The Interplay Between Information and Estimation Measures

Dongning Guo
Northwestern University
Evanston, IL 60208, U.S.A.

Shlomo Shamai (Shitz)
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel

Sergio Verdú
Princeton University
Princeton, NJ 08544, U.S.A.

now
the essence of knowledge
Boston — Delft

Full text available at: http://dx.doi.org/10.1561/2000000018

Foundations and Trends® in Signal Processing

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

D. Guo, S. Shamai, S. Verdú. *The Interplay Between Information and Estimation Measures*. Foundations and Trends® in Signal Processing, vol. 6, no. 4, pp. 243–429, 2012.

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

ISBN: 978-1-60198-749-5
© 2013 D. Guo, S. Shamai, S. Verdú

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Foundations and Trends® in Signal Processing

Volume 6, Issue 4, 2012

Editorial Board

Editor-in-Chief

Editors

Editorial Scope

Topics

Foundations and Trends® in Signal Processing publishes survey and tutorial articles in the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital and multirate signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing

Information for Librarians

Foundations and Trends® in Signal Processing, 2012, Volume 6, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Foundations and Trends® in Signal Processing
Vol. 6, No. 4 (2012) 243–429

DOI: 10.1561/2000000018

The Interplay Between Information and Estimation Measures

Dongning Guo
Northwestern University
Evanston, IL 60208, U.S.A.

Shlomo Shamai (Shitz)
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel

Sergio Verdú
Princeton University
Princeton, NJ 08544, U.S.A.

Contents

Abstract

This monograph surveys the interactions between information measures and estimation measures as well as their applications. The emphasis is on formulas that express the major information measures, such as entropy, mutual information and relative entropy in terms of the minimum mean square error achievable when estimating random variables contaminated by Gaussian noise. These relationships lead to wide applications ranging from a universal relationship in continuous-time nonlinear filtering to optimal power allocation in communication systems, to the simplified proofs of important results in information theory such as the entropy power inequality and converses in multiuser information theory.

D. Guo, S. Shamai, S. Verdú. *The Interplay Between Information and Estimation Measures*. Foundations and Trends® in Signal Processing, vol. 6, no. 4, pp. 243–429, 2012.

DOI: 10.1561/2000000018.

1

Introduction

If *information theory* and *estimation theory* are thought of as two scientific languages, then their key vocabularies are information measures and estimation measures, respectively. The basic information measures are *entropy*, *mutual information* and *relative entropy*. Among the most important estimation measures are *mean square error (MSE)* and *Fisher information*. Playing a paramount role in information theory and estimation theory, those measures are akin to mass, force and velocity in classical mechanics, or energy, entropy and temperature in thermodynamics.

The theory about how to and how well one can estimate an unknown quantity based on related observations has taken shape over the course of several centuries with contributions from towering figures such as Bayes, Legendre, Gauss, Laplace, Bernoulli, and Fisher. A rigorous foundation of the theory of probability and random processes was built by Kolmogorov in the 1930s, which allowed parametric estimation theory to assume full generality on a sound mathematical footing. The estimation of analog signals in additive noise (see, e.g., Kotelnikov [98]) is a key step in statistics, signal processing, communications, control, economics and financial engineering. Shortly after the solution by Wiener and Kolmogorov of linear minimum mean square estimators of wide-sense stationary random processes, 1948 saw the inception of Shannon's information theory with "A Mathematical Theory of Communication" [155]. Among several problems, Shannon was concerned with optimal ways of combatting noise, not only by estimation techniques at the receiver but by the design of the information-carrying signals sent by the transmitter. Shannon's theorems for lossless compression, transmission via noisy channels [155] and lossy source coding [157] characterize the best achievable rate in the limit of infinite dimensions. As a branch of information theory, coding theory studies how to approach the fundamental limits using practical algorithms, many of which involve optimal or near-optimal signal estimation.

Under several important probabilistic models, some surprisingly simple formulas have been found to connect basic information measures and estimation measures. These relations are the subject of this monograph. An early example, dating back to 1959, is de Bruijn's identity, which states that the derivative (with respect to the noise standard deviation) of the differential entropy of a Gaussian noise contaminated random variable is equal to the Fisher information of the variable. In 2004, the authors of this monograph found a general differential relationship commonly referred to as the I–MMSE formula. Let X be an arbitrary scalar real-valued random variable and N be a standard Gaussian (aka standard normal) random variable independent of X. The I–MMSE formula states that, for every $s > 0$, the input–output mutual information of a scalar Gaussian model with signal-to-noise ratio (SNR) gain s can be expressed as:[1]

$$I(X; \sqrt{s}\, X + N) = \frac{1}{2} \int_0^s \mathsf{mmse}\left(X | \sqrt{\gamma}\, X + N\right) \mathrm{d}\gamma \tag{1.1}$$

where $\mathsf{mmse}\left(X | \sqrt{\gamma}\, X + N\right)$ stands for the minimum mean square error (MMSE) of estimating X using the noisy observation $\sqrt{\gamma}\, X + N$. As a consequence of (1.1), if X takes its values in a discrete real number

---

[1] In general, the unit of the information measures depends on the base of the logarithm with which they are defined—called bits if the base is 2, or nats if the base is e. For convenience, we adopt natural logarithms and use nats as the unit of all information measures throughout this monograph.

set, then

$$H(X) = \frac{1}{2}\int_0^\infty \mathsf{mmse}\left(X|\sqrt{\gamma}\,X+N\right)\mathrm{d}\gamma. \tag{1.2}$$

Such formulas have proven useful for establishing important results, such as the entropy power inequality and capacity bounds for certain multiuser channels.

