

References

- [ACH97] Auslender, A., Cominetti, R., and Haddou, M., 1997. “Asymptotic Analysis for Penalty and Barrier Methods in Convex and Linear Programming,” *Math. of Operations Research*, Vol. 22, pp. 43-62.
- [ALS14] Abernethy, J., Lee, C., Sinha, A., and Tewari, A., 2014. “Online Linear Optimization via Smoothing,” arXiv preprint arXiv:1405.6076.
- [AgB14] Agarwal, A., and Bottou, L., 2014. “A Lower Bound for the Optimization of Finite Sums,” arXiv preprint arXiv:1410.0723.
- [AgD11] Agarwal, A., and Duchi, J. C., 2011. “Distributed Delayed Stochastic Optimization,” In *Advances in Neural Information Processing Systems (NIPS 2011)*, pp. 873-881.
- [AlG03] Alizadeh, F., and Goldfarb, D., 2003. “Second-Order Cone Programming,” *Math. Programming*, Vol. 95, pp. 3-51.
- [AnH13] Andersen, M. S., and Hansen, P. C., 2013. “Generalized Row-Action Methods for Tomographic Imaging,” *Numerical Algorithms*, Vol. 67, pp. 1-24.
- [Arm66] Armijo, L., 1966. “Minimization of Functions Having Continuous Partial Derivatives,” *Pacific J. Math.*, Vol. 16, pp. 1-3.
- [Ash72] Ash, R. B., 1972. *Real Analysis and Probability*, Academic Press, NY.
- [AtV95] Atkinson, D. S., and Vaidya, P. M., 1995. “A Cutting Plane Algorithm for Convex Programming that Uses Analytic Centers,” *Math. Programming*, Vol. 69, pp. 1-44.
- [AuE76] Aubin, J. P., and Ekeland, I., 1976. “Estimates of the Duality Gap in Nonconvex Optimization,” *Math. of Operations Research*, Vol. 1, pp. 255- 245.
- [AuT03] Auslender, A., and Teboulle, M., 2003. *Asymptotic Cones and Functions in Optimization and Variational Inequalities*, Springer, NY.
- [AuT04] Auslender, A., and Teboulle, M., 2004. “Interior Gradient and Epsilon-Subgradient Descent Methods for Constrained Convex Minimization,” *Math. of Operations Research*, Vol. 29, pp. 1-26.
- [Aus76] Auslender, A., 1976. *Optimization: Methodes Numeriques*, Mason, Paris.
- [Aus92] Auslender, A., 1992. “Asymptotic Properties of the Fenchel Dual

- Functional and Applications to Decomposition Problems,” *J. of Optimization Theory and Applications*, Vol. 73, pp. 427-449.
- [BBC11] Bertsimas, D., Brown, D. B., and Caramanis, C., 2011. “Theory and Applications of Robust Optimization,” *SIAM Review*, Vol. 53, pp. 464-501.
- [BBG09] Bordes, A., Bottou, L., and Gallinari, P., 2009. “SGD-QN: Careful Quasi-Newton Stochastic Gradient Descent,” *J. of Machine Learning Research*, Vol. 10, pp. 1737-1754.
- [BBL07] Berry, M. W., Browne, M., Langville, A. N., Pauca, V. P., and Plemmons, R. J., 2007. “Algorithms and Applications for Approximate Nonnegative Matrix Factorization,” *Computational Statistics and Data Analysis*, Vol. 52, pp. 155-173.
- [BBY12] Borwein, J. M., Burachik, R. S., and Yao, L., 2012. “Conditions for Zero Duality Gap in Convex Programming,” *arXiv preprint arXiv:1211.4953*.
- [BCK06] Bauschke, H. H., Combettes, P. L., and Kruk, S. G., 2006. “Extrapolation Algorithm for Affine-Convex Feasibility Problems,” *Numer. Algorithms*, Vol. 41, pp. 239-274.
- [BGI95] Burachik, R., Grana Drummond, L. M., Iusem, A. N., and Svaiter, B. F., 1995. “Full Convergence of the Steepest Descent Method with Inexact Line Searches,” *Optimization*, Vol. 32, pp. 137-146.
- [BGL06] Bonnans, J. F., Gilbert, J. C., Lemaréchal, C., and Sagastizábal, S. C., 2006. *Numerical Optimization: Theoretical and Practical Aspects*, Springer, NY.
- [BGN09] Ben-Tal, A., El Ghaoui, L., and Nemirovski, A., 2009. *Robust Optimization*, Princeton Univ. Press, Princeton, NJ.
- [BHG08] Blatt, D., Hero, A. O., Gauchman, H., 2008. “A Convergent Incremental Gradient Method with a Constant Step Size,” *SIAM J. Optimization*, Vol. 18, pp. 29-51.
- [BHT87] Bertsekas, D. P., Hossein, P., and Tseng, P., 1987. “Relaxation Methods for Network Flow Problems with Convex Arc Costs,” *SIAM J. on Control and Optimization*, Vol. 25, pp. 1219-1243.
- [BJM12] Bach, F., Jenatton, R., Mairal, J., and Obozinski, G., 2012. “Optimization with Sparsity-Inducing Penalties,” *Foundations and Trends in Machine Learning*, Vol. 4, pp. 1-106.
- [BKM14] Burachik, R. S., Kaya, C. Y., and Majeed, S. N., 2014. “A Duality Approach for Solving Control-Constrained Linear-Quadratic Optimal Control Problems,” *SIAM J. on Control and Optimization*, Vol. 52, pp. 1423-1456.
- [BLY14] Bragin, M. A., Luh, P. B., Yan, J. H., Yu, N., and Stern, G. A., 2014. “Convergence of the Surrogate Lagrangian Relaxation Method,” *J. of Optimization Theory and Applications*, on line.

- [BMN01] Ben-Tal, A., Margalit, T., and Nemirovski, A., 2001. “The Ordered Subsets Mirror Descent Optimization Method and its Use for the Positron Emission Tomography Reconstruction,” in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications* (D. Butnariu, Y. Censor, and S. Reich, eds.), Elsevier, Amsterdam, Netherlands.
- [BMR00] Birgin, E. G., Martinez, J. M., and Raydan, M., 2000. “Non-monotone Spectral Projected Gradient Methods on Convex Sets,” *SIAM J. on Optimization*, Vol. 10, pp. 1196-1211.
- [BMS99] Boltyanski, V., Martini, H., and Soltan, V., 1999. *Geometric Methods and Optimization Problems*, Kluwer, Boston.
- [BNO03] Bertsekas, D. P., Nedić, A., and Ozdaglar, A. E., 2003. *Convex Analysis and Optimization*, Athena Scientific, Belmont, MA.
- [BOT06] Bertsekas, D. P., Ozdaglar, A. E., and Tseng, P., 2006. “Enhanced Fritz John Optimality Conditions for Convex Programming,” *SIAM J. on Optimization*, Vol. 16, pp. 766-797.
- [BPC11] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J., 2011. *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*, Now Publishers Inc, Boston, MA.
- [BPP13] Bhatnagar, S., Prasad, H., and Prashanth, L. A., 2013. *Stochastic Recursive Algorithms for Optimization*, Lecture Notes in Control and Information Sciences, Springer, NY.
- [BPT97a] Bertsekas, D. P., Polymenakos, L. C., and Tseng, P., 1997. “An ϵ -Relaxation Method for Separable Convex Cost Network Flow Problems,” *SIAM J. on Optimization*, Vol. 7, pp. 853-870.
- [BPT97b] Bertsekas, D. P., Polymenakos, L. C., and Tseng, P., 1997. “Epsilon-Relaxation and Auction Methods for Separable Convex Cost Network Flow Problems,” in *Network Optimization*, Pardalos, P. M., Hearn, D. W., and Hager, W. W., (Eds.), Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, NY, pp. 103-126.
- [BSL14] Bergmann, R., Steidl, G., Laus, F., and Weinmann, A., 2014. “Second Order Differences of Cyclic Data and Applications in Variational Denoising,” arXiv preprint arXiv:1405.5349.
- [BSS06] Bazararaa, M. S., Sherali, H. D., and Shetty, C. M., 2006. *Nonlinear Programming: Theory and Algorithms*, 3rd Edition, Wiley, NY.
- [BST14] Bolte, J., Sabach, S., and Teboulle, M., 2014. “Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems,” *Math. Programming*, Vol. 146, pp. 1-36.
- [BaB88] Barzilai, J., and Borwein, J. M., 1988. “Two Point Step Size Gradient Methods,” *IMA J. of Numerical Analysis*, Vol. 8, pp. 141-148.
- [BaB96] Bauschke, H. H., and Borwein, J. M., 1996. “On Projection Algorithms for Solving Convex Feasibility Problems,” *SIAM Review*, Vol. 38, pp. 367-426.

- [BaC11] Bauschke, H. H., and Combettes, P. L., 2011. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, NY.
- [BaM11] Bach, F., and E. Moulines, E., 2011. “Non-Asymptotic Analysis of Stochastic Approximation Algorithms for Machine Learning,” *Advances in Neural Information Processing Systems (NIPS 2011)*.
- [BaT85] Balas, E., and Toth, P., 1985. “Branch and Bound Methods,” in *The Traveling Salesman Problem*, Lawler, E., Lenstra, J. K., Rinnoy Kan, A. H. G., and Shmoys, D. B., (Eds.), Wiley, NY, pp. 361-401.
- [BaW75] Balinski, M., and Wolfe, P., (Eds.), 1975. *Nondifferentiable Optimization*, Math. Programming Study 3, North-Holland, Amsterdam.
- [Bac14] Bacak, M., 2014. “Computing Medians and Means in Hadamard Spaces,” arXiv preprint arXiv:1210.2145v3.
- [Bau01] Bauschke, H. H., 2001. “Projection Algorithms: Results and Open Problems,” in *Inherently Parallel Algorithms in Feasibility and Optimization and their Applications* (D. Butnariu, Y. Censor, and S. Reich, eds.), Elsevier, Amsterdam, Netherlands.
- [BeF12] Becker, S., and Fadili, J., 2012. “A Quasi-Newton Proximal Splitting Method,” in *Advances in Neural Information Processing Systems (NIPS 2012)*, pp. 2618-2626.
- [BeG83] Bertsekas, D. P., and Gafni, E., 1983. “Projected Newton Methods and Optimization of Multicommodity Flows,” *IEEE Trans. Automat. Control*, Vol. AC-28, pp. 1090-1096.
- [BeG92] Bertsekas, D. P., and Gallager, R. G., 1992. *Data Networks*, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ.
On line at <http://web.mit.edu/dimitrib/www/datanets.html>.
- [BeL88] Becker, S., and LeCun, Y., 1988. “Improving the Convergence of Back-Propagation Learning with Second Order Methods,” in *Proceedings of the 1988 Connectionist Models Summer School*, San Matteo, CA.
- [BeL07] Bengio, Y., and LeCun, Y., 2007. “Scaling Learning Algorithms Towards AI,” *Large-Scale Kernel Machines*, Vol. 34, pp. 1-41.
- [BeM71] Bertsekas, D. P., and Mitter, S. K., 1971. “Steepest Descent for Optimization Problems with Nondifferentiable Cost Functionals,” *Proc. 5th Annual Princeton Confer. Inform. Sci. Systems*, Princeton, NJ, pp. 347-351.
- [BeM73] Bertsekas, D. P., and Mitter, S. K., 1973. “A Descent Numerical Method for Optimization Problems with Nondifferentiable Cost Functionals,” *SIAM J. on Control*, Vol. 11, pp. 637-652.
- [BeN01] Ben-Tal, A., and Nemirovskii, A., 2001. *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, SIAM, Philadelphia, PA.
- [BeO02] Bertsekas, D. P., Ozdaglar, A. E., 2002. “Pseudonormality and a Lagrange Multiplier Theory for Constrained Optimization,” *J. of Opti-*

- mization Theory and Applications, Vol. 114, pp. 287-343.
- [BeS82] Bertsekas, D. P., and Sandell, N. R., 1982. "Estimates of the Duality Gap for Large-Scale Separable Nonconvex Optimization Problems," Proc. 1982 IEEE Conf. Decision and Control, pp. 782-785.
- [BeT88] Bertsekas, D. P., and Tseng, P., 1988. "Relaxation Methods for Minimum Cost Ordinary and Generalized Network Flow Problems," Operations Research, Vol. 36, pp. 93-114.
- [BeT89a] Bertsekas, D. P., and Tsitsiklis, J. N., 1989. Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, Englewood Cliffs, NJ; republished by Athena Scientific, Belmont, MA, 1997. On line at <http://web.mit.edu/dimitrib/www/pdc.html>.
- [BeT89b] Ben-Tal, A., and Teboulle, M., 1989. "A Smoothing Technique for Nondifferentiable Optimization Problems," Optimization, Lecture Notes in Mathematics, Vol. 1405, pp. 1-11.
- [BeT91] Bertsekas, D. P., and Tsitsiklis, J. N., 1991. "Some Aspects of Parallel and Distributed Iterative Algorithms - A Survey," Automatica, Vol. 27, pp. 3-21.
- [BeT94a] Bertsekas, D. P., and Tseng, P., 1994. "Partial Proximal Minimization Algorithms for Convex Programming," SIAM J. on Optimization, Vol. 4, pp. 551-572.
- [BeT94b] Bertsekas, D. P., and Tseng, P., 1994. "RELAX-IV: A Faster Version of the RELAX Code for Solving Minimum Cost Flow Problems," Massachusetts Institute of Technology, Laboratory for Information and Decision Systems Report LIDS-P-2276, Cambridge, MA.
- [BeT96] Bertsekas, D. P., and Tsitsiklis, J. N., 1996. Neuro-Dynamic Programming, Athena Scientific, Belmont, MA.
- [BeT97] Bertsimas, D., and Tsitsiklis, J. N., 1997. Introduction to Linear Optimization, Athena Scientific, Belmont, MA.
- [BeT00] Bertsekas, D. P., and Tsitsiklis, J. N., 2000. "Gradient Convergence of Gradient Methods with Errors," SIAM J. on Optimization, Vol. 36, pp. 627-642.
- [BeT03] Beck, A., and Teboulle, M., 2003. "Mirror Descent and Nonlinear Projected Subgradient Methods for Convex Optimization," Operations Research Letters, Vol. 31, pp. 167-175.
- [BeT09a] Beck, A., and Teboulle, M., 2009. "Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems," IEEE Trans. on Image Processing, Vol. 18, pp. 2419-2434.
- [BeT09b] Beck, A., and Teboulle, M., 2009. "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," SIAM J. on Imaging Sciences, Vol. 2, pp. 183-202.
- [BeT10] Beck, A., and Teboulle, M., 2010. "Gradient-Based Algorithms with Applications to Signal-Recovery Problems," in Convex Optimization

- in *Signal Processing and Communications* (Y. Eldar and D. Palomar, eds.), Cambridge Univ. Press, pp. 42-88.
- [BeT13] Beck, A., and Tetrushvili, L., 2013. "On the Convergence of Block Coordinate Descent Type Methods," *SIAM J. on Optimization*, Vol. 23, pp. 2037-2060.
- [BeY09] Bertsekas, D. P., and Yu, H., 2009. "Projected Equation Methods for Approximate Solution of Large Linear Systems," *J. of Computational and Applied Mathematics*, Vol. 227, pp. 27-50.
- [BeY10] Bertsekas, D. P., and Yu, H., 2010. "Asynchronous Distributed Policy Iteration in Dynamic Programming," *Proc. of Allerton Conf. on Communication, Control and Computing*, Allerton Park, Ill, pp. 1368-1374.
- [BeY11] Bertsekas, D. P., and Yu, H., 2011. "A Unifying Polyhedral Approximation Framework for Convex Optimization," *SIAM J. on Optimization*, Vol. 21, pp. 333-360.
- [BeZ97] Ben-Tal, A., and Zibulevsky, M., 1997. "Penalty/Barrier Multiplier Methods for Convex Programming Problems," *SIAM J. on Optimization*, Vol. 7, pp. 347-366.
- [Ben09] Bengio, Y., 2009. "Learning Deep Architectures for AI," *Foundations and Trends in Machine Learning*, Vol. 2, pp. 1-127.
- [Ber71] Bertsekas, D. P., 1971. "Control of Uncertain Systems With a Set-Membership Description of the Uncertainty," Ph.D. Thesis, Dept. of EECS, MIT; may be downloaded from <http://web.mit.edu/dimitrib/www/publ.html>.
- [Ber72] Bertsekas, D. P., 1972. "Stochastic Optimization Problems with Nondifferentiable Cost Functionals with an Application in Stochastic Programming," *Proc. 1972 IEEE Conf. Decision and Control*, pp. 555-559.
- [Ber73] Bertsekas, D. P., 1973. "Stochastic Optimization Problems with Nondifferentiable Cost Functionals," *J. of Optimization Theory and Applications*, Vol. 12, pp. 218-231.
- [Ber75a] Bertsekas, D. P., 1975. "Necessary and Sufficient Conditions for a Penalty Method to be Exact," *Math. Programming*, Vol. 9, pp. 87-99.
- [Ber75b] Bertsekas, D. P., 1975. "Nondifferentiable Optimization via Approximation," *Math. Programming Study 3*, Balinski, M., and Wolfe, P., (Eds.), North-Holland, Amsterdam, pp. 1-25.
- [Ber75c] Bertsekas, D. P., 1975. "Combined Primal-Dual and Penalty Methods for Constrained Optimization," *SIAM J. on Control*, Vol. 13, pp. 521-544.
- [Ber75d] Bertsekas, D. P., 1975. "On the Method of Multipliers for Convex Programming," *IEEE Transactions on Aut. Control*, Vol. 20, pp. 385-388.
- [Ber76a] Bertsekas, D. P., 1976. "On the Goldstein-Levitin-Poljak Gradient Projection Method," *IEEE Trans. Automat. Control*, Vol. 21, pp. 174-184.

