Adaptation, Learning, and Optimization over Networks

Ali H. Sayed
University of California at Los Angeles

the essence of knowledge
Boston — Delft

Foundations and Trends® in Machine Learning

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

A. H. Sayed. *Adaptation, Learning, and Optimization over Networks.* Foundations and Trends® in Machine Learning, vol. 7, no. 4-5, pp. 311–801, 2014.

This Foundations and Trends® issue was typeset in LaTeX using a class file designed by Neal Parikh. Printed on acid-free paper.

ISBN: 978-1-60198-851-5

Editorial Scope

Topics

Foundations and Trends® in Machine Learning publishes survey and tutorial articles on the theory, algorithms and applications of machine learning, including the following topics:

- Adaptive control and signal processing
- Applications and case studies
- Behavioral, cognitive, and neural learning
- Bayesian learning
- Classification and prediction
- Clustering
- Data mining
- Dimensionality reduction
- Evaluation
- Game theoretic learning
- Graphical models
- Independent component analysis
- Inductive logic programming
- Kernel methods
- Markov chain Monte Carlo
- Model choice
- Nonparametric methods
- Online learning
- Optimization
- Reinforcement learning
- Relational learning
- Robustness
- Spectral methods
- Statistical learning theory
- Variational inference
- Visualization

Information for Librarians

Foundations and Trends® in Machine Learning, 2014, Volume 7, 6 issues. ISSN paper version 1935-8237. ISSN online version 1935-8245. Also available as a combined paper and online subscription.

Foundations and Trends® in Machine Learning
Vol. 7, No. 4-5 (2014) 311–801

DOI: 10.1561/2200000051

Adaptation, Learning, and Optimization over Networks

Ali H. Sayed
University of California at Los Angeles

Contents

Abstract

This work deals with the topic of information processing over graphs. The presentation is largely self-contained and covers results that relate to the analysis and design of multi-agent networks for the distributed solution of optimization, adaptation, and learning problems from streaming data through localized interactions among agents. The results derived in this work are useful in comparing network topologies against each other, and in comparing networked solutions against centralized or batch implementations. There are many good reasons for the peaked interest in distributed implementations, especially in this day and age when the word "network" has become commonplace whether one is referring to social networks, power networks, transportation networks, biological networks, or other types of networks. Some of these reasons have to do with the benefits of cooperation in terms of improved performance and improved resilience to failure. Other reasons deal with privacy and secrecy considerations where agents may not be comfortable sharing their data with remote fusion centers. In other situations, the data may already be available in dispersed locations, as happens with cloud computing. One may also be interested in learning through data mining from big data sets. Motivated by these considerations, this work examines the limits of performance of distributed stochastic-gradient solutions and discusses procedures that help bring forth their potential more fully. The presentation adopts a useful statistical framework and derives performance results that elucidate the mean-square stability, convergence, and steady-state behavior of the learning networks. The work also illustrates how distributed processing over graphs gives rise to some revealing phenomena due to the coupling effect among the agents. These phenomena are discussed in the context of adaptive networks, along with examples from a variety of areas including distributed sensing, intrusion detection, distributed estimation, online adaptation, network system theory, and machine learning.

A. H. Sayed. *Adaptation, Learning, and Optimization over Networks.* Foundations and Trends® in Machine Learning, vol. 7, no. 4-5, pp. 311–801, 2014.

DOI: 10.1561/2200000051.

1

Motivation and Notation

1.1 Introduction

Network science is a fascinating field that is evolving rapidly across many domains [15, 19, 92, 121, 155, 179, 208]. As remarked in [208], and for long, classical system and learning theories have focused on optimizing stand-alone systems or learners with great success. Nevertheless, progress in recent decades in the biological sciences [16, 50, 131, 147], animal behavior studies [7, 50, 79, 90, 188, 220], and the neuroscience of the brain [20, 49, 226], has revealed remarkable patterns of organization and structured complexity in the behavior of biological networks, animal groups, and in the dynamics of brain connectivity. These studies have brought forward notable examples of complex systems that derive their sophistication from coordination among simpler units and from the aggregation and processing of decentralized pieces of information. While each unit in these systems is not capable of sophisticated behavior on its own, it is the interaction among the constituents that leads to systems that are resilient to failure and that are capable of adjusting their behavior in response to changes in their environment.

These discoveries have motivated diligent efforts towards a deeper understanding of information processing, adaptation, and learning over complex networks in several disciplines including machine learning, optimization, control, economics, biological sciences, information sciences, and the social sciences. A common goal in these investigations has been to develop theory and tools that enable the design of networks with sophisticated learning and processing abilities, such as networks that are able to solve important inference and optimization tasks in a distributed manner by relying on agents that interact locally and do not rely on fusion centers to collect and process their information.

1.2 Biological Networks

Examples abound for the viability of such designs in the realm of biological networks. Nature is laden with examples of networks exhibiting sophisticated behavior that arises from interactions among agents of limited abilities. For example, fish schools are unusually skilled at navigating their environment with remarkable discipline and at configuring the topology of their school in the face of danger from predators [79, 188]; when a predator is sighted or sensed, the entire school of fish adjusts its configuration to let the predator through and then coalesces again to continue its schooling behavior. It is reasonable to assume that this complex behavior is the result of sensing information spreading fast across the school of fish through local interactions among adjacent members of the school. Likewise, in bee swarms, it is observed that only a small fraction of the agents (about 5%) are informed and this small fraction of agents is still capable of guiding an entire swarm of bees to their new hive [12, 22, 125, 220]. It is a remarkable property of biological networks and animal groups that sophisticated behavior is able to arise from simple interactions among limited agents [119, 200, 229].

1.3 Distributed Processing

Motivated by these observations, this work deals with the topic of information processing over graphs and how collaboration among agents in a network can lead to superior adaptation and learning performance. The presentation covers results and tools that relate to the analysis and design of networks that are able to solve optimization, adaptation, and learning problems in an efficient and distributed manner from streaming data through localized interactions among their agents.

The treatment extends the presentation from [208] in several directions[1] and covers three intertwined topics: (a) how to perform distributed *optimization* over networks; (b) how to perform distributed *adaptation* over networks; and (c) how to perform distributed *learning* over networks. In these three domains, we examine and compare the advantages and limitations of non-cooperative, centralized, and distributed *stochastic-gradient* solutions. In the non-cooperative mode of operation, agents act independently of each other in their pursuit of their desired objective. In the centralized mode of operation, agents transmit their (collected or processed) data to a fusion center, which is capable of processing the data centrally. The fusion center then shares the results of the analysis back with the distributed agents. While centralized solutions can be powerful, they still suffer from some limitations. First, in real-time applications where agents collect data continuously, the repeated exchange of information back and forth between the agents and the fusion center can be costly especially when these exchanges occur over wireless links or require nontrivial routing resources. Second, in some sensitive applications, agents may be reluctant to share their data with remote centers for various reasons including privacy and secrecy considerations. More importantly perhaps, centralized solutions have a critical point of failure: if the central processor fails, then this solution method collapses altogether.

Distributed implementations, on the other hand, pursue the desired objective through *localized* interactions among the agents. In the distributed mode of operation, agents are connected by a topology and they are permitted to share information only with their immediate neighbors. There are many good reasons for the peaked interest in such distributed solutions, especially in this day and age when the word "network" has become commonplace whether one is referring to social networks, power networks, transportation networks, biological networks, or other types of networks. Some of these reasons have to do

[1] The author is grateful to IEEE for allowing reproduction of material from [208] in this work.

with the benefits of cooperation in terms of improved performance and improved robustness and resilience to failure. Other reasons deal with privacy and secrecy considerations where agents may not be comfortable sharing their data with remote fusion centers. In other situations, the data may already be available in dispersed locations, as happens with cloud computing. One may also be interested in learning and extracting information through data mining from large data sets. Decentralized learning procedures offer an attractive approach to dealing with such large data sets. Decentralized mechanisms can also serve as important enablers for the design of robotic swarms, which can assist in the exploration of disaster areas.

