Universitatea Politehnica Bucureşti
Facultatea de Automatică şi Calculatoare
Departamentul de Automatică şi Ingineria Sistemelor

TEZĂ DE ABILITARE

Metode de Descreştere pe Coordonate pentru Optimizare Rară

(Coordinate Descent Methods for Sparse Optimization)

Ion Necoară

2013

Chapter 2

Summary

2.1 Contributions of this thesis

The main optimization problem of interest considered in this thesis has the following form:

$$\min_{x\in\mathbb{R}^n} F(x) \quad (= f(x) + \Psi(x)) \tag{2.1}$$
$$\text{s.t.:} \quad Ax = b,$$

where f is a smooth function (i.e. with Lipschitz gradient), Ψ is a simple convex function (i.e. minimization of the sum of this function with a quadratic term is easy) and matrix $A \in \mathbb{R}^{m\times n}$ is usually sparse according to some graph structure. Another characteristic of this problem is its very large dimension, i.e. n is very large, in particular we deal with millions or even billions of variables. We further assume that the decision variable x can be decomposed in (block) components $x = [x_1^T \; x_2^T \dots x_N^T]^T$, where $x_i \in \mathbb{R}^{n_i}$ and $\sum_i n_i = n$. Note that this problem is very general and appears in many engineering applications:

- Ψ is the indicator function of some convex set X that can be written usually as a Cartesian product $X = X_1 \times X_2 \times \cdots \times X_N$, where $X_i \subseteq \mathbb{R}^{n_i}$. This problem is known in the literature as *separable optimization problem with linear coupling constraints* and appears in many applications from distributed control and estimation [13,62,65,100,112], network optimization [9,22,82,98,110,121], computer vision [10,44], etc.

- Ψ is either the indicator function of some convex set $X = X_1 \times X_2 \times \cdots \times X_N$ or 1-norm $\|x\|_1$ (in order to induce sparsity in the solution) and matrix $A = 0$. This problem appears in distributed model predictive control [61,103], image processing [14,21,47,105], classification [99,123,124], data mining [16,86,119], etc.

- Ψ is the indicator function of some convex set $X = X_1 \times X_2 \times \cdots \times X_N$ and $A = a^T$, i.e. a single linear coupled constraint. This problem appears is page ranking (also known as Google problem) [59,76], control [39,83,84,104], learning [16–18,109,111], truss topology design [42], etc.

We notice that (2.1) belongs to the class of large-scale optimization problems with sparse data/solutions. The standard approach for solving the large-scale optimization problem (2.1) is to use decomposition. Decomposition methods represent a powerful tool for solving these type of problems due to their ability of dividing the original large-scale problem into smaller subproblems which are coordinated by a master problem. Decomposition methods can be divided in two main classes: primal and dual decomposition. While in the primal decomposition methods the optimization problem is solved using the original formulation and variables, in dual decomposition the constraints are moved into the cost using the Lagrange multipliers and the dual problem is solved. In the last 7 years I have pursued both approaches in my research. From my knowledge I am one of the first researchers that used smoothing techniques in Lagrangian dual decomposition in order to obtain faster convergence rates for the corresponding algorithms (see e.g. the papers [64, 65, 71, 72, 90, 91, 110]). In this thesis however, I have opted to present some of my recent results on primal decomposition, namely coordinate descent methods (see e.g. the papers [59–61, 65, 67, 70, 84]). The main contributions of this thesis, by chapters, are as follows:

Chapter 1: In this chapter we develop random (block) coordinate descent methods for minimizing large-scale convex problems with linearly coupled constraints and prove that it obtains in expectation an ϵ-accurate solution in at most $\mathcal{O}(\frac{1}{\epsilon})$ iterations. Since we have coupled constraints in the problem, we need to devise an algorithm that updates randomly two (block) components per iteration. However, the numerical complexity per iteration of the new methods is usually much cheaper than that of methods based on full gradient information. We focus on how to choose the probabilities to make the randomized algorithm to converge as fast as possible and we arrive at solving sparse SDPs. Analysis for rate of convergence in probability is also provided. For strongly convex functions the new methods converge linearly. We also extend the main algorithm, where we update two (block) components per iteration, to a parallel random coordinate descent algorithm, where we update more than two (block) components per iteration and we show that for this parallel version the convergence rate depends linearly on the number of (block) components updated. Numerical tests confirm that on large optimization problems with cheap coordinate derivatives the new methods are much more efficient than methods based on full gradient. This chapter is based on papers [67,68].