- [Ber76b] Bertsekas, D. P., 1976. "Multiplier Methods: A Survey," *Automatica*, Vol. 12, pp. 133-145.
- [Ber76c] Bertsekas, D. P., 1976. "On Penalty and Multiplier Methods for Constrained Optimization," *SIAM J. on Control and Optimization*, Vol. 14, pp. 216-235.
- [Ber77] Bertsekas, D. P., 1977. "Approximation Procedures Based on the Method of Multipliers," *J. Optimization Theory and Applications*, Vol. 23, pp. 487-510.
- [Ber79] Bertsekas, D. P., 1979. "A Distributed Algorithm for the Assignment Problem," Lab. for Information and Decision Systems Working Paper, MIT, Cambridge, MA.
- [Ber81] Bertsekas, D. P., 1981. "A New Algorithm for the Assignment Problem," *Mathematical Programming*, Vol. 21, pp. 152-171.
- [Ber82a] Bertsekas, D. P., 1982. *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, NY; republished in 1996 by Athena Scientific, Belmont, MA. On line at <http://web.mit.edu/dimitrib/www/lagrmult.html>.
- [Ber82b] Bertsekas, D. P., 1982. "Projected Newton Methods for Optimization Problems with Simple Constraints," *SIAM J. on Control and Optimization*, Vol. 20, pp. 221-246.
- [Ber82c] Bertsekas, D. P., 1982. "Distributed Dynamic Programming," *IEEE Trans. Aut. Control*, Vol. AC-27, pp. 610-616.
- [Ber83] Bertsekas, D. P., 1983. "Distributed Asynchronous Computation of Fixed Points," *Math. Programming*, Vol. 27, pp. 107-120.
- [Ber85] Bertsekas, D. P., 1985. "A Unified Framework for Primal-Dual Methods in Minimum Cost Network Flow Problems," *Mathematical Programming*, Vol. 32, pp. 125-145.
- [Ber91] Bertsekas, D. P., 1991. *Linear Network Optimization: Algorithms and Codes*, MIT Press, Cambridge, MA.
- [Ber92] Bertsekas, D. P., 1992. "Auction Algorithms for Network Problems: A Tutorial Introduction," *Computational Optimization and Applications*, Vol. 1, pp. 7-66.
- [Ber96] Bertsekas, D. P., 1996. "Incremental Least Squares Methods and the Extended Kalman Filter," *SIAM J. on Optimization*, Vol. 6, pp. 807-822.
- [Ber97] Bertsekas, D. P., 1997. "A New Class of Incremental Gradient Methods for Least Squares Problems," *SIAM J. on Optimization*, Vol. 7, pp. 913-926.
- [Ber98] Bertsekas, D. P., 1998. *Network Optimization: Continuous and Discrete Models*, Athena Scientific, Belmont, MA.
- [Ber99] Bertsekas, D. P., 1999. *Nonlinear Programming: 2nd Edition*, Athena Scientific, Belmont, MA.

- [Ber07] Bertsekas, D. P., 2007. *Dynamic Programming and Optimal Control*, Vol. I, 3rd Edition, Athena Scientific, Belmont, MA.
- [Ber09] Bertsekas, D. P., 2009. *Convex Optimization Theory*, Athena Scientific, Belmont, MA.
- [Ber10a] Bertsekas, D. P., 2010. “Extended Monotropic Programming and Duality,” Lab. for Information and Decision Systems Report LIDS-P-2692, MIT, March 2006, Revised in Feb. 2010; a version appeared in *J. of Optimization Theory and Applications*, Vol. 139, pp. 209-225.
- [Ber10b] Bertsekas, D. P., 2010. “Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey,” Lab. for Information and Decision Systems Report LIDS-P-2848, MIT.
- [Ber11] Bertsekas, D. P., 2011. “Incremental Proximal Methods for Large Scale Convex Optimization,” *Math. Programming*, Vol. 129, pp. 163-195.
- [Ber12] Bertsekas, D. P., 2012. *Dynamic Programming and Optimal Control*, Vol. II, 4th Edition: *Approximate Dynamic Programming*, Athena Scientific, Belmont, MA.
- [Ber13] Bertsekas, D. P., 2013. *Abstract Dynamic Programming*, Athena Scientific, Belmont, MA.
- [BiF07] Bioucas-Dias, J., and Figueiredo, M. A. T., 2007. “A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration,” *IEEE Trans. Image Processing*, Vol. 16, pp. 2992-3004.
- [BiL97] Birge, J. R., and Louveaux, 1997. *Introduction to Stochastic Programming*, Springer-Verlag, New York, NY.
- [Bis95] Bishop, C. M., 1995. *Neural Networks for Pattern Recognition*, Oxford Univ. Press, NY.
- [BoL00] Borwein, J. M., and Lewis, A. S., 2000. *Convex Analysis and Nonlinear Optimization*, Springer-Verlag, NY.
- [BoL05] Bottou, L., and LeCun, Y., 2005. “On-Line Learning for Very Large Datasets,” *Applied Stochastic Models in Business and Industry*, Vol. 21, pp. 137-151.
- [BoS00] Bonnans, J. F., and Shapiro, A., 2000. *Perturbation Analysis of Optimization Problems*, Springer-Verlag, NY.
- [BoV04] Boyd, S., and Vandenbergue, L., 2004. *Convex Optimization*, Cambridge Univ. Press, Cambridge, UK.
- [Bor08] Borkar, V. S., 2008. *Stochastic Approximation: A Dynamical Systems Viewpoint*, Cambridge Univ. Press.
- [Bot05] Bottou, L., 2005. “SGD: Stochastic Gradient Descent,” <http://leon.bottou.org/projects/sgd>.
- [Bot09] Bottou, L., 2009. “Curiously Fast Convergence of Some Stochastic Gradient Descent Algorithms,” unpublished open problem offered to the attendance of the SLDS 2009 conference.

- [Bot10] Bottou, L., 2010. “Large-Scale Machine Learning with Stochastic Gradient Descent,” In Proc. of COMPSTAT 2010, pp. 177-186.
- [BrL78] Brezis, H., and Lions, P. L., 1978. “Produits Infinis de Resolvantes,” Israel J. of Mathematics, Vol. 29, pp. 329-345.
- [BrS13] Brown, D. B., and Smith, J. E., 2013. “Information Relaxations, Duality, and Convex Stochastic Dynamic Programs,” Working Paper, Fuqua School of Business, Durham, NC, USA.
- [Bre73] Brezis, H., 1973. “Operateurs Maximaux Monotones et Semi-Groupes de Contractions Dans les Espaces de Hilbert,” North-Holland, Amsterdam.
- [BuM13] Burachik, R. S., and Majeed, S. N., 2013. “Strong Duality for Generalized Monotropic Programming in Infinite Dimensions,” J. of Mathematical Analysis and Applications, Vol. 400, pp. 541-557.
- [BuQ98] Burke, J. V., and Qian, M., 1998. “A Variable Metric Proximal Point Algorithm for Monotone Operators,” SIAM J. on Control and Optimization, Vol. 37, pp. 353-375.
- [Bur91] Burke, J. V., 1991. “An Exact Penalization Viewpoint of Constrained Optimization,” SIAM J. on Control and Optimization, Vol. 29, pp. 968-998.
- [CDS01] Chen, S. S., Donoho, D. L., and Saunders, M. A., 2001. “Atomic Decomposition by Basis Pursuit,” SIAM Review, Vol. 43, pp. 129-159.
- [CFM75] Camerini, P. M., Fratta, L., and Maffioli, F., 1975. “On Improving Relaxation Methods by Modified Gradient Techniques,” Math. Programming Studies, Vol. 3, pp. 26-34.
- [CGT00] Conn, A. R., Gould, N. I., and Toint, P. L., 2000. Trust Region Methods, SIAM, Philadelphia, PA.
- [CHY13] Chen, C., He, B., Ye, Y., and Yuan, X., 2013. “The Direct Extension of ADMM for Multi-Block Convex Minimization Problems is not Necessarily Convergent,” Optimization Online.
- [CPR14] Chouzenoux, E., Pesquet, J. C., and Repetti, A., 2014. “Variable Metric Forward-Backward Algorithm for Minimizing the Sum of a Differentiable Function and a Convex Function,” J. of Optimization Theory and Applications, Vol. 162, pp. 107-132.
- [CPS92] Cottle, R. W., Pang, J.-S., and Stone, R. E., 1992. The Linear Complementarity Problem, Academic Press, NY.
- [CRP12] Chandrasekaran, V., Recht, B., Parrilo, P. A., and Willsky, A. S., 2012. “The Convex Geometry of Linear Inverse Problems,” Foundations of Computational Mathematics, Vol. 12, pp. 805-849.
- [CaC68] Canon, M. D., and Cullum, C. D., 1968. “A Tight Upper Bound on the Rate of Convergence of the Frank-Wolfe Algorithm,” SIAM J. on Control, Vol. 6, pp. 509-516.
- [CaG74] Cantor, D. G., Gerla, M., 1974. “Optimal Routing in Packet

- Switched Computer Networks,” *IEEE Trans. on Computing*, Vol. C-23, pp. 1062-1068.
- [CaR09] Candés, E. J., and Recht, B., 2009. “Exact Matrix Completion via Convex Optimization,” *Foundations of Computational Math.*, Vol. 9, pp. 717-772.
- [CaT10] Candés, E. J., and Tao, T., 2010. “The Power of Convex Relaxation: Near-Optimal Matrix Completion,” *IEEE Trans. on Information Theory*, Vol. 56, pp. 2053-2080.
- [CeH87] Censor, Y., and Herman, G. T., 1987. “On Some Optimization Techniques in Image Reconstruction from Projections,” *Applied Numer. Math.*, Vol. 3, pp. 365-391.
- [CeS08] Cegielski, A., and Suchocka, A., 2008. “Relaxed Alternating Projection Methods,” *SIAM J. Optimization*, Vol. 19, pp. 1093-1106.
- [CeZ92] Censor, Y., and Zenios, S. A., 1992. “The Proximal Minimization Algorithm with D-Functions,” *J. Opt. Theory and Appl.*, Vol. 73, pp. 451-464.
- [CeZ97] Censor, Y., and Zenios, S. A., 1997. *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford Univ. Press, NY.
- [ChG59] Cheney, E. W., and Goldstein, A. A., 1959. “Newton’s Method for Convex Programming and Tchebycheff Approximation,” *Numer. Math.*, Vol. I, pp. 253-268.
- [ChR97] Chen, G. H., and Rockafellar, R. T., 1997. “Convergence Rates in Forward-Backward Splitting,” *SIAM J. on Optimization*, Vol. 7, pp. 421-444.
- [ChT93] Chen, G., and Teboulle, M., 1993. “Convergence Analysis of a Proximal-Like Minimization Algorithm Using Bregman Functions,” *SIAM J. on Optimization*, Vol. 3, pp. 538-543.
- [ChT94] Chen, G., and Teboulle, M., 1994. “A Proximal-Based Decomposition Method for Convex Minimization Problems,” *Mathematical Programming*, Vol. 64, pp. 81-101.
- [Cha04] Chambolle, A., 2004. “An Algorithm for Total Variation Minimization and Applications,” *J. of Mathematical Imaging and Vision*, Vol. 20, pp. 89-97.
- [Che07] Chen, Y., 2007. “A Smoothing Inexact Newton Method for Minimax Problems,” *Advances in Theoretical and Applied Mathematics*, Vol. 2, pp. 137-143.
- [Cla10] Clarkson, K. L., 2010. “Coresets, Sparse Greedy Approximation, and the Frank-Wolfe Algorithm,” *ACM Transactions on Algorithms*, Vol. 6, pp. 63.
- [CoT13] Couellan, N. P., and Trafalis, T. B., 2013. “On-line SVM Learning via an Incremental Primal-Dual Technique,” *Optimization Methods and Software*, Vol. 28, pp. 256-275.

- [CoV13] Combettes, P. L., and Vu, B. C., 2013. “Variable Metric Quasi-Fejér Monotonicity,” *Nonlinear Analysis: Theory, Methods and Applications*, Vol. 78, pp. 17-31.
- [Com01] Combettes, P. L., 2001. “Quasi-Fejérian Analysis of Some Optimization Algorithms,” *Studies in Computational Mathematics*, Vol. 8, pp. 115-152.
- [Cry71] Cryer, C. W., 1971. “The Solution of a Quadratic Programming Problem Using Systematic Overrelaxation,” *SIAM J. on Control*, Vol. 9, pp. 385-392.
- [DBW12] Duchi, J. C., Bartlett, P. L., and Wainwright, M. J., 2012. “Randomized Smoothing for Stochastic Optimization,” *SIAM J. on Optimization*, Vol. 22, pp. 674-701.
- [DCD14] Defazio, A. J., Caetano, T. S., and Domke, J., 2014. “Finito: A Faster, Permutable Incremental Gradient Method for Big Data Problems,” *Proceedings of the 31st ICML*, Beijing.
- [DHS06] Dai, Y. H., Hager, W. W., Schittkowski, K., and Zhang, H., 2006. “The Cyclic Barzilai-Borwein Method for Unconstrained Optimization,” *IMA J. of Numerical Analysis*, Vol. 26, pp. 604-627.
- [DHS11] Duchi, J., Hazan, E., and Singer, Y., 2011. “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” *J. of Machine Learning Research*, Vol. 12, pp. 2121-2159.
- [DMM06] Drineas, P., Mahoney, M. W., and Muthukrishnan, S., 2006. “Sampling Algorithms for L2 Regression and Applications,” *Proc. 17th Annual SODA*, pp. 1127-1136.
- [DMM11] Drineas, P., Mahoney, M. W., Muthukrishnan, S., and Sarlos, T., 2011. “Faster Least Squares Approximation,” *Numerische Mathematik*, Vol. 117, pp. 219-249.
- [DRT11] Dhillon, I. S., Ravikumar, P., and Tewari, A., 2011. “Nearest Neighbor Based Greedy Coordinate Descent,” in *Advances in Neural Information Processing Systems 24*, (NIPS 2011), pp. 2160-2168.
- [DaW60] Dantzig, G. B., and Wolfe, P., 1960. “Decomposition Principle for Linear Programs,” *Operations Research*, Vol. 8, pp. 101-111.
- [DaY14a] Davis, D., and Yin, W., 2014. “Convergence Rate Analysis of Several Splitting Schemes,” *arXiv preprint arXiv:1406.4834*.
- [DaY14b] Davis, D., and Yin, W., 2014. “Convergence Rates of Relaxed Peaceman-Rachford and ADMM Under Regularity Assumptions,” *arXiv preprint arXiv:1407.5210*.
- [Dan67] Danskin, J. M., 1967. *The Theory of Max-Min and its Application to Weapons Allocation Problems*, Springer, NY.
- [Dav76] Davidon, W. C., 1976. “New Least Squares Algorithms,” *J. Optimization Theory and Applications*, Vol. 18, pp. 187-197.