For these various reasons, we devote some good effort in this work towards quantifying the limits of performance of distributed solutions and towards discussing design procedures that can bring forth their potential more fully. Our emphasis is on solutions that are able to learn from streaming data. In particular, we shall study three families of distributed strategies: (a) incremental strategies, (b) consensus strategies, and (c) diffusion strategies — see Chapter 7. We shall derive expressions that quantify the behavior of the distributed algorithms and use the expressions to compare their performance and to illustrate under what conditions network cooperation is beneficial to the learning and adaptation process. While the social benefit, defined as the average performance across the network, generally improves through cooperation, it is not necessarily the case that the individual agents will always benefit from cooperation: some agents may see their performance degrade relative to the non-cooperative mode of operation [215, 277]. This observation will motivate us to seek optimized combination policies that enable all agents in a network to enhance their performance through cooperation.

1.4 Adaptive Networks

We shall study distributed solutions in the context of adaptive networks [208, 209, 215], which consist of a collection of agents *with* adaptation and learning abilities. The agents are linked together through a topology and they interact with each other through localized *in-network* processing to solve inference and optimization problems in a fully distributed and online manner. The continuous sharing and diffusion of information across the network enables the agents to respond in real-time to drifts in the data and to changes in the network topology. Such networks are scalable, robust to node and link failures, and are particularly suitable for learning from big data sets by tapping into the power of collaboration among distributed agents. The networks are also endowed with cognitive abilities [108, 208] due to the sensing abilities of their agents, their interactions with their neighbors, and an embedded feedback mechanism for acquiring and refining information. Each agent is not only capable of experiencing the environment directly, but it also receives information through interactions with its neighbors and processes this information to drive its learning process.

Adaptive networks are well-suited to perform decentralized information processing tasks. They are also well-suited to model several forms of complex behavior exhibited by biological [16, 50, 131, 147] and social networks [15, 77, 92, 121, 230] such as fish schooling [188], prey-predator maneuvers [105, 171], bird formations [110, 119], bee swarming [12, 22, 125, 220], bacteria motility [25, 189, 258], and social and economic interactions [98, 103]. Examples of references that discuss applications of the *diffusion* distributed algorithms studied in this work to problems involving biological and social networks include [56, 65, 156, 213, 215, 246, 247, 250, 276]. Examples of references that discuss applications of *consensus* implementations include [2, 18, 64, 80, 118, 122, 123, 181, 184, 185, 199, 200, 255]. We do not discuss biological networks in this work and refer the reader instead to the above references; the survey article [215] provides some further motivation.

1.5 Organization

This work is largely self-contained. It provides an extended treatment of topics presented in condensed form in the survey [208], and of several other additional topics. For maximal benefit, readers may review first the background material in Appendices A through G on complex gradient vectors and Hessian matrices, convex functions, mean-value theorems, Lipschitz conditions, matrix theory, and logistic regression.

In preparation for the study of multi-agent networks, Chapters 2–4 review some fundamental results on optimization, adaptation, and learning by *single* stand-alone agents. The emphasis is on stochastic-gradient constructions. The presentation in these chapters provides insights that will be useful in our subsequent study of adaptation and learning by a collection of networked agents. This latter study is more demanding due to the coupling among interacting agents, and due to the fact that networks are generally sparsely connected. The results in this work will help clarify the effect of network topology on performance and will develop tools that enable designers to compare various strategies against each other and against the centralized solution.

1.6 Notation and Symbols

All vectors are column vectors, with the exception of the regression vector (denoted by the letters u or $\boldsymbol{u}$), which will be taken to be a row vector for convenience of presentation. Table 1.1 lists the main conventions used in our exposition. In particular, note that we use ***boldface*** letters to refer to random quantities and *normal* font to refer to their realizations or deterministic quantities. We also use T for matrix or vector transposition and $*$ for complex-conjugate transposition.

Moreover, for generality, we treat the case in which the variables of interest are generally *complex-valued*; when necessary, we show how the results simplify in the real case. Some subtle differences in the analysis arise when dealing with complex data. These differences would be masked if we focus exclusively on real-valued data. Moreover, studying design problems with complex data is relevant for many fields, especially in the domain of signal processing and communications problems.

Table 1.1: List of notation and symbols used in the text and appendices.

$\mathbb{R}$	Field of real numbers.
$\mathbb{C}$	Field of complex numbers.
$\mathbb{1}$	Column vector with all its entries equal to one.
I_M	Identity matrix of size $M \times M$.
$\boldsymbol{d}$	Boldface notation denotes random variables.
d	Normal font denotes realizations of random variables.
A	Capital letters denote matrices.
a	Small letters denote vectors or scalars.
α	Greek letters denote scalars.
$d(i)$	Small letters with parenthesis denote scalars.
d_i	Small letters with subscripts denote vectors.
T	Matrix transposition.
$*$	Complex-conjugate transposition.
$\mathrm{Re}(z)$	Real part of complex number z.
$\mathrm{Im}(z)$	Imaginary part of complex number z.
$\mathrm{col}\{a, b\}$	Column vector with entries a and b.
$\mathrm{diag}\{a, b\}$	Diagonal matrix with entries a and b.
$\mathrm{vec}\{A\}$	Vector obtained by stacking the columns of A.
$\mathrm{bvec}\{\mathcal{A}\}$	Vector obtained by vectorizing and stacking blocks of $\mathcal{A}$.
$\|x\|$	Euclidean norm of its vector argument.
$\|x\|^2_\Sigma$	Weighted square value $x^*\Sigma x$.
$\|A\|$	Two-induced norm of matrix A, also equal to $\sigma_{\max}(A)$.
$\|A\|_1$	Maximum absolute column sum of matrix A.
$\|A\|_\infty$	Maximum absolute row sum of matrix A.
$A \geq 0$	Matrix A is non-negative definite.
$A > 0$	Matrix A is positive-definite.
$\rho(A)$	Spectral radius of matrix A.
$\lambda_{\max}(A)$	Maximum eigenvalue of the Hermitian matrix A.
$\lambda_{\min}(A)$	Minimum eigenvalue of the Hermitian matrix A.
$\sigma_{\max}(A)$	Maximum singular value of A.
$A \otimes B$	Kronecker product of A and B.
$\mathcal{A} \otimes_b \mathcal{B}$	Block Kronecker product of block matrices $\mathcal{A}$ and $\mathcal{B}$.
$a \preceq b$	Element-wise comparison of the entries of vectors a and b.
$\delta_{k,\ell}$	Kronecker delta sequence: 1 when $k = \ell$ and 0 when $k \neq \ell$.
$\alpha = O(\mu)$	Signifies that $\|\alpha\| \leq c\|\mu\|$ for some constant $c > 0$.
$\alpha = o(\mu)$	Signifies that $\alpha/\mu \to 0$ as $\mu \to 0$.
$\alpha(\mu) \doteq \beta(\mu)$	Signifies that $\alpha(\mu)$ and $\beta(\mu)$ agree to first order in μ.
$\limsup_{n\to\infty} a(n)$	Limit superior of the sequence $a(n)$.
$\liminf_{n\to\infty} a(n)$	Limit inferior of the sequence $a(n)$.

References

[1] R. Abdolee, B. Champagne, and A. H. Sayed. Diffusion LMS strategies for parameter estimation over fading wireless channels. In *Proc. IEEE ICC*, pages 1926–1930. Budapest, Hungary, June 2013.

[2] D. Acemoglu and A. Ozdaglar. Opinion dynamics and learning in social networks. *Dyn. Games Appl.*, 1(1):3–49, Mar. 2011.

[3] T. Adali, P. J. Schreier, and L. L. Scharf. Complex-valued signal processing: The proper way to deal with impropriety. *IEEE Trans. Signal Process.*, 59(11):5101–5125, Nov. 2011.

