Workshop on Mathematical Data Science (MDS) 2019

October 13–15, 2019

Schloss Dürnstein, Austria

Booklet of Abstracts

Table of Contents

Session 5 Chair: R. F. H. Fischer **Tue 08:30 – 10:00**

Session 6 Chair: A. Lapidoth **Tue 10:30 – 12:30**

Session 7 Chair: A. Ephremides **Tue 14:00 – 16:00**

Elementary algebraic topology via codes on graphs

G. David Forney, Jr.
forney@mit.edu
Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, MA 02139

This paper aims to introduce the concepts of elementary algebraic topology using concepts from the field of codes on graphs, namely normal realizations and normal factor graphs, and thereby to show their close relationship. As in prior work (*e.g.,* Molkaraie–Loeliger, Al-Bashabsheh–Vontobel), the main application is to Ising-type models of statistical physics.

Generalized Random Gilbert-Varshamov Codes

Anelia Somekh-Baruch
Bar-Ilan University
somekha@biu.ac.il

Jonathan Scarlett
National University of Singapore
scarlett@comp.nus.edu.sg

Albert Guillén i Fàbregas
ICREA & Universitat Pompeu Fabra
University of Cambridge
guillen@ieee.org

Abstract

We introduce a random coding technique for transmission over discrete memoryless channels reminiscent of the basic construction attaining the Gilbert-Varshamov bound for codes in Hamming spaces. The code construction is based on drawing codewords recursively from a fixed type class, in such a way that a newly generated codeword must be at a certain minimum distance from all previously chosen codewords, according to some generic distance function. We show that the random coding scheme attains an error exponent that is at least as high as both the random-coding exponent and the expurgated exponent, recovering the Csiszár and Körner exponent as a special case. We show that cost-constrained version of the proposed random coding scheme yields the dual expression of the above error exponent and extends its validity for general alphabets, possibly non-finite.

This work was supported in part by the Israel Science Foundation under grant 631/17, the European Research Council under Grant 725411, by the Spanish Ministry of Economy and Competitiveness under Grant TEC2016-78434-C3-1-R, and by an NUS Early Career Research Award.

"On Sliding Window Decoding of Convolutional Codes for Streaming Applications"

by Daniel J. Costello Jr.[1], Min Zhu[2], and David G. M. Mitchell[3]

The increasing demand for high-speed data streaming services over the Internet has focused attention on the need for reliable, low latency error control coding strategies suitable for continuous transmission. Convolutional codes lend themselves naturally to this environment, and conventional decoding methods, such as Viterbi decoding and sequential decoding, can provide moderate coding gains with low decoder complexity and latency. However, as channel resources become stressed by increasing traffic, the use of capacity-approaching code designs will be needed to meet service requirements. In this paper, we discuss two such approaches - braided convolutional codes (BCCs) and spatially coupled low-density parity-check (SC-LDPC) codes - that, combined with sliding window decoding (SWD), promise capacity-approaching performance with moderate decoder complexity and latency requirements.

It has been noted, however, that decoder error propagation is a significant problem that arises in the use of these capacity-approaching methods in a streaming environment. Particularly for small window sizes, i. e., low latency operation, an initial burst of decoding errors can trigger additional decoding errors that continue indefinitely and result in unacceptably high decoded error rates. Also, we have noted that it can be difficult to assess the extent of the damage with traditional computer simulation techniques, which require that "jobs" be submitted in frames of a fixed length, thus masking the full effect of decoder error propagation.

In this paper, we discuss several methods, based on monitoring the log-likelihood ratios (LLRs) of decoded symbols, designed to combat decoder error propagation in BCCs, including extending the size of the decoding window (increasing latency), a resynchronization mechanism, and a retransmission strategy. We also discuss how such methods may be similarly applied to combating decoder error propagation in SWD of SC-LDPC codes.

1 Dept. of Electrical Engineering, University of Notre Dame, IN, USA

2 State Key Lab. of ISN, Xidian University, Xi'an, China

3 Dept. of Electrical and Computer Engineering, New Mexico State University, NM, USA

Big Ideas in Wireless Inside 4/5G Mobile Networks

Arogyaswami Paulraj
Stanford University
apaualraj@stanford.edu

Abstract— **There are many big ideas proposed for wireless technology. Those related to transmit are usually called out in the standards and those related to receive only are usually not explicit in standards. We focus on 4G/5G systems.**

Though big ideas are well defined at the time of original invention and initial academic research, their impact in practical systems get complicated and intertwined with many other concepts.

