

# Bibliography

- [1] Agarwal, R. P., O'Regan, D., and Sahu, D. R. (2007). Iterative construction of fixed points of nearly asymptotically nonexpansive mappings. *Journal of nonlinear and convex analysis*, 8(1):61–79.
- [2] Aharon, M., Elad, M., and Bruckstein, A. (2006).  $k$ -svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11):4311–4322.
- [3] Ando, R. K. and Zhang, T. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *J. Mach. Learn. Res.*, 6:1817–1853.
- [4] Argyriou, A., Cléménçon, S., and Zhang, R. (2013). Learning the Graph of Relations Among Multiple Tasks. Research report.
- [5] Argyriou, A., Evgeniou, T., and Pontil, M. (2008). Convex multi-task feature learning. *Mach. Learn.*, 73(3):243–272.
- [6] Argyriou, A., Micchelli, C. A., Pontil, M., Shen, L., and Xu, Y. (2011). Efficient first order methods for linear composite regularizers. *CoRR*, abs/1104.1436.
- [7] Armijo, L. (1966). Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3.
- [8] Attouch, H., Czarnecki, M., and Peypouquet, J. (2011a). Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities. *SIAM Journal on Optimization*, 21(4):1251–1274.
- [9] Attouch, H., Czarnecki, M., and Peypouquet, J. (2011b). Prox-penalization and splitting methods for constrained variational problems. *SIAM Journal on Optimization*, 21(1):149–173.
- [10] Attouch, H., Peypouquet, J., and Redont, P. (2014). A dynamical approach to an inertial forward-backward algorithm for convex minimization. *SIAM Journal on Optimization*, 24(1):232–256.
- [11] Attouch, H., Peypouquet, J., and Redont, P. (2015). Fast convergence of an inertial gradient-like system with vanishing viscosity. *ArXiv e-prints*.

- [12] Bach, F., Jenatton, R., Mairal, J., and Obozinski, G. (2012a). Optimization with sparsity-inducing penalties. *Found. Trends Mach. Learn.*, 4(1):1–106.
- [13] Bach, F., Jenatton, R., Mairal, J., and Obozinski, G. (2012b). Structured sparsity through convex optimization. *Statist. Sci.*, 27(4):450–468.
- [14] Bach, F. R. (2008). Consistency of the group lasso and multiple kernel learning. *J. Mach. Learn. Res.*, 9:1179–1225.
- [15] Banach, S. (1922). Sur les operations dans les ensembles abstraits et leur applications aux equations integrales. *Fund. Math.*, (3):133–181.
- [16] Barzilai, J. and Borwein, J. M. (1988). Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148.
- [17] Bauschke, H. H. and Borwein, J. M. (1996). On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426.
- [18] Bauschke, H. H. and Combettes, P. L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer Publishing Company, Incorporated, 1st edition.
- [19] Bauschke, H. H., Cruz, J. B., Nghia, T. T., Phan, H. M., and Wang, X. (2014). The rate of linear convergence of the douglas–rachford algorithm for subspaces is the cosine of the friedrichs angle. *Journal of Approximation Theory*, 185:63 – 79.
- [20] Bauschke, H. H., Goebel, R., Lucet, Y., and Wang, X. (2008). The proximal average: Basic theory. *SIAM Journal on Optimization*, 19(2):766–785.
- [21] Beck, A. and Teboulle, M. (2003). Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, 31(3):167 – 175.
- [22] Beck, A. and Teboulle, M. (2009a). Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434.
- [23] Beck, A. and Teboulle, M. (2009b). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202.
- [24] Beck, A. and Teboulle, M. (2012). Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization*, 22(2):557–580.
- [25] Becker, S., Bobin, J., and Candès, E. J. (2011). NESTA: A fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences*, 4(1):1–39.
- [26] Ben-Tal, A., Margalit, T., and Nemirovski, A. (2001). The ordered subsets mirror descent optimization method with applications to tomography. *SIAM Journal on Optimization*, 12(1):79–108.