Chapter 2: In this chapter we develop randomized block-coordinate descent methods for minimizing multi-agent convex optimization problems with a single linear coupled constraint over networks and prove that they obtain in expectation an ϵ accurate solution in at most $\mathcal{O}(\frac{1}{\lambda_2(Q)\epsilon})$ iterations, where $\lambda_2(Q)$ is the second smallest eigenvalue of a matrix Q that is defined in terms of the probabilities and the number of blocks. However, the computational complexity per iteration of our methods is much simpler than the one of a method based on full gradient information and each iteration can be computed in a completely distributed way. We focus on how to choose the probabilities to make these randomized algorithms to converge as fast as possible and we arrive at solving a sparse SDP. Analysis for rate of convergence in probability is also provided. For strongly convex functions our distributed algorithms converge linearly. We also extend the main algorithm to a parallel random coordinate descent method and to problems with more general linearly coupled constraints for which we also derive rate of convergence. The new algorithms were implemented in Matlab and applied for solving the Google problem, and the simulation results show the superiority of our approach compared to methods based on full gradient. This chapter is based on papers [58, 59, 69].

Chapter 3: In this chapter we propose a variant of the random coordinate descent method for solving linearly constrained convex optimization problems with composite objective functions. If the smooth part of the objective function has Lipschitz continuous gradient, then we prove that our method obtains an ϵ-optimal solution in $\mathcal{O}(N^2/\epsilon)$ iterations, where N is the number of blocks. For the class of problems with cheap coordinate derivatives we show that the new method is faster than methods based on full-gradient information. Analysis for the rate of convergence in probability is also provided. For strongly convex functions our method converges linearly. The proposed algorithm was implemented in code C and tested on real data from SVM and on the problem of finding the Chebyshev center for a set of points. Extensive numerical tests confirm that on very large problems, our method is much more numerically efficient than methods based on full gradient information or coordinate descent methods based on greedy index selection. This chapter is based on paper [70].

Chapter 4: In this chapter we analyze several new methods for solving nonconvex optimization problems with the objective function formed as a sum of two terms: one is nonconvex and smooth, and another is convex but simple and its structure is known. Further, we consider both cases: unconstrained and linearly constrained nonconvex problems. For optimization problems of the above structure, we propose random coordinate descent algorithms and analyze their convergence properties. For the general case, when the objective function is nonconvex and composite we prove asymptotic convergence for the sequences generated by our algorithms to stationary points and sublinear rate of convergence in expectation for some optimality measure. Additionally, if the objective function satisfies an error bound condition we derive a local linear rate of convergence for the expected values of the objective function. We also present extensive numerical experiments on the eigenvalue complementarity problem for evaluating the performance of our algorithms in comparison with state-of-the-art methods. From the numerical experiments we can observe that on large optimization problems the new methods are much more efficient than methods based on full gradient. This chapter is based on papers [84,85].

Chapter 5: In this chapter we propose a distributed version of a randomized (block) coordinate descent method for minimizing the sum of a partially separable smooth convex function and a fully separable non-smooth convex function. Under the assumption of block Lipschitz continuity of the gradient of the smooth function, this method is shown to have a sublinear convergence rate. Linear convergence rate of the method is obtained for the newly introduced class of generalized error bound functions. We prove that the new class of generalized error bound functions encompasses both global/local error bound functions and smooth strongly convex functions. We also show that the theoretical estimates on the convergence rate depend on the number of blocks chosen randomly and a natural measure of separability of the objective function. The new algorithm was implemented in code C and tested on the constrained lasso problem. Numerical experiments show that by parallelization we can accelerate substantially the rate of convergence of the classical random coordinate descent method. This chapter is based on paper [60].

Chapter 6: In this chapter we propose a parallel coordinate descent algorithm for solving smooth convex optimization problems with separable constraints that may arise e.g. in distributed model predictive control (MPC) for linear networked systems. Our algorithm is based on block coordinate descent updates in parallel and has a very simple iteration. We prove (sub)linear rate of convergence for the new algorithm under standard assumptions for smooth convex optimization. Further, our algorithm uses local information and thus is suitable for distributed implementations. Moreover, it has low iteration complexity, which makes it appropriate for embedded control. An MPC scheme based on this new parallel algorithm is derived, for which every subsystem in the network can compute feasible and stabilizing control inputs using distributed and cheap computations. For ensuring stability of the MPC scheme, we use a terminal cost formulation derived from a distributed synthesis. The proposed control method was q implemented on a PLC Siemens for controlling a four tank process. This chapter is based on paper [61].