In this talk we list some big ideas and comment on which ideas are used, not yet used or no longer used (and if there is time – why?)

AGE OF INFORMATION AND DATA SEMANTICS

Anthony Ephremides

University of Maryland

ABSTRACT

One of the aspects of the emerging discipline of Data Science concerns the reasons that underlie the transmission of information. That is, instead of focusing only on duly executing as fast as possible the transfer of as many messages as possible (which is the traditional objective of the communication process), attention is paid also to the purpose and/or value of the transfer of information. Without considering message "meaning", we are interested in other measures of "significance" for the transmitted messages. Such measures can be referred to as the "Semantics" of the data. For example, we may want to transmit messages for the purpose of executing some computation task or for implementing a control function.

One such example of semantics is the "freshness" of the data, better known as the Age of Information (AoI), which has been experiencing fast-growing interest. In this presentation, we review the basics of AoI and identify some of the potential fundamental aspects of freshness. To maintain the Age of Information at the lowest possible level, it is not sufficient to minimize the delay that accompanies transmission. It is also important to consider the sampling pattern of the signal. It is through this connection to sampling that interesting connections between Information Theory and Signal Processing may arise. In fact, Age alone may not be the most meaningful measure of freshness. Rather, a notion of "effective" Age is needed.

A step in that direction is the consideration of the Age of Incorrect Information (AoII), which captures the intrinsic value of freshness which, in turn, is to minimize the error in estimating and predicting the value of an evolving signal. If the objective of the transmission is indeed to enable the receiver to estimate with minimum error, the AoI (or, better, the AoII) is a suitable surrogate for the error itself. Rather than minimizing the error directly, it may be simpler to minimize the Age.

The objective of this talk will be to elaborate these points and identify potential challenges and benefits from the use of this "Semantic" measure.

On Least Squares with NUV Priors

Hans-Andrea Loeliger
ETH Zurich

Normal priors with unknown variance (NUV) include a large class of convex and nonconvex sparsity promoting priors and blend well with Gaussian message passing. Such priors and algorithms can be put to many uses: least-squares with sparsity and outliers, priors for imaging problems, clustering, dictionary learning, linear-Gaussian dynamical systems with jumps, blind signal parsing, sparse control of dynamical systems, and more.

References

[1] H.-A. Loeliger, Boxiao Ma, H. Malmberg, and F. Wadehn, "Factor graphs with NUV priors and iteratively reweighted descent for sparse least squares and more", *Int. Symposium on Turbo Codes and Iterative Information Processing (ISTC) 2018,* Hongkong, China, Dec. 3–7, 2018.

[2] H.-A. Loeliger, L. Bruderer, H. Malmberg, F. Wadehn, and N. Zalmai "On sparsity by NUV-EM, Gaussian message passing, and Kalman smoothing," *2016 Information Theory & Applications Workshop (ITA),* San Diego, CA, Feb. 2016.

[3] N. Zalmai, *A State Space World for Detecting and Estimating Events and Learning Sparse Signal Decompositions,* PhD thesis 24360 at ETH Zurich, 2017.

Rényi Entropy and Guessing: Old and New Results

Igal Sason

Abstract

We introduce in this talk the Rényi entropy and the Arimoto-Rényi conditional entropy, together with asymptotically tight bounds on the guessing moments by Arikan (1996) which are expressed as a function of these Rényi information measures. For a discrete random variable X which takes a finite n of possible values, the problem of maximizing the Rényi entropy of a function of X over all functions which are mappings from a set of cardinality n to a set of a *smaller* cardinality m (with fixed values of $m < n$) is *strongly NP-hard*. We provide an upper bound on this maximal Rényi entropy with a guarantee on its largest possible gap from the exact value, together with a simple algorithm to construct this function. This work was inspired by the recently published paper by Cicalese et al. (IEEE Trans. on IT, 2018), which is focused on the Shannon entropy, and it strengthens and generalizes the results of that paper to Rényi entropies of arbitrary positive orders by the use of majorization theory. We discuss the implications of these results in the context of the guessing problem.

The new findings in this talk are based on the paper: I. Sason, "Tight bounds on the Renyi entropy via majorization with applications to guessing and compression," *Entropy*, vol. 20, no. 12, paper 896, pp. 1–25, November 2018.