[4] A. Agarwal and J. Duchi. Distributed delayed stochastic optimization. In *Proc. Neural Information Processing Systems (NIPS)*, pages 873–881. Granada, Spain, Dec. 2011.

[5] L. V. Ahlfors. *Complex Analysis.* McGraw Hill, NY, 3rd edition, 1979.

[6] T. Y. Al-Naffouri and A. H. Sayed. Transient analysis of data-normalized adaptive filters. *IEEE Trans. Signal Process.*, 51(3):639–652, Mar. 2003.

[7] J. Alcock. *Animal Behavior: An Evolutionary Approach.* Sinauer Associates, 9th edition, 2009.

[8] P. Alriksson and A. Rantzer. Distributed Kalman filtering using weighted averaging. In *Proc. Int. Symp. Math. Thy Net. Sys (MTNS)*, pages 1–6. Kyoto, Japan, 2006.

[9] R. Arablouei, S. Werner, Y.-F. Huang, and K. Dogancay. Distributed least-mean-square estimation with partial diffusion. *IEEE Trans. Signal Process.*, 62(2):472–484, Jan. 2014.

[10] J. Arenas-Garcia, A. R. Figueiras-Vidal, and A. H. Sayed. Mean-square performance of a convex combination of two adaptive filters. *IEEE Trans. Signal Process.*, 54(3):1078–1090, March 2006.

[11] A. Auslender and M. Teboulle. Interior gradient and proximal methods for convex and conic optimization. *SIAM J. Optim.*, 16(3):697–725, 2006.

[12] A. Avitabile, R. A. Morse, and R. Boch. Swarming honey bees guided by pheromones. *Ann. Entomol. Soc. Am.*, 68:1079–1082, 1975.

[13] T. C. Aysal, M. J. Coates, and M. G. Rabbat. Distributed average consensus with dithered quantization. *IEEE Trans. Signal Process.*, 56(10):4905–4918, October 2008.

[14] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione. Broadcast gossip algorithms for consensus. *IEEE Trans. Signal Process.*, 57(7):2748–2761, July 2009.

[15] A.-L. Barabási. *Linked: How Everything Is Connected to Everything Else and What It Means.* Plume, NY, 2003.

[16] A.-L. Barabási and Z. N. Oltvai. Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101–113, 2004.

[17] R. Baraniuk. Compressive sensing. *IEEE Signal Processing Magazine*, 25:21–30, Mar. 2007.

[18] S. Barbarossa and G. Scutari. Bio-inspired sensor network design. *IEEE Signal Processing Magazine*, 24(3):26–35, May 2007.

[19] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical Processes on Complex Networks.* Cambridge University Press, 2008.

[20] M. F. Bear, B. W. Connors, and M. A. Paradiso. *Neuroscience: Exploring the Brain.* Lippincott, Williams & Wilkins, 3rd edition, 2006.

[21] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Img. Sci.*, 2:183–202, March 2009.

[22] M. Beekman, R. L. Fathke, and T. D. Seeley. How does an informed minority of scouts guide a honey bee swarm as it flies to its new home? *Animal Behavior*, 71:161–171, 2006.

[23] F. Benezit, V. Blondel, P. Thiran, J. Tsitsiklis, and M. Vetterli. Weighted gossip: Distributed averaging using non-doubly stochastic matrices. In *Proc. IEEE Int. Symp. Inf. Thy*, pages 1753–1757. Austin, TX, Jun. 2010.

[24] F. Benezit, A. G. Dimakis, P. Thiran, and M. Vetterli. Order-optimal consensus through randomized path averaging. *IEEE Trans. Inf. Theory*, 56(10):5150–5167, Oct. 2010.

[25] H. Berg. Motile behavior of bacteria. *Physics Today*, 53(1):24–29, 2000.

[26] R. L. Berger. A necessary and sufficient condition for reaching a consensus using DeGroot's method. *J. Amer. Stat. Assoc.*, 76(374):415–418, Jun. 1981.

[27] A. Berman and R. J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences*. SIAM, PA, 1994.

[28] A. Bertrand, M. Moonen, and A. H. Sayed. Diffusion bias-compensated RLS estimation over adaptive networks. *IEEE Trans. Signal Process.*, 59(11):5212–5224, Nov. 2011.

[29] D. Bertsekas. *Convex Analysis and Optimization*. Athena Scientific, 2003.

[30] D. P. Bertsekas. A new class of incremental gradient methods for least squares problems. *SIAM J. Optim.*, 7(4):913–926, 1997.

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

[32] D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, Singapore, 1st edition, 1997.

[33] D. P. Bertsekas and J. N. Tsitsiklis. Gradient convergence in gradient methods with errors. *SIAM J. Optim.*, 10(3):627–642, 2000.

[34] P. Bianchi, G. Fort, W. Hachem, and J. Jakubowicz. Convergence of a distributed parameter estimator for sensor networks with local averaging of the estimates. In *Proc. IEEE ICASSP*, pages 3764–3767. Prague, Czech, May 2011.

[35] L. Billera and P. Diaconis. A geometric interpretation of the metropolis-hastings algorithm. *Statist. Sci.*, 16:335–339, 2001.

[36] K. Binmore and J. Davies. *Calculus Concepts and Methods*. Cambridge University Press, 2007.

[37] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007.

[38] D. Blatt, A. O. Hero, and H. Gauchman. A convergent incremental gradient method with a constant step size. *SIAM J. Optim.*, 18:29–51, 2008.

[39] V. D. Blondel, J. M. Hendrickx, A. Olshevsky, and J. N. Tsitsiklis. Convergence in multiagent coordination, consensus, and flocking. In *Proc. IEEE Conf. Dec. Control (CDC)*, pages 2996–3000. Seville, Spain, Dec. 2005.

[40] J. R. Blum. Multidimensional stochastic approximation methods. *Ann. Math. Stat.*, 25:737–744, 1954.

[41] B. Bollobas. *Modern Graph Theory.* Springer, 1998.

[42] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing Markov chain on a graph. *SIAM Review*, 46(4):667–689, Dec. 2004.

[43] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah. Randomized gossip algorithms. *IEEE Trans. Inf. Theory*, 52(6):2508–2530, Jun. 2006.

[44] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning,* NOW Publishers, 3(1):1–122, 2010.

[45] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[46] P. Braca, S. Marano, and V. Matta. Running consensus in wireless sensor networks. In *Proc. 11th International Conference on Information Fusion*, pages 1–6. Cologne, Germany, June 2008.

[47] D. H. Brandwood. A complex gradient operator and its apphcation in adaptive array theory. *IEE Proc.*, 130 parts F and H(1):11–16, 1983.

[48] R. A. Brualdi and S. Mellendorf. Regions in the complex plane containing the eigenvalues of a matrix. *Amer. Math. Monthly*, 101:975–985, 1994.

[49] G. Buzsaki. *Rythms of the Brain.* Oxford University Press, 2011.

[50] S. Camazine, J. L. Deneubourg, N. R. Franks, J. Sneyd, G. Theraulaz, and E. Bonabeau. *Self-Organization in Biological Systems.* Princeton University Press, 2003.

[51] E. J. Candes, M. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.*, 14:877–905, 2007.

[52] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri. Distributed Kalman filtering using consensus strategies. *IEEE J. Sel. Areas Communications*, 26(4):622–633, Sep. 2008.

[53] F. Cattivelli and A. H. Sayed. Diffusion distributed Kalman filtering with adaptive weights. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, pages 908–912. Pacific Grove, CA, Nov. 2009.

[54] F. Cattivelli and A. H. Sayed. Diffusion strategies for distributed Kalman filtering and smoothing. *IEEE Trans. Autom. Control*, 55(9):2069–2084, Sep. 2010.