- [27] Berinde, V. (2002). On the stability of some fixed point procedures. *Bul. Ştiinţ. - Univ. BaiaMare, Ser. B, Matematica-Informatica*, 18(1):7–14.
- [28] Berinde, V. (2007). *Iterative Approximation of Fixed Points*. Springer Berlin Heidelberg.
- [29] Bertsekas, D. (1995). *Nonlinear Programming*. Athena Scientific.
- [30] Bertsekas, D. P. (1999). *Nonlinear Programming*. Athena Scientific, 2nd edition.
- [31] Bertsekas, D. P. and Tsitsiklis, J. N. (2000). Gradient convergence in gradient methods with errors. *SIAM Journal on Optimization*, 10(3):627–642.
- [32] Bickel, S., Bogojeska, J., Lengauer, T., and Scheffer, T. (2008). Multi-task learning for hiv therapy screening. In *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, pages 56–63, New York, NY, USA. ACM.
- [33] Boţ, R. I. and Csetnek, E. R. (2015). A hybrid proximal-extragradient algorithm with inertial effects. *Numerical Functional Analysis and Optimization*, 36(8):951–963.
- [34] Boţ, R. I., Csetnek, E. R., and Hendrich, C. (2015). Inertial douglas–rachford splitting for monotone inclusion problems. *Applied Mathematics and Computation*, 256:472 – 487.
- [35] Bosede, A. O. and Rhoades, B. (2010). Stability of picard and mann iteration for a general class of functions. *Journal of Advanced Mathematical Studies*, 3(2):23–26.
- [36] Boţ, Csetnek, E. R., and Heinrich, A. (2013). A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators. *SIAM Journal on Optimization*, 23(4):2011–2036.
- [37] Boţ, R. I. and Csetnek, E. R. (2014). Forward-backward and tseng’s type penalty schemes for monotone inclusion problems. *Set-Valued and Variational Analysis*, 22(2):313–331.
- [38] Bot, R. I. and Csetnek, E. R. (2014). An inertial alternating direction method of multipliers. *arXiv:1404.4582*.
- [39] Boţ, R. I., Csetnek, E. R., and László, S. C. (2016). An inertial forward–backward algorithm for the minimization of the sum of two nonconvex functions. *EURO Journal on Computational Optimization*, 4(1):3–25.
- [40] Boţ, R. I. and Hendrich, C. (2013). A double smoothing technique for solving unconstrained nondifferentiable convex optimization problems. *Computational Optimization and Applications*, 54(2):239–262.
- [41] Bottou, L., Curtis, F. E., and Nocedal, J. (2016). Optimization methods for large-scale machine learning. *arXiv preprint arXiv:1606.04838*.

- [42] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122.
- [43] Boyd, S. P., Xiao, L., and Mutapcic, A. (2003). Subgradient Methods. In *stanford.edu*.
- [44] Byrne, C. (2004). A unified treatment of some iterative algorithms in signal processing and image reconstruction. *Inverse Problems*, 20:103–120.
- [45] Cai, Y., Tang, Y., and Liu, L. (2012). Iterative algorithms for minimum-norm fixed point of non-expansive mapping in hilbert space. *Fixed Point Theory and Applications*, 2012(1):49.
- [46] Candes, E. J., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509.
- [47] Chambolle, A. and Dossal, C. (2015). On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm". *Journal of Optimization Theory and Applications*, 166(3):968–982.
- [48] Chambolle, A. and Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145.
- [49] Chambolle, A. and Pock, T. (2016). An introduction to continuous optimization for imaging. *Acta Numerica*, 25:161–319.
- [50] Chang, S., Qi, G.-J., Yang, Y., Aggarwal, C. C., Zhou, J., Wang, M., and Huang, T. S. (2016). Large-scale supervised similarity learning in networks. *Knowledge and Information Systems*, 48(3):707–740.
- [51] Chapelle, O., Shivaswamy, P., Vadrevu, S., Weinberger, K., Zhang, Y., and Tseng, B. (2010). Multi-task learning for boosting with application to web search ranking. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '10, pages 1189–1198, New York, NY, USA. ACM.
- [52] Chen, X., Lin, Q., Kim, S., Carbonell, J. G., and Xing, E. P. (2012). Smoothing proximal gradient method for general structured sparse regression. *Ann. Appl. Stat.*, 6(2):719–752.
- [53] Cheung, Y.-m. and Lou, J. (2017). Proximal average approximated incremental gradient descent for composite penalty regularized empirical risk minimization. *Machine Learning*, 106(4):595–622.
- [54] Chidume, C. and Mutangadura, S. (2001). An example on the mann iteration method for lipschitz pseudocontractions. *Proceedings of the American Mathematical Society*, 129(8):2359–2363.