Flash Helping for Additive-Noise Channels

Amos Lapidoth and Gian Marti

Abstract

Flash helping is proposed as a technique for a helper to assist the encoder by providing it with a rate-limited description of the noise. It is optimal on the single-user Gaussian channel, the multiple-access Gaussian channel, and on the Exponential channel. It is applicable and optimal irrespective of whether the helper observes the noise causally or noncausally.

I. Introduction

To gain some insight into the promise of cooperative communications, we consider an additive Gaussian noise channel with an altruistic helper that observes the noise and wishes to help the encoder by providing it with a rate-limited description of said noise. Since the description is rate limited, the helper must quantize the noise, and it seeks to do so in a way that will provide the encoder with the greatest possible information-theoretic benefit. Here we show that "flash helping," which was originally proposed for decoder assistance [1] (with extensions to multi-terminal settings in [2]) is optimal. In fact, it is optimal irrespective of whether the assistance is provided noncausally or causally. Extensions to multiple-access channels and to the Exponential channel are discussed in [3].

The time-k output Y_k of the additive Gaussian noise channel is

$$Y_k = x_k + Z_k, \tag{1}$$

where $x_k \in \mathbb{R}$ is the time-k input, and the noise sequence $\{Z_k\}$ is IID $\sim \mathcal{N}(0, \mathsf{N})$, where $\mathsf{N} > 0$. A rate-R blocklength-n communication scheme for our setting can be described as follows. A message m is picked from the message set $\mathcal{M} = \{1, \dots, 2^{nR}\}$. Since the decoder receives no help, it guesses the message based on the channel output sequence $\mathbf{y}$ alone. It is thus a mapping $\psi_{\text{dec}} \colon \mathbb{R}^n \to \mathcal{M}$ that maps $\mathbf{y} \in \mathbb{R}^n$ to the decoder's guess $\hat{m}$.

The description of the encoder and the helper depends on whether or not we impose a causality constraint. A *noncausal* helper $\phi_{\text{nc-help}}$ is a mapping $\phi_{\text{nc-help}} \colon \mathbb{R}^n \to \mathcal{T}$, where $\mathcal{T} = \{1, \dots, 2^{nR_{\text{h}}}\}$ is the set of possible descriptions. Applying $\phi_{\text{nc-help}}$ to the noise sequence Z^n produces its description T. A noncausal encoder $\phi_{\text{nc-enc}} \colon \mathcal{M} \times \mathcal{T} \to \mathbb{R}^n$ maps the message m and the noise's description t to the transmitted length-n sequence $\mathbf{x}(m, t)$, which is required to satisfy for every $m \in \mathcal{M}$ the average power constraint $\mathbb{E}\big[\|\mathbf{x}(m, T)\|^2\big] \le n\mathsf{P}$, where $\|\cdot\|$ denotes the Euclidean norm, and $\mathsf{P} > 0$.

A *causal* helper describes the noise sequence z^n causally using an n-tuple $(t_1, \dots, t_n)$, where t_k takes values in the set $\mathcal{T}_k$ and must be computable from the noise samples z^k. (We use A^k to denote $A_1, \dots, A_k$, and we use A^n and $\mathbf{A}$ interchangeably.) It is thus described by n mappings $\{\phi^{(k)}_{\text{c-help}}\}^{n}_{k=1}$, with the k-th mapping $\phi^{(k)}_{\text{c-help}} \colon \mathbb{R}^k \to \mathcal{T}_k$ mapping z^k to t_k. To satisfy the description-rate constraint, we impose the cardinality constraint

$$\prod_{k=1}^{n} |\mathcal{T}_k| \leq 2^{nR_{\mathsf{h}}}. \tag{2}$$

Wishing to convey the message m and having observed the descriptions $t_1, \ldots, t_k$, the encoder emits the time-k channel input $x_k(m, t_1, \ldots, t_k)$. A causal encoder thus comprises n mappings $\{\phi^{(k)}_{\text{c-enc}}\}^{n}_{k=1}$,

$$\phi^{(k)}_{\text{c-enc}} \colon \mathcal{M} \times \mathcal{T}_1 \times \cdots \times \mathcal{T}_k \to \mathbb{R}, \tag{3}$$

satisfying the power constraint

$$\frac{1}{n} \sum_{k=1}^{n} \mathbb{E}\Big[x_k\big(m, t_1(Z^1), t_2(Z^2) \ldots, t_k(Z^k)\big)^2\Big] \leq \mathsf{P}, \quad \forall m \in \mathcal{M}. \tag{4}$$

The capacity is defined as the supremum of rates that allow for arbitrarily small probability of error. Our main result on the single-user channel is given in the following theorem. It relates the capacity $C(R_{\mathsf{h}})$ of the channel with a helper to its capacity $C(0)$ witout one.