Bibliography

[1] I. Alvarado, D. Limon, D. Munoz de la Pena, J.M. Maestre, M.A. Ridao, H. Scheu, W. Marquardt, R.R. Negenborn, B. De Schutter, F. Valencia, and J. Espinosa. A comparative analysis of distributed mpc techniques applied to the hd-mpc four-tank benchmark. *Journal of Process Control*, 21(5):800 –815, 2011.

[2] A. Auslender. *Optimisation Methodes Numeriques*. Masson, 1976.

[3] R.O. Barr and E.G. Gilbert. Some effcient algortihms for a class of abstract optimization problems arising in optimal control. *IEEE Transaction on Automatic Control*, 14:640–652, 1969.

[4] H. Bauschke and J.M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.

[5] A. Beck. The 2-coordinate descent method for solving double-sided simplex constrained minimization problems. Technical report, Israel Institute of Technology, Haifa, Israel, dec 2012.

[6] A. Beck and L. Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM Journal on Optimization*, 23(4):2037–2060, 2013.

[7] P. Berman, N. Kovoor, and P. M. Pardalos. *Algorithms for least distance problem, Complexity in Numerical Optimization*. World Scientific, 1993.

[8] D. P. Bertsekas. *Nonlinear Programming, 2nd edition*. Athena Scientific, Belmont, MA, 1999.

[9] D.P. Bertsekas and J. Tsitsiklis. *Paralel and Distributed Computation: Numerical Methods*. Prentice Hall, 1989.

[10] C.M. Bishop. *Pattern Recognition and Machine Learning*. Springer-Verlag, New York, 2006.

[11] S. Bonettini. Inexact block coordinate descent methods with application to nonnegative matrix factorization. *Journal of Numerical Analysis*, 22:1431–1452, 2011.

[12] P.H. Calamai and J.J. More. Projected gradient methods for linearly constrained problems. *Mathematical Programming*, 39:93–116, 1987.

[13] E. Camponogara and H.F. Scherer. Distributed optimization for model predictive control of linear dynamic networks with control-input and output constraints. *IEEE Transactions on Automation Science and Engineering*, 8(1):233–242, 2011.

[14] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52:489–509, 2006.

[15] A. Canutescu and R.L. Dunbrack. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Science*, 12:963–972, 2003.

[16] C. C. Chang and C. J. Lin. Libsvm: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 27:1–27, 2011.

[17] K.W. Chang, C.J. Hsieh, and C.J. Lin. Coordinate descent method for large-scale l_2-loss linear support vector machines. *Journal of Machine Learning Research*, 9:1369–1398, 2008.

[18] O. Chapelle, V. Sindhwani, and S. Keerthi. Optimization techniques for semi-supervised support vector machines. *Journal of Machine Learning Research*, 2:203–233, 2008.

[19] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Review*, 43:129–159, 2001.

[20] X. Chen, M. K. Ng, and C. Zhang. Non-lipschitz ℓ_p -regularization and box constrained model for image restoration. *IEEE Transactions on Image Processing*, 21(12):4709–4721, 2012.

[21] P. L. Combettes. The convex feasibility problem in image recovery. In P. Hawkes, editor, *Advances in Imaging and Electron Physics*, pages 155–270. Academic Press, 1996.

[22] A.J. Connejo, R. Minguez, E. Castillo, and R. Garcia-Bertrand. *Decomposition Techniques in Mathematical Programming: Engineering and Science Applications*. Springer-Verlag, 2006.

[23] J.R. Correa, A.S. Schulz, and N.E. Stier Moses. Selfish routing in capacitated networks. *Mathematics of Operations Research*, pages 961–976, 2004.

[24] F. Deutsch and H. Hundal. The rate of convergence for the cyclic projections algorithm i: Angles between convex sets. *Journal of Approximation Theory*, 142:36–55, 2006.