The simplicity of (1.1), (1.2) and many other related formulas to be introduced in later chapters is remarkable and intriguing. The link between estimation theory and information theory that underlies such results is the likelihood function. In particular, optimal estimation often involves a test or optimization of certain likelihoods or likelihood ratios. Entropy, mutual information and relative entropy are expressed as the expectations of certain log-likelihoods or log-likelihood ratios. The information measures find operational meaning as the answers to problems such as the maximum number of distinguishable messages through a communication system or the minimum number of codewords required for encoding a source within a certain distortion, assuming essentially optimal estimation or reconstruction.

The goals of this monograph are to provide a handbook of known formulas which directly relate information measures and estimation measures, to provide intuition and draw connections between these formulas, to highlight some important applications, and to motivate further explorations. Our main focus is on such formulas in the context of the additive Gaussian noise model, with lesser treatment of others such as the Poisson point process channel. Also included are a number of new results which are published here for the first time. We provide proofs of some basic results, whereas many more technical proofs already available in the literature are omitted. A new, complete proof for the I–MMSE formula (1.1) is developed, which includes some technical details omitted in the original papers. It is our hope that collecting many related results and elucidating their relationships will facilitate in-depth understanding and further applications.

References

[1] R. Ahlswede. Multi-way communication channels. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 23–52, 1971.

[2] R. Ahlswede. The final form of Tao's inequality relating conditional expectation and conditional mutual information. *Advances in Mathematics of Communications*, 1(2):239, 2007.

[3] M. S. Ali and D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Ser. B*, 28(1):131–142, 1966.

[4] A. Alvarado, F. Brännström, E. Agrell, and T. Koch. High-SNR asymptotics of mutual information for discrete constellations. In *Proc. IEEE Int. Symp. Inform. Theory.* Istanbul, Turkey, 2013.

[5] A. Ashikhmin, G. Kramer, and S. ten Brink. Extrinsic information transfer functions: Model and erasure channel properties. *IEEE Trans. Inform. Theory*, 50(11):2657–2673, Nov. 2004.

[6] R. Atar and T. Weissman. Mutual information, relative entropy, and estimation in the Poisson channel. *IEEE Trans. Inform. Theory*, 58(3):1302–1318, Mar. 2012.

[7] V. Balakirsky. A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels. *IEEE Trans. Inform. Theory*, 41(6):1889–1902, November 1995.

[8] B. Bandemer and A. El-Gamal. Communication with disturbance constraints. In *Proc. IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, 2011.

[9] P. Banelli. Non-linear transformation of Gaussians and Gaussian-mixtures with implications on estimation and information theory. *arXiv:1111.5950v3*, 2013.

[10] A. Banerjee, X. Guo, and H. Wang. On the optimality of conditional expectation as a Bregman predictor. *IEEE Trans. Inform. Theory*, 51:2664–2669, July 2005.

[11] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh. Clustering with Bregman Divergences. *J. Mach. Learn. Res.*, 6:1705–1749, Dec. 2005.

[12] I. Bar-David and S. Shamai. On information transfer by envelope constrained signals over the AWGN channel. *IEEE Trans. Inform. Theory*, 34(3):371–379, May 1988.

[13] A. R. Barron. Entropy and the central limit theorem. *Annals of Probability*, 14(1):336–342, 1986.

[14] T. Bayes and R. Price. An essay towards solving a problem in the doctrine of chances. (By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS). *Philosophical Transactions of the Royal Society of London (1683-1775)*, pages 370–418, 1763.

[15] A. Bennatan, S. Shamai, and A. R. Calderbank. Soft-decoding-based strategies for relay and interference channels: Analysis and achievable rates using LDPC codes. *submitted to IEEE Trans. Inform. Theory*, 2013. arXiv:1008.1766v3.

[16] P. P. Bergmans. A simple converse for broadcast channels with additive white Gaussian noise. *IEEE Trans. Inform. Theory*, 20:279–280, Mar. 1974.

[17] K. Bhattad and K. Narayanan. An MSE-based transfer chart for analyzing iterative decoding schemes using a Gaussian approximation. *IEEE Trans. Inform. Theory*, 53(1):22–38, 2007.

[18] J. Binia. Divergence and minimum mean-square error in continuous-time additive white Gaussian noise channels. *IEEE Trans. Inform. Theory*, 52(3):1160–1163, Mar. 2006.

[19] R. E. Blahut. *Principles and Practice of Information Theory*. Addison-Wesley, 1987.

[20] G. Blower. *Random Matrices: High Dimensional Phenomena*. London Mathematical Society Lecture Notes. Cambridge University Press, Cambridge, U.K., 2009.

[21] B. Z. Bobrovsky and M. Zakai. A lower bound on the estimation error for certaindiffusion processes. *IEEE Trans. Inform. Theory*, 22(1):45–52, Jan. 1976.