Theorem 1. *The capacity of the encoder-assisted average-power constrained additive Gaussian noise channel is*

$$C(R_{\mathsf{h}}) = \frac{1}{2} \log\left(1 + \frac{\mathsf{P}}{\mathsf{N}}\right) + R_{\mathsf{h}} \tag{5}$$
$$= C(0) + R_{\mathsf{h}} \tag{6}$$

irrespective of whether the help is provided causally or noncausally.

- *No rate exceeding the RHS of* (5) *can be achieved even if the noise's description that is provided to the encoder is also provided to the decoder.*
- *If the noise is not Gaussian but satisfies the hypotheses that guarantee Bennett's high-resolution quantization result [4], [5, Theorem 6.2], then* (6) *is achievable with causal help even for non-Gaussian noise, provided that in* (6) *we interpret* $C(0)$ *as the capacity of the non-Gaussian noise channel in the absence of help.*

References

[1] S. I. Bross and A. Lapidoth, "The additive noise channel with a helper," in *2019 IEEE Information Theory Workshop (ITW)*. Visby, Sweden: IEEE, Aug. 2019.

[2] I. S. Bross, A. Lapidoth, and G. Marti, "Decoder-assisted communications over additive noise channels," *submitted*, 2019.

[3] A. Lapidoth and G. Marti, "Encoder-assisted communications over additive noise channels," *submitted*, 2019.

[4] W. Bennett, "Spectra of quantized signals," *Bell Systems Technical Journal*, vol. 27, pp. 446–472, July 1948.

[5] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*. Springer, 2007.

Data Privacy for a ρ-Recoverable Function

Ajaykrishnan Nageswaran and Prakash Narayan
Department of Electrical and Computer Engineering
and Institute for Systems Research
University of Maryland, College Park, MD 20742, USA
E-mail: {ajayk, prakash}@umd.edu

A user's data is represented by a finite-valued random variable. Given a function of the data, a querier is required to recover, with at least a prescribed probability ρ, the value of the function based on a query response provided by the user. The user devises the query response, subject to the recoverability requirement, so as to maximize privacy of the data from the querier. Privacy is measured by the probability of error incurred by the querier in estimating the data from the query response, and is nonincreasing in ρ. We discuss single and multiple independent query responses, with each response satisfying the recoverability requirement, that provide maximum privacy to the user. Achievability schemes with explicit randomization mechanisms for query responses are given and their privacy compared with converse upper bounds. Also mentioned will be more stringent forms of privacy, viz. predicate privacy and list privacy, and the notion of divergence privacy of the probability distribution of the data.

Conferencing in Arbitrarily Varying Channels: New Models and Observations

Yossef Steinberg
Dept. of Electrical Engineering
Technion - IIT
Haifa 32000, ISRAEL
Email: ysteinbe@technion.ac.il

***Abstract*—This talk focuses on recent advances in the study of multi-user communication systems, that suffer from high degree of uncertainty in the channel model. Specifically, we study arbitrarily varying (AV) network models with cooperation links between users. It has long been observed that cooperation between users in a regular (not necessarily AV) communication network can considerably increase the network capacity, as it can be used to generate dependence between transmission signals of remotely located users. In AV settings, the task of the cooperation link is twofold: besides its traditional function of creating correlation between different users, it can be used to distribute small amounts of common randomness between users, thus facilitating the use of deterministic codes for users whose channels are symmetrisable. Thus a small amount of cooperation can have little effect on the capacity of a regular (non-AV) network, but a dramatic effect on the capacity of an AV system. This fact has already been observed by Wiese and Boche in the context of the multiple access channel.**

Another source of variability in modern communication channels, independent of whether the channel is AV or not, is the uncertainty of cooperation. In modern ad-hoc networks, users come and go, and their willingness to serve as relays or helpers is not guaranteed a priori. Moreover, in complex, non-centralised ad hoc networks, part of the users cannot be informed about the situation of the cooperation link, thus coding schemes cannot be swapped per scenario. Traditional cooperation schemes, developed so far in the IT literature, rely heavily on the cooperation - if the link is absent, decoding cannot be performed. A new approach to uncertain cooperation, explored in recent works in the context of regular (non-AV) systems, is to devise coding scheme that are robust in the following sense: the decoders exploit the cooperation when it is present, but can still operate when it is absent, possibly leading to lower decoding rates.