[25] F. Deutsch and H. Hundal. The rate of convergence for the cyclic projections algorithm ii: Regularity of convex sets. *Journal of Approximation Theory*, 155:155–184, 2008.

[26] L. Fainshil and M. Margaliot. A maximum principle for positive bilinear control systems with applications to positive linear switched systems. *SIAM Journal of Control and Optimization*, 50:2193–2215, 2012.

[27] M. Farina and R. Scattolini. Distributed predictive control: a non-cooperative algorithm with neighbor-to-neighbor communication for linear systems. *Automatica*, 2012.

[28] C. Godsil and G. Royle. *Algebraic graph theory*. Springer, 2001.

[29] D. Goldfarb and S. Ma. Fast multiple splitting algorithms for convex optimization. Technical report, Department of IEOR, Columbia University, 2010.

[30] L.G. Gubin, B.T. Polyak, and E.V. Raik. The method of projections for finding the common point of convex sets. *Computational Mathematics and Mathematical Physics*, 7(6):1211–1228, 1967.

[31] G.M. Heal. Planning without prices. *Review of Economic Studies*, 36:347–362, 1969.

[32] Y.C. Ho, L.D. Servi, and R. Suri. A class of center-free resource allocation algorithms. *Large Scale Systems*, 1:51–62, 1980.

[33] B. Hu and A. Linnemann. Toward infinite-horizon optimality in nonlinear model predictive control. *IEEE Transactions on Automatic Control*, 47(4):679–682, 2002.

[34] L. Hurwicz. The design of mechanisms for resource allocation. *American Economic Review*, 63:1–30, 1973.

[35] D. Hush, P. Kelly, C. Scovel, and I. Steinwart. Qp algorithms with guaranteed accuracy and run time for support vector machines. *Journal of Machine Learning Research*, 7:733–769, 2006.

[36] G.M. James, C. Paulson, and P. Rusmevichientong. The constrained lasso. Technical report, University of Southern California, 2013.

[37] J.L. Jerez, K.V. Ling, G.A. Constantinides, and E.C. Kerrigan. Model predictive control for deeply pipelined field-programmable gate array implementation: algorithms and circuitry. *IET Control Theory and Applications*, 6(8):1029–1041, 2012.

[38] A. Jokic and M. Lazar. On decentralized stabilization of discrete-time nonlinear systems. In *Proceedings of American Control Conference*, pages 5777–5782, 2009.

[39] J. Judice, M. Raydan, S. Rosa, and S. Santos. On the solution of the symmetric eigenvalue complementarity problem by the spectral projected gradient algorithm. *Numerical Algorithms*, 47:391–407, 2008.

[40] K. C. Kiwiel. On linear-time algorithms for the continuous quadratic knapsack problem. *Journal of Optimization Theory and Applications*, 134:549–554, 2007.

[41] D. Knuth. *The art of computer programming*. Addison-Wesley, Boston, USA, 1981.

[42] M. Kocvara and J. Outrata. Effective reformulations of the truss topology design problem. *Optimization and Engineering*, 2006.

[43] P. Komarek and A. Moore. Fast robust logistic regression for large sparse datasets with binary outputs. In *Artificial Intelligence and Statistics*, 2003.

[44] N. Komodakis, N. Paragios, and G. Tziritas. Mrf energy minimization & beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):531–552, 2011.

[45] J. Kurose and R. Simha. Microeconomic approach to optimal resource allocation in distributed computer systems. *IEEE Transactions on Computers*, 38:705–717, 1989.

[46] D. Leventhal and A.S. Lewis. Randomized methods for linear constraints: Convergence rates and conditioning. Technical report, Cornell University, 2008. http://arxiv.org/abs/0806.3015.

[47] Y. Li and S. Osher. Coordinate descent optimization for l_1 minimization with application to compressed sensing; a greedy algorithm. *Inverse Problems and Imaging*, 3:487–503, 2009.

[48] D. Limon, T. Alamo, and E.F. Camacho. Stable constrained mpc without terminal constraint. In *Proceedings of American Control Conference*, pages 4893–4898, 2003.

[49] C. J. Lin, S. Lucidi, L. Palagi, A. Risi, and M. Sciandrone. A decomposition algorithm model for singly linearly constrained problems subject to lower and upper bounds. *Journal of Optimization Theory and Applications*, 141:107–126, 2009.