[55] F. Cattivelli and A. H. Sayed. Analysis of spatial and incremental LMS processing for distributed estimation. *IEEE Trans. Signal Process.*, 59(4):1465–1480, April 2011.

[56] F. Cattivelli and A. H. Sayed. Modeling bird flight formations using diffusion adaptation. *IEEE Trans. Signal Process.*, 59(5):2038–2051, May 2011.

[57] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed. A diffusion RLS scheme for distributed estimation over adaptive networks. In *Proc. IEEE Work. Signal Process. Adv. Wireless Comm. (SPAWC)*, pages 1–5. Helsinki, Finland, June 2007.

[58] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed. Diffusion recursive least-squares for distributed estimation over adaptive networks. *IEEE Trans. Signal Process.*, 56(5):1865–1877, May 2008.

[59] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed. Diffusion strategies for distributed Kalman filtering: Formulation and performance analysis. In *Proc. Int. Work. Cogn. Inform. Process. (CIP)*, pages 36–41. Santorini, Greece, June 2008.

[60] F. S. Cattivelli and A. H. Sayed. Diffusion LMS algorithms with information exchange. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, pages 251–255. Pacific Grove, CA, Nov. 2008.

[61] F. S. Cattivelli and A. H. Sayed. Diffusion mechanisms for fixed-point distributed Kalman smoothing. In *Proc. EUSIPCO*, pages 1–4. Lausanne, Switzerland, Aug. 2008.

[62] F. S. Cattivelli and A. H. Sayed. Diffusion LMS strategies for distributed estimation. *IEEE Trans. Signal Process.*, 58(3):1035–1048, Mar. 2010.

[63] R. Cavalcante, I. Yamada, and B. Mulgrew. An adaptive projected subgradient approach to learning in diffusion networks. *IEEE Trans. Signal Process.*, 57(7):2762–2774, July 2009.

[64] C. Chamley, A. Scaglione, and L. Li. Models for the diffusion of beliefs in social networks. *IEEE Signal Processing Magazine*, 30, May 2013.

[65] J. Chen and A. H. Sayed. Bio-inspired cooperative optimization with application to bacteria motility. In *Proc. IEEE ICASSP*, pages 5788–5791. Prague, Czech Republic, May 2011.

[66] J. Chen and A. H. Sayed. Diffusion adaptation strategies for distributed optimization and learning over networks. *IEEE Trans. Signal Process.*, 60(8):4289–4305, Aug. 2012.

[67] J. Chen and A. H. Sayed. Distributed pareto-optimal solutions via diffusion adaptation. In *Proc. IEEE Work. Stat. Signal Process. (SSP)*, pages 648–651. Ann Arbor, MI, Aug. 2012.

[68] J. Chen and A. H. Sayed. On the limiting behavior of distributed optimization strategies. In *Proc. 50th Annual Allerton Conference on Communication, Control, and Computing*, pages 1535–1542. Monticello, IL, Oct. 2012.

[69] J. Chen and A. H. Sayed. Distributed pareto optimization via diffusion strategies. *IEEE J. Selected Topics in Signal Processing*, 7(2):205–220, April 2013.

[70] J. Chen and A. H. Sayed. On the learning behavior of adaptive networks — Part I: Transient analysis. *Submitted for publication.* Also available as arXiv:1312.7581 [cs.MA], Dec. 2013.

[71] J. Chen and A. H. Sayed. On the learning behavior of adaptive networks — Part II: Performance analysis. *Submitted for publication.* Also available as arXiv:1312.7580 [cs.MA], Dec. 2013.

[72] J. Chen and A. H. Sayed. Controlling the limit point of left-stochastic policies over adaptive networks. *Submitted for publication*, 2014.

[73] Y. Chen, Y. Gu, and A. O. Hero. Sparse LMS for system identification. In *Proc. IEEE ICASSP*, pages 3125–3128. Taipei, Taiwan, May 2009.

[74] S. Chouvardas, G. Mileounis, N. Kalouptsidis, and S. Theodoridis. A greedy sparsity-promoting LMS for distributed adaptive learning in diffusion networks. In *Proc. IEEE ICASSP*, pages 5415–5419. Vancouver, BC, Canada, 2013.

[75] S. Chouvardas, K. Slavakis, Y. Kopsinis, and S. Theodoridis. A sparsity-promoting adaptive algorithm for distributed learning. *IEEE Trans. Signal Process.*, 60(10):5412–5425, Oct. 2012.

[76] S. Chouvardas, K. Slavakis, and S. Theodoridis. Adaptive robust distributed learning in diffusion sensor networks. *IEEE Trans. Signal Process.*, 59(10):4692–4707, Oct. 2011.

[77] N. Christakis and J. Fowler. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives.* Little, Brown and Company, 2009.

[78] F. Iutzeler P. Ciblat and W. Hachem. Analysis of sum-weight-like algorithms for averaging in wireless sensor networks. *IEEE Trans. Signal Process.*, 6(11):2802–2814, Jun. 2013.

[79] I. D. Couzin. Collective cognition in animal groups. *Trends in Cognitive Sciences*, 13:36–43, Jan. 2009.

[80] I. D. Couzin, J. Krause, R. James, G. D. Ruxton, and N. R. Franks. Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology*, 218:1–11, 2002.

[81] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, NJ, 1991.

[82] D. M. Cvetković, M. Doob, and H. Sachs. *Spectra of Graphs: Theory and Applications.* Wiley, NY, 1998.

[83] A. Das and M. Mesbahi. Distributed linear parameter estimation in sensor networks based on laplacian dynamics consensus algorithm. In *Proc. IEEE SECON*, volume 2, pages 440–449. Reston, VA, Sep. 2006.

[84] M. H. DeGroot. Reaching a consensus. *J. Amer. Stat. Assoc.*, 69(345):118–121, 1974.

[85] O. Dekel, R. Gilad-Bachrach, O. Shamir, and L. Xiao. Optimal distributed online prediction. In *Proc. International Conference on Machine Learning (ICML)*, pages 713–720. Bellevue, WA, Jun. 2011.

[86] P. Di Lorenzo and A. H. Sayed. Sparse distributed learning based on diffusion adaptation. *IEEE Trans. Signal Process.*, 61(6):1419–1433, March 2013.

[87] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione. Gossip algorithms for distributed signal processing. *Proceedings of the IEEE*, 98(11):1847–1864, Nov. 2010.

[88] P. M. Djuric and Y. Wang. Distributed Bayesian learning in multiagent systems. *IEEE Signal Processing Magazine*, 29(2):65–76, Mar. 2012.

[89] R. M. Dudley. *Real Analysis and Probability.* Cambridge Univ. Press, 2nd edition, 2003.

[90] L. A. Dugatkin. *Principles of Animal Behavior.* W. W. Norton & Company, 2nd edition, 2009.

[91] R. Durret. *Probability Theory and Examples.* Duxbury Press, 2nd edition, 1996.

[92] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World.* Cambridge University Press, 2010.

[93] C. H. Edwards Jr. *Advanced Calculus of Several Variables*. Dover Publications, NY, 1995.

[94] D. G. Feingold and R. S. Varga. Block diagonally dominant matrices and generalizations of the gerschgorin circle theorem. *Pacific J. Math.*, 12:1241–1250, 1962.

[95] J. Fernandez-Bes, J. Arenas-Garcia, and A. H. Sayed. Adjustment of combination weights over adaptive diffusion networks. In *Proc. IEEE ICASSP*, pages 1–5. Florence, Italy, May 2014.

[96] A. Feuer and E. Weinstein. Convergence analysis of LMS filters with uncorrelated Gaussian data. *IEEE Trans. Acoust., Speech, Signal Process.*, 33(1):222–230, Feb. 1985.

[97] J. B. Foley and F. M. Boland. A note on the convergence analysis of LMS adaptive filters with Gaussian data. *IEEE Trans. Acoust., Speech, Signal Process.*, 36(7):1087–1089, Jul. 1988.