An interesting family of problems arise in networks with these two sources of uncertainty: the channel statistics is AV, and cooperation links are not guaranteed to exist a priori. Although this seems a complex model, it is realistic in modern networks, as statistical properties of channels do change in time in an arbitrary manner (due to fading, jammers, malicious attacks, and more), and the dynamic nature of modern networks induces uncertainty not only in their statistics, but also in their topology. Interestingly, it turns out that closed form results can be obtained for these involved scenarios.

In this talk I will give an overview of the state of the art, and describe recent works and results on the topics described above. The goal is to study networks with the highest possible degree of uncertainty, that can still yield meaningful models and results.

The presentation is based on joint works with Dor Itzhak, Wasim Huleihel and Uzi Pereg.

Dimensions of Uncertainty in Communication Theory

Ezio Biglieri†

The uncertainty concept usually dealt with in communication theory is associated with unknown outcomes that differ each time one runs an experiment under similar conditions. The standard tool used in this situation is probability theory, as all forms of uncertainty are treated in terms of the single dimension of probability. However, one should realize that uncertainty can take multiple aspects: specifically, the "aleatory" uncertainty caused by the randomness of system behavior, and the "epistemic" uncertainty due to ignorance. Aleatory uncertainty is attributed to outcomes that for practical purposes cannot be predicted and are therefore treated as stochastic (e.g., the result of a coin flip), whereas epistemic uncertainty is attributed to missing information or expertise or inadequacy of one's model of aleatory uncertainty. For example, in performance evaluation of wireless communication no single fading-channel model can be fully accurate for a wide variety of channels. Performance can be computed once a model is chosen, but the actual physical channel model pertains to epistemic uncertainty.

Under epistemic uncertainty, citing verbating from [4], "*it is better to have an analysis which is correct and honestly distinguishes between variability and incertitude than an analysis that depends on unjustified assumptions and wishful thinking. $(\cdots)$ If the price of a correct assessment is broad uncertainty as a recognition or admission of limitations in our scientific knowledge, then we must pay that price.*"

To deal with epistemic uncertainty, we need appropriate mathematical tools: a calculus by which this type of uncertainty can be properly manipulated, a meaningful way of measuring the amount of relevant uncertainty in any situation that is formalizable in the theory, and a way to develop methodological aspects of the theory, including procedures of making the various uncertainty principles operational within the theory [5]. We advocate the use of "probability boxes," which are interval-type bounds on cumulative distribution functions that can handle a great deal of model uncertainties, imprecisely specified distributions, and poorly known or unknown dependences of random variables.

Bibliography

[1] C. Alsina, M. J. Frank, and B. Schweizer, *Associative Functions: Triangular Norms and Copulas.* Singapore: World Scientific Publ. Co., 2006.

[2] B. M. Ayyub and G. J. Klir, *Uncertainty Modeling and Analysis in Engineering and the Sciences.* Boca Raton, FL: Chapman & Hall/CRC, 2006.

[3] S. Ferson, "Model uncertainty in risk analysis," *Proceedings of the 6th International Workshop of Reliable Engineering Computing: Reliability and Computations of Infrastructures*, Chicago, IL, pp. 27–43, May 25–28, 2014.

[4] S. Ferson, V. Kreinovich, L. Ginzburg, D. S. Myers, and K. Sentz, *Constructing Probability Boxes and Dempster–Shafer Structures.* Albuquerque, NM: Sandia National Laboratory Report SAND 2002âĂŞ4015, January 2003.

[5] G. J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory.* Hoboken, NJ: J. Wiley & Sons, 2006.

[6] R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis.* Philadelphia, PA: Society for Industrial and Applied Mathematics, 2009.

[7] R. B. Nelsen, *An Introduction to Copulas* (2nd Ed.). New York, NY: Springer, 2006.

