

Bibliography for

ICML 2014 Tutorial on Frank-Wolfe and Greedy Convex Optimization

June 21, 2014

See also the references within the two recent overview papers [Jag13, HJN14].

1 About the Frank-Wolfe Method

The following references represent work about the Frank-Wolfe method itself:

Pioneers: [FW56] original paper

[Wol61] convex optimization duality

[LP66] strongly convex sets and functions, conditional gradient terminology

[DR67, DR70] banach space setting

[DH78, Dun79] approximate oracles

[Hol74, Hoh77] fully corrective variant (simplicial decomposition)

More Recent Work: [Jag13, HJN14, FG13, Lan13] modern overview papers

[LP66, GM86, AST08, GH13, LJJ13] linear convergence under additional assumptions

[SSSZ10, Jag11, Zha11] accuracy - sparsity trade-off

[GH14] $1/t^2$ rate for strongly convex domains

[Wol70, MDM74, GM86, AFNS14, RSW14, LJJ13] away steps and swap steps

[HJN14, ZYS12] FW for penalized (instead of constrained) problems

[HJN14, Lan13, ASS14, HK12] non-smooth objectives

[MZWG14] mixing FW and proximal methods

[LJJSP13] block-version of FW

[Cla10, Jag11] duality gap with a rate, and (affine invariant) curvature

[DR67, Hea82] duality gap without a rate

[Pat93] different alternatives to linear subproblems

[DHM11, Jag11, TRD11],[DHM12, Appendix D] FW for atomic domains

[Tem12] related greedy methods from the perspective of convex analysis

[Bac13b] duality between subgradient and FW methods

[Mai13] FW as a case of optimization with first-order surrogate functions

[GN13] information theoretic lower bounds for FW

2 Applications of Frank-Wolfe Methods

A very incomplete collection of some applications:

[Fuk84] transportation problems

[Bac13c] submodular optimization

[Haz08, JS10] trace-norm application and SDPs

[WJW05] spanning tree polytope

- [dBG12, LT13] sparse PCA
- [HJN14] total-variation norm for image processing
- [TKC05, GJ09, Cla10, OG10, AFÑS14, NKT14] kernel SVM training
- [LJJSP13, BSM13] structured prediction
- [BLJO12] herding (learning distributions)
- [Bac13a, Jag13] matrix factorizations
- [LSSS13] greedy optimization for training deep neural networks
- [TBSR13] infinite number of atoms, learning a sparse Fourier decomposition

3 Additional References used in the Tutorial Slides

Among others:

- [MZ93] matching pursuit
- [CRPW12] atomic norms

References

- [AFÑS14] Hector Allende, Emanuele Frandi, Ricardo Ñanculef, and Claudio Sartori. A novel Frank–Wolfe algorithm. Analysis and applications to large-scale SVM training. *Information Sciences*, March 2014.
- [ASS14] Andreas Argyriou, Marco Signoretto, and Johan Suykens. Hybrid Conditional Gradient - Smoothing Algorithms with Applications to Sparse and Low Rank Regularization. *arXiv.org*, April 2014.
- [AST08] Selin Damla Ahipaaoğlu, Peng Sun, and Michael Todd. Linear Convergence of a Modified Frank–Wolfe Algorithm for Computing Minimum-Volume Enclosing Ellipsoids. *Optimization Methods and Software*, 23(1):5–19, February 2008.
- [Bac13a] Francis Bach. Convex relaxations of structured matrix factorizations. *arXiv.org*, September 2013.
- [Bac13b] Francis Bach. Duality between subgradient and conditional gradient methods. *arXiv.org*, 2013.
- [Bac13c] Francis Bach. Learning with Submodular Functions: A Convex Optimization Perspective. *Foundations and Trends in Machine Learning*, 6(2-3):145–373, 2013.
- [BLJO12] Francis Bach, Simon Lacoste-Julien, and Guillaume Obozinski. On the Equivalence between Herding and Conditional Gradient Algorithms. In *ICML*, 2012.
- [BSM13] David Belanger, Dan Sheldon, and Andrew McCallum. Marginal Inference in MRFs using Frank-Wolfe. In *NIPS 2013 Workshop on Greedy Algorithms and Frank-Wolfe*, pages 1–5, December 2013.
- [Cla10] Kenneth L Clarkson. Coresets, Sparse Greedy Approximation, and the Frank-Wolfe Algorithm. *ACM Transactions on Algorithms*, 6(4), August 2010.
- [CRPW12] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The Convex Geometry of Linear Inverse Problems. *Foundations of Computational Mathematics*, 12(6):805–849, October 2012.
- [dBG12] Alexandre d’Aspremont, Francis Bach, and Laurent El Ghaoui. Approximation Bounds for Sparse Principal Component Analysis. *arXiv.org*, May 2012.
- [DH78] Joseph C Dunn and S Harshbarger. Conditional Gradient Algorithms with Open Loop Step Size Rules. *Journal of Mathematical Analysis and Applications*, 62(2):432–444, February 1978.
- [DHM11] Miroslav Dudk, Zaid Harchaoui, and Jerome Malick. Learning with matrix gauge regularizers. In *NIPS Optimization Workshop*, 2011.
- [DHM12] Miroslav Dudk, Zaid Harchaoui, and Jerome Malick. Lifted coordinate descent for learning with trace-norm regularization. In *AISTATS*, March 2012.