[22] B. Z. Bobrovsky and M. Zakai. Asymptotic a priori estimates for the error in thenonlinear filtering problem. *IEEE Trans. Inform. Theory*, 28(2):371–376, Mar. 1982.

[23] R. K. Boel and V. E. Benes. Recursive nonlinear estimation of a diffusion acting as therate of an observed Poisson process. *IEEE Trans. Inform. Theory*, 26(5):561–575, Sept. 1980.

[24] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford Univerisity Press, 2013.

[25] L. M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics*, 7:200–217, 1967.

[26] P. Brémaud. *Point Processes and Queues: Martingale Dynamics*. Springer-Verlag, New York, 1981.

[27] L. D. Brown. Admissible estimators, recurrent diffusions, and insoluble boundary value problems. *The Annals of Mathematical Statistics*, pages 855–903, 1971.

[28] V. V. Buldygin and Y. V. Kozachenko. Sub-Gaussian random variables. *Ukrainian Mathematical Journal*, 32:483–489, 1980.

[29] R. Bustin, R. Liu, H. V. Poor, and S. Shamai. An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel. *EURASIP Journal on Wireless Communications and Networking, Special Issue on Wireless Physical Security*, 2009, Nov. 2009.

[30] R. Bustin, M. Payaró, D. P. Palomar, and S. Shamai. On MMSE crossing properties and implications in parallel vector Gaussian channels. *IEEE Trans. Inform. Theory*, 59(2):818–844, Feb. 2013.

[31] R. Bustin and S. Shamai. The I-MMSE approach on the weak Gaussian Z-interference channel and the type I Gaussian broadcast-Z-interference channels. In *IEEE 26th Convention of Electrical and Electronic Engineers in Israel*, Israel, November 17–20 2010.

[32] R. Bustin and S. Shamai. On Gaussian channels with MMSE interference. In *International Zurich Seminar on Communications*, Zurich, Switzerland, 2012.

[33] R. Bustin and S. Shamai. MMSE of “bad” codes. *IEEE Trans. Inform. Theory*, 59(2):733–743, 2013.

[34] G. Caire and K. Kumar. Information theoretic foundations of adaptive coded modulation. *Proc. IEEE*, 95(12):2274–2298, Dec. 2007.

[35] W. R. Carson, M. Chen, M. R. Rodrigues, R. Calderbank, and L. Carin. Communications-inspired projection design with application to compressive sensing. *SIAM Journal on Imaging Sciences*, 5(4):1185–1212, 2012.

[36] G. Chechik, A. Globerson, N. Tishby, and Y. Weiss. Information bottleneck for Gaussian variables. *J. Machine Learning Research*, 6:165–188, 2005.

[37] M. Chen and J. Lafferty. Mishmatched estimation and relative entropy in vector Gaussian channels. In *Proc. IEEE Int. Symp. Inform. Theory.* Istanbul, Turkey, 2013.

[38] Y.-K. Chia, R. Soundararajan, and T. Weissman. Estimation with a helper who knows the interference. *IEEE Trans. Inform. Theory*, 59(11):7097–7117, 2013.

[39] S. S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi. Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design. *IEEE Trans. Wireless Commun.*, 7(12):4792–4799, December 2008.

[40] A. Cohen, T. Weissman, and N. Merhav. Scanning and sequential decision making for multidimensional data—Part II: The noisy case. *IEEE Trans. Inform. Theory*, 54(12):5609–5631, 2008.

[41] M. H. M. Costa. A new entropy power inequality. *IEEE Trans. Inform. Theory*, 31:751–760, Nov. 1985.

[42] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, 2nd edition, 2006.

[43] I. Csiszár. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19:2032–2066, 1991.

[44] I. Csiszár. Axiomatic characterizations of information measures. *Entropy*, 10:261–273, 2008.

[45] I. Csiszár and J. Körner. Graph decomposition: A new key to coding theorems. *IEEE Trans. Inform. Theory*, IT–27(1):5–12, January 1981.

[46] I. Csiszár and P. Narayan. Channel capacity for a given decoding metric. *IEEE Trans. Inform. Theory*, 41(1):35–43, January 1995.

[47] M. H. A. Davis. Capacity and cutoff rate for Poisson-type channels. *IEEE Trans. Inform. Theory*, 26(6):710–715, Nov. 1980.

[48] A. Del Coso and S. Simoens. Distributed compression for MIMO coordinated networks with a backhaul constraint. *IEEE Trans. Wireless Commun.*, 8(9):4698–4709, 2009.

[49] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Trans. Inform. Theory*, 37(6):1501–1518, Nov. 1991.

[50] R. L. Dobrushin. Passage to the limit under the information and entropy signs. *Theory Probab. Appl.*, 5:25–32, 1960.

[51] T. E. Duncan. On the calculation of mutual information. *SIAM Journal of Applied Mathematics*, 19:215–220, July 1970.

[52] T. E. Duncan. Mutual information for stochastic differential equations. *Information and Control*, 19:265–271, 1971.

[53] T. E. Duncan. Mutual information for stochastic signals and fractional Brownian motion. *IEEE Trans. Inform. Theory*, 54(10):4432–4438, October 2008.

[54] T. E. Duncan. Mutual information for stochastic signals and Levy processes. *IEEE Trans. Inform. Theory*, 56(1):18–24, January 2010.