[8] H. M. Regan, S. Ferson, and D. Berleant, "Equivalence of methods for uncertainty propagation of real-valued random variables," *International Journal of Approximate Reasoning*, vol. 36, pp. 1–30, 2004.

†Universitat Pompeu Fabra, Barcelona, Spain. E-mail: e.biglieri@ieee.org

This work was supported by the European Research Council under the H2020 Framework Programme/ERC grant agreement 694974.

Channel Capacity: From Waves to Particles and Back Again

Richard E. Blahut

Abstract—**The well-known Shannon capacity expression for the bandlimited additive gaussian-noise channel is correct from the point-of-view of the mathematics, but incomplete from the point-of-view of the physics. This is because of the emergent granularity of a lightwave at low signal levels. Quantum information theory, though formally the proper tool to study this, is a tool too sharp to obtain insightful answers. We propose an intermediate semiclassical information theory based on the Poisson transform of Mandel and Wolf. The wave and particle views of a lightwave are seen to be the two sides of the Poisson transform. The capacity of a photon channel conjectured by Gordon (1962) and Forney (1963) based on maximum-entropy considerations is proved to be correct, and the Shannon capacity of the bandlimited gaussian noise channel is shown to be the emergent capacity formula as the number of photons becomes large.**

Convolutions, Fourier Transforms, and Rigged Hilbert Spaces

H. G. Feichtinger

Abstract—**It is the purpose of this presentation to recall some of the reasons why engineers and mathematicians are interested in the Fourier transform. As a student, I was told, that the Fourier transform is important because it transforms the complicated operation of convolution into simple pointwise multiplication. But why should we be interested in convolution? A valid answer came to me from the engineers, who are teaching linear time-invariant systems, impulse response and transfer functions in their first courses.**

Comparing the two sides of Fourier Analysis, the applied and the theoretical one (which in fact via the theory of tempered distributions by Laurent Schwartz, with important applications to partial differential equations), one observes that they have not too much to do. I will provide a few striking examples in the talk.

Given this situation, I suggest a reconciliation of the two worlds, based on long-term cooperation with engineers, both in research and teaching. The theory of Banach Gelfand Triples, also known as Rigged Hilbert Spaces, provides such a possibility. The modern approach to time-frequency analysis (TFA) allows a simple description of the Segal algebra $S_0(R^d)$, which forms an algebra of test-functions (via integrability of the STFT). The dual space (or equivalently distributional completion), also called space of ``mild distributions'' can be described as the space of all tempered distributions which have a bounded spectrogram (STFT). As time permits we will give a few examples, showing that within this setting a mathematically justified treatment of most expressions arising in engineering applications (such as Shannon's Sampling Theorem, representation of systems as convolution operators, etc.) is possible.

The material presented is part of a long-term project by the speaker, and there is a list of talks and papers available from the NuHAG web-page, e.g. www.nuhag.eu/talks (access via ``visitor'' and ``nuhagtalks'').

Data Science by TV on the Graph

Gerald Matz

Abstract—**Graph signal processing is a modern paradigm to deal with large data sets. It captures the intrinsic structure of the data via the topology of a graph. By capitalizing on the graph structure, diverse large-scale learning and inference problems can be tackled. Graph signal processing is promising in applications like sensor networks, social networks, infrastructure networks, or biological networks.**

In this talk I will report some of our recent work in which we build on the notion of graph total variation to formulate a consistent theoretical framework and efficient distributed algorithms for data reconstruction, network structure inference, and clustering.

Capacity of dynamical storage systems

Ohad Elishco Alexander Barg

Abstract

We introduce a dynamical model of node repair in distributed storage systems wherein the storage nodes are subjected to failures according to independent Poisson processes. The main parameter that we study is the time-average capacity of the network in the scenario where a fixed subset of the nodes support a higher repair bandwidth than the other nodes. The sequence of node failures generates random permutations of the nodes in the encoded block, and we model the state of the network as a Markov random walk on permutations of n elements. As our main result we show that the capacity of the network can be increased compared to the static (worst-case) model of the storage system, while maintaining the same (average) repair bandwidth, and we derive estimates of the increase. We also quantify the capacity increase in the case that the repair center has information about the sequence of the recently failed storage nodes.

Preprint of the full paper is available as arXiv:1908.09900.