[50] N. List and H. U. Simon. General polynomial time decomposition algorithms. *Lecture Notes in Computer Science*, 3559:308–322, 2005.

[51] Z.Q. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: A general approach. *Annals of Operations Research*, 46–47(1):157–178, 1993.

[52] Z.Q. Luo and P. Tseng. On the convergence rate of dual ascent methods for linearly constrained convex minimization. *Mathematics of Operations Research*, 18(2):846–867, 1993.

[53] Z.Q. Luo and P. Tseng. A coordinate gradient descent method for nonsmooth separable minimization. *Journal of Optimization Theory and Applications*, 72(1), 2002.

[54] S. Ma and S. Zhang. An extragradient-based alternating direction method for convex minimization. Technical report, Chinese University of Hong Kong, January 2013.

[55] O.L. Mangasarian. Computable numerical bounds for lagrange multipliers of stationary points of non-convex differentiable non-linear programs. *Operations Research Letters*, 4(2):47–48, 1985.

[56] M. Mongeau and M. Torki. Computing eigenelements of real symmetric matrices via optimization. *Computational Optimization and Applications*, 29:263–287, 2004.

[57] M.V. Nayakkankuppam. Solving large-scale semidefinite programs in parallel. *Mathematical Programming*, 109:477–504, 2007.

[58] I. Necoara. A random coordinate descent method for large-scale resource allocation problems. In *Proceedings of 51th IEEE Conference on Decision and Control*, 2012.

[59] I. Necoara. Random coordinate descent algorithms for multi-agent convex optimization over networks. *IEEE Transactions on Automatic Control*, 58(8):2001–2012, 2013.

[60] I. Necoara and D. Clipici. Distributed random coordinate descent methods for composite optimization. *SIAM Journal of Optimization*, partially accepted:1–41, 2013.

[61] I. Necoara and D. Clipici. Efficient parallel coordinate descent algorithm for convex optimization problems with separable constraints: application to distributed MPC. *Journal of Process Control*, 23(3):243–253, 2013.

[62] I. Necoara, D. Doan, and J. A. K. Suykens. Application of the proximal center decomposition method to distributed model predictive control. In *Proceedings of the Conference on Decision and Control*, pages 2900–2905, 2008.

[63] I. Necoara, L. Ferranti, and T. Keviczky. An adaptive constraint tightening approach to linear mpc based on approximation algorithms for optimization. *Optimal Control Applications and Methods*, in press:1–18, 2014.

[64] I. Necoara and V. Nedelcu. Rate analysis of inexact dual first order methods: application to dual decomposition. *IEEE Trans. Automatic Control*, DOI: 10.1109/TAC.2013.2294614:1–12, 2013.

[65] I. Necoara, V. Nedelcu, and I. Dumitrache. Parallel and distributed optimization methods for estimation and control in networks. *Journal of Process Control*, 21(5):756–766, 2011.

[66] I. Necoara, V. Nedelcu, T. Keviczky, M. D. Doan, and B. de Schutter. Stability of linear model predictive control based on tightening and approximate optimal control inputs. In *Proceedings of 52nd Conference on Decision and Control*, 2013.

[67] I. Necoara, Y. Nesterov, and F. Glineur. A random coordinate descent method on large optimization problems with linear constraints. In *Int. Conference on Continuous Optimization*, 2013.

[68] I. Necoara, Y. Nesterov, and F. Glineur. A random coordinate descent method on large-scale optimization problems with linear constraints. Technical report, University Politehnica Bucharest, June 2013.

[69] I. Necoara and A. Patrascu. A random coordinate descent algorithm for singly linear constrained smooth optimization. In *Proceedings of 20th Mathematical Theory of Network and Systems*, 2012.

[70] I. Necoara and A. Patrascu. A random coordinate descent algorithm for optimization problems with composite objective function and linear coupled constraints. *Computational Optimization and Applications*, DOI:10.1007/s10589-013-9598-8, 2013.

[71] I. Necoara and J.A.K. Suykens. Application of a smoothing technique to decomposition in convex optimization. *IEEE Transactions on Automatic Control*, 53(11):2674–2679, 2008.

[72] I. Necoara and J.A.K. Suykens. An interior-point lagrangian decomposition method for separable convex optimization. *J. Optimization Theory and Applications*, 143(3):567–588, 2009.