- [DR67] Vladimir Fedorovich Demyanov and Aleksandr Moiseevich Rubinov. The Minimization of a Smooth Convex Functional on a Convex Set. *SIAM Journal on Control*, 5(2), 1967.
- [DR70] Vladimir Fedorovich Demyanov and Aleksandr Moiseevich Rubinov. *Approximate methods in optimization problems*. Elsevier Publishing Company, 1970.
- [Dun79] Joseph C Dunn. Rates of Convergence for Conditional Gradient Algorithms Near Singular and Nonsingular Extremals. *SIAM Journal on Control and Optimization*, 17(2):187–211, 1979.
- [FG13] Robert M Freund and Paul Grigas. New Analysis and Results for the Frank-Wolfe Method. *arXiv.org*, July 2013.
- [Fuk84] Masao Fukushima. A modified Frank-Wolfe algorithm for solving the traffic assignment problem. *Transportation Research Part B: Methodological*, 18(2):169–177, April 1984.
- [FW56] Marguerite Frank and Philip Wolfe. An Algorithm for Quadratic Programming. *Naval Research Logistics Quarterly*, 3:95–110, 1956.
- [GH13] Dan Garber and Elad Hazan. Playing Non-linear Games with Linear Oracles. *FOCS 2013 - 54th Annual Symposium on Foundations of Computer Science*, pages 420–428, 2013.
- [GH14] Dan Garber and Elad Hazan. Faster Rates for the Frank-Wolfe Method over Strongly-Convex Sets. *arXiv.org*, June 2014.
- [GJ09] Bernd Gärtner and Martin Jaggi. Coresets for polytope distance. *SCG '09: Proceedings of the 25th annual symposium on Computational geometry*, June 2009.
- [GM86] Jacques Guélat and Patrice Marcotte. Some Comments on Wolfe’s ‘Away Step’. *Mathematical Programming*, 35(1):110–119, 1986.
- [GN13] Cristobal Guzman and Arkadi Nemirovski. On Lower Complexity Bounds for Large-Scale Smooth Convex Optimization. *arXiv.org*, July 2013.
- [Haz08] Elad Hazan. Sparse Approximate Solutions to Semidefinite Programs. In *LATIN 2008*, pages 306–316. Springer Berlin Heidelberg, 2008.
- [Hea82] Donald W Hearn. The gap function of a convex program. *Operations Research Letters*, 1(2):67–71, April 1982.
- [HJN14] Zaid Harchaoui, Anatoli Juditsky, and Arkadi Nemirovski. Conditional gradient algorithms for norm-regularized smooth convex optimization. *Mathematical Programming*, April 2014.
- [HK12] Elad Hazan and Satyen Kale. Projection-free Online Learning. In *ICML*, 2012.
- [Hoh77] Balder Hohenbalken. Simplicial decomposition in nonlinear programming algorithms. *Mathematical Programming*, 13(1):49–68, December 1977.
- [Hol74] Charles A Holloway. An Extension of the Frank and Wolfe Method of Feasible Directions. *Mathematical Programming*, 6(1):14–27, December 1974.
- [Jag11] Martin Jaggi. *Sparse Convex Optimization Methods for Machine Learning*. PhD thesis, ETH Zürich, October 2011.
- [Jag13] Martin Jaggi. Revisiting Frank-Wolfe: Projection-Free Sparse Convex Optimization. In *ICML 2013 - Proceedings of the 30th International Conference on Machine Learning*, 2013.
- [JS10] Martin Jaggi and Marek Sulovský. A Simple Algorithm for Nuclear Norm Regularized Problems. *ICML 2010: Proceedings of the 27th international conference on Machine learning*, 2010.
- [Lan13] Guanghui Lan. The Complexity of Large-scale Convex Programming under a Linear Optimization Oracle. *optimization-online.org*, 2013.
- [LJJ13] Simon Lacoste-Julien and Martin Jaggi. An Affine Invariant Linear Convergence Analysis for Frank-Wolfe Algorithms. *arXiv.org*, December 2013.