The authors are with Institute for Systems Research and Department of ECE, University of Maryland, College Park, MD 20742, USA, emails {ohadeli,abarg}@umd.edu. Research supported by NSF grants CCF1618603 and CCF1814487.

Expectation-Consistent Approximate Inference with Vector-Valued Diagonalization — An Unbiasing Interpretation

Robert F.H. Fischer
Institut für Nachrichtentechnik, Universität Ulm, Ulm, Germany, Email: robert.fischer@uni-ulm.de

Abstract

In 2005, Opper and Winther [12] presented a general framework for solving inference problems, called *expectation consistent (EC) approximate inference*. The main idea is to replace a non-tractable, high-dimensional probability density function (pdf) by a suitably chosen one from an exponential family. They argue that their approximation is good, as long as the moments of the original and substitute pdf match. Moreover, a simple "single loop" algorithm was given, which realizes loopy belief propagation. Later, Fletcher et al. [6] presented a generalization (basically to maximum-a-posteriori (MAP) estimation) and studied the convergence behavior.

It is of interest to apply the general framework to the *standard linear regression problem* where noisy linear measurements of the form

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$$

are present. Thereby, $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is a known (sensing) matrix, the elements of $\boldsymbol{x} \in \mathbb{R}^N$ are i.i.d. with known marginal pdf $\mathsf{f}_{\mathsf{x}}(x)$, and $\boldsymbol{n}$ is zero-mean Gaussian with variance σ_n^2 per component.

For $M < N$ and choosing the pdf $\mathsf{f}_{\mathsf{x}}(x)$ suitable, the *compressed sensing (CS)* problem [2], [4] is included. A huge bunch of recovery algorithms exists; among them the powerful family of *approximate message passing (AMP)* [5], [10]. Recently, *vector approximate message passing (VAMP)* [13], [14] has been proposed; it can be straightforwardly derived from the EC framework. Similar (but not identical) approaches are *orthogonal AMP* [9] and *iterative MMSE estimation and soft feedback (IMS)* [15].

One degree of freedom when applying the EC approach is to define which moments have to match—this step is also called *diagonalization*. Often (e.g., in VAMP) "uniform diagonalization" (diagonal restricted in [12]) is chosen. Here, besides the first-order moments (means of the elements of $\boldsymbol{x}$) the average second-order moment (and hence variance) should match—we call this strategy *average variance*. As a consequence, within the algorithm $m_i = \mathrm{E}\{x_i\}$, $i = 1, \ldots, N$, and $\sigma^2_{\text{avg}} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{E}\{(x_i - m_i)^2\}$ are tracked.

Alternatively, a "vector-valued diagonalization" (diagonal in [12]) can be employed where the first- and second-order moments of all elements of $\boldsymbol{x}$ should match. Here, $m_i = \mathrm{E}\{x_i\}$, $i = 1, \ldots, N$, and $\sigma_i^2 = \mathrm{E}\{(x_i - m_i)^2\}$, $i = 1, \ldots, N$, are tracked. Hence, *individual variances* characterize the reliabilities of the N unknowns in $\boldsymbol{x}$—in contrast to a single average reliability over the entire vector. However, in numerical simulations it turns out that applying the EC approach to the CS setup, using an average variance leads to better performance than using individual variances (which, unfortunately, has higher numerical complexity).

In the present contribution, we enlighten the main reasons for this effect. To that end, the single-loop EC algorithm is reformulated as iterating over two minimum mean-squared error (MMSE) estimation problems (cf. [14]). Noteworthy, since both diagonalization strategies consider the two first moments, inherently Gaussian distributions are treated. The relation between the natural parameters [1], [11] of the three pdfs involved in the EC algorithm are shown to be nothing else than the removal of the bias inherent in any MMSE solution [3], [7]—which, in turn, is nothing else than the calculation of extrinsic information [8].

For the step of non-linear, per-component estimation of the unknowns obeying the signal pdf (the "q to r" step in [12]; the "denoising" block in [14, Alg. 3]), following [16], two versions of unbiasing are presented, namely a *"signal-oriented"* and a *"noise-oriented"* approach. The adequate unbiasing relations (in particular for the conditional variances) are derived.

Numerical simulations are provided for a discrete CS setting—the elements of $\boldsymbol{x}$ are drawn from the finite alphabet $\{-1, 0, +1\}$. The results cover that using individual variances in combination with "noise-oriented" unbiasing provides the best results; in all cases VAMP, i.e., utilizing an average variance, is outperformed.

