

# Bibliographic Notes

**Chapter 1** For a comprehensive treatment of multidimensional calculus and linear algebra, the reader can refer to [24, 29, 33, 34, 35] and also Appendix A of [9].

**Chapter 2** The topic of optimality conditions in Sections 2.1–2.3 is classical and can also be found in many other books such as [1, 30]. The principal minors criterion is also known as “Sylvester’s criterion,” and its proof can be found for example in [24].

**Chapter 3** A comprehensive study of least squares methods and generalizations can be found in [11]. The discussion on circle fitting follows [4]. An excellent guide for MATLAB is the book [20].

**Chapter 4** The gradient method is discussed in many books; see, for example, [9, 27, 31]. More background and convergence results on the Gauss–Newton method can be found in the book [28]. The original paper of Weiszfeld appeared in [39]. The analysis in Section 4.6 follows [6].

**Chapter 5** More details and further extensions of Newton’s method can be found, for example, in [9, 15, 17, 27, 28]. The hybrid Newton’s method can be found in [38]. An excellent reference for an abundant of optimization methods is the book [28].

**Chapters 6 and 7** A classical reference for convex analysis is [32]. The book [10] also contains a comprehensive treatment of the subject.

**Chapter 8** A large variety of examples of convex optimization problems can be found in [13] and also in [8]. The original paper of Markowitz describing the portfolio optimization model is [23]. The CVX MATLAB software as well as a user guide can be found in [19]. The CVX software uses two conic optimization solvers: SeDuMi [36] and SDPT3 [37]. The reformulation of the trust region subproblem as a convex problem described in Section 8.2.7 follows the paper [11].

**Chapter 9** Stationarity is a basic concept in optimization that can be found in many sources such as [9]. The gradient mapping is extensively studied in [27]. The analysis of the convergence gradient projection method follows [5, 6]. The discussion on sparsity constrained problems is based on [3]. The iterative hard thresholding method was initially presented and studied in [12].

**Chapter 10** Generalization and extensions of separation and alternative theorems can be found in [32, 10, 21]. The discussion on the orthogonal regression problem follows [4].

**Chapter 11** A comprehensive treatment of the KKT conditions, including variants that were not presented in this book, can be found in [1, 9]. The derivation of the second order necessary conditions follows the paper [7]. Optimality conditions and algorithms for solving the trust region subproblem and its generalizations can be found in [26, 25, 16]. The total least squares problem was initially presented in [18] and was extensively studied and generalized by many authors; see the review paper [22] and references therein. The derivation of the reduced form of the total least squares problem via the KKT conditions follows [2].

**Chapter 12** Duality is a classic topic and is covered in many other optimization books; see, for example, [1, 9, 10, 13, 21, 32]. The dual algorithm for solving the denoising problem is based on Chambolle's algorithm for solving two-dimensional denoising problems with total variation regularization [14]. More on duality in geometric programming can be found in [30].

# Bibliography

- [1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear programming. Theory and algorithms*. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition, 2006.
- [2] A. Beck and A. Ben-Tal. On the solution of the Tikhonov regularization of the total least squares problem. *SIAM J. Optim.*, 17(1):98–118, 2006.
- [3] A. Beck and Y. C. Eldar. Sparsity constrained nonlinear optimization: Optimality conditions and algorithms. *SIAM J. Optim.*, 23(3):1480–1509, 2013.
- [4] A. Beck and D. Pan. On the solution of the GPS localization and circle fitting problems. *SIAM J. Optim.*, 22(1):108–134, 2012.
- [5] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.
- [6] A. Beck and M. Teboulle. Gradient-based algorithms with applications to signal-recovery problems. In *Convex optimization in signal processing and communications*, pages 42–88. Cambridge University Press, Cambridge, 2010.
- [7] A. Ben-Tal. Second-order and related extremality conditions in nonlinear programming. *J. Optim. Theory Appl.*, 31(2):143–165, 1980.
- [8] A. Ben-Tal and A. Nemirovski. *Lectures on modern convex optimization. Analysis, algorithms, and engineering applications*. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.
- [9] D. P. Bertsekas. *Nonlinear programming*. Athena Scientific, Belmont, MA, second edition, 1999.
- [10] D. P. Bertsekas. *Convex analysis and optimization*. Athena Scientific, Belmont, MA, 2003. With Angelia Nedić and Asuman E. Ozdaglar.
- [11] A. Björck. *Numerical methods for least squares problems*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.
- [12] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. *J. Fourier Anal. Appl.*, 14(5-6):629–654, 2008.
- [13] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.
- [14] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004. Special issue on mathematics and image analysis.
- [15] R. Fletcher. *Practical methods of optimization. Vol. 1. Unconstrained optimization*. John Wiley & Sons, Ltd., Chichester, 1980. A Wiley-Interscience Publication.