- [55] Chidume, C. E. (1994). Approximation of fixed points of strongly pseudocontractive mappings. *Proceedings of the American Mathematical Society*, 120(2):545–551.
- [56] Cho, S. Y., Qin, X., and Wang, L. (2014). Strong convergence of a splitting algorithm for treating monotone operators. *Fixed Point Theory and Applications*, 2014(1):94.
- [57] Cho, Y. J., Zhou, H., and Guo, G. (2004). Weak and strong convergence theorems for three-step iterations with errors for asymptotically nonexpansive mappings. *Computers and Mathematics with Applications*, 47(4):707 – 717.
- [58] Chua, T.-S., Tang, J., Hong, R., Li, H., Luo, Z., and Zheng, Y. (2009). Nus-wide: A real-world web image database from national university of singapore. In *Proceedings of the ACM International Conference on Image and Video Retrieval, CIVR '09*, pages 48:1–48:9, New York, NY, USA. ACM.
- [59] Ćirić, L. B. (1974). A generalization of banach’s contraction principle. *Proceedings of the American Mathematical Society*, 45(2):267–273.
- [60] Collobert, R. and Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th International Conference on Machine Learning, ICML '08*, pages 160–167, New York, NY, USA. ACM.
- [61] Combettes, P. L. (2004). Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6):475–504.
- [62] Combettes, P. L. and Pesquet, J. C. (2007). A douglas-rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):564–574.
- [63] Combettes, P. L. and Pesquet, J.-C. (2011). *Proximal Splitting Methods in Signal Processing*, pages 185–212. Springer New York, New York, NY.
- [64] Combettes, P. L. and Wajs, V. R. (2005). Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200.
- [65] Cominetti, R., Soto, J. A., and Vaisman, J. (2014). On the rate of convergence of krasnosel’skiĭ-mann iterations and their connection with sums of bernoullis. *Israel Journal of Mathematics*, 199(2):757–772.
- [66] Condat, L. (2013). A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, 158(2):460–479.
- [67] Covitz, H. and Nadler, S. B. (1970). Multi-valued contraction mappings in generalized metric spaces. *Israel Journal of Mathematics*, 8(1):5–11.

- [68] Daumé, III, H., Kumar, A., and Saha, A. (2010). Frustratingly easy semi-supervised domain adaptation. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing, DANLP 2010*, pages 53–59, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [69] Davis, D. and Yin, W. (2014). Faster convergence rates of relaxed peaceman-rachford and admm under regularity assumptions. *Mathematics of Operations Research*, 0(0):null.
- [70] Davis, D. and Yin, W. (2016). *Convergence Rate Analysis of Several Splitting Schemes*, pages 115–163. Springer International Publishing, Cham.
- [71] Davis, D. and Yin, W. (2017). A three-operator splitting scheme and its optimization applications. *Set-Valued and Variational Analysis*.
- [72] Defazio, A., Bach, F., and Lacoste-Julien, S. (2014). Saga: A fast incremental gradient method with support for non-strongly convex composite objectives. In *Advances in Neural Information Processing Systems*, pages 1646–1654.
- [73] Demanet, L. and Zhang, X. (2016). Eventual linear convergence of the douglas-rachford iteration for basis pursuit. *Math. Comput.*, 85(297):209–238.
- [74] Devolder, O., Glineur, F., and Nesterov, Y. (2012). Double smoothing technique for large-scale linearly constrained convex optimization. *SIAM Journal on Optimization*, 22(2):702–727.
- [75] Dinevari, T. and Frigon, M. (2013). Fixed point results for multivalued contractions on a metric space with a graph. *Journal of Mathematical Analysis and Applications*, 405(2):507 – 517.
- [76] Douglas, J. and Rachford, H. H. (1956). On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439.
- [77] Duchi, J. and Singer, Y. (2009). Efficient online and batch learning using forward backward splitting. *J. Mach. Learn. Res.*, 10:2899–2934.
- [78] Duchi, J. C., Gould, S., and Koller, D. (2008). Projected subgradient methods for learning sparse gaussians. In McAllester, D. A. and Myllymäki, P., editors, *UAI*, pages 145–152. AUAI Press.
- [79] Dunn, J. and Harshbarger, S. (1978). Conditional gradient algorithms with open loop step size rules. *Journal of Mathematical Analysis and Applications*, 62(2):432 – 444.
- [80] Dutta, P. and Choudhury, B. S. (2008). A generalisation of contraction principle in metric spaces. *Fixed Point Theory and Applications*.