- [DeM71] Demjanov, V. F., and Malozemov, V. N., 1971. "On the Theory of Non-Linear Minimax Problems," *Russian Math. Surveys*, Vol. 26, p. 57.
- [DeR70] Demjanov, V. F., and Rubinov, A. M., 1970. *Approximate Methods in Optimization Problems*, American Elsevier, NY.
- [DeS96] Dennis, J. E., and Schnabel, R. B., 1996. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, SIAM, Philadelphia, PA.
- [DeT91] Dennis, J. E., and Torczon, V., 1991. "Direct Search Methods on Parallel Machines," *SIAM J. on Optimization*, Vol. 1, pp. 448-474.
- [Dem66] Demjanov, V. F., 1966. "The Solution of Several Minimax Problems," *Kibernetika*, Vol. 2, pp. 58-66.
- [Dem68] Demjanov, V. F., 1968. "Algorithms for Some Minimax Problems," *J. of Computer and Systems Science*, Vol. 2, pp. 342-380.
- [DoE03] Donoho, D. L., Elad, M., 2003. "Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via ℓ_1 Minimization," *Proc. of the National Academy of Sciences*, Vol. 100, pp. 2197-2202.
- [DrH04] Drezner, Z., and Hamacher, H. W., 2004. *Facility Location: Applications and Theory*, Springer, NY.
- [DuS83] Dunn, J. C., and Sachs, E., 1983. "The Effect of Perturbations on the Convergence Rates of Optimization Algorithms," *Appl. Math. Optim.*, Vol. 10, pp. 143-157.
- [DuS09] Duchi, J., and Singer, Y., 2009. "Efficient Online and Batch Learning Using Forward Backward Splitting," *J. of Machine Learning Research*, Vol. 10, pp. 2899-2934.
- [Dun79] Dunn, J. C., 1979. "Rates of Convergence for Conditional Gradient Algorithms Near Singular and Nonsingular Extremals," *SIAM J. on Control and Optimization*, Vol. 17, pp. 187-211.
- [Dun80] Dunn, J. C., 1980. "Convergence Rates for Conditional Gradient Sequences Generated by Implicit Step Length Rules," *SIAM J. on Control and Optimization*, Vol. 18, pp. 473-487.
- [Dun81] Dunn, J. C., 1981. "Global and Asymptotic Convergence Rate Estimates for a Class of Projected Gradient Processes," *SIAM J. on Control and Optimization*, Vol. 19, pp. 368-400.
- [Dun87] Dunn, J. C., 1987. "On the Convergence of Projected Gradient Processes to Singular Critical Points," *J. of Optimization Theory and Applications*, Vol. 55, pp. 203-216.
- [Dun91] Dunn, J. C., 1991. "A Subspace Decomposition Principle for Scaled Gradient Projection Methods: Global Theory," *SIAM J. on Control and Optimization*, Vol. 29, pp. 219-246.
- [EcB92] Eckstein, J., and Bertsekas, D. P., 1992. "On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators," *Math. Programming*, Vol. 55, pp. 293-318.

- [EcS13] Eckstein, J., and Silva, P. J. S., 2013. "A Practical Relative Error Criterion for Augmented Lagrangians," *Math. Programming*, Vol. 141, Ser. A, pp. 319-348.
- [Eck94] Eckstein, J., 1994. "Nonlinear Proximal Point Algorithms Using Bregman Functions, with Applications to Convex Programming," *Math. of Operations Research*, Vol. 18, pp. 202-226.
- [Eck03] Eckstein, J., 2003. "A Practical General Approximation Criterion for Methods of Multipliers Based on Bregman Distances," *Math. Programming*, Vol. 96, Ser. A, pp. 61-86.
- [Eck12] Eckstein, J., 2012. "Augmented Lagrangian and Alternating Direction Methods for Convex Optimization: A Tutorial and Some Illustrative Computational Results," RUTCOR Research Report RRR 32-2012, Rutgers, Univ.
- [EkT76] Ekeland, I., and Temam, R., 1976. *Convex Analysis and Variational Problems*, North-Holland Publ., Amsterdam.
- [ElM75] Elzinga, J., and Moore, T. G., 1975. "A Central Cutting Plane Algorithm for the Convex Programming Problem," *Math. Programming*, Vol. 8, pp. 134-145.
- [Erm76] Ermoliev, Yu. M., 1976. *Stochastic Programming Methods*, Nauka, Moscow.
- [Eve63] Everett, H., 1963. "Generalized Lagrange Multiplier Method for Solving Problems of Optimal Allocation of Resources," *Operations Research*, Vol. 11, pp. 399-417.
- [FGW02] Forsgren, A., Gill, P. E., and Wright, M. H., 2002. "Interior Methods for Nonlinear Optimization," *SIAM Review*, Vol. 44, pp. 525-597.
- [FHT10] Friedman, J., Hastie, T., and Tibshirani, R., 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent," *J. of Statistical Software*, Vol. 33, pp. 1-22.
- [FJS98] Facchinei, F., Judice, J., and Soares, J., 1998. "An Active Set Newton Algorithm for Large-Scale Nonlinear Programs with Box Constraints," *SIAM J. on Optimization*, Vol. 8, pp. 158-186.
- [FLP02] Facchinei, F., Lucidi, S., and Palagi, L., 2002. "A Truncated Newton Algorithm for Large Scale Box Constrained Optimization," *SIAM J. on Optimization*, Vol. 12, pp. 1100-1125.
- [FLT02] Fukushima, M., Luo, Z.-Q., and Tseng, P., 2002. "Smoothing Functions for Second-Order-Cone Complementarity Problems," *SIAM J. Optimization*, Vol. 12, pp. 436-460.
- [FaP03] Facchinei, F., and Pang, J.-S., 2003. *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer Verlag, NY.
- [Fab73] Fabian, V., 1973. "Asymptotically Efficient Stochastic Approximation: The RM Case," *Ann. Statist.*, Vol. 1, pp. 486-495.

- [FeM91] Ferris, M. C., and Mangasarian, O. L., 1991. "Finite Perturbation of Convex Programs," *Appl. Math. Optim.*, Vol. 23, pp. 263-273.
- [FeM02] Ferris, M. C., and Munson, T. S., 2002. "Interior-Point Methods for Massive Support Vector Machines," *SIAM J. on Optimization*, Vol. 13, pp. 783-804.
- [FeR12] Fercoq, O., and Richtarik, P., 2012. "Accelerated, Parallel, and Proximal Coordinate Descent," arXiv preprint arXiv:1312.5799.
- [Fen51] Fenchel, W., 1951. *Convex Cones, Sets, and Functions*, Mimeographed Notes, Princeton Univ.
- [FIH95] Florian, M. S., and Hearn, D., 1995. "Network Equilibrium Models and Algorithms," *Handbooks in OR and MS*, Ball, M. O., Magnanti, T. L., Monma, C. L., and Nemhauser, G. L., (Eds.), Vol. 8, North-Holland, Amsterdam, pp. 485-550.
- [FiM68] Fiacco, A. V., and McCormick, G. P., 1968. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, NY.
- [FiN03] Figueiredo, M. A. T., and Nowak, R. D., 2003. "An EM Algorithm for Wavelet-Based Image Restoration," *IEEE Trans. Image Processing*, Vol. 12, pp. 906-916.
- [Fle00] Fletcher, R., 2000. *Practical Methods of Optimization*, 2nd edition, Wiley, NY.
- [FoG83] Fortin, M., and Glowinski, R., 1983. "On Decomposition-Coordination Methods Using an Augmented Lagrangian," in: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, North-Holland, Amsterdam.
- [FrG13] Friedlander, M. P., and Goh, G., 2013. "Tail Bounds for Stochastic Approximation," arXiv preprint arXiv:1304.5586.
- [FrG14] Freund, R. M., and Grigas, P., 2014. "New Analysis and Results for the Frank-Wolfe Method," arXiv preprint arXiv:1307.0873, to appear in *Math. Programming*.
- [FrS00] Frommer, A., and Szyld, D. B., 2000. "On Asynchronous Iterations," *J. of Computational and Applied Mathematics*, Vol. 123, pp. 201-216.
- [FrS12] Friedlander, M. P., and Schmidt, M., 2012. "Hybrid Deterministic-Stochastic Methods for Data Fitting," *SIAM J. Sci. Comput.*, Vol. 34, pp. A1380-A1405.
- [FrT07] Friedlander, M. P., and Tseng, P., 2007. "Exact Regularization of Convex Programs," *SIAM J. on Optimization*, Vol. 18, pp. 1326-1350.
- [FrW56] Frank, M., and Wolfe, P., 1956. "An Algorithm for Quadratic Programming," *Naval Research Logistics Quarterly*, Vol. 3, pp. 95-110.
- [Fra02] Frangioni, A., 2002. "Generalized Bundle Methods," *SIAM J. on Optimization*, Vol. 13, pp. 117-156.
- [Fri56] Frisch, M. R., 1956. "La Resolution des Problemes de Programme

- Lineaire par la Methode du Potential Logarithmique,” *Cahiers du Seminaire D’Econometrie*, Vol. 4, pp. 7-20.
- [FuM81] Fukushima, M., and Mine, H., 1981. “A Generalized Proximal Point Algorithm for Certain Non-Convex Minimization Problems,” *Internat. J. Systems Sci.*, Vol. 12, pp. 989-1000.
- [Fuk92] Fukushima, M., 1992. “Application of the Alternating Direction Method of Multipliers to Separable Convex Programming Problems,” *Computational Optimization and Applications*, Vol. 1, pp. 93-111.
- [GBY12] Grant, M., Boyd, S., and Ye, Y., 2012. “CVX: Matlab Software for Disciplined Convex Programming, Version 2.0 Beta,” *Recent Advances in Learning and Control*, cvx.com.
- [GGM06] Gaudioso, M., Giallombardo, G., and Miglionico, G., 2006. “An Incremental Method for Solving Convex Finite Min-Max Problems,” *Math. of Operations Research*, Vol. 31, pp. 173-187.
- [GHV92] Goffin, J. L., Haurie, A., and Vial, J. P., 1992. “Decomposition and Nondifferentiable Optimization with the Projective Algorithm,” *Management Science*, Vol. 38, pp. 284-302.
- [GKX10] Gupta, M. D., Kumar, S., and Xiao, J. 2010. “L1 Projections with Box Constraints,” *arXiv preprint arXiv:1010.0141*.
- [GLL86] Grippo, L., Lampariello, F., and Lucidi, S., 1986. “A Nonmonotone Line Search Technique for Newton’s Method,” *SIAM J. on Numerical Analysis*, Vol. 23, pp. 707-716.
- [GLY94] Goffin, J. L., Luo, Z.-Q., and Ye, Y., 1994. “On the Complexity of a Column Generation Algorithm for Convex or Quasiconvex Feasibility Problems,” in *Large Scale Optimization: State of the Art*, Hager, W. W., Hearn, D. W., and Pardalos, P. M., (Eds.), Kluwer, Boston.
- [GLY96] Goffin, J. L., Luo, Z.-Q., and Ye, Y., 1996. “Complexity Analysis of an Interior Cutting Plane Method for Convex Feasibility Problems,” *SIAM J. on Optimization*, Vol. 6, pp. 638-652.
- [GMW81] Gill, P. E., Murray, W., and Wright, M. H., 1981. *Practical Optimization*, Academic Press, NY.
- [GNS08] Griva, I., Nash, S. G., and Sofer, A., 2008. *Linear and Nonlinear Optimization*, 2nd Edition, SIAM, Philadelphia, PA.
- [GOP14] Gurbuzbalaban, M., Ozdaglar, A., and Parrilo, P., 2014. “A Globally Convergent Incremental Newton Method,” *arXiv preprint arXiv:1410.5284*.
- [GPR67] Gubin, L. G., Polyak, B. T., and Raik, E. V., 1967. “The Method of Projection for Finding the Common Point in Convex Sets,” *USSR Comput. Math. Phys.*, Vol. 7, pp. 1-24.
- [GSW12] Gamarnik, D., Shah, D., and Wei, Y., 2012. “Belief Propagation for Min-Cost Network Flow: Convergence and Correctness,” *Operations Research*, Vol. 60, pp. 410-428.

- [GaB84] Gafni, E. M., and Bertsekas, D. P., 1984. "Two-Metric Projection Methods for Constrained Optimization," *SIAM J. on Control and Optimization*, Vol. 22, pp. 936-964.
- [GaM76] Gabay, D., and Mercier, B., 1976. "A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite-Element Approximations," *Comp. Math. Appl.*, Vol. 2, pp. 17-40.
- [Gab79] Gabay, D., 1979. *Méthodes Numeriques pour l'Optimization Non Lineaire*, These de Doctorat d'Etat et Sciences Mathematiques, Univ. Pierre at Marie Curie (Paris VI).
- [Gab83] Gabay, D., 1983. "Applications of the Method of Multipliers to Variational Inequalities," in M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, North-Holland, Amsterdam.
- [GeM05] Gerlach, S., and Matzenmiller, A., 2005. "Comparison of Numerical Methods for Identification of Viscoelastic Line Spectra from Static Test Data," *International J. for Numerical Methods in Engineering*, Vol. 63, pp. 428-454.
- [Geo72] Geoffrion, A. M., 1972. "Generalized Benders Decomposition," *J. of Optimization Theory and Applications*, Vol. 10, pp. 237-260.
- [Geo77] Geoffrion, A. M., 1977. "Objective Function Approximations in Mathematical Programming," *Math. Programming*, Vol. 13, pp. 23-27.
- [GiB14] Giselsson, P., and Boyd, S., 2014. "Metric Selection in Douglas-Rachford Splitting and ADMM," arXiv preprint arXiv:1410.8479.
- [GiM74] Gill, P. E., and Murray, W., (Eds.), 1974. *Numerical Methods for Constrained Optimization*, Academic Press, NY.
- [GIM75] Glowinski, R. and Marrocco, A., 1975. "Sur l' Approximation par Elements Finis d' Ordre un et la Resolution par Penalisation-Dualite d'une Classe de Problemes de Dirichlet Non Lineaires" *Revue Francaise d'Automatique Informatique Recherche Operationnelle, Analyse Numerique*, R-2, pp. 41-76.
- [GoK13] Gonzaga, C. C., and Karas, E. W., 2013. "Fine Tuning Nesterov's Steepest Descent Algorithm for Differentiable Convex Programming," *Math. Programming*, Vol. 138, pp. 141-166.
- [GoO09] Goldstein, T., and Osher, S., 2009. "The Split Bregman Method for L1-Regularized Problems," *SIAM J. on Imaging Sciences*, Vol. 2, pp. 323-343.
- [GoS10] Goldstein, T., and Setzer, S., 2010. "High-Order Methods for Basis Pursuit," *UCLA CAM Report*, 10-41.
- [GoV90] Goffin, J. L., and Vial, J. P., 1990. "Cutting Planes and Column Generation Techniques with the Projective Algorithm," *J. Opt. Th. and Appl.*, Vol. 65, pp. 409-429.
- [GoV02] Goffin, J. L., and Vial, J. P., 2002. "Convex Nondifferentiable

- Optimization: A Survey Focussed on the Analytic Center Cutting Plane Method,” *Optimization Methods and Software*, Vol. 17, pp. 805-867.
- [GoZ12] Gong, P., and Zhang, C., 2012. “Efficient Nonnegative Matrix Factorization via Projected Newton Method,” *Pattern Recognition*, Vol. 45, pp. 3557-3565.
- [Gol64] Goldstein, A. A., 1964. “Convex Programming in Hilbert Space,” *Bull. Amer. Math. Soc.*, Vol. 70, pp. 709-710.
- [Gol85] Golshtein, E. G., 1985. “A Decomposition Method for Linear and Convex Programming Problems,” *Matecon*, Vol. 21, pp. 1077-1091.
- [Gon00] Gonzaga, C. C., 2000. “Two Facts on the Convergence of the Cauchy Algorithm,” *J. of Optimization Theory and Applications*, Vol. 107, pp. 591-600.
- [GrS99] Grippo, L., and Sciandrone, M., 1999. “Globally Convergent Block-Coordinate Techniques for Unconstrained Optimization,” *Optimization Methods and Software*, Vol. 10, pp. 587-637.
- [GrS00] Grippo, L., and Sciandrone, M., 2000. “On the Convergence of the Block Nonlinear Gauss-Seidel Method Under Convex Constraints,” *Operations Research Letters*, Vol. 26, pp. 127-136.
- [Gri94] Grippo, L., 1994. “A Class of Unconstrained Minimization Methods for Neural Network Training,” *Optim. Methods and Software*, Vol. 4, pp. 135-150.
- [Gri00] Grippo, L., 2000. “Convergent On-Line Algorithms for Supervised Learning in Neural Networks,” *IEEE Trans. Neural Networks*, Vol. 11, pp. 1284-1299.
- [Gul92] Guler, O., 1992. “New Proximal Point Algorithms for Convex Minimization,” *SIAM J. on Optimization*, Vol. 2, pp. 649-664.
- [HCW14] Hong, M., Chang, T. H., Wang, X., Razaviyayn, M., Ma, S., and Luo, Z. Q., 2014. “A Block Successive Upper Bound Minimization Method of Multipliers for Linearly Constrained Convex Optimization,” *arXiv preprint arXiv:1401.7079*.
- [HJN14] Harchaoui, Z., Juditsky, A., and Nemirovski, A., 2014. “Conditional Gradient Algorithms for Norm-Regularized Smooth Convex Optimization,” *Math. Programming*, pp. 1-38.
- [HKR95] den Hertog, D., Kaliski, J., Roos, C., and Terlaky, T., 1995. “A Path-Following Cutting Plane Method for Convex Programming,” *Annals of Operations Research*, Vol. 58, pp. 69-98.
- [HLV87] Hearn, D. W., Lawphongpanich, S., and Ventura, J. A., 1987. “Restricted Simplicial Decomposition: Computation and Extensions,” *Math. Programming Studies*, Vol. 31, pp. 119-136.
- [HMT10] Halko, N., Martinsson, P.-G., and Tropp, J. A., 2010. “Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions,” *arXiv preprint arXiv:0909.4061*.