[98] J. Fowler and N. Christakis. Cooperative behavior cascades in human social networks. *Proc. Nat. Acad. Sciences*, 107(12):5334–5338, 2010.

[99] F. R. Gantmacher. *The Theory of Matrices*. Chelsea Publishing Company, NY, 1959.

[100] W. A. Gardner. Learning characterisitcs of stochastic-gradient-descent algorithms: A general study, analysis, and critique. *Signal Process.*, 6(2):113–133, Apr. 1984.

[101] S. Gerschgorin. Über die abgrenzung der eigenwerte einer matrix. *Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk*, 7:749–754, 1931.

[102] O. N. Gharehshiran, V. Krishnamurthy, and G. Yin. Distributed energy-aware diffusion least mean squares: Game-theoretic learning. *IEEE J. Sel. Top. Signal Process.*, 7(5):1–16, Oct. 2013.

[103] B. Golub and M. O. Jackson. Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2:112–149, 2010.

[104] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, 3rd edition, 1996.

[105] W. D. Hamilton. Geometry for the selfish herd. *Journal of Theoretical Biology*, 31:295–311, 1971.

[106] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, Apr. 1970.

[107] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, NJ, 2002.

[108] S. Haykin. *Cognitive Dynamic Systems*. Cambridge University Press, 2012.

[109] E. S. Helou and A. R. De Pierro. Incremental subgradients for constrained convex optimization: A unified framework and new methods. *SIAM J. Optim.*, 20:1547–1572, 2009.

[110] F. H. Heppner. Avian flight formations. *Bird-Banding*, 45(2):160–169, 1974.

[111] A. Hjorungnes. *Complex-Valued Matrix Derivatives*. Cambridge University Press, 2011.

[112] O. Hlinka, O. Sluciak, F. Hlawatsch, and P. M. Djuric. Likelihood consensus and its application to distributed particle filtering. *IEEE Trans. Signal Process.*, 60(8):4334–4349, August 2012.

[113] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2003.

[114] L. Horowitz and K. Senne. Performance advantage of complex LMS for controlling narrow-band adaptive arrays. *IEEE Trans. Acoust., Speech, Signal Process.*, 29(3):722–736, Jun. 1981.

[115] D. W. Hosmer and S. Lemeshow. *Applied Logistic Regression*. Wiley, NJ, 2nd edition, 2000.

[116] K. Hreutz-Delgado. The complex gradient operator and the cr-calculus. Available online as manuscript arXiv:0906.4835 [math.OC], June 2009.

[117] J. Hu, L. Xie, and C. Zhang. Diffusion Kalman filtering based on covariance intersection. *IEEE Trans. Signal Process.*, 60(2):891–902, Feb. 2012.

[118] S. Hubbard, P. Babak, S. T. Sigurdsson, and K. G. Magnusson. A model of the formation of fish schools and migrations of fish. *Ecological Modeling*, 174:359–374, June 2004.

[119] D. Hummel. Aerodynamic aspects of formation flight in birds. *J. Theor. Biol.*, 104(3):321–347, 1983.

[120] M. D. Intriligator. *Mathematical Optimization and Economic Theory*. Prentice-Hall, NJ, 1971.

[121] M. Jackson. *Social and Economic Networks*. Princeton University Press, Princeton, NJ, 2008.

[122] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. Autom. Control*, 48(6):988–1001, Jun. 2003.

[123] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi. Non-bayesian social learning. *Game. Econ. Behav.*, 76(1):210–225, Sep. 2012.

[124] D. Jakovetic, J. Xavier, and J. M. F. Moura. Cooperative convex optimization in netowrked systems: Augmented lagranian algorithms with directed gossip communication. *IEEE Trans. Signal Process.*, 59(8):3889–3902, Aug. 2011.

[125] S. Janson, M. Middendorf, and M. Beekman. Honeybee swarms: How do scouts guide a swarm of uninformed bees? *Animal Behavior*, 70:349–358, 2005.

[126] J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.

[127] C. Jiang, Y. Chen, and K. J. Ray Liu. Distributed adaptive networks: A graphical evolutionary game-theoretic view. *IEEE Trans. Signal Process.*, 61(22):5675–5688, Nov. 2013.

[128] B. Johansson, T. Keviczky, M. Johansson, and K. Johansson. Subgradient methods and consensus algorithms for solving convex optimization problems. In *Proc. IEEE Conf. Dec. Control (CDC)*, pages 4185–4190. Cancun, Mexico, December 2008.

[129] B. Johansson, M. Rabi, and M. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM J. Optim.*, 20:1157–1170, 2009.

[130] S. Jones, R. C. III, and W. Reed. Analysis of error-gradient adaptive linear estimators for a class of stationary dependent processes. *IEEE Trans. Inf. Theory*, 28(2):318–329, Mar. 1982.

[131] B. H. Junker and F. Schreiber. *Analysis of Biological Networks*. Wiley, NJ, 2008.

[132] T. Kailath. *Linear Systems*. Prentice Hall, NJ, 1980.

[133] T. Kailath, A. H. Sayed, and B. Hassibi. *Linear Estimation*. Prentice Hall, NJ, 2000.

[134] S. Kar and J. M. F. Moura. Sensor networks with random links: Topology design for distributed consensus. *IEEE Trans. Signal Process.*, 56(7):3315–3326, July 2008.

[135] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks: Link failures and channel noise. *IEEE Trans. Signal Process.*, 57(1):355–369, Jan. 2009.

[136] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor networks: Link failures and channel noise. *IEEE Trans. Signal Process.*, 57(1):355–369, Jan. 2009.

[137] S. Kar and J. M. F. Moura. Distributed consensus algorithms in sensor netowrks: quantized data and random link failures. *IEEE Trans. Signal Process.*, 58(3):1383–1400, Mar. 2010.

[138] S. Kar and J. M. F. Moura. Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs. *IEEE Journal on Selected Topics in Signal Processing*, 5(4):674–690, Aug. 2011.

[139] S. Kar, J. M. F. Moura, and K. Ramanan. Distributed parameter estimation in sensor networks: Nonlinear observation models and imperfect communication. *IEEE Trans. Inf. Theory*, 58(6):3575–3605, Jun. 2012.

[140] R. M. Karp. Reducibility among combinational problems. In R. E. Miller and J. W. Thatcher, editors, *Complexity of Computer Computations*, pages 85–104. Plenum Press, NY, 1972.

[141] D. Kempe, A Dobra, and J. Gehrke. Gossip-based computation of aggregate information. In *Proc. Annual IEEE Symp. Found. Computer Sci.*, pages 482–491. Cambridge, MA, Oct. 2003.

[142] A. Khalili, M. A. Tinati, A. Rastegarnia, and J. A. Chambers. Steady state analysis of diffusion LMS adaptive networks with noisy links. *IEEE Trans. Signal Process.*, 60(2):974–979, Feb. 2012.

[143] U. A. Khan and J. M. F. Moura. Distributing the Kalman filter for large-scale systems. *IEEE Trans. Signal Process.*, 56(10):4919–4935, Oct. 2008.

[144] W. Kocay and D. L. Kreher. *Graphs, Algorithms and Optimization*. Chapman & Hall/CRC Press, Boca Raton, 2005.

[145] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Dover Publications, 1975.

[146] R. H. Koning, H. Neudecker, and T. Wansbeek. Block kronecker products and the vecb operator. *Linear Algebra Appl.*, 149:165–184, Apr. 1991.

[147] F. Kopos, editor. *Biological Networks*. World Scientific Publishing Company, 2007.

[148] Y. Kopsinis, K. Slavakis, and S. Theodoridis. Online sparse system identification and signal reconstruction using projections onto weighted balls. *IEEE Trans. Signal Process.*, 59(3):936–952, Mar. 2010.