- [81] Eckstein, J. and Bertsekas, D. P. (1992). On the douglas—rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1):293–318.
- [82] Ferris, M. C. and Munson, T. S. (2002). Interior-point methods for massive support vector machines. *SIAM J. on Optimization*, 13(3):783–804.
- [83] Figueiredo, M. A. T., Nowak, R. D., and Wright, S. J. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586–597.
- [84] Frank, M. and Wolfe, P. (1956). An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110.
- [85] Ghadimi, S. and Lan, G. (2016). Accelerated gradient methods for nonconvex non-linear and stochastic programming. *Mathematical Programming*, 156(1):59–99.
- [86] Gilbert, J. C. and Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization*, 2(1):21–42.
- [87] Giselsson, P. and Boyd, S. (2017). Linear convergence and metric selection for douglas-rachford splitting and admm. *IEEE Transactions on Automatic Control*, 62(2):532–544.
- [88] Goldstein, A. A. (1965). On steepest descent. *Journal of the Society for Industrial and Applied Mathematics Series A Control*, 3(1):147–151.
- [89] Gong, P., Zhang, C., Lu, Z., Huang, J. Z., and Ye, J. (2013). A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28, ICML’13*, pages II–37–II–45. JMLR.org.
- [90] Guler, O. (1991). On the convergence of the proximal point algorithm for convex minimization. *SIAM Journal on Control and Optimization*, 29(2):403–419.
- [91] Hager, W. W. and Zhang, H. (2005). A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J. on Optimization*, 16(1):170–192.
- [92] Hager, W. W. and Zhang, H. (2006). A survey of nonlinear conjugate gradient methods. *Pacific journal of Optimization*, 2(1):35–58.
- [93] Hale, E. T., Yin, W., and Zhang, Y. (2008). Fixed-point continuation for  $\ell_1$ -minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130.
- [94] Han, L. and Zhang, Y. (2015). Learning tree structure in multi-task learning. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’15*, pages 397–406, New York, NY, USA. ACM.

- [95] Hastie, T., Tibshirani, R., and Friedman, J. (2003). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, corrected edition.
- [96] Hicks, T. L. and Kubicek, J. D. (1977). On the mann iteration process in a hilbert space. *Journal of Mathematical Analysis and Applications*, 59(3):498 – 504.
- [97] Hong, M. and Luo, Z.-Q. (2017). On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*, 162(1):165–199.
- [98] Hu, C., Pan, W., and Kwok, J. T. (2009). Accelerated gradient methods for stochastic optimization and online learning. In Bengio, Y., Schuurmans, D., Lafferty, J. D., Williams, C. K. I., and Culotta, A., editors, *Advances in Neural Information Processing Systems 22*, pages 781–789. Curran Associates, Inc.
- [99] Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyraas, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., and Clamp, M. (2002). The ensembl genome database project. *Nucleic Acids Research*, 30(1):38–41.
- [100] Ishikawa, S. (1974). Fixed points by a new iteration method. *Proceedings of the American Mathematical Society*, 44(1):147–150.
- [101] Ishikawa, S. (1976). Fixed points and iteration of a nonexpansive mapping in a banach space. *Proceedings of the American Mathematical Society*, 59(1):65–71.
- [102] Ito, N., Takeda, A., and Toh, K.-C. (2017). A unified formulation and fast accelerated proximal gradient method for classification. *Journal of Machine Learning Research*, 18(16):1–49.
- [103] Jacob, L., Obozinski, G., and Vert, J.-P. (2009). Group lasso with overlap and graph lasso. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 433–440, New York, NY, USA. ACM.
- [104] Jaggi, M. (2013). Revisiting frank-wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages I–427–I–435. JMLR.org.
- [105] Jenatton, R., Mairal, J., Obozinski, G., and Bach, F. (2011). Proximal methods for hierarchical sparse coding. *J. Mach. Learn. Res.*, 12:2297–2334.
- [106] Juditsky, A. and Nemirovski, A. (2012). *First-Order Methods for Nonsmooth Convex Large-Scale Optimization, I: General Purpose Methods*, pages 121–148. MIT Press.
- [107] Kim, S. and Xing, E. P. (2012). Tree-guided group lasso for multi-response regression with structured sparsity, with an application to eqtl mapping. *Ann. Appl. Stat.*, 6(3):1095–1117.