- [16] C. Fortin and H. Wolkowicz. The trust region subproblem and semidefinite programming. *Optim. Methods Softw.*, 19(1):41–67, 2004.
- [17] P. E. Gill, W. Murray, and M. H. Wright. *Practical optimization*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1981.
- [18] G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM J. Numer. Anal.*, 17(6):883–893, 1980.
- [19] M. Grant and S. Boyd. CVX: MATLAB software for disciplined convex programming, version 2.0 beta. <http://cvxr.com/cvx>, September 2013.
- [20] D. J. Higham and N. J. Higham. *MATLAB guide*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2005.
- [21] O. L. Mangasarian. *Nonlinear programming*. McGraw-Hill Book Co., New York, 1969.
- [22] I. Markovsky and S. Van Huffel. Overview of total least-squares methods. *Signal Process.*, 87(10):2283–2302, October 2007.
- [23] H. Markowitz. Portfolio selection. *J. Finance*, 7:77–91, 1952.
- [24] C. Meyer. *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. With 1 CD-ROM (Windows, Macintosh and UNIX) and a solutions manual (iv+171 pp.).
- [25] J. J. Moré. Generalizations of the trust region subproblem. *Optim. Methods Software*, 2:189–209, 1993.
- [26] J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM J. Sci. Statist. Comput.*, 4(3):553–572, 1983.
- [27] Y. Nesterov. *Introductory lectures on convex optimization*, volume 87 of Applied Optimization. Kluwer Academic Publishers, Boston, MA, 2004. A basic course.
- [28] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.
- [29] J. M. Ortega and W. C. Rheinboldt. *Iterative solution of nonlinear equations in several variables*, volume 30 of Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. Reprint of the 1970 original.
- [30] A. L. Peressini, F. E. Sullivan, and J. J. Uhl, Jr. *The mathematics of nonlinear programming*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1988.
- [31] B. T. Polyak. *Introduction to optimization*. Translations Series in Mathematics and Engineering. Optimization Software Inc. Publications Division, New York, 1987. Translated from the Russian, With a foreword by Dimitri P. Bertsekas.
- [32] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.
- [33] W. Rudin. *Real analysis*. McGraw-Hill, N. Y., 1976.
- [34] S. Strang. *Calculus*. Wellesley-Cambridge Press, 1991.
- [35] S. Strang. *Introduction to linear algebra*. Wellesley-Cambridge Press, fourth edition, 2009.
- [36] J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11-12:625–653, 1999.

- [37] K. C. Toh, M. J. Todd, and R. H. Tütüncü. SDPT3—a MATLAB software package for semi-definite programming, version 1.3. *Optim. Methods Softw.*, 11/12(1–4):545–581, 1999. Interior point methods.
- [38] L. Vandenberghe. *Applied numerical computing*. 2011. Lecture notes, in collaboration with S. Boyd.
- [39] E. V. Weiszfeld. Sur le point pour lequel la somme des distances de  $n$  points donnees est minimum. *The Tohoku Mathematical Journal*, 43:355–386, 1937.

# Index

- active constraints, 195, 208
- affine function, 118
- argmax, 14
- argmin, 14
- arithmetic geometric mean inequality, 139
- backtracking, 51, 178
- basic feasible solution, 107
- bisection, 219
- boundary point, 7
- boundary set, 7
- bounded below, 36
- bounded set, 8
- Carathéodory theorem, 102
- Cauchy–Schwarz, 3
- Chebyshev ball, 153
- Chebyshev center, 153, 162, 256
- Cholesky factorization, 89
- circle fitting, 45
- classification, 150, 266
- closed ball, 6
- closed line segment, 2
- closed set, 7
- closure, 8
- coercive, 25
- compact set, 8
- complementary slackness, 196, 246
- concave function, 118
- condition number, 58, 60
- cone, 104
- conic combination, 106
- conic duality theorem, 260
- conic hull, 106
- conic problem, 254
- conic representation theorem, 106
- constant stepsize, 51
- constrained least squares, 218
- constraint qualification, 210
- continuously differentiable, 9
- contour set, 7
- convex combination, 101
- convex function, 117
- convex hull, 102
- convex optimization, 147
- convex polytope, 100
- convex problem, 147
- convex set, 97
- CVX, 158
- damped Gauss–Newton method, 67, 68
- damped Newton’s method, 88
- data fitting, 39
- denoising, 42
- descent direction, 49
- descent directions method, 50
- descent lemma, 74
- descent method, 50
- diagonally dominant matrix, 22
- directional derivative, 8
- distance function, 130, 157
- dot product, 2
- dual cone, 114, 205
- dual objective function, 238
- dual problem, 237
- effective domain, 135
- eigenvalue, 5
- eigenvector, 5
- ellipsoid, 99
- epigraph, 136
- Euclidean norm, 3
- exact line search, 51
- extended real-valued function, 135
- extreme point, 111
- facility location, 69
- Farkas lemma, 192
- feasible descent direction, 207
- feasible set, 13
- feasible solution, 13
- Fejér monotonicity, 183
- Fermat’s theorem, 16
- Fermat–Weber problem, 68, 80
- first projection theorem, 157
- Fritz–John conditions, 208
- Frobenius norm, 4
- full column rank, 37
- function
  - distance, 130, 157
  - dual objective, 238
  - Huber, 189
  - quadratic, 32
  - Rosenbrock, 60
- generalized Slater’s condition, 216
- geometric programming, 267
- global maximum and minimum, 13
- global optimum, 13
- Gordan’s theorem, 194, 205, 208
- gradient, 9
- gradient inequality, 119
- gradient mapping, 177
- gradient method, 52
- gradient projection method, 175
- Hessian, 10
- hidden convexity, 155, 165
- Hölder’s inequality, 140
- Huber function, 189
- hyperplane, 98
- identity matrix, 2
- ill-conditioned matrices, 60
- indefinite matrix, 18
- indicator function, 135
- induced matrix norm, 4
- inequality
  - Jensen’s, 118
  - Kantorovich, 59
  - linear matrix, 254