[149] P. Lancaster and L. Rodman. *Algebraic Riccati Equations*. Oxford University Press, NY, 1995.

[150] P. Lancaster and M. Tismenetsky. *Theory of Matrices with Applications*. Academic Press, NY, 2nd edition, 1985.

[151] R. Larson and B. H. Edwards. *Calculus*. Brooks Cole, 9th edition, 2009.

[152] J-W. Lee, S-E. Kim, W.-J. Song, and A. H. Sayed. Spatio-temporal diffusion mechanisms for adaptation over networks. In *Proc. EUSIPCO*, pages 1040–1044. Barcelona, Spain, Aug.–Sep. 2011.

[153] J.-W. Lee, S.-E. Kim, W.-J. Song, and A. H. Sayed. Spatio-temporal diffusion strategies for estimation and detection over networks. *IEEE Trans. Signal Process.*, 60(8):4017–4034, August 2012.

[154] S. Lee and A. Nedic. Distributed random projection algorithm for convex optimization. *IEEE J. Selected Topics in Signal Processing*, 7(2):221–229, Apr. 2013.

[155] T. G. Lewis. *Network Science: Theory and Applications*. Wiley, NJ, 2009.

[156] J. Li and A. H. Sayed. Modeling bee swarming behavior through diffusion adaptation with asymmetric information sharing. *EURASIP Journal on Advances in Signal Processing*, 2012. 2012:18, doi:10.1186/1687-6180-2012-18.

[157] L. Li, C. G. Lopes, J. Chambers, and A. H. Sayed. Distributed estimation over an adaptive incremental network based on the affine projection algorithm. *IEEE Trans. Signal Process.*, 58(1):151–164, Jan. 2010.

[158] Y. Liu, C. Li, and Z. Zhang. Diffusion sparse least-mean squares over networks. *IEEE Trans. Signal Process.*, 60(8):4480–4485, Aug. 2012.

[159] C. Lopes and A. H. Sayed. Diffusion adaptive networks with changing topologies. In *Proc. IEEE ICASSP*, pages 3285–3288. Las Vegas, April 2008.

[160] C. G. Lopes and A. H. Sayed. Distributed processing over adaptive networks. In *Proc. Adaptive Sensor Array Processing Workshop*, pages 1–5. MIT Lincoln Laboratory, MA, June 2006.

[161] C. G. Lopes and A. H. Sayed. Diffusion least-mean-squares over adaptive networks. In *Proc. IEEE ICASSP*, volume 3, pages 917–920. Honolulu, Hawaii, April 2007.

[162] C. G. Lopes and A. H. Sayed. Incremental adaptive strategies over distributed networks. *IEEE Trans. Signal Process.*, 55(8):4064–4077, Aug. 2007.

[163] C. G. Lopes and A. H. Sayed. Steady-state performance of adaptive diffusion least-mean squares. In *Proc. IEEE Work. Stat. Signal Process. (SSP)*, pages 136–140. Madison, WI, Aug. 2007.

[164] C. G. Lopes and A. H. Sayed. Diffusion least-mean squares over adaptive networks: Formulation and performance analysis. *IEEE Trans. Signal Process.*, 56(7):3122–3136, July 2008.

[165] O. Macchi. *Adaptive Processing: The Least Mean Squares Approach with Applications in Transmission.* Wiley, NY, 1995.

[166] G. Mateos, Gonzalo, I. D. Schizas, and G. B. Giannakis. Distributed recursive least-squares for consensus-based in-network adaptive estimation. *IEEE Trans. Signal Process.*, 57(11):4583–4599, Nov. 2009.

[167] G. Mateos, I. D. Schizas, and G. B. Giannakis. Performance analysis of the consensus-based distributed LMS algorithm. *EURASIP J. Adv. Signal Process.*, pages 1–19, 2009. 10.1155/2009/981030, Article ID 981030.

[168] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[169] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra.* SIAM, PA, 2001.

[170] S. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Cambridge Univ. Press, 2nd edition, 2009.

[171] H. Milinski and R. Heller. Influence of a predator on the optimal foraging behavior of sticklebacks. *Nature*, 275:642–644, 1978.

[172] D. S. Mitrinović. *Elementary Inequalities.* P. Noordhoff Ltd., Netherlands, 1964.

[173] A. Nedic and D. P. Bertsekas. Incremental subgradient methods for nondifferentiable optimization. *SIAM J. Optim.*, 12(1):109–138, 2001.

[174] A. Nedic and A. Olshevsky. Distributed optimization over time-varying directed graphs. *Submitted for publication.* Also available as arXiv:1303.2289 [math.OC], Mar. 2014.

[175] A. Nedic and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *IEEE Trans. Autom. Control*, 54(1):48–61, Jan. 2009.

[176] A. Nedic and A. Ozdaglar. Cooperative distributed multi-agent optimization. In Y. Eldar and D. Palomar, editors, *Convex Optimization in Signal Processing and Communications*, pages 340–386. Cambridge University Press, 2010.

[177] Y. Nesterov. A method for solving the convex programming problem with convergence rate $o(1/k^2)$. *Dokl. Akad. Nauk SSSR*, 269(3):543–547, 1983.

[178] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, 2004.

[179] M. Newman. *Networks: An Introduction*. Oxford University Press, 2010.

[180] R. Olfati-Saber. Distributed Kalman filter with embedded consensus filters. In *Proc. IEEE Conf. Dec. Control (CDC)*, pages 8179–8184. Seville, Spain, Dec. 2005.

[181] R. Olfati-Saber. Flocking for multi-agent dynamic systems: Algorithms and theory. *IEEE Trans. Autom. Control*, 51:401–420, Mar. 2006.

[182] R. Olfati-Saber. Distributed Kalman filtering for sensor networks. In *Proc. 46th IEEE Conf. Decision Control*, pages 5492–5498. New Orleans, LA, Dec. 2007.

[183] R. Olfati-Saber. Kalman-consensus filter: Optimality, stability, and performance. In *Proc. IEEE Conf. Dec. Control (CDC)*, pages 7036–7042. Shangai, China, 2009.

[184] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, Jan. 2007.

[185] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time-delays. *IEEE Trans. Autom. Control*, 49:1520–1533, Sep. 2004.

[186] R. Olfati-Saber and J. Shamma. Consensus filters for sensor networks and distributed sensor fusion. In *Proc. IEEE Conf. Dec. Control (CDC)*, pages 6698–6703. Seville, Spain, Dec. 2005.

[187] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, NY, 4th edition, 2002.

[188] B. L. Partridge. The structure and function of fish schools. *Scientific American*, 246(6):114–123, June 1982.

[189] K. Passino. Biomimicry of bacterial foraging for distributed optimization and control. *IEEE Control Systems Magazine*, 22(6):52–67, 2002.

[190] S. U. Pillai, T. Suel, and S. Cha. The perron–frobenius theorem: Some of its applications. *IEEE Signal Process. Mag.*, 22(2):62–75, Mar. 2005.

[191] B. Poljak. *Introduction to Optimization*. Optimization Software, NY, 1987.

[192] B. T. Poljak and Y. Z. Tsypkin. Pseudogradient adaptation and training algorithms. *Autom. Remote Control*, 12:83–94, 1973.

[193] J. B. Predd, S. B. Kulkarni, and H. V. Poor. Distributed learning in wireless sensor networks. *IEEE Signal Processing Magazine*, 23(4):56–69, Jul. 2006.

[194] J. B. Predd, S. R. Kulkarni, and H. V. Poor. A collaborative training algorithm for distributed learning. *IEEE Trans. Inf. Theory*, 55(4):1856–1871, April 2009.

[195] M. G. Rabbat and R. D. Nowak. Quantized incremental algorithms for distributed optimization. *IEEE J. Sel. Areas Commun.*, 23(4):798–808, 2005.