Acknowledgment

The author thanks Carmen Sippel (Ulm University) and Norbert Görtz (Technical University Vienna) for valuable discussions.

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) within the framework "Compressed Sensing in der Informationsverarbeitung (CoSIP)" — grant FI 982/16-1.

References

[1] L.D. Brown. Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. *Lecture Notes – Monograph Series*, Volume 9, Institute of Math. Statistics, Hayward, USA, 1986.

[2] E.J. Candes, J. Romberg, T. Tao. Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] J.M. Cioffi, G.P. Dudevoir, M.V. Eyuboglu, G.D. Forney. MMSE Decision-Feedback Equalizers and Coding. I. Equalization Results. *IEEE Transactions on Communications*, vol. 43, no. 10, pp. 2582–2594, Oct. 1995.

[4] D.L. Donoho. Compressed Sensing. *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] D.L. Donoho, A. Maleki, A. Montanari. Message Passing Algorithms for Compressed Sensing: I. Motivation and Construction. In *IEEE Information Theory Workshop (ITW)*, Cairo, Egypt, Jan. 2010.

[6] A. Fletcher, M. Sahraee-Ardakan, S. Rangan, P. Schniter. Expectation Consistent Approximate Inference: Generalizations and Convergence. In *IEEE International Symposium on Information Theory*, Barcelona, Spain, July 2016.

[7] G.D. Forney. On the Role of MMSE Estimation in Approaching the Information-Theoretic Limits of Linear Gaussian Channels: Shannon Meets Wiener. In *Allerton Conference on Communication, Control, and Computing*, Monticello, IL, USA, Oct. 2003.

[8] F.R. Kschischang, B.J. Frey, H.-A. Loeliger. Factor Graphs and the Sum-Product Algorithm. *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[9] J. Ma, L. Ping. Orthogonal AMP. *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[10] A. Maleki. Approximate Message Passing Algorithms for Compressed Sensing. PhD Thesis, Stanford University, Stanford, USA, Sep. 2011.

[11] P. Moulin, V.V. Veeravalli. *Statistical Inference for Engineers and Data Scientists*. Cambridge University Press, 2018.

[12] M. Opper, O. Winther. Expectation Consistent Approximate Inference. *Journal of Machine Learning Research*, vol. 6, pp. 2177–2204, Dec. 2005.

[13] S. Rangan, P. Schniter, A.K. Fletcher. Vector Approximate Message Passing. In *IEEE International Symposium on Information Theory*, Aachen, Germany, June 2017.

[14] S. Rangan, P. Schniter, A.K. Fletcher. Vector Approximate Message Passing. *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6664–6684, Oct. 2019.

[15] S. Sparrer, R.F.H. Fischer. Algorithms for the Iterative Estimation of Discrete-Valued Sparse Vectors. In *International ITG Conference on Systems, Communication, and Coding*, Hamburg, Germany, Feb. 2017.

[16] S. Sparrer, R.F.H. Fischer. Bias Compensation in Iterative Soft-Feedback Algorithms with Application to (Discrete) Compressed Sensing. In *IEEE Statistical Signal Processing Workshop 2018*, Freiburg, Germany, June 2018.

Analysis of Information Content in Dynamic Networks

Wojciech Szpankowski
Department of Computer Science
Purdue University
W. Lafayette, IN 47907, U.S.A.
http://www.cs.purdue.edu/homes/spa

Abstract—**Shannon information theory has served as a bedrock for advances in communication and storage systems over the past six decades. However, this theory does not handle well higher order structures (e.g., graphs, geometric structures), temporal aspects (e.g., real-time considerations), or semantics, which are essential aspects of data and information that underlie a broad class of current and emerging data science applications. In this talk, we present some recent results on structural and temporal information in dynamic networks/graphs, in which nodes and edges are added and removed over time.**

We focus on two related problems: (i) compression of structures -- for a given graph model, we exhibit an efficient algorithm for invertibly mapping network structures (i.e., graph isomorphism types) to bit strings of minimum expected length, and (ii) node arrival order inference -- for a dynamic graph model, we determine the extent to which the order of node arrivals can be inferred from a snapshot of the graph structure. For both problems, we apply analytic combinatorics, probabilistic, and information-theoretic methods to find statistical limits and efficient algorithms for achieving those limits.