[55] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai. Noncoherent capacity of underspread fading channels. *IEEE Trans. Inform. Theory*, 56(1):367–395, January 2010.

[56] E. Ekrem and S. Ulukus. The secrecy capacity region of the Gaussian MIMO multi-receiver wiretap channel. *IEEE Trans. Inform. Theory*, 57(4):2083–2114, 2011.

[57] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, 2011.

[58] Y. Ephraim and B. L. Mark. *Bivariate Markov Processes and Their Estimation*, volume 6 (1) of *Foundations and Trends in Signal Processing*. now Publishers, 2013.

[59] Y. Ephraim and N. Merhav. Lower and upper bounds on the minimum mean-square error in composite source signal estimation. *IEEE Trans. Inform. Theory*, 38(6):1709–1724, 1992.

[60] R. Esposito. On a relation between detection and estimation indecision theory. *Information and Control*, 12:116–120, 1968.

[61] J. Evans and V. Krishnamurthy. Exact filters for doubly stochastic AR models with conditionally Poisson observations. *IEEE Trans. Inform. Theory*, 44(4):794–798, Apr. 1999.

[62] G. L. Fillmore and G. Lachs. Information rates for photocount detection systems. *IEEE Trans. Inform. Theory*, 15(4):465–468, July 1969.

[63] R. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society*, 222-A:309–368, 1922.

[64] G. D. Forney, Jr. Shannon meets Wiener II: On MMSE estimation in successive decoding schemes. In *Proc. Allerton Conf. Commun., Control, & Computing*. Monticello, IL, USA, 2004.

[65] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968.

[66] R. G. Gallager. Capacity and coding for degraded broadcast channels. *Problemy Peredachi Informatsii*, 10(3):3014, July-Sept. 1974. Translated: Probl. Inform. Transmission., pp. 185-193, July-Sept. 1974.

[67] A. Ganti, A. Lapidoth, and E. Telatar. Mishmatched decoding revisited: general alphabets, channels with memory, and the wide-band limit. *IEEE Trans. Inform. Theory*, 10(7):2315–2329, November 2000.

[68] K. F. Gauss. *Teoria Motus Corporum Coelestium in Sectionibus Conicus Solem Ambientieum*. Dover, 1809.

[69] Y. Geng and C. Nair. The capacity region of the two-receiver vector Gaussian broadcast channel with private and common messages. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 586–590, 2012.

[70] C. Gil Taborda, F. Pérez-Cruz, and D. Guo. Information–estimation relationships over the binomial and negative binomial models. *submitted to IEEE Trans. Inform. Theory*, 2013.

[71] I. V. Girsanov. On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory Probab. Appl.*, 5(3):285–301, 1960.

[72] J. Grandell. *Mixed Poisson Processes*. Chapman & Hall, 2nd edition, 1997.

[73] R. M. Gray. *Entropy and information theory*. Springer-Verlag, 1990.

[74] T. Guess and M. K. Varanasi. An information-theoretic framework for deriving canonical decision-feedback receivers in Gaussian channels. *IEEE Trans. Inform. Theory*, 51(1):173–187, Jan. 2005.

[75] A. Guillén i Fàbregas, A. Martinez, and G. Caire. *Bit-interleaved coded modulation*, volume 5 (1–2) of *Foundations and Trends in Communications and Information Theory*. now Publishers, 2008.

[76] D. Guo. *Gaussian Channels: Information, Estimation and Multiuser Detection*. PhD thesis, Department of Electrical Engineering, Princeton University, 2004.

[77] D. Guo. Relative entropy and score function: New information-estimation relationships through arbitrary additive perturbation. In *Proc. IEEE Int. Symp. Inform. Theory*. Seoul, Korea, 2009.

[78] D. Guo, S. Shamai, and S. Verdú. Additive non-Gaussian noise channels: Mutual information and conditional mean estimation. In *Proc. IEEE Int. Symp. Inform. Theory*. Adelaide, Australia, Sept. 2005.

[79] D. Guo, S. Shamai, and S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory*, 51(4):1261–1282, Apr. 2005.

[80] D. Guo, S. Shamai (Shitz), and S. Verdú. Mutual information and conditional mean estimation in Poisson channels. *IEEE Trans. Inform. Theory*, 54(5):1837–1849, May 2008.

[81] D. Guo and T. Tanaka. Generic multiuser detection and statistical physics. In M. Honig, editor, *Advances in Multiuser Detection*, chapter 5. Wiley-IEEE Press, 2009.

[82] D. Guo and S. Verdú. Randomly spread CDMA: Asymptotics via statistical physics. *IEEE Trans. Inform. Theory*, 51(6):1982–2010, June 2005.

[83] D. Guo, Y. Wu, S. Shamai, and S. Verdú. Estimation in Gaussian noise: Properties of the minimum mean-square error. *IEEE Trans. Inform. Theory*, 57(4):2371–2385, Apr. 2011.

[84] T. S. Han. *Information-Spectrum Methods in Information Theory*. Springer, 2003.

[85] R. V. L. Hartley. Transmission of information. *Bell System Technical Journal*, pages 535–563, 1928.

[86] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, 3rd edition, 1996.

[87] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[88] J. Y. N. Hui. *Fundamental Issues of Multiple Accessing*. Ph.D. dissertation, Mass. Inst. Technol., Cambridge, MA, USA, 1983.