[196] M. G. Rabbat, R. D. Nowak, and J. A. Bucklew. Generalized consensus computation in networked systems with erasure links. In *Proc. IEEE Work. Signal Process. Adv. Wireless Comm. (SPAWC)*, pages 1088–1092. New York, NY, June 2005.

[197] S. S. Ram, A. Nedic, and V. V. Veeravalli. Distributed stochastic subgradient projection algorithms for convex optimization. *J. Optim. Theory Appl.*, 147(3):516–545, 2010.

[198] R. Remmert. *Theory of Complex Functions*. Springer-Verlag, 1991.

[199] W. Ren and R. W. Beard. Consensus seeking in multi-agent systems under dynamically changing interaction topologies. *IEEE Trans. Autom. Control*, 50:655–661, May 2005.

[200] C. W. Reynolds. Flocks, herds, and schools: A distributed behavior model. *ACM Proc. Comput. Graphs Interactive Tech.*, pages 25–34, 1987.

[201] H. Robbins and S. Monro. A stochastic approximation method. *Ann. Math. Stat.*, 22:400–407, 1951.

[202] O. L. Rortveit, J. H. Husoy, and A. H. Sayed. Diffusion LMS with communications constraints. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, pages 1645–1649. Pacific Grove, CA, Nov. 2010.

[203] H. L. Royden. *Real Analysis*. Prentice-Hall, NJ, 3rd edition, 1988.

[204] V. Saligrama, M. Alanyali, and O. Savas. Distributed detection in sensor networks with packet losses and finite capacity links. *IEEE Trans. Signal Process.*, 54:4118–4132, 2006.

[205] S. Sardellitti, M. Giona, and S. Barbarossa. Fast distributed average consensus algorithms based on advection-diffusion processes. *IEEE Trans. Signal Process.*, 58(2):826–842, Feb. 2010.

[206] A. H. Sayed. *Fundamentals of Adaptive Filtering.* Wiley, NJ, 2003.

[207] A. H. Sayed. *Adaptive Filters.* Wiley, NJ, 2008.

[208] A. H. Sayed. Adaptive networks. *Proceedings of the IEEE*, 102(4):460–497, April 2014.

[209] A. H. Sayed. Diffusion adaptation over networks. In R. Chellapa and S. Theodoridis, editors, *E-Reference Signal Processing*, vol. 3, pages 323–454. Academic Press, 2014. Also available as arXiv:1205.4220v1 [cs.MA], May 2012.

[210] A. H. Sayed and F. Cattivelli. Distributed adaptive learning mechanisms. In S. Haykin and K. J. Ray Liu, editors, *Handbook on Array Processing and Sensor Networks*, pages 695–722. Wiley, NJ, 2009.

[211] A. H. Sayed and C. Lopes. Distributed recursive least-squares strategies over adaptive networks. In *Proc. Asilomar Conf. Signals, Syst. Comput.*, pages 233–237. Pacific Grove, CA, Oct.-Nov. 2006.

[212] A. H. Sayed and C. G. Lopes. Adaptive processing over distributed networks. *IEICE Trans. Fund. of Electron., Commun. and Comput. Sci.*, E90-A(8):1504–1510, 2007.

[213] A. H. Sayed and F. A. Sayed. Diffusion adaptation over networks of particles subject to brownian fluctuations. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, pages 685–690. Pacific Grove, CA, Nov. 2011.

[214] A. H. Sayed, S-Y. Tu, and J. Chen. Online learning and adaptation over networks: More information is not necessarily better. In *Proc. Information Theory and Applications Workshop (ITA)*, pages 1–8. San Diego, Feb. 2013.

[215] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. Towfic. Diffusion strategies for adaptation and learning over networks. *IEEE Signal Processing Magazine*, 30(3):155–171, May 2013.

[216] D. S. Scherber and H. C. Papadopoulos. Locally constructed algorithms for distributed computations in ad-hoc networks. In *Proc. Information Processing in Sensor Networks (IPSN)*, pages 11–19. Berkeley, CA, April 2004.

[217] I. D. Schizas, G. Mateos, and G. B. Giannakis. Distributed LMS for consensus-based in-network adaptive processing. *IEEE Trans. Signal Process.*, 57(6):2365–2382, June 2009.

[218] L. Schmetterer. Stochastic approximation. *Proc. Berkeley Symp. Math. Statist. Probab.*, pages 587–609, 1961.

[219] P. J. Schreier and L. L. Scharf. *Statistical Signal Processing of Complex-Valued Data*. Cambridge University Press, 2010.

[220] T. D. Seeley, R. A. Morse, and P. K. Visscher. The natural history of the flight of honey bee swarms. *Psyche.*, 86:103–114, 1979.

[221] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, 2nd edition, 2007.

[222] D. Shah. Gossip algorithms. *Found. Trends Netw.*, 3:1–125, 2009.

[223] K. Slavakis, Y. Kopsinis, and S. Theodoridis. Adaptive algorithm for sparse system identification using projections onto weighted ℓ_1 balls. In *Proc. IEEE ICASSP*, pages 3742–3745. Dallas, TX, Mar. 2010.

[224] S. Sonnenburg, V. Franc, E. Yom-Tov, and M. Sebag. Pascal large scale learning challenge. Online site at http://largescale.ml.tu-berlin.de, 2008.

[225] A. Speranzon, C. Fischione, and K. H. Johansson. Distributed and collaborative estimation over wireless sensor networks. In *Proc. IEEE Conf. Dec. Control (CDC)*, pages 1025–1030. San Dieog, USA, Dec. 2006.

[226] O. Sporns. *Networks of the Brain*. MIT Press, 2010.

[227] K. Srivastava and A. Nedic. Distributed asynchronous constrained stochastic optimization. *IEEE J. Sel. Topics. Signal Process.*, 5(4):772–790, Aug. 2011.

[228] S. S. Stankovic, M. S. Stankovic, and D. S. Stipanovic. Decentralized parameter estimation by consensus based stochastic approximation. *IEEE Trans. Autom. Control*, 56(3):531–543, Mar. 2011.

[229] D. J. T. Sumpter and S. C. Pratt. A modeling framework for understanding social insect foraging. *Behavioral Ecology and Sociobiology*, 53:131–144, 2003.

[230] J. Surowiecki. *The Wisdom of the Crowds*. Doubleday, 2004.

[231] N. Takahashi and I. Yamada. Parallel algorithms for variational inequalities over the cartesian product of the intersections of the fixed point sets of nonexpansive mappings. *J. Approx. Theory*, 153(2):139–160, Aug. 2008.

[232] N. Takahashi and I. Yamada. Link probability control for probabilistic diffusion least-mean squares over resource-constrained networks. In *Proc. IEEE ICASSP*, pages 3518–3521. Dallas, TX, Mar. 2010.

[233] N. Takahashi, I. Yamada, and A. H. Sayed. Diffusion least-mean-squares with adaptive combiners: Formulation and performance analysis. *IEEE Trans. Signal Process.*, 58(9):4795–4810, Sep. 2010.

[234] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 4th edition, 2008.

[235] S. Theodoridis, K. Slavakis, and I. Yamada. Adaptive learning in a world of projections: A unifying framework for linear and nonlinear classification and regression tasks. *IEEE Signal Processing Magazine*, 28(1):97–123, Jan. 2011.

[236] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal Statistical Society: Series B*, 58:267–288, 1996.

[237] Z. Towfic, J. Chen, and A. H. Sayed. On the generalization ability of distributed online learners. In *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. Santander, Spain, Sep. 2012.

[238] Z. Towfic and A. H. Sayed. Adaptive stochastic convex optimization over networks. In *Proc. 51th Annual Allerton Conference on Communication, Control, and Computing*, pages 1–6. Monticello, IL, Oct. 2013.

[239] Z. Towfic and A. H. Sayed. Adaptive penalty-based distributed stochastic convex optimization. *IEEE Trans. Signal Process.*, 62(15):3924–3938, Aug. 2014.