- [HTF09] Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition, Springer, NY. On line at <http://statweb.stanford.edu/tibs/ElemStatLearn/>
- [HYS15] Hu, Y., Yang, X., and Sim, C. K., 2015. "Inexact Subgradient Methods for Quasi-Convex Optimization Problems," *European Journal of Operational Research*, Vol. 240, pp. 315-327.
- [HYS13] Hou, K., Zhou, Z., So, A. M. C., and Luo, Z. Q., 2013. "On the Linear Convergence of the Proximal Gradient Method for Trace Norm Regularization," in *Advances in Neural Information Processing Systems (NIPS 2013)*, pp. 710-718.
- [Ha90] Ha, C. D., 1990. "A Generalization of the Proximal Point Algorithm," *SIAM J. on Control and Optimization*, Vol. 28, pp. 503-512.
- [HaB70] Haarhoff, P. C., and Buys, J. D., 1970. "A New Method for the Optimization of a Nonlinear Function Subject to Nonlinear Constraints," *Computer J.*, Vol. 13, pp. 178-184.
- [HaH93] Hager, W. W., and Hearn, D. W., 1993. "Application of the Dual Active Set Algorithm to Quadratic Network Optimization," *Computational Optimization and Applications*, Vol. 1, pp. 349-373.
- [HaM79] Han, S. P., and Mangasarian, O. L., 1979. "Exact Penalty Functions in Nonlinear Programming," *Math. Programming*, Vol. 17, pp. 251-269.
- [Hay08] Haykin, S., 2008. *Neural Networks and Learning Machines*, (3rd Ed.), Prentice Hall, Englewood Cliffs, NJ.
- [HeD09] Helou, E. S., and De Pierro, A. R., 2009. "Incremental Subgradients for Constrained Convex Optimization: A Unified Framework and New Methods," *SIAM J. on Optimization*, Vol. 20, pp. 1547-1572.
- [HeL89] Hearn, D. W., and Lawphongpanich, S., 1989. "Lagrangian Dual Ascent by Generalized Linear Programming," *Operations Res. Letters*, Vol. 8, pp. 189-196.
- [HeM11] Henrion, D., and Malick, J., 2011. "Projection Methods for Conic Feasibility Problems: Applications to Polynomial Sum-of-Squares Decompositions," *Optimization Methods and Software*, Vol. 26, pp. 23-46.
- [HeM12] Henrion, D., and Malick, J., 2012. "Projection Methods in Conic Optimization," In *Handbook on Semidefinite, Conic and Polynomial Optimization*, Springer, NY, pp. 565-600.
- [Her09] Gabor, H., 2009. *Fundamentals of Computerized Tomography: Image Reconstruction from Projection*, (2nd ed.), Springer, NY.
- [Hes69] Hestenes, M. R., 1969. "Multiplier and Gradient Methods," *J. Opt. Th. and Appl.*, Vol. 4, pp. 303-320.
- [Hes75] Hestenes, M. R., 1975. *Optimization Theory: The Finite Dimensional Case*, Wiley, NY.

- [Hil93] Hiriart-Urruty, J.-B., and Lemarechal, C., 1993. *Convex Analysis and Minimization Algorithms*, Vols. I and II, Springer-Verlag, Berlin and NY.
- [Hil57] Hildreth, C., 1957. "A Quadratic Programming Procedure," *Naval Res. Logist. Quart.*, Vol. 4, pp. 79-85. See also "Erratum," *Naval Res. Logist. Quart.*, Vol. 4, p. 361.
- [HoK71] Hoffman, K., and Kunze, R., 1971. *Linear Algebra*, Pearson, Englewood Cliffs, NJ.
- [HoL13] Hong, M., and Luo, Z. Q., 2013. "On the Linear Convergence of the Alternating Direction Method of Multipliers," arXiv preprint arXiv:1208.3922.
- [Hoh77] Hohenbalken, B. von, 1977. "Simplicial Decomposition in Nonlinear Programming," *Math. Programming*, Vol. 13, pp. 49-68.
- [Hol74] Holloway, C. A., 1974. "An Extension of the Frank and Wolfe Method of Feasible Directions," *Math. Programming*, Vol. 6, pp. 14-27.
- [IPS03] Iusem, A. N., Pennanen, T., and Svaiter, B. F., 2003. "Inexact Variants of the Proximal Point Algorithm Without Monotonicity," *SIAM J. on Optimization*, Vol. 13, pp. 1080-1097.
- [IST94] Iusem, A. N., Svaiter, B. F., and Teboulle, M., 1994. "Entropy-Like Proximal Methods in Convex Programming," *Math. of Operations Research*, Vol. 19, pp. 790-814.
- [IbF96] Ibaraki, S., and Fukushima, M., 1996. "Partial Proximal Method of Multipliers for Convex Programming Problems," *J. of Operations Research Society of Japan*, Vol. 39, pp. 213-229.
- [IuT95] Iusem, A. N., and Teboulle, M., 1995. "Convergence Rate Analysis of Nonquadratic Proximal Methods for Convex and Linear Programming," *Math. of Operations Research*, Vol. 20, pp. 657-677.
- [Ius99] Iusem, A. N., 1999. "Augmented Lagrangian Methods and Proximal Point Methods for Convex Minimization," *Investigacion Operativa*, Vol. 8, pp. 11-49.
- [Ius03] Iusem, A. N., 2003. "On the Convergence Properties of the Projected Gradient Method for Convex Optimization," *Computational and Applied Mathematics*, Vol. 22, pp. 37-52.
- [JFY09] Joachims, T., Finley, T., and Yu, C.-N. J., 2009. "Cutting-Plane Training of Structural SVMs," *Machine Learning*, Vol. 77, pp. 27-59.
- [JRJ09] Johansson, B., Rabi, M., and Johansson, M., 2009. "A Randomized Incremental Subgradient Method for Distributed Optimization in Networked Systems," *SIAM J. on Optimization*, Vol. 20, pp. 1157-1170.
- [Jag13] Jaggi, M., 2013. "Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization," *Proc. of ICML 2013*.
- [JiZ14] Jiang, B., and Zhang, S., 2014. "Iteration Bounds for Finding the ϵ -Stationary Points for Structured Nonconvex Optimization," arXiv preprint

arXiv:1410.4066.

[JoY09] Joachims, T., and Yu, C.-N. J., 2009. “Sparse Kernel SVMs via Cutting-Plane Training,” *Machine Learning*, Vol. 76, pp. 179-193.

[JoZ13] Johnson, R., and Zhang, T., 2013. “Accelerating Stochastic Gradient Descent Using Predictive Variance Reduction,” *Advances in Neural Information Processing Systems 26 (NIPS 2013)*.

[Joa06] Joachims, T., 2006. “Training Linear SVMs in Linear Time,” *International Conference on Knowledge Discovery and Data Mining*, pp. 217-226.

[JuN11a] Juditsky, A., and Nemirovski, A., 2011. “First Order Methods for Nonsmooth Convex Large-Scale Optimization, I: General Purpose Methods,” in *Optimization for Machine Learning*, by Sra, S., Nowozin, S., and Wright, S. J. (eds.), MIT Press, Cambridge, MA, pp. 121-148.

[JuN11b] Juditsky, A., and Nemirovski, A., 2011. “First Order Methods for Nonsmooth Convex Large-Scale Optimization, II: Utilizing Problem’s Structure,” in *Optimization for Machine Learning*, by Sra, S., Nowozin, S., and Wright, S. J. (eds.), MIT Press, Cambridge, MA, pp. 149-183.

[KaW94] Kall, P., and Wallace, S. W., 1994. *Stochastic Programming*, Wiley, Chichester, UK.

[Kac37] Kaczmarz, S., 1937. “Approximate Solution of Systems of Linear Equations,” *Bull. Acad. Pol. Sci., Lett. A* 35, pp. 335-357 (in German); English transl.: *Int. J. Control*, Vol. 57, pp. 1269-1271, 1993.

[Kar84] Karmarkar, N., 1984. “A New Polynomial-Time Algorithm for Linear Programming,” In *Proc. of the 16th Annual ACM Symp. on Theory of Computing*, pp. 302-311.

[Kel60] Kelley, J. E., 1960. “The Cutting-Plane Method for Solving Convex Programs,” *J. Soc. Indust. Appl. Math.*, Vol. 8, pp. 703-712.

[Kel99] Kelley, C. T., 1999. *Iterative Methods for Optimization*, Siam, Philadelphia, PA.

[Kib80] Kibardin, V. M., 1980. “Decomposition into Functions in the Minimization Problem,” *Automation and Remote Control*, Vol. 40, pp. 1311-1323.

[Kiw04] Kiwiel, K. C., 2004. “Convergence of Approximate and Incremental Subgradient Methods for Convex Optimization,” *SIAM J. on Optimization*, Vol. 14, pp. 807-840.

[KoB72] Kort, B. W., and Bertsekas, D. P., 1972. “A New Penalty Function Method for Constrained Minimization,” *Proc. 1972 IEEE Confer. Decision Control*, New Orleans, LA, pp. 162-166.

[KoB76] Kort, B. W., and Bertsekas, D. P., 1976. “Combined Primal-Dual and Penalty Methods for Convex Programming,” *SIAM J. on Control and Optimization*, Vol. 14, pp. 268-294.

[KoN93] Kortanek, K. O., and No, H., 1993. “A Central Cutting Plane

- Algorithm for Convex Semi-Infinite Programming Problems,” *SIAM J. on Optimization*, Vol. 3, pp. 901-918.
- [Kor75] Kort, B. W., 1975. “Combined Primal-Dual and Penalty Function Algorithms for Nonlinear Programming,” Ph.D. Thesis, Dept. of Engineering-Economic Systems, Stanford Univ., Stanford, Ca.
- [Kra55] Krasnosel’skii, M. A., 1955. “Two Remarks on the Method of Successive Approximations,” *Uspehi Mat. Nauk*, Vol. 10, pp. 123-127.
- [KuC78] Kushner, H. J., and Clark, D. S., 1978. *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, NY.
- [KuY03] Kushner, H. J., and Yin, G., 2003. *Stochastic Approximation and Recursive Algorithms and Applications*, Springer-Verlag, NY.
- [LBB98] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., 1998. “Gradient-Based Learning Applied to Document Recognition,” *Proceedings of the IEEE*, Vol. 86, pp. 2278-2324.
- [LJS12] Lacoste-Julien, S., Jaggi, M., Schmidt, M., and Pletscher, P., 2012. “Block-Coordinate Frank-Wolfe Optimization for Structural SVMs,” arXiv preprint arXiv:1207.4747.
- [LLS12] Lee, J., Sun, Y., and Saunders, M., 2012. “Proximal Newton-Type Methods for Convex Optimization,” *NIPS 2012*.
- [LLS14] Lee, J., Sun, Y., and Saunders, M., 2014. “Proximal Newton-Type Methods for Minimizing Composite Functions,” arXiv preprint arXiv:1206.1623.
- [LLX14] Lin, Q., Lu, Z., and Xiao, L., 2014. “An Accelerated Proximal Coordinate Gradient Method and its Application to Regularized Empirical Risk Minimization,” arXiv preprint arXiv:1407.1296.
- [LLZ09] Langford, J., Li, L., and Zhang, T., 2009. “Sparse Online Learning via Truncated Gradient,” In *Advances in Neural Information Processing Systems (NIPS 2009)*, pp. 905-912.
- [LMS92] Lustig, I. J., Marsten, R. E., and Shanno, D. F., 1992. “On Implementing Mehrotra’s Predictor-Corrector Interior-Point Method for Linear Programming,” *SIAM J. on Optimization*, Vol. 2, pp. 435-449.
- [LMY12] Lu, Z., Monteiro, R. D. C., and Yuan, M., 2012. “Convex Optimization Methods for Dimension Reduction and Coefficient Estimation in Multivariate Linear Regression,” *Mathematical Programming*, Vol. 131, pp. 163-194.
- [LPS98] Larsson, T., Patriksson, M., and Stromberg, A.-B., 1998. “Ergodic Convergence in Subgradient Optimization,” *Optimization Methods and Software*, Vol. 9, pp. 93-120.
- [LRW98] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E., 1998. “Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions,” *SIAM J. on Optimization*, Vol. 9, pp. 112-147.

- [LVB98] Lobo, M. S., Vandenberghe, L., Boyd, S., and Lebret, H., 1998. “Applications of Second-Order Cone Programming,” *Linear Algebra and Applications*, Vol. 284, pp. 193-228.
- [LWS14] Liu, J., Wright, S. J., and Sridhar, S., 2014. “An Asynchronous Parallel Randomized Kaczmarz Algorithm,” Univ. of Wisconsin Report, arXiv preprint arXiv:1401.4780.
- [LaD60] Land, A. H., and Doig, A. G., 1960. “An Automatic Method for Solving Discrete Programming Problems,” *Econometrica*, Vol. 28, pp. 497-520.
- [LaS87] Lawrence, J., and Spingarn, J. E., 1987. “On Fixed Points of Non-expansive Piecewise Isometric Mappings,” *Proc. London Math. Soc.*, Vol. 55, pp. 605-624.
- [LaT85] Lancaster, P., and Tismenetsky, M., 1985. *The Theory of Matrices*, Academic Press, NY.
- [Lan14] Landi, G., 2014. “A Modified Newton Projection Method for ℓ_1 -Regularized Least Squares Image Deblurring,” *J. of Mathematical Imaging and Vision*, pp. 1-14.
- [LeL10] Leventhal, D., and Lewis, A. S., 2010. “Randomized Methods for Linear Constraints: Convergence Rates and Conditioning,” *Math. of Operations Research*, Vol. 35, pp. 641-654.
- [LeP65] Levitin, E. S., and Poljak, B. T., 1965. “Constrained Minimization Methods,” *Ž. Vychisl. Mat. i Mat. Fiz.*, Vol. 6, pp. 787-823.
- [LeS93] Lemaréchal, C., and Sagastizábal, C., 1993. “An Approach to Variable Metric Bundle Methods,” in *Systems Modelling and Optimization*, Proc. of the 16th IFIP-TC7 Conference, Compiègne, Henry, J., and Yvon, J.-P., (Eds.), *Lecture Notes in Control and Information Sciences* 197, pp. 144-162.
- [LeS99] Lee, D., and Seung, H., 1999. “Learning the Parts of Objects by Non-Negative Matrix Factorization,” *Nature*, Vol. 401, pp. 788-791.
- [LeS13] Lee, Y. T., and Sidford, A., 2013. “Efficient Accelerated Coordinate Descent Methods and Faster Algorithms for Solving Linear Systems,” Proc. 2013 IEEE 54th Annual Symposium on Foundations of Computer Science (FOCS), pp. 147-156.
- [LeW11] Lee, S., and Wright, S. J., 2011. “Approximate Stochastic Subgradient Estimation Training for Support Vector Machines,” Univ. of Wisconsin Report, arXiv preprint arXiv:1111.0432.
- [Lem74] Lemaréchal, C., 1974. “An Algorithm for Minimizing Convex Functions,” in *Information Processing '74*, Rosenfeld, J. L., (Ed.), North-Holland, Amsterdam, pp. 552-556.
- [Lem75] Lemaréchal, C., 1975. “An Extension of Davidon Methods to Non-differentiable Problems,” *Math. Programming Study* 3, Balinski, M., and Wolfe, P., (Eds.), North-Holland, Amsterdam, pp. 95-109.