- [108] Kiwiel, K. C. and Lopuch, B. (1997). Surrogate projection methods for finding fixed points of firmly nonexpansive mappings. *SIAM Journal on Optimization*, 7(4):1084–1102.
- [109] Komodakis, N. and Pesquet, J. C. (2015). Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Processing Magazine*, 32(6):31–54.
- [110] Korpelevich, G. M. (1976). An extragradient method for finding saddle points and for other problems. *Matecon*, 12(4):747–756.
- [111] Lacoste-Julien, S. and Jaggi, M. (2015). On the global linear convergence of frank-wolfe optimization variants. In *Proceedings of the 28th International Conference on Neural Information Processing Systems, NIPS’15*, pages 496–504, Cambridge, MA, USA. MIT Press.
- [112] Lan, G. (2012). An optimal method for stochastic composite optimization. *Mathematical Programming*, 133(1):365–397.
- [113] Lan, G. (2015). Bundle-level type methods uniformly optimal for smooth and nonsmooth convex optimization. *Mathematical Programming*, 149(1):1–45.
- [114] Lehdili, N. and Moudafi, A. (1996). Combining the proximal algorithm and tikhonov regularization. *Optimization*, 37(3):239–252.
- [115] Lemaréchal, C., Nemirovskii, A., and Nesterov, Y. (1995). New variants of bundle methods. *Mathematical Programming*, 69(1):111–147.
- [116] Li, B., Yang, Q., and Xue, X. (2009). Transfer learning for collaborative filtering via a rating-matrix generative model. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML ’09*, pages 617–624, New York, NY, USA. ACM.
- [117] Li, H. and Lin, Z. (2015). Accelerated proximal gradient methods for nonconvex programming. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28*, pages 379–387. Curran Associates, Inc.
- [118] Lin, C.-J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural Comput.*, 19(10):2756–2779.
- [119] Lions, P. L. and Mercier, B. (1979). Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979.
- [120] Liouville, J. (1837). Sur le developpement des fonction ou parties de fonctions de fonctions en series. *Second Memoire. Journ. de Math.*, (2):16–35.

- [121] Liu, H., Palatucci, M., and Zhang, J. (2009). Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09*, pages 649–656, New York, NY, USA. ACM.
- [122] Liu, J. and Ye, J. (2010). Efficient  $l_1/l_q$  norm regularization. *CoRR*, abs/1009.4766.
- [123] Lorenz, D. A. and Pock, T. (2015). An inertial forward-backward algorithm for monotone inclusions. *Journal of Mathematical Imaging and Vision*, 51(2):311–325.
- [124] Lounici, K., Pontil, M., Tsybakov, A. B., and van de Geer, S. (2009). Taking advantage of sparsity in multi-task learning. In *Proceedings of the 22nd Conference on Learning Theory (COLT)*, pages 73–82.
- [125] Luo, D., Ding, C., and Huang, H. (2013). Toward structural sparsity: an explicit  $l_2 - l_0$  approach. *Knowledge and Information Systems*, 36(2):411–438.
- [126] Lustig, M., Donoho, D., and Pauly, J. M. (2007). Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195.
- [127] Maingé, P.-E. (2008). Convergence theorems for inertial km-type algorithms. *Journal of Computational and Applied Mathematics*, 219(1):223–236.
- [128] Mann, W. R. (1953). Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4:506–510.
- [129] Marino, G. and Xu, H.-K. (2007). Weak and strong convergence theorems for strict pseudo-contractions in hilbert spaces. *Journal of Mathematical Analysis and Applications*, 329(1):336 – 346.
- [130] Meir, A. and Keeler, E. (1969). A theorem on contraction mappings. *Journal of Mathematical Analysis and Applications*, 28(2):326 – 329.
- [131] Nadler, S. B. (1969). Multi-valued contraction mappings. *Pacific J. Math.*, 30(2):475–488.
- [132] Naikal, N., Yang, A. Y., and Sastry, S. S. (2010). Towards an efficient distributed object recognition system in wireless smart camera networks. In *2010 13th International Conference on Information Fusion*, pages 1–8.
- [133] Nedic, A. and Bertsekas, D. P. (2001). Incremental subgradient methods for non-differentiable optimization. *SIAM J. on Optimization*, 12(1):109–138.
- [134] Negahban, S. and Wainwright, M. J. (2008). Phase transitions for high-dimensional joint support recovery. In Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 21*, pages 1161–1168. Curran Associates, Inc.