[89] W. Huleihel and N. Merhav. Analysis of mismatched estimation errors using gradients of partition functions. *arXiv:1306.0094*, 2013.

[90] I. A. Ibragimov and R. Z. Hasminskii. *Statistical estimation, asymptotic theory*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York, 1981.

[91] J. Jiao, K. Venkat, and T. Weissman. Pointwise relations between information and estimation in the Poisson channel. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 449–453. Istanbul, Turkey, 2013.

[92] Y. M. Kabanov. The capacity of a channel of the Poisson type. *Theory of Probability and Its Applications*, 23:143–147, 1978.

[93] T. T. Kadota, M. Zakai, and J. Ziv. Mutual information of the white Gaussian channel with and without feedback. *IEEE Trans. Inform. Theory*, 17(4):368–371, July 1971.

[94] V. Karunakaran. *Complex Analysis*. Alpha Science International Ltd., 2nd edition, 2005.

[95] A. Khisti and G. Wornell. Secure transmission with multiple antennas - Part I: The MISOME wiretap channel. *IEEE Trans. Inform. Theory*, 55(7):3088–3104, July 2009.

[96] A. Khisti and G. Wornell. Secure transmission with multiple antennas - Part II: The MIMOME wiretap channel. *IEEE Trans. Inform. Theory*, 56(11):5515–5532, November 2010.

[97] A. N. Kolmogorov. On the Shannon theory of information transmission in the case of continuous signals. *IEEE Trans. Inform. Theory*, 2:102–108, Sept. 1956.

[98] V. Kotelnikov. *The Theory of Potential Noise Immunity*. Moscow, 1956.

[99] G. Kramer. Topics in multi-user information theory. *Foundations and Trends in Communications and Information Theory*, 4 (4–5)(now Publishers), 2008.

[100] S. G. Krantz and H. R. Parks. *A Primer of Real Analytic Functions*. Birkhäuser, Boston, 2nd edition, 2002.

[101] V. Krishnamurthy and R. J. Elliott. Filters for estimating Markov modulated Poisson processes and image-based tracking. *Automatica*, 33(5):821–833, 1997.

[102] S. Kullback. *Information Theory and Statistics*. Dover, New York, 1968.

[103] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.

[104] A. Lapidoth and S. Shamai. The Poisson multiple-access channel. *IEEE Trans. Inform. Theory*, 44(2):488–501, Mar. 1998.

[105] A. Lapidoth and S. Shamai. Fading channels: How perfect need ‘perfect side information’ be? *IEEE Trans. Inform. Theory*, 48(5):1118–1134, May 2002.

[106] P. S. Laplace. *Essai philosophique sur les probabilités.* Courcier, 1814.

[107] A. M. Legendre. Nouvelles méthodes pour la détermination des orbites des comètes. *Paris: Courcier, 1806; VIII, 55 p.; in 4.; DCCC. f. 10*, 1, 1806.

[108] N. Letzepis and A. Guillén i Fàbregas. Outage probability of the Gaussian MIMO free-space optical channel with PPM. *IEEE Trans. Commun.*, 57(12):3682–3690, 2009.

[109] S. K. Leung-Yan-Cheong and M. E. Hellman. The Gaussian wire-tap channel. *IEEE Trans. Inform. Theory*, IT-24:451–456, 1978.

[110] E. H. Lieb. Proof of an entropy conjecture of Wehrl. *Commun. Math. Phys.*, 62(1):35–41, 1978.

[111] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes I: General Theory.* Springer, 2nd edition, 2001.

[112] R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes II: Applications.* Springer, 2nd edition, 2001.

[113] J. Liu, J.-K. Zhang, and K. M. Wong. On the design of minimum BER linear space-time block codes for MIMO systems equipped with MMSE receivers. *IEEE Trans. Signal Processing*, 54(8):3147–3158, 2006.

[114] R. Liu, T. Liu, H. Poor, and S. Shamai. A vector generalization of costa entropy-power inequality and applications. *IEEE Trans. Inform. Theory*, 56(4):1865–1879, April 2010.

[115] R. Liu, T. Liu, V. H. Poor, and S. Shamai. MIMO Gaussian broadcast channels with common messages. In *Information Theory and Applications Workshop.* La Jolla, CA, USA, 2010.

[116] T. Liu and S. Shamai. A note on the secrecy capacity of the multiple-antenna wiretap channel. *IEEE Trans. Inform. Theory*, 55(6):2547–2553, June 2009.

[117] A. Lozano, A. M. Tulino, and S. Verdú. Optimum power allocation for parallel Gaussian channels with arbitrary input distributions. *IEEE Trans. Inform. Theory*, 52(7):3033–3051, July 2006.

[118] J. D. Markel and A. H. Gray. *Linear Prediction of Speech*, volume 3. Springer-Verlag, 1976.

[119] N. C. Martins, M. A. Dahleh, and J. C. Doyle. Fundamental limitations of disturbance attenuation in the presence of side information. *IEEE Trans. Automat. Contr.*, 52(1):56–66, 2007.

[120] E. Mayer-Wolf and M. Zakai. Some relations between mutual information and estimation error in Wiener space. *Annals of Applied Probability*, 17:1102–1116, June 2007.