- [Lem89] Lemaire, B., 1989. "The Proximal Algorithm," in *New Methods in Optimization and Their Industrial Uses*, J.-P. Penot, (ed.), Birkhauser, Basel, pp. 73-87.
- [LiM79] Lions, P. L., and Mercier, B., 1979. "Splitting Algorithms for the Sum of Two Nonlinear Operators," *SIAM J. on Numerical Analysis*, Vol. 16, pp. 964-979.
- [LiP87] Lin, Y. Y., and Pang, J.-S., 1987. "Iterative Methods for Large Convex Quadratic Programs: A Survey," *SIAM J. on Control and Optimization*, Vol. 18, pp. 383-411.
- [LiW14] Liu, J., and Wright, S. J., 2014. "Asynchronous Stochastic Coordinate Descent: Parallelism and Convergence Properties," Univ. of Wisconsin Report, arXiv preprint arXiv:1403.3862.
- [Lin07] Lin, C. J., 2007. "Projected Gradient Methods for Nonnegative Matrix Factorization," *Neural Computation*, Vol. 19, pp. 2756-2779.
- [Lit66] Litvakov, B. M., 1966. "On an Iteration Method in the Problem of Approximating a Function from a Finite Number of Observations," *Avtom. Telemekh.*, No. 4, pp. 104-113.
- [Lju77] Ljung, L., 1977. "Analysis of Recursive Stochastic Algorithms," *IEEE Trans. on Automatic Control*, Vol. 22, pp. 551-575.
- [LuT91] Luo, Z. Q., and Tseng, P., 1991. "On the Convergence of a Matrix-Splitting Algorithm for the Symmetric Monotone Linear Complementarity Problem," *SIAM J. on Control and Optimization*, Vol. 29, pp. 1037-1060.
- [LuT92] Luo, Z. Q., and Tseng, P., 1992. "On the Convergence of the Coordinate Descent Method for Convex Differentiable Minimization," *J. Optim. Theory Appl.*, Vol. 72, pp. 7-35.
- [LuT93a] Luo, Z. Q., and Tseng, P., 1993. "On the Convergence Rate of Dual Ascent Methods for Linearly Constrained Convex Minimization," *Math. of Operations Research*, Vol. 18, pp. 846-867.
- [LuT93b] Luo, Z. Q., and Tseng, P., 1993. "Error Bound and Reduced-Gradient Projection Algorithms for Convex Minimization over a Polyhedral Set," *SIAM J. on Optimization*, Vol. 3, pp. 43-59.
- [LuT93c] Luo, Z. Q., and Tseng, P., 1993. "Error Bounds and Convergence Analysis of Feasible Descent Methods: A General Approach," *Annals of Operations Research*, Vol. 46, pp. 157-178.
- [LuT94a] Luo, Z. Q., and Tseng, P., 1994. "Analysis of an Approximate Gradient Projection Method with Applications to the Backpropagation Algorithm," *Optimization Methods and Software*, Vol. 4, pp. 85-101.
- [LuT94b] Luo, Z. Q., and Tseng, P., 1994. "On the Rate of Convergence of a Distributed Asynchronous Routing Algorithm," *IEEE Trans. on Automatic Control*, Vol. 39, pp. 1123-1129.
- [LuT13] Luss, R., and Teboulle, M., 2013. "Conditional Gradient Algorithms for Rank-One Matrix Approximations with a Sparsity Constraint,"

- SIAM Review, Vol. 55, pp. 65-98.
- [LuY08] Luenberger, D. G., and Ye, Y., 2008. *Linear and Nonlinear Programming*, 3rd Edition, Springer, NY.
- [Lue84] Luenberger, D. G., 1984. *Introduction to Linear and Nonlinear Programming*, 2nd Edition, Addison-Wesley, Reading, MA.
- [Luo91] Luo, Z. Q., 1991. "On the Convergence of the LMS Algorithm with Adaptive Learning Rate for Linear Feedforward Networks," *Neural Computation*, Vol. 3, pp. 226-245.
- [Luq84] Luque, F. J., 1984. "Asymptotic Convergence Analysis of the Proximal Point Algorithm," *SIAM J. on Control and Optimization*, Vol. 22, pp. 277-293.
- [MRS10] Mosk-Aoyama, D., Roughgarden, T., and Shah, D., 2010. "Fully Distributed Algorithms for Convex Optimization Problems," *SIAM J. on Optimization*, Vol. 20, pp. 3260-3279.
- [MSQ98] Mifflin, R., Sun, D., and Qi, L., 1998. "Quasi-Newton Bundle-Type Methods for Nondifferentiable Convex Optimization," *SIAM J. on Optimization*, Vol. 8, pp. 583-603.
- [MYF03] Moriyama, H., Yamashita, N., and Fukushima, M., 2003. "The Incremental Gauss-Newton Algorithm with Adaptive Step-size Rule," *Computational Optimization and Applications*, Vol. 26, pp. 107-141.
- [MaM01] Mangasarian, O. L., Musicant, D. R., 2001. "Lagrangian Support Vector Machines," *J. of Machine Learning Research*, Vol. 1, pp. 161-177.
- [MaS94] Mangasarian, O. L., and Solodov, M. V., 1994. "Serial and Parallel Backpropagation Convergence Via Nonmonotone Perturbed Minimization," *Opt. Methods and Software*, Vol. 4, pp. 103-116.
- [Mai13] Mairal, J., 2013. "Optimization with First-Order Surrogate Functions," arXiv preprint arXiv:1305.3120.
- [Mai14] Mairal, J., 2014. "Incremental Majorization-Minimization Optimization with Application to Large-Scale Machine Learning," arXiv preprint arXiv:1402.4419.
- [Man53] Mann, W. R., 1953. "Mean Value Methods in Iteration," *Proc. Amer. Math. Soc.*, Vol. 4, pp. 506-510.
- [Mar70] Martinet, B., 1970. "Regularisation d'Inéquations Variationnelles par Approximations Successives," *Revue Fran. d'Automatique et Infomatique Rech. Opérationnelle*, Vol. 4, pp. 154-159.
- [Mar72] Martinet, B., 1972. "Determination Approché d'un Point Fixe d'une Application Pseudo-Contractante. Cas de l'Application Prox," *Comptes Rendus de l'Academie des Sciences, Paris, Serie A 274*, pp. 163-165.
- [Meh92] Mehrotra, S., 1992. "On the Implementation of a Primal-Dual Interior Point Method," *SIAM J. on Optimization*, Vol. 2, pp. 575-601.
- [Mey07] Meyn, S., 2007. *Control Techniques for Complex Networks*, Cambridge Univ. Press, NY.

- [MiF81] Mine, H., Fukushima, M. 1981. "A Minimization Method for the Sum of a Convex Function and a Continuously Differentiable Function," *J. of Optimization Theory and Applications*, Vol. 33, pp. 9-23.
- [Mif96] Mifflin, R., 1996. "A Quasi-Second-Order Proximal Bundle Algorithm," *Math. Programming*, Vol. 73, pp. 51-72.
- [Min62] Minty, G. J., 1962. "Monotone (Nonlinear) Operators in Hilbert Space," *Duke J. of Math.*, Vol. 29, pp. 341-346.
- [Min64] Minty, G. J., 1964. "On the Monotonicity of the Gradient of a Convex Function," *Pacific J. of Math.*, Vol. 14, pp. 243-247.
- [Min86] Minoux, M., 1986. *Math. Programming: Theory and Algorithms*, Wiley, NY.
- [MoT89] Moré, J. J., and Toraldo, G., 1989. "Algorithms for Bound Constrained Quadratic Programming Problems," *Numer. Math.*, Vol. 55, pp. 377-400.
- [NBB01] Nedić, A., Bertsekas, D. P., and Borkar, V., 2001. "Distributed Asynchronous Incremental Subgradient Methods," *Proc. of 2000 Haifa Workshop "Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications,"* by D. Butnariu, Y. Censor, and S. Reich, Eds., Elsevier, Amsterdam.
- [NJB09] Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A., 2009. "Robust Stochastic Approximation Approach to Stochastic Programming," *SIAM J. on Optimization*, Vol. 19, pp. 1574-1609.
- [NSW14] Needell, D., Srebro, N., and Ward, R., 2014. "Stochastic Gradient Descent and the Randomized Kaczmarz Algorithm," *arXiv preprint arXiv:1310.5715v3*.
- [NaT02] Nazareth, L., and Tseng, P., 2002. "Gilding the Lily: A Variant of the Nelder-Mead Algorithm Based on Golden-Section Search," *Computational Optimization and Applications*, Vol. 22, pp. 133-144.
- [NaZ05] Narkiss, G., and Zibulevsky, M., 2005. "Sequential Subspace Optimization Method for Large-Scale Unconstrained Problems," *Technion-IIT, Department of Electrical Engineering*.
- [NeB00] Nedić, A., and Bertsekas, D. P., 2000. "Convergence Rate of Incremental Subgradient Algorithms," *Stochastic Optimization: Algorithms and Applications*, S. Uryasev and P. M. Pardalos, Eds., Kluwer, pp. 263-304.
- [NeB01] Nedić, A., and Bertsekas, D. P., 2001. "Incremental Subgradient Methods for Nondifferentiable Optimization," *SIAM J. on Optimization*, Vol. 12, pp. 109-138.
- [NeB10] Nedić, A., and Bertsekas, D. P., 2010. "The Effect of Deterministic Noise in Subgradient Methods," *Math. Programming, Ser. A*, Vol. 125, pp. 75-99.

- [NeC13] Necoara, I., and Clipici, D., 2013. “Distributed Coordinate Descent Methods for Composite Minimization,” arXiv preprint arXiv:1312.5302.
- [NeN94] Nesterov, Y., and Nemirovskii, A., 1994. Interior Point Polynomial Algorithms in Convex Programming, SIAM, Studies in Applied Mathematics 13, Philadelphia, PA.
- [NeO09a] Nedić, A., and Ozdaglar, A., 2009. “Distributed Subgradient Methods for Multi-Agent Optimization,” IEEE Trans. on Aut. Control, Vol. 54, pp. 48-61.
- [NeO09b] Nedić, A., and Ozdaglar, A., 2009. “Subgradient Methods for Saddle-Point Problems,” J. of Optimization Theory and Applications, Vol. 142, pp. 205-228.
- [NeW88] Nemhauser, G. L., and Wolsey, L. A., 1988. Integer and Combinatorial Optimization, Wiley, NY.
- [NeY83] Nemirovsky, A., and Yudin, D. B., 1983. Problem Complexity and Method Efficiency, Wiley, NY.
- [Ned10] Nedić, A., 2010. “Random Projection Algorithms for Convex Set Intersection Problems,” Proc. 2010 IEEE Conference on Decision and Control, Atlanta, Georgia, pp. 7655-7660.
- [Ned11] Nedić, A., 2011. “Random Algorithms for Convex Minimization Problems,” Math. Programming, Ser. B, Vol. 129, pp. 225-253.
- [Nee10] Needell, D., 2010. “Randomized Kaczmarz Solver for Noisy Linear Systems,” BIT Numerical Mathematics, Vol. 50, pp. 395-403.
- [Nes83] Nesterov, Y., 1983. “A Method for Unconstrained Convex Minimization Problem with the Rate of Convergence $O(1/k^2)$,” Doklady AN SSSR, Vol. 269, pp. 543-547; translated as Soviet Math. Dokl.
- [Nes95] Nesterov, Y., 1995. “Complexity Estimates of Some Cutting Plane Methods Based on Analytic Barrier,” Math. Programming, Vol. 69, pp. 149-176.
- [Nes04] Nesterov, Y., 2004. Introductory Lectures on Convex Optimization, Kluwer Academic Publisher, Dordrecht, The Netherlands.
- [Nes05] Nesterov, Y., 2005. “Smooth Minimization of Nonsmooth Functions,” Math. Programming, Vol. 103, pp. 127-152.
- [Nes12] Nesterov, Y., 2012. “Efficiency of Coordinate Descent Methods on Huge-Scale Optimization Problems,” SIAM J. on Optimization, Vol. 22, pp. 341-362.
- [Nev75] Neveu, J., 1975. Discrete Parameter Martingales, North-Holland, Amsterdam, The Netherlands.
- [NoW06] Nocedal, J., and Wright, S. J., 2006. Numerical Optimization, 2nd Edition, Springer, NY.
- [Noc80] Nocedal, J., 1980. “Updating Quasi-Newton Matrices with Limited Storage,” Math. of Computation, Vol. 35, pp. 773-782.
- [OBG05] Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W., 2005.

- “An Iterative Regularization Method for Total Variation-Based Image Restoration,” *Multiscale Modeling and Simulation*, Vol. 4, pp. 460-489.
- [OJW05] Olafsson, A., Jeraj, R., and Wright, S. J., 2005. Optimization of Intensity-Modulated Radiation Therapy with Biological Objectives,” *Physics in Medicine and Biology*, Vol. 50, pp. 53-57.
- [OMV00] Ouorou, A., Mahey, P., and Vial, J. P., 2000. “A Survey of Algorithms for Convex Multicommodity Flow Problems,” *Management Science*, Vol. 46, pp. 126-147.
- [OrR70] Ortega, J. M., and Rheinboldt, W. C., 1970. *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, NY.
- [OvG14] Ovcharova, N., and Gwinner, J., 2014. “A Study of Regularization Techniques of Nondifferentiable Optimization in View of Application to Hemivariational Inequalities,” *J. of Optimization Theory and Applications*, Vol. 162, pp. 754-778.
- [OzB03] Ozdaglar, A. E., and Bertsekas, D. P., 2003. “Routing and Wavelength Assignment in Optical Networks,” *IEEE Trans. on Networking*, Vol. 11, pp. 259-272.
- [PKP09] Predd, J. B., Kulkarni, S. R., and Poor, H. V., 2009. “A Collaborative Training Algorithm for Distributed Learning,” *IEEE Transactions on Information Theory*, Vol. 55, pp. 1856-1871.
- [PaE10] Palomar, D. P., and Eldar, Y. C., (Eds.), 2010. *Convex Optimization in Signal Processing and Communications*, Cambridge Univ. Press, NY.
- [PaT94] Paatero, P., and Tapper, U., 1994. “Positive Matrix Factorization: A Non-Negative Factor Model with Optimal Utilization of Error Estimates of Data Values,” *Environmetrics*, Vol. 5, pp. 111-126.
- [PaY84] Pang, J. S., Yu, C. S., 1984. “Linearized Simplicial Decomposition Methods for Computing Traffic Equilibria on Networks,” *Networks*, Vol. 14, pp. 427-438.
- [Paa97] Paatero, P., 1997. “Least Squares Formulation of Robust Non-Negative Factor Analysis,” *Chemometrics and Intell. Laboratory Syst.*, Vol. 37, pp. 23-35.
- [Pan84] Pang, J. S., 1984. “On the Convergence of Dual Ascent Methods for Large-Scale Linearly Constrained Optimization Problems,” Unpublished manuscript, The Univ. of Texas at Dallas.
- [Pap81] Papavassilopoulos, G., 1981. “Algorithms for a Class of Nondifferentiable Problems,” *J. of Optimization Theory and Applications*, Vol. 34, pp. 41-82.
- [Pas79] Passty, G. B., 1979. “Ergodic Convergence to a Zero of the Sum of Monotone Operators in Hilbert Space,” *J. Math. Anal. Appl.*, Vol. 72, pp. 383-390.
- [Pat93] Patriksson, M., 1993. “Partial Linearization Methods in Nonlinear