- [135] Nesterov, Y. (1983). A method of solving a convex programming problem with convergence rate  $o(1/\sqrt{k})$ . *Soviet Mathematics Doklady*, 27(2):372–376.
- [136] Nesterov, Y. (2005). Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152.
- [137] Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems. *Mathematical Programming*, 120(1):221–259.
- [138] Nesterov, Y. (2013). Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161.
- [139] Nesterov, Y. et al. (2007). Gradient methods for minimizing composite objective function.
- [140] Nishihara, R., Lessard, L., Recht, B., Packard, A., and Jordan, M. I. (2015). A general analysis of the convergence of admm. In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15*, pages 343–352. JMLR.org.
- [141] Nocedal, J. and Wright, S. (2006). *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York.
- [142] Noor, M. A. (2000). New approximation schemes for general variational inequalities. *Journal of Mathematical Analysis and Applications*, 251(1):217 – 229.
- [143] Noor, M. A. (2007). General variational inequalities and nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 331(2):810 – 822.
- [144] Noor, M. A. and Huang, Z. (2007). Three-step methods for nonexpansive mappings and variational inequalities. *Applied Mathematics and Computation*, 187(2):680 – 685.
- [145] Ochs, P., Brox, T., and Pock, T. (2015a). ipiasco: Inertial proximal algorithm for strongly convex optimization. *Journal of Mathematical Imaging and Vision*, 53(2):171–181.
- [146] Ochs, P., Chen, Y., Brox, T., and Pock, T. (2014). ipiano: Inertial proximal algorithm for nonconvex optimization. *SIAM Journal on Imaging Sciences*, 7(2):1388–1419.
- [147] Ochs, P., Dosovitskiy, A., Brox, T., and Pock, T. (2015b). On iteratively reweighted algorithms for nonsmooth nonconvex optimization in computer vision. *SIAM Journal on Imaging Sciences*, 8(1):331–372.
- [148] O’Donoghue, B. and Candès, E. (2015). Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15(3):715–732.
- [149] Opial, Z. (1967). Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73(4):591–597.

- [150] Paknazar, M., Eshaghi Gordji, M., De La Sen, M., and Vaezpour, S. M. (2013). N-fixed point theorems for nonlinear contractions in partially ordered metric spaces. *Fixed Point Theory and Applications*, 2013(1):111.
- [151] Parikh, N. and Boyd, S. (2014). Proximal algorithms. *Found. Trends Optim.*, 1(3):127–239.
- [152] Patrinos, P., Stella, L., and Bemporad, A. (2014). Douglas-rachford splitting: Complexity estimates and accelerated variants. In *53rd IEEE Conference on Decision and Control*, pages 4234–4239.
- [153] Peaceman, D. W. and H. H. Rachford, J. (1955). The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, 3(1):28–41.
- [154] Pesquet, J.-C. and Pustelnik, N. (2012). A Parallel Inertial Proximal Optimization Method. *Pacific journal of optimization*, 8(2):273–305.
- [155] Phan, H. M. (2016). Linear convergence of the douglas–rachford method for two closed sets. *Optimization*, 65(2):369–385.
- [156] Picard, E. (1890). Memoire sur la theorie des equations aux derivees partielles et la methode des approximations successives. *J. Math. Pures et Appl.*, (6):145–210.
- [157] Poljak, B. T. (1987). *Introduction to optimization*. Optimization Software.
- [158] Polyak, B. T. (1964). Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4(5):1–17.
- [159] Qian, N. (1999). On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145 – 151.
- [160] Quattoni, A., Carreras, X., Collins, M., and Darrell, T. (2009). An efficient projection for  $l_1$  &  $l_\infty$  regularization. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML '09*, pages 857–864, New York, NY, USA. ACM.
- [161] Raguét, H., Fadili, J., and Peyré, G. (2013). A generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226.
- [162] Rao, N., Nowak, R., Cox, C., and Rogers, T. (2016). Classification with the sparse group lasso. *IEEE Transactions on Signal Processing*, 64(2):448–463.
- [163] Rasiwasia, N., Costa Pereira, J., Coviello, E., Doyle, G., Lanckriet, G. R., Levy, R., and Vasconcelos, N. (2010). A new approach to cross-modal multimedia retrieval. In *Proceedings of the 18th ACM International Conference on Multimedia, MM '10*, pages 251–260, New York, NY, USA. ACM.