[121] C. Méasson, A. Montanari, T. Richardson, and R. Urbanke. The generalized area theorem and some of its consequences. *IEEE Trans. Inform. Theory*, 55(11):4793–4821, 2009.

[122] N. Merhav. *Statistical Physics and Information Theory*, volume 6 (1–2) of *Foundations and Trends in Communications and Information Theory.* now Publishers, 2009.

[123] N. Merhav, D. Guo, and S. Shamai. Statistical physics of signal estimation in Gaussian noise: Theory and examples of phase transition. *IEEE Trans. Inform. Theory*, 56:1400–1416, Mar. 2010.

[124] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai. On information rates for mismatched decoders. *IEEE Trans. Inform. Theory*, 40(6):1953–1967, November 1994.

[125] R. Narasimhan. Finite-SNR diversity-multiplexing tradeoff for correlated Rayleigh and rician MIMO channels. *IEEE Trans. Inform. Theory*, 52(9):3965–3979, 2006.

[126] K. R. Narayanan and A. R. Srinivasa. On the thermodynamic temperature of a general distribution. *arXiv:0711.1460v2*, 2007.

[127] A. No and T. Weissman. Minimax filtering regret via relations between information and estimation. *submitted to IEEE Trans. Inform. Theory*, 2013. arXiv:1301.5096v3.

[128] F. Oggier and B. Hassibi. The secrecy capacity of the MIMO wiretap channel. *IEEE Trans. Inform. Theory*, 57:4961–4972, Aug. 2011.

[129] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications.* Springer-Verlag Berlin Heidelberg, 6th edition, 2003.

[130] D. P. Palomar and S. Verdú. Gradient of mutual information in linear vector Gaussian channels. *IEEE Trans. Inform. Theory*, 52(1):141–154, Jan. 2006.

[131] D. P. Palomar and S. Verdú. Lautum information. *IEEE Trans. Inform. Theory*, 54(3):964–975, Mar. 2008.

[132] S. Park, E. Serpedin, and K. Qaraqe. On the equivalence between Stein and de Bruijn identities. *arXiv:1202.0015*, 2012.

[133] M. Payaró and D. P. Palomar. Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels. *IEEE Trans. Inform. Theory*, 55:3613–3628, 2009.

[134] M. Peleg, A. Sanderovich, and S. Shamai. On extrinsic information of good binary codes operating on Gaussian channels. *European Trans. Telecommun.*, 18:133–139, 2007.

[135] F. Perez-Cruz, M. R. D. Rodrigues, and S. Verdú. MIMO Gaussian channels with arbitrary inputs: Optimal precoding and power allocation. *IEEE Trans. Inform. Theory*, 56(3):1070–1084, march 2010.

[136] L. Pérez-Freire, P. Comesana, J. R. Troncoso-Pastoriza, and F. Perez-Gonzalez. Watermarking security: a survey. In *Transactions on Data Hiding and Multimedia Security I*, pages 41–72. Springer, 2006.

[137] M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.

[138] M. S. Pinsker, V. V. Prelov, and S. Verdú. Sensitivity of channel capacity. *IEEE Trans. Inform. Theory*, 41(6):1877–1888, Nov. 1995.

[139] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inform. Theory*, 56(5):2307–2359, May 2010.

[140] H. V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1994.

[141] V. Prelov and S. Verdú. Second-order asymptotics of mutual information. *IEEE Trans. Inform. Theory*, 50(8):1567–1580, Aug. 2004.

[142] V. V. Prelov. Communication channel capacity with almost Gaussian noise. *Theory Probab. Appl.*, 33:405–422, Apr. 1987.

[143] V. V. Prelov and E. C. van der Meulen. An asymptotic expression for the information and capacity of a multidimensional channel with weak input signals. *IEEE Trans. Inform. Theory*, 39:1728–1735, Sept. 1993.

[144] M. Raginsky and I. Sason. *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*, volume 10 (1–2) of *Foundations and Trends in Communications and Information Theory*. now Publishers, 2013.

[145] C. Rao and B. Hassibi. Analysis of multiple-antenna wireless links at low snr. *IEEE Trans. Inform. Theory*, 50(9):2123–2130, Sept. 2004.

[146] O. Rioul. Information theoretic proofs of entropy power inequalities. *IEEE Trans. Inform. Theory*, 57(1):33–55, January 2011.

[147] M. R. D. Rodrigues. Multiple-antenna fading channels with arbitrary inputs: characterization and optimization of the information rate. *IEEE Trans. Inform. Theory*, 2013. to appear.

[148] A. Sanderovich, O. Somekh, H. Poor, and S. Shamai. Uplink macro diversity of limited backhaul cellular network. *IEEE Trans. Inform. Theory*, 55(8):3457–3478, Aug. 2009.

[149] A. Segall. Recursive estimation from discrete-time point processes. *IEEE Trans. Inform. Theory*, 22(4):422–431, July 1976.

[150] S. Shamai. Capacity of a pulse amplitude modulated direct detection photon channel. *IEE Proceedings*, 137(6):424–430, Dec. 1990.

[151] S. Shamai. From constrained signaling to network interference alignment via an information-estimation perspective. *IEEE Information Theory Society Newsletter*, 62(7):6–24, 2012.