- Programming,” *J. of Optimization Theory and Applications*, Vol. 78, pp. 227-246.
- [Pat98] Patriksson, M., 1998. “Cost Approximation: A Unified Framework of Descent Algorithms for Nonlinear Programs,” *SIAM J. Optimization*, Vol. 8, pp. 561-582.
- [Pat99] Patriksson, M., 1999. *Nonlinear Programming and Variational Inequality Problems: A Unified Approach*, Springer, NY.
- [Pat01] Patriksson, M., 2001. “Simplicial Decomposition Algorithms,” *Encyclopedia of Optimization*, Springer, pp. 2378-2386.
- [Pat04] Patriksson, M., 2004. “Algorithms for Computing Traffic Equilibria,” *Networks and Spatial Economics*, Vol. 4, pp. 23-38.
- [Pen02] Pennanen, T., 2002. “Local Convergence of the Proximal Point Algorithm and Multiplier Methods Without Monotonicity,” *Math. of Operations Research*, Vol. 27, pp. 170-191.
- [Pfl96] Pflug, G., 1996. *Optimization of Stochastic Models. The Interface Between Simulation and Optimization*, Kluwer, Boston.
- [PiZ94] Pinar, M., and Zenios, S., 1994. “On Smoothing Exact Penalty Functions for Convex Constrained Optimization,” *SIAM J. on Optimization*, Vol. 4, pp. 486-511.
- [PoJ92] Poljak, B. T., and Juditsky, A. B., 1992. “Acceleration of Stochastic Approximation by Averaging,” *SIAM J. on Control and Optimization*, Vol. 30, pp. 838-855.
- [PoT73] Poljak, B. T., and Tsyppkin, Y. Z., 1973. “Pseudogradient Adaptation and Training Algorithms,” *Automation and Remote Control*, Vol. 12, pp. 83-94.
- [PoT74] Poljak, B. T., and Tretjakov, N. V., 1974. “An Iterative Method for Linear Programming and its Economic Interpretation,” *Matecon*, Vol. 10, pp. 81-100.
- [PoT80] Poljak, B. T., and Tsyppkin, Y. Z., 1980. “Adaptive Estimation Algorithms (Convergence, Optimality, Stability),” *Automation and Remote Control*, Vol. 40, pp. 378-389.
- [PoT81] Poljak, B. T., and Tsyppkin, Y. Z., 1981. “Optimal Pseudogradient Adaptation Algorithms,” *Automation and Remote Control*, Vol. 41, pp. 1101-1110.
- [PoT97] Polyak, R., and Teboulle, M., 1997. “Nonlinear Rescaling and Proximal-Like Methods in Convex Optimization,” *Math. Programming*, Vol. 76, pp. 265-284.
- [Pol64] Poljak, B. T., 1964. “Some Methods of Speeding up the Convergence of Iteration Methods,” *Z. Vyčisl. Mat. i Mat. Fiz.*, Vol. 4, pp. 1-17.
- [Pol71] Polak, E., 1971. *Computational Methods in Optimization: A Unified Approach*, Academic Press, NY.

- [Pol78] Poljak, B. T., 1978. “Nonlinear Programming Methods in the Presence of Noise,” *Math. Programming*, Vol. 14, pp. 87–97.
- [Pol79] Poljak, B. T., 1979. “On Bertsekas’ Method for Minimization of Composite Functions,” *Internat. Symp. Systems Opt. Analysis*, Benoussan, A., and Lions, J. L., (Eds.), Springer-Verlag, Berlin and NY, pp. 179-186.
- [Pol87] Poljak, B. T., 1987. *Introduction to Optimization*, Optimization Software Inc., NY.
- [Pol88] Polyak, R. A., 1988. “Smooth Optimization Methods for Minimax Problems,” *SIAM J. on Control and Optimization*, Vol. 26, pp. 1274-1286.
- [Pol92] Polyak, R. A., 1992. “Modified Barrier Functions (Theory and Methods),” *Math. Programming*, Vol. 54, pp. 177-222.
- [Pow69] Powell, M. J. D., 1969. “A Method for Nonlinear Constraints in Minimizing Problems,” in *Optimization*, Fletcher, R., (Ed.), Academic Press, NY, pp. 283-298.
- [Pow73] Powell, M. J. D., 1973. “On Search Directions for Minimization Algorithms,” *Math. Programming*, Vol. 4, pp. 193-201.
- [Pow11] Powell, W. B., 2011. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*, 2nd Ed., Wiley, NY.
- [Pre95] Prekopa, A., 1995. *Stochastic Programming*, Kluwer, Boston.
- [Psh65] Pshenichnyi, B. N., 1965. “Dual Methods in Extremum Problems,” *Kibernetika*, Vol. 1, pp. 89-95.
- [Pyt98] Pytlak, R., 1998. “An Efficient Algorithm for Large-Scale Nonlinear Programming Problems with Simple Bounds on the Variables,” *SIAM J. on Optimization*, Vol. 8, pp. 532-560.
- [QSG13] Qin, Z., Scheinberg, K., and Goldfarb, D., 2013. “Efficient Block-Coordinate Descent Algorithms for the Group Lasso,” *Math. Programming Computation*, Vol. 5, pp. 143-169.
- [RFP10] Recht, B., Fazel, M., and Parrilo, P. A., 2010. “Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization,” *SIAM Review*, Vol. 52, pp. 471-501.
- [RGV14] Richard, E., Gaifas, S., and Vayatis, N., 2014. “Link Prediction in Graphs with Autoregressive Features,” *J. of Machine Learning Research*, Vol. 15, pp. 565-593.
- [RHL13] Razaviyayn, M., Hong, M., and Luo, Z. Q., 2013. “A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization,” *SIAM J. on Optimization*, Vol. 23, pp. 1126-1153.
- [RHW86] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1986. “Learning Internal Representation by Error Backpropagation,” in *Parallel Distributed Processing-Explorations in the Microstructure of Cognition*, by Rumelhart and McClelland, (eds.), MIT Press, Cambridge, MA, pp. 318-362.
- [RHW88] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., 1988.

- “Learning Representations by Back-Propagating Errors,” in *Cognitive Modeling*, by T. A. Polk, and C. M. Seifert, (eds.), MIT Press, Cambridge, MA, pp. 213-220.
- [RHZ14] Razaviyayn, M., Hong, M., and Luo, Z. Q., 2013. “A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization,” *SIAM J. on Optimization*, Vol. 23, pp. 1126-1153.
- [RNV09] Ram, S. S., Nedić, A., and Veeravalli, V. V., 2009. “Incremental Stochastic Subgradient Algorithms for Convex Optimization,” *SIAM J. on Optimization*, Vol. 20, pp. 691-717.
- [RNV10] Ram, S. S., Nedić, A., and Veeravalli, V. V., 2010. “Distributed Stochastic Subgradient Projection Algorithms for Convex Optimization,” *J. of Optimization Theory and Applications*, Vol. 147, pp. 516-545.
- [ROF92] Rudin, L. I., Osher, S., and Fatemi, E., 1992. “Nonlinear Total Variation Based Noise Removal Algorithms,” *Physica D: Nonlinear Phenomena*, Vol. 60, pp. 259-268.
- [RRW11] Recht, B., Re, C., Wright, S. J., and Niu, F., 2011. “Hogwild: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent,” in *Advances in Neural Information Processing Systems (NIPS 2011)*, pp. 693-701.
- [RSW13] Rao, N., Shah, P., Wright, S., and Nowak, R., 2013. “A Greedy Forward-Backward Algorithm for Atomic Norm Constrained Minimization,” in *Proc. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5885-5889.
- [RTV06] Roos, C., Terlaky, T., and Vial, J. P., 2006. *Interior Point Methods for Linear Optimization*, Springer, NY.
- [RXB11] Recht, B., Xu, W., and Hassibi, B., 2011. “Null Space Conditions and Thresholds for Rank Minimization,” *Math. Programming*, Vol. 127, pp. 175-202.
- [RaN04] Rabbat, M. G., and Nowak, R. D., 2004. “Distributed Optimization in Sensor Networks,” in *Proc. Inf. Processing Sensor Networks*, Berkeley, CA, pp. 20-27.
- [RaN05] Rabbat M. G., and Nowak R. D., 2005. “Quantized Incremental Algorithms for Distributed Optimization,” *IEEE J. on Select Areas in Communications*, Vol. 23, pp. 798-808.
- [Ray93] Raydan, M., 1993. “On the Barzilai and Borwein Choice of Steplength for the Gradient Method,” *IMA J. of Numerical Analysis*, Vol. 13, pp. 321-326.
- [Ray97] Raydan, M., 1997. “The Barzilai and Borwein Gradient Method for the Large Scale Unconstrained Minimization Problem,” *SIAM J. on Optimization*, Vol. 7, pp. 26-33.
- [ReR13] Recht, B., and Ré, C., 2013. “Parallel Stochastic Gradient Algorithms for Large-Scale Matrix Completion,” *Math. Programming Computation*, Vol. 5, pp. 201-226.

- [Rec11] Recht, B., 2011. "A Simpler Approach to Matrix Completion," *The J. of Machine Learning Research*, Vol. 12, pp. 3413-3430.
- [RiT14] Richtarik, P., and Takac, M., 2014. "Iteration Complexity of Randomized Block-Coordinate Descent Methods for Minimizing a Composite Function," *Math. Programming*, Vol. 144, pp. 1-38.
- [RoS71] Robbins, H., and Siegmund, D. O., 1971. "A Convergence Theorem for Nonnegative Almost Supermartingales and Some Applications," *Optimizing Methods in Statistics*, pp. 233-257; see "Herbert Robbins Selected Papers," Springer, NY, 1985, pp. 111-135.
- [RoW91] Rockafellar, R. T., and Wets, R. J.-B., 1991. "Scenarios and Policy Aggregation in Optimization Under Uncertainty," *Math. of Operations Research*, Vol. 16, pp. 119-147.
- [RoW98] Rockafellar, R. T., and Wets, R. J.-B., 1998. *Variational Analysis*, Springer-Verlag, Berlin.
- [Rob99] Robinson, S. M., 1999. "Linear Convergence of Epsilon-Subgradient Descent Methods for a Class of Convex Functions," *Math. Programming, Ser. A*, Vol. 86, pp. 41-50.
- [Roc66] Rockafellar, R. T., 1966. "Characterization of the Subdifferentials of Convex Functions," *Pacific J. of Mathematics*, Vol. 17, pp. 497-510.
- [Roc70] Rockafellar, R. T., 1970. *Convex Analysis*, Princeton Univ. Press, Princeton, NJ.
- [Roc73] Rockafellar, R. T., 1973. "A Dual Approach to Solving Nonlinear Programming Problems by Unconstrained Optimization," *Math. Programming*, pp. 354-373.
- [Roc76a] Rockafellar, R. T., 1976. "Monotone Operators and the Proximal Point Algorithm," *SIAM J. on Control and Optimization*, Vol. 14, pp. 877-898.
- [Roc76b] Rockafellar, R. T., 1976. "Augmented Lagrangians and Applications of the Proximal Point Algorithm in Convex Programming," *Math. of Operations Research*, Vol. 1, pp. 97-116.
- [Roc76c] Rockafellar, R. T., 1976. "Solving a Nonlinear Programming Problem by Way of a Dual Problem," *Symp. Matematica*, Vol. 27, pp. 135-160.
- [Roc84] Rockafellar, R. T., 1984. *Network Flows and Monotropic Optimization*, Wiley, NY; republished by Athena Scientific, Belmont, MA, 1998.
- [Rud76] Rudin, W., 1976. *Real Analysis*, McGraw-Hill, NY.
- [Rup85] Ruppert, D., 1985. "A Newton-Raphson Version of the Multivariate Robbins-Monro Procedure," *The Annals of Statistics*, Vol. 13, pp. 236-245.
- [Rus86] Ruszczyński, A., 1986. "A Regularized Decomposition Method for Minimizing a Sum of Polyhedral Functions," *Math. Programming*, Vol. 35, pp. 309-333.

- [Rus06] Ruszczyński, A., 2006. *Nonlinear Optimization*, Princeton Univ. Press, Princeton, NJ.
- [SBC91] Saarinen, S., Bramley, R. B., and Cybenko, G., 1991. “Neural Networks, Backpropagation and Automatic Differentiation,” in *Automatic Differentiation of Algorithms*, by A. Griewank and G. F. Corliss, (eds.), SIAM, Philadelphia, PA, pp. 31-42.
- [SBK64] Shah, B., Buehler, R., and Kempthorne, O., 1964. “Some Algorithms for Minimizing a Function of Several Variables,” *J. Soc. Indust. Appl. Math.*, Vol. 12, pp. 74-92.
- [SBT12] Shah, P., Bhaskar, B. N., Tang, G., and Recht, B., 2012. “Linear System Identification via Atomic Norm Regularization,” arXiv preprint arXiv:1204.0590.
- [SDR09] Shapiro, A., Dentcheva, D., and Ruszczyński, A., 2009. *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Phila., PA.
- [SFR09] Schmidt, M., Fung, G., and Rosales, R., 2009. “Optimization Methods for ℓ_1 -Regularization,” Univ. of British Columbia, Technical Report TR-2009-19.
- [SKS12] Schmidt, M., Kim, D., and Sra, S., 2012. “Projected Newton-Type Methods in Machine Learning,” in *Optimization for Machine Learning*, by Sra, S., Nowozin, S., and Wright, S. J., (eds.), MIT Press, Cambridge, MA, pp. 305-329.
- [SLB13] Schmidt, M., Le Roux, N., and Bach, F., 2013. “Minimizing Finite Sums with the Stochastic Average Gradient,” arXiv preprint arXiv:1309.2388.
- [SNW12] Sra, S., Nowozin, S., and Wright, S. J., 2012. *Optimization for Machine Learning*, MIT Press, Cambridge, MA.
- [SRB11] Schmidt, M., Roux, N. L., and Bach, F. R., 2011. “Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization,” In *Advances in Neural Information Processing Systems*, pp. 1458-1466.
- [SSS07] Shalev-Shwartz, S., Singer, Y., Srebro, N., and Cotter, A., 2007. “Pegasos: Primal Estimated Subgradient Solver for SVM,” in *ICML 07*, New York, NY, pp. 807-814.
- [SaB96] Savari, S. A., and Bertsekas, D. P., 1996. “Finite Termination of Asynchronous Iterative Algorithms,” *Parallel Computing*, Vol. 22, pp. 39-56.
- [SaT13] Saha, A., and Tewari, A., 2013. “On the Nonasymptotic Convergence of Cyclic Coordinate Descent Methods,” *SIAM J. on Optimization*, Vol. 23, pp. 576-601.
- [Sak66] Sakrison, D. T., 1966. “Stochastic Approximation: A Recursive Method for Solving Regression Problems,” in *Advances in Communication Theory and Applications*, 2, A. V. Balakrishnan, ed., Academic Press, NY, pp. 51-106.

- [Say14] Sayed, A. H., 2014. "Adaptation, Learning, and Optimization over Networks," *Foundations and Trends in Machine Learning*, Vol. 7, no. 4-5, pp. 311-801.
- [ScF14] Schmidt, M., and Friedlander, M. P., 2014. "Coordinate Descent Converges Faster with the Gauss-Southwell Rule than Random Selection," *Advances in Neural Information Processing Systems 27 (NIPS 2014)*.
- [Sch82] Schnabel, R. B., 1982. "Determining Feasibility of a Set of Non-linear Inequality Constraints," *Math. Programming Studies*, Vol. 16, pp. 137-148.
- [Sch86] Schrijver, A., 1986. *Theory of Linear and Integer Programming*, Wiley, NY.
- [Sch10] Schmidt, M., 2010. "Graphical Model Structure Learning with L1-Regularization," PhD Thesis, Univ. of British Columbia.
- [Sch14a] Schmidt, M., 2014. "Convergence Rate of Stochastic Gradient with Constant Step Size," Computer Science Report, Univ. of British Columbia.
- [Sch14b] Schmidt, M., 2014. "Convergence Rate of Proximal Gradient with General Step-Size," Dept. of Computer Science, Unpublished Note, Univ. of British Columbia.
- [ShZ12] Shamir, O., and Zhang, T., 2012. "Stochastic Gradient Descent for Non-Smooth Optimization: Convergence Results and Optimal Averaging Schemes," arXiv preprint arXiv:1212.1824.
- [Sha79] Shapiro, J. E., 1979. *Mathematical Programming Structures and Algorithms*, Wiley, NY.
- [Sho85] Shor, N. Z., 1985. *Minimization Methods for Nondifferentiable Functions*, Springer-Verlag, Berlin.
- [Sho98] Shor, N. Z., 1998. *Nondifferentiable Optimization and Polynomial Problems*, Kluwer Academic Publishers, Dordrecht, Netherlands.
- [SmS04] Smola, A. J., and Scholkopf, B., 2004. "A Tutorial on Support Vector Regression," *Statistics and Computing*, Vol. 14, pp. 199-222.
- [SoZ98] Solodov, M. V., and Zavriev, S. K., 1998. "Error Stability Properties of Generalized Gradient-Type Algorithms," *J. Opt. Theory and Appl.*, Vol. 98, pp. 663-680.
- [Sol98] Solodov, M. V., 1998. "Incremental Gradient Algorithms with Step-sizes Bounded Away from Zero," *Computational Optimization and Applications*, Vol. 11, pp. 23-35.
- [Spa03] Spall, J. C., 2003. *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, J. Wiley, Hoboken, NJ.
- [Spa12] Spall, J. C., 2012. "Cyclic Seesaw Process for Optimization and Identification," *J. of Optimization Theory and Applications*, Vol. 154, pp. 187-208.
- [Spi83] Spingarn, J. E., 1983. "Partial Inverse of a Monotone Operator," *Applied Mathematics and Optimization*, Vol. 10, pp. 247-265.