- [164] Raydan, M. (1997). The barzilai and borwein gradient method for the large scale unconstrained minimization problem. *SIAM Journal on Optimization*, 7(1):26–33.
- [165] Reich, S. and Zaslavski, A. (2000). Convergence of krasnoselskii-mann iterations of nonexpansive operators. *Mathematical and Computer Modelling*, 32(11):1423 – 1431.
- [166] Rhoades, B. E. (1977). A comparison of various definitions of contractive mappings. *Transactions of the American Mathematical Society*, 226:257–290.
- [167] Rockafellar, R., Wets, M., and Wets, R. (2009). *Variational Analysis*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 1 edition.
- [168] Rockafellar, R. T. (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898.
- [169] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chapter Learning Internal Representations by Error Propagation, pages 318–362. MIT Press, Cambridge, MA, USA.
- [170] Ryu, E. K. and Boyd, S. (2015). A primer on monotone operator methods.
- [171] Ryu, E. K. and Boyd, S. (2016). Primer on monotone operator methods.
- [172] Sahu, D., Xu, H.-K., and Yao, J.-C. (2009). Asymptotically strict pseudocontractive mappings in the intermediate sense. *Nonlinear Analysis: Theory, Methods and Applications*, 70(10):3502 – 3511.
- [173] Sahu, D. R. (2011). Applications of the s-iteration process to constrained minimization problems and split feasibility problems. *Fixed Point Theory*, 12(1):187–204.
- [174] Sahu, D. R. and Ansari, Q. H. (2016). *Fixed Point Theory and Graph Theory*. Academic Press, 1 edition.
- [175] Sahu, D. R., Ansari, Q. H., and Yao, J. C. (2015). The prox-tikhonov-like forward-backward method and applications. *Taiwanese Journal of Mathematics*, 19(2):481–503.
- [176] Sahu, D. R. and Yao, J. C. (2011). The prox-tikhonov regularization method for the proximal point algorithm in banach spaces. *Journal of Global Optimization*, 51(4):641–655.
- [177] Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal*.
- [178] Schmidt, M., Roux, N. L., and Bach, F. (2011). Convergence rates of inexact proximal-gradient methods for convex optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, NIPS’11, pages 1458–1466, USA. Curran Associates Inc.

- [179] Schu, J. (1991). Iterative construction of fixed points of asymptotically nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 158(2):407 – 413.
- [180] Shor, N. Z., Kiwiel, K. C., and Ruszcayński, A. (1985). *Minimization Methods for Non-differentiable Functions*. Springer-Verlag New York, Inc., New York, NY, USA.
- [181] Smola, A. J., Vishwanathan, S. V. N., and Le, Q. V. (2007). Bundle methods for machine learning. In *Proceedings of the 20th International Conference on Neural Information Processing Systems*, NIPS’07, pages 1377–1384, USA. Curran Associates Inc.
- [182] Sonnenburg, S., Rätsch, G., Henschel, S., Widmer, C., Behr, J., Zien, A., Bona, F. d., Binder, A., Gehl, C., and Franc, V. (2010). The shogun machine learning toolbox. *J. Mach. Learn. Res.*, 11:1799–1802.
- [183] Sra, S. (2011). *Fast Projections onto  $\ell_{1,q}$ -Norm Balls for Grouped Feature Selection*, pages 305–317. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [184] Sra, S., Nowozin, S., and Wright, S. J. (2011). *Optimization for Machine Learning*. The MIT Press.
- [185] Suantai, S. (2005). Weak and strong convergence criteria of noor iterations for asymptotically nonexpansive mappings. *Journal of Mathematical Analysis and Applications*, 311(2):506 – 517.
- [186] Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–15550.
- [187] Suzuki, T. (2008). A generalized banach contraction principle that characterizes metric completeness. *Proceedings of the American Mathematical Society*, 136(5):1861–1869.
- [188] Tan, K. and Xu, H. (1993). Approximating fixed points of nonexpansive mappings by the ishikawa iteration process. *Journal of Mathematical Analysis and Applications*, 178(2):301 – 308.
- [189] Tan, Q. F. and Narayanan, S. S. (2012). Novel variations of group sparse regularization techniques with applications to noise robust automatic speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(4):1337–1346.
- [190] Taskar, B., Lacoste-Julien, S., and Jordan, M. I. (2005). Structured prediction via the extragradient method. In *Proceedings of the 18th International Conference on Neural Information Processing Systems*, NIPS’05, pages 1345–1352, Cambridge, MA, USA. MIT Press.