[73] V. Nedelcu, I. Necoara, and D. Q. Quoc. Computational complexity of inexact gradient augmented lagrangian methods: application to constrained mpc. *SIAM Journal on Control and Optimization*, 52(5):1–26, 2014.

[74] A. Nedic, A. Ozdaglar, and A.P. Parrilo. Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, 55(4):922–938, 2010.

[75] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer, Boston, USA, 2004.

[76] Y. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.

[77] Y. Nesterov. Subgradient methods for huge-scale optimization problems. *CORE DIscussion Paper*, 2012/02, 2012.

[78] Y. Nesterov. Gradient methods for minimizing composite objective functions. *Mathematical Programming*, 140:125–161, 2013.

[79] Y. Nesterov. Gradient methods for minimizing composite objective functions. *Mathematical Programming*, 140(1):125–161, 2013.

[80] Y. Nesterov and S. Shpirko. Primal-dual subgradient method for huge-scale linear conic problems. Technical report, CORE, UCL, Louvain, Belgium, 2012. `http://www.optimization-online.org/DB_FILE/2012/08/3590.pdf`.

[81] F. Niu, B. Recht, C. Re, and S. Wright. Hogwild!: A lock-free approach to parallelizing stochastic gradient descent. *NIPS*, 2012.

[82] D.P. Palomar and M. Chiang. A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications*, 24(8):1439–1451, 2006.

[83] B.N. Parlett. *The Symmetric Eigenvalue Problem*. SIAM, 1997.

[84] A. Patrascu and I. Necoara. Efficient random coordinate descent algorithms for large-scale structured nonconvex optimization. *Journal of Global Optimization*, DOI: 10.1007/s10898-014-0151-9:1–31, 2013.

[85] A. Patrascu and I. Necoara. A random coordinate descent algorithm for large-scale sparse nonconvex optimization. In *Proceedings of 12th European Control Conferencel*, 2013.

[86] J. C. Platt. Fast training of support vector machines using sequential minimal optimization. *Advances in Kernel Methods: Support Vector Learning*, MIT Press, 1999.

[87] B.T. Poliak. *Introduction to Optimization*. Optimization Software, 1987.

[88] M.J.D. Powell. On search directions for minimization algorithms. *Mathematical Programming*, 4:193–201, 1973.

[89] Z. Qin, K. Scheinberg, and D. Goldfarb. Efficient block-coordinate descent algorithms for the group lasso. *Mathematical Programming Computation*, 5(2):143–169, 2013.

[90] D.Q. Quoc, I. Necoara, and M. Diehl. Path-following gradient-based decomposition algorithms for separable convex optimization. *Journal of Global Optimization*, DOI: 10.1007/s10898-013-0085-7:1–25, 2013.

[91] D.Q. Quoc, I. Necoara, I. Savorgnan, and M. Diehl. An inexact perturbed path-following method for lagrangian decomposition in large-scale separable convex optimization. *SIAM Journal of Optimization*, 23(1):95–125, 2013.

[92] J.B. Rawlings and D.Q. Mayne. *Model Predictive Control: Theory and Design*. Nob Hill Publishing, 2009.

[93] P. Richtarik and M. Takac. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming*, 2012.

[94] P. Richtarik and M. Takac. Parallel coordinate descent methods for big data optimization. Technical report, University of Edinburgh, Scotland, December 2012.

[95] S. M. Robinson. Bounds for error in the solution set of a perturbed linear program. *Linear Algebra and its Applications*, 6:69–81, 1973.

[96] R.T. Rockafellar and R.J. Wets. *Variational Analysis*. Springer-Verlag, New York, 1998.

[97] R.T. Rockafeller. The elementary vectors of a subspace in r^n. In *Combinatorial Mathematics and its Applications*, pages 104–127, Chapel Hill, North Carolina, 1969.

[98] R.T. Rockafeller. *Network Flows and Monotropic Optimization*. Wiley-Interscience, 1984.

[99] S. Ryali, K. Supekar, D. A. Abrams, and V. Menone. Sparse logistic regression for whole-brain classication of fmri data. *NeuroImage*, 51(2):752–764, 2010.

[100] S. Samar, S. Boyd, and D. Gorinevsky. Distributed estimation via dual decomposition. In *Proceedings European Control Conference (ECC)*, pages 1511–1516, 2007.