- [Spi85] Spingarn, J. E., 1985. "Applications of the Method of Partial Inverses to Convex Programming: Decomposition," *Math. Programming*, Vol. 32, pp. 199-223.
- [StV09] Strohmer, T., and Vershynin, R., 2009. "A Randomized Kaczmarz Algorithm with Exponential Convergence," *J. Fourier Anal. Appl.*, Vol. 15, pp. 262-278.
- [StW70] Stoer, J., and Witzgall, C., 1970. *Convexity and Optimization in Finite Dimensions*, Springer-Verlag, Berlin.
- [StW75] Stephanopoulos, G., and Westerberg, A. W., 1975. "The Use of Hestenes' Method of Multipliers to Resolve Dual Gaps in Engineering System Optimization," *J. Optimization Theory and Applications*, Vol. 15, pp. 285-309.
- [Str76] Strang, G., 1976. *Linear Algebra and Its Applications*, Academic Press, NY.
- [Str97] Stromberg, A-B., 1997. *Conditional Subgradient Methods and Ergodic Convergence in Nonsmooth Optimization*, Ph.D. Thesis, Univ. of Linköping, Sweden.
- [SuB98] Sutton, R. S., and Barto, A. G., 1998. *Reinforcement Learning*, MIT Press, Cambridge, MA.
- [TBA86] Tsitsiklis, J. N., Bertsekas, D. P., and Athans, M., 1986. "Distributed Asynchronous Deterministic and Stochastic Gradient Optimization Algorithms," *IEEE Trans. on Automatic Control*, Vol. AC-31, pp. 803-812.
- [TBT90] Tseng, P., Bertsekas, D. P., and Tsitsiklis, J. N., 1990. "Partially Asynchronous, Parallel Algorithms for Network Flow and Other Problems," *SIAM J. on Control and Optimization*, Vol. 28, pp. 678-710.
- [TVS10] Teo, C. H., Vishwanathan, S. V. N., Smola, A. J., and Le, Q. V., 2010. "Bundle Methods for Regularized Risk Minimization," *The J. of Machine Learning Research*, Vol. 11, pp. 311-365.
- [TaP13] Talischi, C., and Paulino, G. H., 2013. "A Consistent Operator Splitting Algorithm and a Two-Metric Variant: Application to Topology Optimization," arXiv preprint arXiv:1307.5100.
- [Teb92] Teboulle, M., 1992. "Entropic Proximal Mappings with Applications to Nonlinear Programming," *Math. of Operations Research*, Vol. 17, pp. 1-21.
- [Teb97] Teboulle, M., 1997. "Convergence of Proximal-Like Algorithms," *SIAM J. Optim.*, Vol. 7, pp. 1069-1083.
- [Ter96] Terlaky, T. (Ed.), 1996. *Interior Point Methods of Mathematical Programming*, Springer, NY.
- [Tib96] Tibshirani, R., 1996. "Regression Shrinkage and Selection via the Lasso," *J. of the Royal Statistical Society, Series B (Methodological)*, Vol. 58, pp. 267-288.

- [Tod01] Todd, M. J., 2001. "Semidefinite Optimization," *Acta Numerica*, Vol. 10, pp. 515-560.
- [TsB87] Tseng, P., and Bertsekas, D. P., 1987. "Relaxation Methods for Problems with Strictly Convex Separable Costs and Linear Constraints," *Math. Programming*, Vol. 38, pp. 303-321.
- [TsB90] Tseng, P., and Bertsekas, D. P., 1990. "Relaxation Methods for Monotropic Programs," *Math. Programming*, Vol. 46, pp. 127-151.
- [TsB91] Tseng, P., and Bertsekas, D. P., 1991. "Relaxation Methods for Problems with Strictly Convex Costs and Linear Constraints," *Math. of Operations Research*, Vol. 16, pp. 462-481.
- [TsB93] Tseng, P., and Bertsekas, D. P., 1993. "On the Convergence of the Exponential Multiplier Method for Convex Programming," *Math. Programming*, Vol. 60, pp. 1-19.
- [TsB00] Tseng, P., and Bertsekas, D. P., 2000. "An Epsilon-Relaxation Method for Separable Convex Cost Generalized Network Flow Problems," *Math. Programming*, Vol. 88, pp. 85-104.
- [TsY09] Tseng, P. and Yun S., 2009. "A Coordinate Gradient Descent Method for Nonsmooth Separable Minimization," *Math. Programming*, Vol. 117, pp. 387-423.
- [Tse91a] Tseng, P., 1991. "Decomposition Algorithm for Convex Differentiable Minimization," *J. of Optimization Theory and Applications*, Vol. 70, pp. 109-135.
- [Tse91b] Tseng, P., 1991. "Applications of a Splitting Algorithm to Decomposition in Convex Programming and Variational Inequalities," *SIAM J. on Control and Optimization*, Vol. 29, pp. 119-138.
- [Tse93] Tseng, P., 1993. "Dual Coordinate Ascent Methods for Non-Strictly Convex Minimization," *Math. Programming*, Vol. 59, pp. 231-247.
- [Tse95] Tseng, P., 1995. "Fortified-Descent Simplicial Search Method," Report, Dept. of Math., Univ. of Washington, Seattle, Wash.; also in *SIAM J. on Optimization*, Vol. 10, 1999, pp. 269-288.
- [Tse98] Tseng, P., 1998. "Incremental Gradient(-Projection) Method with Momentum Term and Adaptive Stepsize Rule," *SIAM J. on Control and Optimization*, Vol. 8, pp. 506-531.
- [Tse00] Tseng, P., 2000. "A Modified Forward-Backward Splitting Method for Maximal Monotone Mappings," *SIAM J. on Control and Optimization*, Vol. 38, pp. 431-446.
- [Tse01a] Tseng, P., 2001. "Convergence of Block Coordinate Descent Methods for Nondifferentiable Minimization," *J. Optim. Theory Appl.*, Vol. 109, pp. 475-494.
- [Tse01b] Tseng, P., 2001. "An Epsilon Out-of-Kilter Method for Monotropic Programming," *Math. of Operations Research*, Vol. 26, pp. 221-233.
- [Tse04] Tseng, P., 2004. "An Analysis of the EM Algorithm and Entropy-

- Like Proximal Point Methods,” *Math. Operations Research*, Vol. 29, pp. 27-44.
- [Tse08] Tseng, P., 2008. “On Accelerated Proximal Gradient Methods for Convex-Concave Optimization,” Report, Math. Dept., Univ. of Washington.
- [Tse09] Tseng, P., 2009. “Some Convex Programs Without a Duality Gap,” *Math. Programming*, Vol. 116, pp. 553-578.
- [Tse10] Tseng, P., 2010. “Approximation Accuracy, Gradient Methods, and Error Bound for Structured Convex Optimization,” *Math. Programming*, Vol. 125, pp. 263-295.
- [VKG14] Vetterli, M., Kovacevic, J., and Goyal, V. K., 2014. *Foundations of Signal Processing*, Cambridge Univ. Press, NY.
- [VMR88] Vogl, T. P., Mangis, J. K., Rigler, A. K., Zink, W. T., and Alkon, D. L., 1988. “Accelerating the Convergence of the Back-Propagation Method,” *Biological Cybernetics*, Vol. 59, pp. 257-263.
- [VaF08] Van Den Berg, E., and Friedlander, M. P., 2008. “Probing the Pareto Frontier for Basis Pursuit Solutions,” *SIAM J. on Scientific Computing*, Vol. 31, pp. 890-912.
- [Van01] Vanderbei, R. J., 2001. *Linear Programming: Foundations and Extensions*, Springer, NY.
- [VeH93] Ventura, J. A., and Hearn, D. W., 1993. “Restricted Simplicial Decomposition for Convex Constrained Problems,” *Math. Programming*, Vol. 59, pp. 71-85.
- [Ven67] Venter, J. H., 1967. “An Extension of the Robbins-Monro Procedure,” *Ann. Math. Statist.*, Vol. 38, pp. 181-190.
- [WDS13] Weinmann, A., Demaret, L., and Storath, M., 2013. “Total Variation Regularization for Manifold-Valued Data,” arXiv preprint arXiv:1312.7710.
- [WFL14] Wang, M., Fang, E., and Liu, H., 2014. “Stochastic Compositional Gradient Descent: Algorithms for Minimizing Compositions of Expected-Value Functions,” *Optimization Online*.
- [WHM13] Wang, X., Hong, M., Ma, S., Luo, Z. Q., 2013. “Solving Multiple-Block Separable Convex Minimization Problems Using Two-Block Alternating Direction Method of Multipliers,” arXiv preprint arXiv:1308.5294.
- [WSK14] Wytock, M., Suvrit S., and Kolter, J. K., 2014. “Fast Newton Methods for the Group Fused Lasso,” *Proc. of 2014 Conf. on Uncertainty in Artificial Intelligence*.
- [WSV00] Wolkowicz, H., Saigal, R., and Vanderbergue, L., (eds), 2000. *Handbook of Semidefinite Programming*, Kluwer, Boston.
- [WaB13a] Wang, M., and Bertsekas, D. P., 2013. “Incremental Constraint Projection-Proximal Methods for Nonsmooth Convex Optimization,” *Lab. for Information and Decision Systems Report LIDS-P-2907*, MIT, to appear

in SIAM J. on Optimization.

[WaB13b] Wang, M., and Bertsekas, D. P., 2013. “Convergence of Iterative Simulation-Based Methods for Singular Linear Systems,” *Stochastic Systems*, Vol. 3, pp. 38-95.

[WaB13c] Wang, M., and Bertsekas, D. P., 2013. “Stabilization of Stochastic Iterative Methods for Singular and Nearly Singular Linear Systems,” *Math. of Operations Research*, Vol. 39, pp. 1-30.

[WaB14] Wang, M., and Bertsekas, D. P., 2014. “Incremental Constraint Projection Methods for Variational Inequalities,” *Mathematical Programming*, pp. 1-43.

[Was04] Wasserman, L., 2004. *All of Statistics: A Concise Course in Statistical Inference*, Springer, NY.

[Wat92] Watson, G. A., 1992. “Characterization of the Subdifferential of Some Matrix Norms,” *Linear Algebra and its Applications*, Vol. 170, pp. 33-45.

[WeO13] Wei, E., and Ozdaglar, A., 2013. “On the $O(1/k)$ Convergence of Asynchronous Distributed Alternating Direction Method of Multipliers,” arXiv preprint arXiv:1307.8254.

[WiH60] Widrow, B., and Hoff, M. E., 1960. “Adaptive Switching Circuits,” *Institute of Radio Engineers, Western Electronic Show and Convention, Convention Record, Part 4*, pp. 96-104.

[Wol75] Wolfe, P., 1975. “A Method of Conjugate Subgradients for Minimizing Nondifferentiable Functions,” *Math. Programming Study 3*, Balinski, M., and Wolfe, P., (Eds.), North-Holland, Amsterdam, pp. 145-173.

[Wri97] Wright, S. J., 1997. *Primal-Dual Interior Point Methods*, SIAM, Philadelphia, PA.

[Wri14] Wright, S. J., 2014. “Coordinate Descent Algorithms,” *Optimization Online*.

[XiZ14] Xiao, L., and Zhang, T., 2014. “A Proximal Stochastic Gradient Method with Progressive Variance Reduction,” arXiv preprint arXiv:1403.4699.

[Xia10] Xiao L., 2010. “Dual Averaging Methods for Regularized Stochastic Learning and Online Optimization,” *J. of Machine Learning Research*, Vol. 11, pp. 2534-2596.

[YBR08] Yu, H., Bertsekas, D. P., and Rousu, J., 2008. “An Efficient Discriminative Training Method for Generative Models,” *Extended Abstract, the 6th International Workshop on Mining and Learning with Graphs (MLG)*.

[YGT93] Ye, Y., Guler, O., Tapia, R. A., and Zhang, Y., 1993. “A Quadratically Convergent $O(\sqrt{nL})$ -Iteration Algorithm for Linear Programming,” *Math. Programming*, Vol. 59, pp. 151-162.

[YNS10] Yousefian, F., Nedić, A., and Shanbhag, U. V., 2010. “Convex

- Nondifferentiable Stochastic Optimization: A Local Randomized Smoothing Technique,” Proc. American Control Conference (ACC), pp. 4875-4880.
- [YNS12] Yousefian, F., Nedić, A., and Shanbhag, U. V., 2012. “On Stochastic Gradient and Subgradient Methods with Adaptive Steplength Sequences,” *Automatica*, Vol. 48, pp. 56-67.
- [YOG08] Yin, W., Osher, S., Goldfarb, D., and Darbon, J., 2008. “Bregman Iterative Algorithms for ℓ_1 -Minimization with Applications to Compressed Sensing,” *SIAM J. on Imaging Sciences*, Vol. 1, pp. 143-168.
- [YSQ14] You, K., Song, S., and Qiu, L., 2014. “Randomized Incremental Least Squares for Distributed Estimation Over Sensor Networks,” Preprints of the 19th World Congress The International Federation of Automatic Control Cape Town, South Africa.
- [Ye92] Ye, Y., 1992. “A Potential Reduction Algorithm Allowing Column Generation,” *SIAM J. on Optimization*, Vol. 2, pp. 7-20.
- [Ye97] Ye, Y., 1997. *Interior Point Algorithms: Theory and Analysis*, Wiley Interscience, NY.
- [YuR07] Yu, H., and Rousu, J., 2007. “An Efficient Method for Large Margin Parameter Optimization in Structured Prediction Problems,” Technical Report C-2007-87, Univ. of Helsinki.
- [ZJL13] Zhang, H., Jiang, J., and Luo, Z. Q., 2013. “On the Linear Convergence of a Proximal Gradient Method for a Class of Nonsmooth Convex Minimization Problems,” *J. of the Operations Research Society of China*, Vol. 1, pp. 163-186.
- [ZLW99] Zhao, X., Luh, P. B., and Wang, J., 1999. “Surrogate Gradient Algorithm for Lagrangian Relaxation,” *J. Optimization Theory and Applications*, Vol. 100, pp. 699-712.
- [ZMJ13] Zhang, L., Mahdavi, M., and Jin, R., 2013. “Linear Convergence with Condition Number Independent Access of Full Gradients,” *Advances in Neural Information Processing Systems 26 (NIPS 2013)*, pp. 980-988.
- [ZTD92] Zhang, Y., Tapia, R. A., and Dennis, J. E., 1992. “On the Superlinear and Quadratic Convergence of Primal-Dual Interior Point Linear Programming Algorithms,” *SIAM J. on Optimization*, Vol. 2, pp. 304-324.
- [Zal02] Zalinescu, C., 2002. *Convex Analysis in General Vector Spaces*, World Scientific, Singapore.
- [Zan69] Zangwill, W. I., 1969. *Nonlinear Programming*, Prentice-Hall, Englewood Cliffs, NJ.
- [Zou60] Zoutendijk, G., 1960. *Methods of Feasible Directions*, Elsevier Publ. Co., Amsterdam.
- [Zou76] Zoutendijk, G., 1976. *Mathematical Programming Methods*, North Holland, Amsterdam.