[152] S. Shamai and A. Lapidoth. Bounds on the capacity of a spectrally constrained Poisson channel. *IEEE Trans. Inform. Theory*, 39(1):19–29, Jan. 1993.

[153] S. Shamai and I. Sason. Variations on the Gallager bounds, connections and applications. *IEEE Trans. Inform. Theory*, 48(12):3029–3051, December 2002.

[154] S. Shamai and A. Steiner. A broadcast approach for a single-user slowly fading mimo channel. *IEEE Trans. Inform. Theory*, 49(10):2617–2635, October 2003.

[155] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and Oct. 1948.

[156] C. E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37:10–21, 1949.

[157] C. E. Shannon. Coding theorems for a discrete source with a fidelity criterion. *IRE National Convention Record, Part 4*, pages 142–163, 1959.

[158] O. Shental, N. Shental, S. Shamai, I. Kanter, A. J. Weiss, and Y. Weiss. Discrete-input two-dimensional gaussian channels with memory: Estimation and information via graphical models and statistical mechaniscs. *IEEE Trans. Inform. Theory*, 54(4):1500–1513, April 2008.

[159] S. Shi, M. Schubert, and H. Boche. Downlink MMSE transceiver optimization for multiuser MIMO systems: Duality and sum-MSE minimization. *IEEE Trans. Signal Processing*, 55(11):5436–5446, 2007.

[160] S. Shi, M. Schubert, and H. Boche. Downlink MMSE transceiver optimization for multiuser MIMO systems: MMSE balancing. *IEEE Trans. Signal Processing*, 56(8):3702–3712, 2008.

[161] N. J. A. Sloane and A. D. Wyner, editors. *Claude Elwood Shannon: Collected Papers*. IEEE Press: NY, 1993.

[162] D. L. Snyder. Filtering and detection for doubly stochastic Poisson processes. *IEEE Trans. Inform. Theory*, 18:91–102, Jan. 1972.

[163] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2:101–112, 1959.

[164] Y. Steinberg, B. Z. Bobrovsky, and Z. Schuss. Fixed-point smoothing of scalar diffusions II: The error of the optimal smoother. *SIAM J. Appl. Math*, 61:1431–1444, 2001.

[165] I. Sutskover, S. Shamai, and J. Ziv. Extremes of information combining. *IEEE Trans. Inform. Theory*, 51(4):1313–1325, 2005.

[166] T. Tao. Szeméredi's regularity lemma revisited. *Contrib. Discr. Math.*, 1(1):8–28, 2006.

[167] C. Tian and J. Chen. Remote vector Gaussian source coding with decoder side information under mutual information and distortion constraints. *IEEE Trans. Inform. Theory*, 55(10):4676–4680, 2009.

[168] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52:479–487, 1988.

[169] D. N. C. Tse. It's easier to approximate. *Information Theory Society Newsletter*, 60(1):6–11, March 2010.

[170] D. N. C. Tse and R. D. Yates. Fading broadcast channels with state information at the receivers. *IEEE Trans. Inform. Theory*, 58:3453–3471, 2012.

[171] A. M. Tulino, A. Lozano, and S. Verdú. Impact of antenna correlation on the capacity of multiantenna channels. *IEEE Trans. Inform. Theory*, 51(7):2491–2509, 2005.

[172] A. M. Tulino and S. Verdú. Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof. *IEEE Trans. Inform. Theory*, 52:4295–4297, Sept. 2006.

[173] D. Tuninetti, S. Shamai, and G. Caire. Scalar fading Gaussian broadcast channels with perfect receiver CSI: Is Gaussian input optimal? In *Proc. Workshop Inform. Theory Applications*. La Jolla, CA, USA, Jan. 2007.

[174] J. H. Van Schuppen. Filtering, prediction and smoothing for counting process observations, a martingale approach. *SIAM J. Appl. Math.*, 32(3):552–570, May 1977.

[175] M. K. Varanasi and T. Guess. Optimum decision feedback multiuser equalization with successive decoding achieves the total capacity of the Gaussian multiple-access channel. In *Proc. Asilomar Conf. on Signals, Systems and Computers*, pages 1405–1409. Monterey, CA, USA, Nov. 1997.

[176] K. Venkat and T. Weissman. Pointwise relations between information and estimation in Gaussian noise. *IEEE Trans. Inform. Theory*, 58:6264–6281, Oct. 2012.

[177] K. Venkat, T. Weissman, Y. Carmon, and S. Shamai. On information, estimation, and lookahead. In *Proc. Allerton Conf. Commun., Control, & Computing*, 2012.

[178] S. Verdú. Capacity region of Gaussian CDMA channels: The symbol-synchronous case. In *Proc. Allerton Conf. Commun., Control, & Computing*, pages 1025–1034, Oct. 1986.

[179] S. Verdú. On channel capacity per unit cost. *IEEE Trans. Inform. Theory*, 36(5):1019–1030, Sept. 1990.

[180] S. Verdú. The exponential distribution in information theory. *Problems of Information Transmission*, 32:86–95, 1996. Translated from Problemy Peredachi Informatsii, Vol. 32,No. 1, pp. 100–111, January–March, 1996.

[181] S. Verdú. Poisson communication theory. *Invited talk in the International TechnionCommunication Day in honor of Israel Bar-David*, Mar. 1999. [Online] Available at: http://www.princeton.edu/∼verdu.