[240] Z. J. Towfic, J. Chen, and A. H. Sayed. Collaborative learning of mixture models using diffusion adaptation. In *Proc. IEEE Workshop Mach. Learn. Signal Process. (MLSP)*, pages 1–6. Beijing, China, Sep. 2011.

[241] K. I. Tsianos, S. Lawlor, and M. G. Rabbat. Push-sum distributed dual averaging for convex optimization. In *Proc. IEEE Conf. Dec. Control (CDC)*, pages 5453–5458. Hawaii, Dec. 2012.

[242] J. Tsitsiklis and M. Athans. Convergence and asymptotic agreement in distributed decision problems. *IEEE Trans. Autom. Control*, 29(1):42–50, Jan. 1984.

[243] J. Tsitsiklis, D. Bertsekas, and M. Athans. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. *IEEE Trans. Autom. Control*, 31(9):803–812, Sep. 1986.

[244] Y. Z. Tsypkin. *Adaptation and Learning in Automatic Systems*. Academic Press, NY, 1971.

[245] S-Y. Tu and A. H. Sayed. Adaptive networks with noisy links. In *Proc. IEEE Globecom*, pages 1–5. Houston, TX, December 2011.

[246] S-Y. Tu and A. H. Sayed. Cooperative prey herding based on diffusion adaptation. In *Proc. IEEE ICASSP*, pages 3752–3755. Prague, Czech Republic, May 2011.

[247] S.-Y. Tu and A. H. Sayed. Mobile adaptive networks. *IEEE J. Sel. Topics. Signal Process.*, 5(4):649–664, Aug. 2011.

[248] S-Y. Tu and A. H. Sayed. On the effects of topology and node distribution on learning over complex adaptive networks. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, pages 1166–1171. Pacific Grove, CA, Nov. 2011.

[249] S.-Y. Tu and A. H. Sayed. Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks. *IEEE Trans. Signal Process.*, 60(12):6217–6234, Dec. 2012.

[250] S-Y. Tu and A. H. Sayed. Effective information flow over mobile adaptive networks. In *Proc. Int. Work. Cogn. Inform. Process. (CIP)*, pages 1–6. Parador de Baiona, Spain, May 2012.

[251] S-Y. Tu and A. H. Sayed. On the influence of informed agents on learning and adaptation over networks. *IEEE Trans. Signal Process.*, 61(6):1339–1356, Mar. 2013.

[252] A. van den Bos. Complex gradient and hessian. *IEE Proc. Vis. Image Signal Process.*, 141(6):380–382, 1994.

[253] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer, NY, 2000.

[254] R. S. Varga. *Gersgorin and His Circles.* Springer-Verlag, Berlin, 2004.

[255] T. Vicsek, A. Czirook, E. Ben-Jacob, O. Cohen, and I. Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75:1226–1229, Aug. 1995.

[256] R. von Mises and H. Pollaczek-Geiringer. Praktische verfahren der gleichungs-auflösung. *Z. Agnew. Math. Mech.*, 9:152–164, 1929.

[257] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1–2):1–305, 2008.

[258] C. Waters and B. Bassler. Quorum sensing: cell-to-cell communication in bacteria. *Annual Review of Cell and Developmental Biology*, 21:319–346, 2005.

[259] G. B. Wetherhill. *Sequential Methods in Statistics.* Methuen, London, 1966.

[260] H. Weyl. Über beschrankte quadratiche formen, deren differenz vollsteig ist. *Rend. Circ. Mat. Palermo*, 27:373–392, 1909.

[261] B. Widrow and M. E. Hoff, Jr. Adaptive switching circuits. *IRE WESCON Conv. Rec.*, Pt. 4:96–104, 1960.

[262] B. Widrow, J. M. McCool, M. G. Larimore, and C. R. Johnson Jr. Stationary and nonstationary learning characterisitcs of the LMS adaptive filter. *Proceedings of the IEEE*, 64(8):1151–1162, Aug. 1976.

[263] B. Widrow and S. D. Stearns. *Adaptive Signal Processing.* Prentice Hall, NJ, 1985.

[264] J. H. Wilkinson. *The Algebraic Eigenvalue Problem.* Oxford University Press, 1965.

[265] W. Wirtinger. Zur formalen theorie der funktionen von mehr komplexen veränderlichen. *Math. Ann.*, 97:357–375, 1927.

[266] L. Xiao and S. Boyd. Fast linear iterations for distributed averaging. *Syst. Control Lett.*, 53(1):65–78, Sep. 2004.

[267] L. Xiao, S. Boyd, and S. Lall. A scheme for robust distributed sensor fusion based on average consensus. In *Proc. IPSN, 2005*, pages 63–70. Los Angeles, CA, April 2005.

[268] L. Xiao, S. Boyd, and S. Lall. A space-time diffusion scheme peer-to-peer least-squares-estimation. In *Proc. Information Processing in Sensor Networks (IPSN)*, pages 168–176. Nashville, TN, April 2006.

[269] F. Yan, S. Sundaram, S. V. N. Vishwanathan, and Y. Qi. Distributed autonomous online learning: Regrets and intrinsic privacy-preserving properties. *IEEE Trans. Knowledge and Data Engineering*, 25(11):2483–2493, Nov. 2013.

[270] N. R. Yousef and A. H. Sayed. A unified approach to the steady-state and tracking analysis of adaptive filters. *IEEE Trans. Signal Process.*, 49(2):314–324, February 2001.

[271] C.-K. Yu and A. H. Sayed. A strategy for adjusting combination weights over adaptive networks. In *Proc. IEEE ICASSP*, pages 4579–4583. Vancouver, Canada, May 2013.

[272] C.-K. Yu, M. van der Schaar, and A. H. Sayed. Reputation design for adaptive networks with selfish agents. In *Proc. IEEE Work. Signal Process. Adv. Wireless Comm. (SPAWC)*, pages 160–164. Darmstadt, Germany, June 2013.

[273] L. A. Zadeh. Optimality and non-scalar-valued performance criteria. *IEEE Trans. Autom. Control*, 8:59–60, Jan. 1963.

[274] X. Zhao and A. H. Sayed. Clustering via diffusion adaptation over networks. In *Proc. Int. Work. Cogn. Inform. Process. (CIP)*, pages 1–6. Parador de Baiona, Spain, May 2012.

[275] X. Zhao and A. H. Sayed. Combination weights for diffusion strategies with imperfect information exchange. In *Proc. IEEE ICC*, pages 398–402. Ottawa, Canada, June 2012.

[276] X. Zhao and A. H. Sayed. Learning over social networks via diffusion adaptation. In *Proc. Asilomar Conf. Signals, Syst., Comput.*, pages 709–713. Pacific Grove, CA, Nov. 2012.

[277] X. Zhao and A. H. Sayed. Performance limits for distributed estimation over LMS adaptive networks. *IEEE Trans. Signal Process.*, 60(10):5107–5124, Oct. 2012.

[278] X. Zhao and A. H. Sayed. Asynchronous adaptation and learning over networks — Part I: Modeling and stability analsysis. *Submitted for publication.* Also available as arXiv:1312.5434 [cs.SY], Dec. 2013.

[279] X. Zhao and A. H. Sayed. Asynchronous adaptation and learning over networks – Part II: Performance analysis. *Submitted for publication.* Also available as arXiv:1312.5438 [cs.SY], Dec. 2013.

[280] X. Zhao and A. H. Sayed. Attaining optimal batch performance via distributed processing over networks. In *Proc. IEEE ICASSP*, pages 5214–5218. Vancouver, Canada, May 2013.

[281] X. Zhao, S.-Y. Tu, and A. H. Sayed. Diffusion adaptation over networks under imperfect information exchange and non-stationary data. *IEEE Trans. Signal Process.*, 60(7):3460–3475, July 2012.