INDEX

A

ADMM 111, 280, 292, 295, 298, 337, 427
Affine function 445
Affine hull 472
Affine set 445
Aggregated gradient method 91, 94, 428
Alternating direction method 111, 280
Analytic center 425
Approximation algorithms 36, 54
Armijo rule 69, 123, 125, 317
Asymptotic sequence 481
Asynchronous computation 33, 104, 376
Asynchronous gradient method 104, 106
Atomic norm 35
Auction algorithm 180, 375
Augmented Lagrangian function 260, 283, 290
Augmented Lagrangian method 109, 115, 120, 261, 294, 326, 337, 362, 384, 389, 430
Averaging of iterates 120, 157, 177

B

Backpropagation 119
Backtracking rule 69, 123
Ball center 426
Barrier function 412
Barrier method 413
Basis 446
Basis pursuit 34, 286
Batching 93
Block coordinate descent 75, 268, 281, 429, 438-442

Bolzano-Weierstrass Theorem 453
Boundary of a set 454
Boundary point 454
Bounded sequence 451, 452
Bounded set 453
Branch-and-bound 7
Bregman distance 388
Bundle methods 110, 187, 272, 295, 385

C

Caratheodory's Theorem 472
Cartesian product 445
Cauchy sequence 452
Central cutting plane methods 425, 432
Chain rule 142, 513
Classification 29
Closed ball 453
Closed function 469
Closed halfspace 484
Closed set 453
Closed set intersections 481
Closed sphere 453
Closedness under linear transformations 483
Closedness under vector sums 483
Closure of a function 476
Closure of a set 453
Closure point 453
Co-finite function 411
Coercive function 496
Compact set 453
Complementary slackness 508
Component of a vector 444
Composition of functions 142, 455, 469, 514
Concave closure 502

- Concave function 468
 - Condition number 60, 122, 315
 - Conditional gradient method 71, 107, 191, 192, 374
 - Cone 468
 - Cone decomposition 219
 - Cone generated by a set 472, 489, 492
 - Confusion region 89, 93
 - Conic duality 14, 19, 23, 511
 - Conic programming 13, 217, 224, 231, 232, 423, 432, 511
 - Conjugate Subgradient Theorem 201, 513
 - Conjugacy Theorem 487
 - Conjugate direction method 64, 68, 320
 - Conjugate function 487
 - Conjugate gradient method 64
 - Constancy space 480
 - Constant stepsize rule 56, 59, 69, 153, 304, 349
 - Constraint aggregation 230
 - Constraint qualification 507
 - Continuity 455, 475
 - Continuous differentiability 141, 172, 457
 - Contraction mapping 56, 296, 312, 458
 - Convergent sequence 451, 452
 - Convex closure 476
 - Convex combination 472
 - Convex function 468, 469
 - Convex hull 472
 - Convex programming 4, 507
 - Convex set 468
 - Convexification of a function 476
 - Coordinate 444
 - Coordinate descent method 75, 104, 369, 429, 439-442
 - Crossing function 499
 - Cutting plane method 107, 182, 211, 228, 270
 - Cyclic coordinate descent 76, 370, 376
 - Cyclic incremental method 84, 96, 166, 343
- D**
- Danskin's Theorem 146, 172
 - Dantzig-Wolfe decomposition 107, 111, 229, 295
 - Decomposition algorithm 7, 77, 289, 363
 - Decomposition of a convex set 479
 - Derivative 456
 - Descent algorithms 54
 - Descent direction 58, 71
 - Descent inequality 122, 305
 - Diagonal dominance 106
 - Diagonal scaling 63, 101, 333, 338
 - Differentiability 457
 - Differentiable convex function 470
 - Differentiation theorems 457, 458, 513, 514
 - Dimension of a convex set 472
 - Dimension of a subspace 446
 - Dimension of an affine set 472
 - Diminishing stepsize rule 69, 127, 157, 174, 316, 351
 - Direct search methods 83
 - Direction of recession 478
 - Directional derivative 137, 170-173, 515
 - Distributed computation 8, 33, 104, 365, 376
 - Domain 444
 - Domain one-dimensional 409
 - Double conjugate 487
 - Dual cone 14, 511
 - Dual function 3, 499
 - Dual pair representation 219
 - Dual problem 2, 147, 164, 499, 506, 507
 - Dual proximal algorithm 257, 336, 384
 - Duality gap estimate 9
 - Duality theory 2, 498, 505
 - Dynamic programming 36, 380

Dynamic stepsize rule 159, 175, 177

E

ϵ -complementary slackness 180
 ϵ -descent algorithm 83, 396, 400, 431
 ϵ -descent direction 400
 ϵ -relaxation method 375
 ϵ -subdifferential 162, 397
 ϵ -subgradient 162, 164, 169, 180, 397
 ϵ -subgradient method 162, 179
 EMP 197, 220, 396, 406
 Effective domain 468
 Entropic descent 396
 Entropy function 383
 Entropy minimization algorithm 383
 Epigraph 468
 Essentially one-dimensional 408
 Euclidean norm 450
 Eventually constant stepsize rule 308, 334
 Exact penalty 39, 45, 365, 369
 Existence of dual optimal solutions 503
 Existence of optimal solutions 483, 495
 Exponential augmented Lagrangian method 116, 134, 389, 431
 Exponential loss 30
 Exponential smoothing 116, 134, 391
 Extended Kalman filter 103, 120
 Extended monotropic programming 83, 197, 220, 229, 396, 406, 431
 Extended real number 443
 Extended real-valued function 468
 Exterior penalty method 110
 Extrapolation 63-66, 322, 338, 427, 428
 Extreme point 490

F

Farkas' Lemma 492, 505, 506

Farout region 89
 Feasibility problem 34, 283, 429
 Feasible direction 71
 Feasible direction methods 71
 Feasible solution 2, 494
 Fejér Convergence Theorem 126, 158, 465
 Fejér monotonicity 464
 Fenchel duality 10, 510
 Fenchel inequality 512
 Finitely generated cone 492
 Forward-backward algorithm 427
 Forward image 445
 Frank-Wolfe algorithm 71, 107, 191, 374
 Fritz John optimality conditions 6
 Full rank 447
 Fundamental Theorem of Linear Programming 494

G

GPA algorithm 200-202, 229
 Gauss-Southwell order 376, 441
 Generalized polyhedral approximation 107, 196, 201, 229
 Generalized simplicial decomposition 209, 229
 Generated cone 472, 489, 492
 Geometric convergence 57
 Global maximum 495
 Global minimum 495
 Gradient 456
 Gradient method 56, 59
 Gradient method distributed 104, 106
 Gradient method with momentum 63, 92
 Gradient projection 73, 82, 136, 302, 374, 385, 396, 427, 434

H

Halfspace 484
 Heavy ball method 63, 92
 Hessian matrix 457
 Hierarchical decomposition 77

Hinge loss 30
 Hyperplane 484
 Hyperplane separation 484-487

I

Ill-conditioning 60, 109, 413
 Image 445, 446
 Improper function 468
 Incremental Gauss-Newton method 103, 120
 Incremental Newton method 97, 101, 118, 119
 Incremental aggregated method 91, 94
 Incremental constraint projection method 102, 365, 429
 Incremental gradient method 84, 105, 118, 119, 130-132
 Incremental gradient with momentum 92
 Incremental method 25, 83, 166, 320
 Incremental proximal method 341, 385, 429
 Incremental subgradient method 84, 166, 341, 385, 428
 Indicator function 487
 Inner linearization 107, 182, 188, 194, 296
 Infeasible problem 494
 Infimum 444
 Inner approximation 402
 Inner product 444
 Instability 186, 191, 269
 Integer programming 6
 Interior of a set 412, 453
 Interior point 453
 Interior point method 108, 412, 415, 423, 432
 Interpolated iteration 249, 253, 298, 459
 Inverse barrier 412
 Inverse image 445, 446

J**K**

Kaczmarz method 85, 98, 131
 Krasnosel'skii-Mann Theorem 252, 285, 300, 459

L

ℓ_1 -norm 451
 ℓ_∞ -norm 450
 LMS method 119
 Lagrange multiplier 507
 Lagrangian function 3, 507
 Lasso problem 27
 Least absolute value deviations 27, 288
 Least mean squares method 119
 Left-continuous function 455
 Level set 469
 Limit 451
 Limit point 451, 453
 Limited memory quasi-Newton method 63, 338
 Line minimization 60, 65, 69, 320
 Line segment principle 473
 Lineality space 479
 Linear-conic problems 15, 16
 Linear convergence 57
 Linear equation 445
 Linear function 445
 Linear inequality 445
 Linear programming 16, 415, 434
 Linear programming duality 506
 Linear regularity 369
 Linear transformation preservation of closedness 483
 Linearly independent vectors 446
 Lipschitz continuity 141, 455, 512
 Local convergence 68
 Local maximum 495
 Local minimum 495
 Location theory 32
 Logarithmic barrier 412, 416
 Logistic loss 30
 Lower limit 452

Lower semicontinuous function 455, 469

M

Majorization-maximization algorithm 392

Matrix completion 28, 35

Matrix factorization 30, 373

Max crossing problem 499

Max function 470

Maximal monotone mapping 255

Maximum likelihood 31

Maximum norm 450

Maximum point 444, 495

Mean Value Theorem 457, 458

Merit function 54, 417

Min common problem 499

Min common/max crossing framework 499

Minimax duality 502, 516

Minimax duality gap 9

Minimax duality theorems 516

Minimax equality 12, 516-518

Minimax problems 9, 12, 35, 113, 147, 164, 215, 217

Minimax theory 498, 502, 516

Minimizer 495

Minimum point 444, 495

Minkowski-Weyl Theorem 492

Minkowski-Weyl representation 492

Mirror descent 82, 385, 395

Momentum term 63, 92

Monotone mapping 255

Monotonically nondecreasing sequence 451

Monotonically nonincreasing sequence 451

Monotropic programming 83, 197, 208, 431

Multicommodity flow 38, 193, 217

Multiplier method 109, 261, 267

N

Negative halfspace 484

Neighborhood 453

Nelder-Mead algorithm 83

Nested sequence 481

Network optimization 37, 189, 193, 208, 217, 375

Newton's method 67, 74, 75, 97, 338, 416, 424

Nonexpansive mapping 57, 249, 459

Nonlinear Farkas' Lemma 505

Nonmonotonic stepsize rule 70

Nonnegative combination 472

Nonquadratic regularization 242, 294, 382, 393

Nonsingular matrix 447

Nonstationary iteration 57, 461

Nonvertical Hyperplane Theorem 486

Nonvertical hyperplane 486

Norm 450

Norm equivalence 454

Normal cone 145

Normal of a hyperplane 484

Nuclear norm 28, 35

Null step 212

Nullspace 447

O

Open ball 453

Open halfspace 484

Open set 453

Open sphere 453

Optimality conditions 3, 144, 470, 508-511, 514

Orthogonal complement 446

Orthogonal vectors 444

Outer approximation 402

Outer linearization 107, 182, 194

Overrelaxation 253

P

PARTAN 64

Parallel subspace 445

Parallel projections method 76, 438

Parallel tangents method 64

Partitioning 9

Partial cutting plane method 187

Partial derivative 456
 Partial minimization 496
 Partial proximal algorithm 297
 Partially asynchronous algorithm 105, 377
 Penalty method 38, 108, 120, 326
 Penalty parameter 38, 40, 109
 Perturbation function 501
 Polar Cone Theorem 488, 492
 Polar cone 488, 491
 Polyhedral Proper Separation Theorem 486
 Polyhedral approximation 107, 182, 196, 217, 385
 Polyhedral cone 492
 Polyhedral function 493
 Polyhedral set 468, 489
 Positive combination 472
 Positive definite matrix 449
 Positive halfspace 484
 Positive semidefinite matrix 449
 Positively homogeneous function 488, 489
 Power regularization 393
 Predictor-corrector method 421, 432
 Primal-dual method 416, 420
 Primal function 501
 Primal problem 2, 499
 Projection Theorem 471
 Prolongation Lemma 473
 Proper Separation Theorem 486
 Proper function 468
 Properly separating hyperplane 485, 486
 Proximal Newton algorithm 338, 428
 Proximal algorithm 80, 110, 120, 234, 293, 307, 374, 384, 393, 439
 Proximal cutting plane method 270
 Proximal gradient algorithm 82, 112, 133, 330, 336, 385, 428, 436-438
 Proximal inner linearization 278
 Proximal operator 248, 296
 Proximal simplicial decomposition 280
 Pythagorean Theorem 450

Q

Quadratic penalty function 39
 Quadratic programming 21, 483
 Quasi-Newton method 68, 74, 338

R

Randomized coordinate descent 376, 440
 Randomized incremental method 84, 86, 131, 344, 353, 429
 Range 444
 Range space 447
 Rank 447
 Recession Cone Theorem 478
 Recession cone of a function 480
 Recession cone of a set 478
 Recession direction 478
 Recession function 480
 Reflection operator 249
 Regression 26
 Regularization 26-31, 117, 133, 232, 235, 287, 361, 382, 393
 Relative boundary 473
 Relative boundary point 473
 Relative interior 473
 Relative interior point 473
 Relatively open 473
 Relaxation method 375
 Reshuffling 96
 Restricted simplicial decomposition 192, 204
 Retractive sequence 481
 Retractive set 482
 Right-continuous function 455
 Robust optimization 20, 47

S

Saddle Point Theorem 518
 Saddle point 498, 517-518
 Scaling 61
 Schwarz inequality 450
 Second order cone programming 17, 49, 50, 230, 423
 Second order expansions 458

- Second order method of multipliers 267
 - Self-dual cone 17
 - Semidefinite programming 17, 22, 424
 - Sensitivity 501
 - Separable problems 7, 289, 362, 405
 - Separating Hyperplane Theorem 485
 - Separating hyperplane 485
 - Sequence definitions 451
 - Serious step 272
 - Set intersection theorems 482
 - Shapley-Folkman Theorem 9
 - Sharp minimum 80, 176, 179, 242-244, 436
 - Shrinkage operation 133, 287, 330, 362
 - Side constraints 213, 217
 - Simplicial decomposition method 72, 107, 182, 188, 193, 209, 221, 228-230, 278, 280, 320
 - Single commodity network flow 38, 208, 375
 - Singular matrix 447
 - Slater condition 8, 505
 - Smoothing 113, 168, 326, 427
 - Span 446
 - Sphere 453
 - Splitting algorithm 113, 120, 427
 - Square root of a matrix 450
 - Stationary iteration 56
 - Steepest descent direction 59, 128
 - Steepest descent method 59, 78, 401
 - Stepsize rules 69, 308, 316
 - Stochastic Newton method 120
 - Stochastic approximation 94
 - Stochastic gradient method 93, 167
 - Stochastic programming 32
 - Stochastic subgradient method 93, 167
 - Strict Separation Theorem 485
 - Strict local minimum 495
 - Strictly convex function 468
 - Strictly separating hyperplane 485
 - Strong convexity 312, 435, 440, 471
 - Strong duality 3, 499
 - Strong monotonicity 296
 - Subdifferential 136, 141, 143, 146, 167, 171, 511-514
 - Subgradient 136, 147, 148, 167, 511-514
 - Subgradient methods 78, 136, 179
 - Subsequence 452
 - Subspace 445
 - Superlinear convergence 68, 100, 338, 393, 437
 - Supermartingale Convergence 462-464
 - Support function 488
 - Support vector machine 30, 48
 - Supporting Hyperplane Theorem 484
 - Supporting hyperplane 484
 - Supremum 444
 - Symmetric matrix 449
- T**
- Theorems of the alternative 52
 - Tikhonov regularization 111
 - Total variation denoising 28
 - Totally asynchronous algorithm 105, 376
 - Triangle inequality 450
 - Trust region 70, 296
 - Two-metric projection method 74, 102, 129, 189, 374
- U**
- Uniform sampling 96, 344
 - Upper limit 452
 - Upper semicontinuous function 455, 469
- V**
- Vector sum preservation of closedness 483
 - Vertical hyperplane 486
- W**
- Weak Duality Theorem 3, 7, 499, 507

Weber point 32, 52
Weierstrass' Theorem 456, 495
Weighted sup-norm 380

Y**Z**

Zero-sum games 9