- [191] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288.
- [192] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108.
- [193] Tseng, P. (2000). A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446.
- [194] Tseng, P. (2008). On accelerated proximal gradient methods for convex-concave optimization. *submitted to SIAM Journal on Optimization*.
- [195] Tseng, P. and Yun, S. (2009). A coordinate gradient descent method for nonsmooth separable minimization. *Mathematical Programming*, 117(1):387–423.
- [196] Turlach, B. A., Venables, W. N., and Wright, S. J. (2005). Simultaneous variable selection. *Technometrics*, 47(3):349–363.
- [197] Vandenberghe, L. (2014). Ee236c-optimization methods for large-scale systems. *Lecture notes*, 9.
- [198] Villa, S., Salzo, S., Baldassarre, L., and Verri, A. (2013). Accelerated and inexact forward-backward algorithms. *SIAM Journal on Optimization*, 23(3):1607–1633.
- [199] Vũ, B. C. (2013). A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, 38(3):667–681.
- [200] Wang, J., Liu, J., and Ye, J. (2013). Efficient mixed-norm regularization: Algorithms and safe screening methods. *CoRR*, abs/1307.4156.
- [201] Wang, J. and Xiao, L. (2017). Exploiting strong convexity from data with primal-dual first-order algorithms. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3694–3702, International Convention Centre, Sydney, Australia. PMLR.
- [202] Wang, K., Wang, W., and Wang, L. (2015). Learning unified sparse representations for multi-modal data. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 3545–3549, Canada. IEEE.
- [203] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.
- [204] Wu, F., Yu, Z., Yang, Y., Tang, S., Zhang, Y., and Zhuang, Y. (2014). Sparse multi-modal hashing. *IEEE Transactions on Multimedia*, 16(2):427–439.
- [205] Xiao, L. (2010). Dual averaging methods for regularized stochastic learning and online optimization. *J. Mach. Learn. Res.*, 11:2543–2596.

- [206] Xu, B. and Noor, M. A. (2002). Fixed-point iterations for asymptotically nonexpansive mappings in banach spaces. *Journal of Mathematical Analysis and Applications*, 267(2):444 – 453.
- [207] Xu, H.-K. (2002). Iterative algorithms for nonlinear operators. *Journal of the London Mathematical Society*, 66(1):240–256.
- [208] Xu, H.-K. (2006). A regularization method for the proximal point algorithm. *Journal of Global Optimization*, 36(1):115–125.
- [209] Yang, A. Y., Zhou, Z., Balasubramanian, A. G., Sastry, S. S., and Ma, Y. (2013a). Fast  $ell_1$ -minimization algorithms for robust face recognition. *IEEE Transactions on Image Processing*, 22(8):3234–3246.
- [210] Yang, H., Lyu, M. R., and King, I. (2013b). Efficient online learning for multitask feature selection. *ACM Trans. Knowl. Discov. Data*, 7(2):6:1–6:27.
- [211] Yao, Y., Zhou, H., and Liou, Y.-C. (2008). Strong convergence of a modified krasnoselski-mann iterative algorithm for non-expansive mappings. *Journal of Applied Mathematics and Computing*, 29(1):383.
- [212] Yin, W., Osher, S., Goldfarb, D., and Darbon, J. (2008). Bregman iterative algorithms for  $\ell_1$ -minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168.
- [213] Yu, J., Rui, Y., and Chen, B. (2014). Exploiting click constraints and multi-view features for image re-ranking. *IEEE Transactions on Multimedia*, 16(1):159–168.
- [214] Yu, J., Vishwanathan, S., Günter, S., and Schraudolph, N. N. (2010). A quasi-newton approach to nonsmooth convex optimization problems in machine learning. *J. Mach. Learn. Res.*, 11:1145–1200.
- [215] Yu, Y.-L. (2013). Better approximation and faster algorithm using the proximal average. In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26*, pages 458–466. Curran Associates, Inc.
- [216] Yuan, L., , J., and Ye, J. (2013). Efficient methods for overlapping group lasso. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(9):2104–2116.
- [217] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.
- [218] Zhang, S., Qian, H., and Gong, X. (2016). An alternating proximal splitting method with global convergence for nonconvex structured sparsity optimization. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI’16, pages 2330–2336. AAAI Press.



- [219] Zhao, P., Rocha, G., and Yu, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *Ann. Statist.*, 37(6A):3468–3497.
- [220] Zhong, L. W. and Kwok, J. T. (2014a). Gradient descent with proximal average for nonconvex and composite regularization. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI’14, pages 2206–2212. AAAI Press.
- [221] Zhong, W. and Kwok, J. (2014b). Accelerated Stochastic Gradient Method for Composite Regularization. In Kaski, S. and Corander, J., editors, *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, volume 33 of *Proceedings of Machine Learning Research*, pages 1086–1094, Reykjavik, Iceland. PMLR.
- [222] Zhou, J., Chen, J., and Ye, J. (2012). *MALSAR: Multi-task Learning via Structural Regularization*. Arizona State University.
- [223] Zhou, J., Yuan, L., Liu, J., and Ye, J. (2011). A multi-task learning formulation for predicting disease progression. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’11, pages 814–822, New York, NY, USA. ACM.