[182] S. Verdú. Spectral efficiency in the wideband regime. *IEEE Trans. Inform. Theory*, 48(6):1319–1343, June 2002.

[183] S. Verdú. Mismatched estimation and relative entropy. *IEEE Trans. Inform. Theory*, 56(8):3712–3720, Aug. 2010.

[184] S. Verdú and D. Guo. A simple proof of the entropy power inequality. *IEEE Trans. Inform. Theory*, pages 2165–2166, May 2006.

[185] S. Verdú and T. S. Han. A general formula for channel capacity. *IEEE Trans. Inform. Theory*, 40(4):1147–1157, July 1994.

[186] S. Verdú and S. Shamai. Spectral efficiency of CDMA with random spreading. *IEEE Trans. Inform. Theory*, 45(2):622–640, Mar. 1999.

[187] S. Verdú and S. Shamai. Variable-rate channel capacity. *IEEE Trans. Inform. Theory*, 56(6):2651–2667, June 2010.

[188] C. Villani. A short proof of the ‘concavity of entropy power’. *IEEE Trans. Inform. Theory*, 46(4):1695–1696, July 2000.

[189] L. Wang, M. Rodrigues, and L. Carin. Generalized Bregman divergence and gradient of mutual information for vector Poisson channels. *arXiv:1301.6648v3*, 2013.

[190] S. Watanabe and Y. Oohama. Broadcast channels with confidential messages by randomness constrained stochastic encoder. *arXiv:1201.6468v2*, 2012.

[191] H. Weingarten, Y. Steinberg, and S. Shamai. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Inform. Theory*, 52:3936–3964, Jun./Jul. 2006.

[192] H. Weingarten, Y. Steinberg, and S. Shamai. On the capacity region of the multi-antenna broadcast channel with common messages. In *Proc. IEEE Int. Symp. Inform. Theory*, Seattle, Washington, July 2006.

[193] T. Weissman. The relationship between causal and noncausal mismatched estimation in continuous-time AWGN channels. *IEEE Trans. Inform. Theory*, 56(9):4256–4273, Sept. 2010.

[194] T. Weissman, Y.-H. Kim, and H. H. Permuter. Directed information, causal estimation, and communication in continuous time. *IEEE Trans. Inform. Theory*, 59(3):1271–1287, 2013.

[195] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary-Time Series, with Engineering Applications.* Wiley, 1949. (Originally published in Feb. 1942 as a classified Nat. Defense Res. Council rep.).

[196] W. M. Wonham. Some applications of stochastic differential equationsto optimal nonlinear filtering. *J. SIAM Control, Ser. A*, 2(3):347–369, 1965.

[197] Y. Wu, D. Guo, and S. Verdú. Derivative of mutual information at zero SNR: the Gaussian-noise case. *IEEE Trans. Inform. Theory*, 57(11):7307–7312, Nov. 2011.

[198] Y. Wu, S. Shamai, and S. Verdú. A formula for the degrees of freedom of the interference channel. *Submitted to IEEE Trans. Inform. Theory*, 2011.

[199] Y. Wu and S. Verdú. MMSE dimension. *IEEE Trans. Inform. Theory*, 57(8):4857–4879, Aug. 2011.

[200] Y. Wu and S. Verdú. Functional properties of minimum mean-square error and mutual information. *IEEE Trans. Inform. Theory*, 58(3):1289–1301, Mar. 2012.

[201] A. D. Wyner. The wire-tap channel. *The Bell System Technical Journal*, 54:1355–1387, Oct. 1975.

[202] T. Yamano. de Bruijn-type identity for systems with flux. *The European Physical Journal B*, 86(8):1–6, 2013.

[203] Y. Yang and R. S. Blum. MIMO radar waveform design based on mutual information and minimum mean-square error estimation. *IEEE Trans. Aerosp. Electron. Syst.*, 43(1):330–343, 2007.

[204] Y.-C. Yao. Estimation of noisy telegraph processes: Nonlinear filtering versus nonlinear smoothing. *IEEE Trans. Inform. Theory*, 31(3):444–446, May 1985.

[205] M. C. Yovits and J. L. Jackson. Linear filter optimization with game theory considerations. *Proc. IRE*, 43(3):376, 1955.

[206] M. Zakai. On mutual information, likelihood ratios, and estimation error for the additive Gaussian channel. *IEEE Trans. Inform. Theory*, 51(9):3017–3024, Sept. 2005.

[207] R. Zamir and M. Feder. A generalization of information theoretic inequalities to linear transformations of independent vector. In *Proceedings of the 6-th Joint Swedish-Russian International Workshop on Information Theory*, pages 254–258. Möle, Sweden, Aug. 1993.

[208] R. Zamir and M. Feder. A generalization of the entropy power inequality with applications. *IEEE Trans. Inform. Theory*, 39(5):1723–1728, Sept. 1993.

[209] Y. Zhu and D. Guo. Ergodic fading Z-interference channels without state information at transmitters. *IEEE Trans. Inform. Theory*, 57, Special Issue on Interference Networks:2627–2647, May 2011.

[210] Y. Zhu and D. Guo. The degrees of freedom of isotropic MIMO interference channels without state information at the transmitters. *IEEE Trans. Inform. Theory*, 58:341–352, Jan. 2012.