- [LJJS13] Simon Lacoste-Julien, Martin Jaggi, Mark Schmidt, and Patrick Pletscher. Block-Coordinate Frank-Wolfe Optimization for Structural SVMs. In *ICML 2013 - Proceedings of the 30th International Conference on Machine Learning*, 2013.
- [LP66] Evgenij S Levitin and Boris T Polyak. Constrained Minimization Methods. *USSR Computational Mathematics and Mathematical Physics*, 6(5):787–823, January 1966.
- [LSS13] Roi Livni, Shai Shalev-Shwartz, and Ohad Shamir. An Algorithm for Training Polynomial Networks. *arXiv.org*, April 2013.
- [LT13] Ronny Luss and Marc Teboulle. Conditional Gradient Algorithms for Rank-One Matrix Approximations with a Sparsity Constraint. *SIAM Review*, 55(1):65–98, February 2013.
- [Mai13] Julien Mairal. Optimization with First-Order Surrogate Functions. In *ICML. INRIA Grenoble Rhône-Alpes / LJK Laboratoire Jean Kuntzmann*, May 2013.
- [MDM74] B Mitchell, Vladimir Fedorovich Demjanov, and V Malozemov. Finding the Point of a Polyhedron Closest to the Origin. *SIAM Journal on Control*, 1974.
- [MZ93] Stéphane Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.
- [MZWG14] Cun Mu, Yuqian Zhang, John Wright, and Donald Goldfarb. Scalable Robust Matrix Recovery: Frank-Wolfe Meets Proximal Methods. *arXiv.org*, March 2014.
- [NKT14] Manu Nandan, Pramod P Khargonekar, and Sachin S Talathi. Fast SVM Training Using Approximate Extreme Points. *Journal of Machine Learning Research*, 15:59–98, 2014.
- [OG10] Hua Ouyang and Alexander Gray. Fast Stochastic Frank-Wolfe Algorithms for Non-linear SVMs. *SDM - SIAM International Conference on Data Mining*, 2010.
- [Pat93] Michael Patriksson. Partial linearization methods in nonlinear programming. *Journal of Optimization Theory and Applications*, 78(2):227–246, August 1993.
- [RSW14] Nikhil Rao, Parikshit Shah, and Stephen J Wright. Forward - Backward Greedy Algorithms for Atomic Norm Regularization. *arXiv.org*, April 2014.
- [SSSZ10] Shai Shalev-Shwartz, Nathan Srebro, and Tong Zhang. Trading Accuracy for Sparsity in Optimization Problems with Sparsity Constraints. *SIAM Journal on Optimization*, 20:2807–2832, 2010.
- [TBSR13] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht. Compressed Sensing Off the Grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, November 2013.
- [Tem12] Vladimir N Temlyakov. Greedy approximation in convex optimization. *arXiv.org*, June 2012.
- [TKC05] Ivor W Tsang, James T Kwok, and Pak-Ming Cheung. Core Vector Machines: Fast SVM Training on Very Large Data Sets. *Journal of Machine Learning Research*, 6:363–392, 2005.
- [TRD11] Ambuj Tewari, Pradeep Ravikumar, and Inderjit S Dhillon. Greedy Algorithms for Structurally Constrained High Dimensional Problems. In *NIPS*, 2011.
- [WJW05] M J Wainwright, T S Jaakkola, and A S Willsky. A New Class of Upper Bounds on the Log Partition Function. *IEEE Transactions on Information Theory*, 51(7):2313–2335, July 2005.
- [Wol61] Philip Wolfe. A duality theorem for nonlinear programming. *Quarterly of applied mathematics*, 19(3):239–244, 1961.
- [Wol70] Philip Wolfe. Convergence Theory in Nonlinear Programming. In J Abadie, editor, *Integer and Nonlinear Programming*, pages 1–23. North-Holland, 1970.
- [Zha11] Tong Zhang. Sparse Recovery With Orthogonal Matching Pursuit Under RIP. *IEEE Transactions on Information Theory*, 57(9):6215–6221, 2011.
- [ZYS12] Xinhua Zhang, Yaoliang Yu, and Dale Schuurmans. Accelerated Training for Matrix-norm Regularization: A Boosting Approach. In *NIPS*, 2012.