- Minkowski's, 140
  - strongly convex, 117
- inner product, 2
- interior, 7
- interior point, 7
- iterative hard-thresholding, 187
- Jacobian, 67
- Jensen's inequality, 118
- Kantorovich inequality, 59
- KKT conditions, 195, 207
- KKT point, 198, 211
- Krein–Milman theorem, 113
- Lagrange multipliers, 196
- Lagrangian function, 198
- least squares, 37
  - linear, 45
  - regularized, 41, 219, 262
  - total, 230
- level set, 7, 130
- line, 97
- line search, 50
- line segment principle, 108
- linear approximation theorem, 10
- linear fitting, 39
- linear function, 118
- linear least squares, 45
- linear matrix inequality, 254
- linear programming, 107, 149, 247
- linear rate, 60
- Lipschitz continuity of the
  - gradient, 73
- local maximum point, 15
- local minimum point, 15
- log-sum-exp, 124
- Lorentz cone, 105
- matrix norm, 3
- maximal value, 13
- maximizer, 13
- minimial value, 13
- minimizer, 13
- Minkowski functional, 113
- Minkowski's inequality, 140
- monomial, 267
- Motzkin's theorem, 205
- negative definite, 18
- negative semidefinite, 18
- neighborhood, 7
- Newton direction, 83
- Newton's method, 64, 83
  - damped, 88
  - pure, 83
- nonexpansive, 174
- nonlinear Farkas' lemma, 242
- nonlinear fitting, 40
- nonlinear least squares, 45, 67
- nonnegative orthant, 1
- nonnegative part, 157
- norm
  - diagonally dominant, 22
  - identity, 2
  - indefinite, 18
  - induced matrix, 4
  - square root of, 21
  - strictly diagonally dominant, 22
- normal cone, 114
- normal system, 37
- open ball, 6
- open line segment, 2
- open set, 7
- optimal set, 148
- orthogonal projection, 156
- orthogonal regression, 203
- overdetermined system, 37
- partial derivative, 8
- pointed cone, 114
- positive definite, 17
- positive orthant, 2
- positive semidefinite, 17
- posynomial, 267
- primal problem, 238
- principal minors criterion, 21
- proper, 135
- pure Newton's method, 83
- Q-norm, 35
- QCQP, 155, 235
- quadratic approximation
  - theorem, 10
- quadratic function, 32
- quadratic problem, 150
- quadratic-over-linear, 125, 127
- quasi-convex function, 131
- Rayleigh quotient, 5, 204, 232
- real-valued, 135
- regularity, 211
- regularization parameter, 41
- regularized least squares, 41, 219, 262
- robust regression, 163
- Rosenbrock function, 60
- saddle point, 24
- scaled gradient method, 63
- second projection theorem, 173
- semidefinite programming, 254
- separation of two convex sets, 242
- separation theorem, 191
- set
  - bounded, 8
  - open, 7
- sgn, 156
- singleton, 187
- Slater's condition, 214
- source localization problem, 80
- sparsity constrained problems, 183
- spectral decomposition, 5
- spectral norm, 4
- square root of a matrix, 21
- standard basis, 1
- stationary point, 17, 169
- strict global maximum, 13
- strict global minimum, 13
- strict local maximum, 15
- strict local minimum, 15
- strict separation theorem, 191
- strictly concave function, 118
- strictly convex function, 117
- strictly diagonally dominant
  - matrix, 22
- strong convexity parameter, 144
- strong duality, 241
- strongly convex function, 144, 190
- sublinear rate, 182
- sufficient decrease lemma, 75, 176
- support, 184
- support function, 136
- supporting hyperplane theorem, 241
- total least squares, 230
- trust region subproblem, 155, 227
- unit-simplex, 2, 254
- unit-sum set, 171
- Vandermonde, 41
- weak duality theorem, 239
- Weierstrass theorem, 25
- well-conditioned matrices, 60
- Young's inequality, 140