[101] H. Schwartz. Uber einen grenzubergang durch alternierendes verfahren. *Vierteljahrsschrift der Naturforschenden Gesselschaft in Zurich*, 15:272–286, 1870.

[102] P.O.M. Scokaert, D.Q. Mayne, and J.B. Rawlings. Suboptimal model predictive control (feasibility implies stability). *IEEE Transactions on Automatic Control*, 44(3):648–654, 1999.

[103] B. T. Stewart, A.N. Venkat, J.B. Rawlings, S. Wright, and G. Pannocchia. Cooperative distributed model predictive control. *Systems & Control Letters*, 59:460–469, 2010.

[104] H.A.L. Thi, M. Moeini, T.P. Dihn, and J. Judice. A dc programming approach for solving the symmetric eigenvalue complementarity problem. *Computational Optimization and Applications*, 51:1097–1117, 2012.

[105] R. Tibshirani and J. Taylor. The solution path of the generalized lasso. *Annals of Statistics*, 39(3):1335–1371, 2011.

[106] P. Tseng. Approximation accuracy, gradient methods and error bound for structured convex optimization. *Mathematical Programming*, 125(2):263–295, 2010.

[107] P. Tseng and S. Yun. A block-coordinate gradient descent method for linearly constrained nonsmooth separable optimization. *Journal of Optimization Theory and Applications*, 140:513–535, 2009.

[108] P. Tseng and S. Yun. A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming*, 117(1–2):387–423, 2009.

[109] P. Tseng and S. Yun. A coordinate gradient descent method for linearly constrained smooth optimization and support vector machine training. *Computational Optimization and Applications*, 47:179–206, 2010.

[110] P. Tsiaflakis, I. Necoara, J. Suykens, and M. Moonen. Improved dual decomposition based optimization for dsl dynamic spectrum management. *IEEE Transactions on Signal Processing*, 58(4):2230–2245, 2010.

[111] V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[112] A.N Venkat, I.A. Hiskens, J.B Rawlings, and S. Wright. Distributed mpc strategies with application to power system automatic generation control. *IEEE Transactions on Control Systems Technology*, 16(6):1192–1206, 2008.

[113] M. Vose. A linear algorithm for generating random numbers with a given distribution. *IEEE Transactions on Software Engineering*, 17(9):972–975, 1991.

[114] M. Wang and D. P. Bertsekas. Incremental constraint projection-proximal methods for nonsmooth convex optimization. Technical report, MIT, July 2013.

[115] P.W. Wang and C.J. Lin. Iteration complexity of feasible descent methods for convex optimization. Technical report, National Taiwan University, 2013.

[116] P.W. Wang and C.J. Lin. Iteration complexity of feasible descent methods for convex optimization. Technical report, Department of Computer Science, National Taiwan University, 2013.

[117] E. Wei, A. Ozdaglar, and A. Jadbabaie. A distributed newton method for network utility maximization. Technical Report 2832, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2011.

[118] A. Wills. *QPC - Quadratic Programming in C*. University of Newcastle, Australia, 2009.

[119] I.H. Witten, E. Frank, and M.A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier, New York, 2011.

[120] S. Wright. Accelerated block coordinate relaxation for regularized optimization. Technical report, University of Wisconsin, 2010.

[121] L. Xiao and S. Boyd. Optimal scaling of a gradient method for distributed resource allocation. *Journal of Optimization Theory and Applications*, 129(3), 2006.

[122] S. Xu, M. Freund, and J. Sun. Solution methodologies for the smallest enclosing circle problem. *Computational Optimization and Applications*, 25:283–292, 2003.

[123] G.X. Yuan, K.W. Chang, C.J. Hsieh, and C.J. Lin. A comparison of optimization methods and software for large-scale l_1-regularized linear classification. *Journal of Machine Learning Research*, 11:3183–3234, 2010.

[124] G.X. Yuan, C.H. Ho, and C.J. Lin. Recent advances of large-scale linear classification. Technical report, Department of Computer Science, National Taiwan University, 2011.

[125] S. Yun and K.C. Toh. A coordinate gradient descent method for l_1-regularized convex minimization. *Computational Optimization and Applications*, 48:273–307, 2011.

[126] G. Zhao. A lagrangian dual method with self-concordant barriers for multistage stochastic convex programming. *Mathematical Programming*, 102:1–24, 2005.