
8.4	Using Approximate Derivatives	280
8.4.1	Concepts and Assumptions	280
8.4.2	Global Convergence	285
8.4.3	Finite-Difference Approximations to Derivatives	297
8.5	Composite Problems and Models	302
9	Conditional Models	307
9.1	Motivation and Formal Description	307
9.2	Convergence to First-Order Critical Points	313
9.3	Convergence to Second-Order Critical Points	319
9.4	Conditional Models and Derivative-Free Optimization	322
9.4.1	Derivative-Free Minimization: Why and How?	322
9.4.2	Basic Concepts in Multivariate Interpolation	324
9.4.3	The Interpolation Error	330
9.5	Conditional Models and Models with Memory	336
9.5.1	Adding Memory to the Model of the Objective Function	336
9.5.2	The Effect of Memory on the Modelling Error	338

10 Algorithmic Extensions	347
10.1 A Nonmonotone Trust-Region Algorithm	347
10.1.1 The Nonmonotone Algorithm	348
10.1.2 Convergence Theory Revisited	350
10.1.3 The Reference Iteration and Other Practicalities	355
10.2 Structured Trust Regions	359
10.2.1 Motivation and Basic Concepts	359
10.2.2 A Trust-Region Algorithm Using Problem Structure	364
10.2.3 Convergence Theory	370
10.3 Trust Regions and Linesearches	376
10.3.1 Linesearch Algorithms as Trust-Region Methods	376
10.3.2 Backtracking at Unsuccessful Iterations	380
10.3.3 Which Steps Are Gradient Related?	383
10.4 Other Measures of Model Accuracy	387
10.4.1 Magical Steps	387
10.4.2 Correction Steps	391
10.4.3 Modified Models	392
10.5 Alternative Trust-Region Management	394
10.5.1 Internal Doubling	394
10.5.2 Basing the Radius Update on the Steplength	396
10.6 Problems with Dynamic Accuracy	399
10.6.1 An Algorithm Using Dynamic Accuracy	400
10.6.2 Conditional Models and Dynamic Accuracy	406
11 Nonsmooth Problems	407
11.1 Algorithms for Nonsmooth Optimization	408
11.2 Convergence to a First-Order Critical Point	414
11.3 Variations on This Theme	423
11.3.1 A Nonmonotonic Variant	423
11.3.2 Correction Steps	424
11.4 Computing the Generalized Gradient	426
11.5 Suitable Models	434
III Trust-Region Methods for Constrained Optimization with Convex Constraints	439
12 Projection Methods for Convex Constraints	441
12.1 Simple Feasible Domains and Their Geometry	441
12.1.1 Simple Bounds on the Variables	442
12.1.2 Other Simple Domains	443
12.1.3 The Projected-Gradient Path	444
12.1.4 A New Criticality Measure	448

12.2	An Algorithm for Problems with Convex Constraints	451
12.2.1	The Generalized Cauchy Point	453
12.2.2	Convergence to First-Order Critical Points	458
12.3	Active Constraints Identification	460
12.3.1	Further Assumptions	460
12.3.2	The Geometry of the Set of Limit Points	461
12.3.3	The Identification of Active Constraints	467
12.4	Convergence to Second-Order Critical Points	474
12.4.1	The Role of the Lagrangian	474
12.4.2	Convex Models	476
12.4.3	Nonconvex Models	481
13	Barrier Methods for Inequality Constraints	491
13.1	Convex Constraints and Barriers	491
13.2	A Trust-Region Method for Barrier Functions	498
13.3	Constrained Cauchy and Eigenpoints	504
13.4	The Primal Log-Barrier Algorithm	511
13.5	Reciprocal Barriers	517
13.6	A Primal-Dual Algorithm	518
13.6.1	The Algorithm	518
13.6.2	Convergence Properties	522
13.6.3	Updating the Vector of Dual Variables	525
13.7	Scaling of Barrier Methods	527
13.7.1	Reintroducing Iteration-Dependent Norms	527
13.7.2	Scaling of the Inner Iterations	528
13.7.3	Scaling of the Outer Iterations	530
13.8	Upper and Lower Bounds on the Variables	535
13.9	Barrier Methods for General Constraints	536
13.10	Adding Linear Equality Constraints	542
13.11	The Affine-Scaling Method	550
13.12	The Method of Coleman and Li	554
13.12.1	The Algorithm	554
13.12.2	Convergence Theory	559
IV	Trust-Region Methods for General Constrained Optimization and Systems of Nonlinear Equations	571
14	Penalty-Function Methods	573
14.1	Penalty Functions and Constrained Optimization	573
14.2	Smooth Penalty Functions	575
14.3	Quadratic Penalty-Function Methods	582
14.4	Augmented Lagrangian Function Methods	593

14.5	Nonsmooth Exact Penalty Functions	610
14.6	Smooth Exact Penalty Functions	616
15	Sequential Quadratic Programming Methods	623
15.1	Introduction	623
15.2	What Is Sequential Quadratic Programming?	624
15.2.1	Methods for Problems with Equality Constraints	624
15.2.2	Methods for Inequality Constraints	631
15.2.3	Quadratic Programming	633
15.3	Merit Functions and SQP Methods	636
15.3.1	The Augmented Lagrangian Penalty Function	637
15.3.2	Nonsmooth Exact Penalty Functions	637
15.3.3	Smooth Exact Penalty Functions	655
15.4	Composite-Step Trust-Region SQP Methods	657
15.4.1	Vardi-Like Approaches	658
15.4.2	Byrd–Omojokun-like Approaches	694
15.4.3	Celis–Dennis–Tapia-like Approaches	711
15.4.4	Inequality Constraints	717
15.5	The Filter Method	721
15.5.1	A Composite-Step Approximate SQP Framework	721
15.5.2	The Notion of a Filter	725
15.5.3	An SQP Filter Algorithm	727
15.5.4	Convergence to First-Order Critical Points	730
15.5.5	An Alternative Step Strategy	742
15.6	Nonquadratic Models	745
15.7	Concluding Remarks	746
16	Nonlinear Equations and Nonlinear Fitting	749
16.1	Nonlinear Equations, Nonlinear Least Squares	749
16.2	Nonlinear Equations in Other Norms	759
16.2.1	Global Convergence	761
16.2.2	Asymptotic Convergence of the Basic Method	761
16.2.3	Second-Order Corrections	766
16.3	Complementarity Problems	770
16.3.1	The Problem and an Associated Merit Function	770
16.3.2	Applying the Basic Nonsmooth Algorithm	772
16.3.3	Regular Complementarity Problems and Their Solutions	774
V	Final Considerations	779
17	Practicalities	781
17.1	Choosing the Algorithmic Parameters	781
17.2	Choosing the Initial Trust-Region Radius	784

17.3	Computing the Generalized Cauchy Point	789
17.4	Other Numerical Considerations	792
17.4.1	Computing the Model Decrease	792
17.4.2	Computing the Value of ρ_k	793
17.4.3	Stopping Conditions	794
17.4.4	Noise and/or Expensive Function Evaluations	798
17.5	Software	799
Afterword		801
Appendix: A Summary of Assumptions		803
Annotated Bibliography		813
Subject and Notation Index		935
Author Index		951