Research overview: Signal processing under communication constraints

Modern challenges in signal processing and statistics impose severe limitations on the amount of memory in the digital representation of the data. My research studies signal processing techniques that explicitly consider bit-restricted digital representation in their operation, as in the following diagram:

![Diagram](image)

This restricted digital representation is the result of constrained internal memory or communication bandwidth between different components or machines, which may be due to limited power or limited capacity compared to the size or amount of data. For example, emerging sensor network technologies in medicine and “smart cities” use many low-cost sensors to collect large amounts of data and transmit it to remote locations. These sensors must operate under severe power restrictions, hence they are limited by their memory and communication bandwidth. In another example, the data is carried by a wideband signal as in mm-waves communication, and the digital representation is obtained through an analog-to-digital conversion system. Another application of my work is in distributed cloud computing, where the size of datasets is disproportionately large compared to the bandwidth available to process information and communicate it across different machines and users. As a result, performance of such systems are dictated by communication, data compression and quantization constraints.

Although the examples above have very different scale in terms of the amount of data involved, the principal challenge in both is similar: compressing raw data in the most efficient way into a limited bit representation of the data. Understanding the fundamental limits of information processing under compression constraints allows us to develop techniques and design systems that support information processing and communication needs at all scales. In addressing this challenge, my research combines signal processing methodologies in the manipulation of signals with an information theoretic approach to the identification of structure within the data. While my work relies on these two well-established fields, my doctoral research, as describe in the following, shows that intriguing new results emerge at the intersection of the two disciplines.

My doctoral research has mainly focused on the way raw high dimensional and possibly analog signals are sampled before they are compressed into their digital representation. In particular, the ability to consider data in its analog form distinguishes my work from other compression and dimensionality reduction techniques, which operates after the analog-to-digital converter stage, i.e., on the data in its digital form. My contribution is in characterizing the minimal distortion that can be attained in this procedure as a function of the sampling rates and the bitrate of the digital representation, as well as the optimal sampling and encoding or compression schemes attaining it. These results provide a unified treatment of time-series processing under the following three information inhibiting phenomena: sampling, compression and additive noise. While the effect of each of these phenomena has been well-understood before, my work shows that the combination of them undermines standard conceptions. For example, it turns out that when the bitrate of the digital representation is restricted, sampling at a rate smaller than Nyquist is optimal.
These findings from my doctoral research have motivated me to study the effect of compression and limited communication on statistical estimation, by developing a unified framework of statistical estimation from encoded or compressed information. This framework is essential in understanding limits of both data intensive tasks such as classification, learning and distributive optimization, as well as applications in systems subject to severe data rate constraints such as low-power sensor networks.

The sections below describe in more detail my past work and future research goals.

**Past research**

In this section I describe the main contributions of my doctorate research. In this work I have characterized the effect of data compression on traditional signal processing techniques. This description highlights how the fundamental performance of these techniques under restricted digital representation undermines previously held conceptions both in signal processing and in information theory.

**Combined Sampling and Compression**

Sampling is a fundamental signal processing technique since it allows continuous-time information sources of the physical world to be converted into a discrete time representation that can be effectively manipulated. In the big data era sampling is also viewed as an effective dimensionality reduction technique widely used in machine learning and statistics. The effort of my doctorate was focused on characterizing the minimal distortion that can be attained in recovering signals from a digital representation of their samples, as in the following diagram:

![Diagram](image)

The results of this work, described in [KGEW16] and [KEG16], is a full characterization of the minimal distortion that can be attained in recovering the original signal as a function of the sampling rate at the sampler and the bitrate at the output of the encoder. Note that without the sampling constraint, this minimal distortion is described by Shannon’s distortion-rate function (DRF) of the original signal. As illustrated in figure below, an important corollary from my characterization is the fact that, for most signal models, this DRF is attained by sampling at a rate smaller than the Nyquist rate.

![Diagram](image)

That is, while Nyquist rate sampling is necessary to attain zero distortion without any constraint on the bitrate of the digital representation, when the latter is imposed the minimal distortion can be attained by sampling at a new critical sampling rate which is typically lower than the Nyquist rate. This new critical sampling rate depends on the bitrate of the digital representation and can be precisely derived in many important cases.

An interesting way to explain this phenomena is by alignment of degrees of freedom reduction between sampling and lossy compression: an optimal compression strategy must discard the part of the signal with the lowest uncertainty. My work showed that a sampler operating at a sub-Nyquist sampling rate can be tuned to discard the exact same part, and therefore achieve the optimal compression performance at a sub-Nyquist sampling rate.
Tradeoffs between Time-Resolution and Bit-resolution

As opposed to ideal source encoding schemes considered in information theory, many practical sampling systems allot a fixed number of bits in their memory to represent each sample. Since memory is limited, high sampling rates would lead to low quantization bit-precision and vice-versa, while reducing sampling rates below the Nyquist rate increase the distortion due to sampling. The result is an interesting trade-off between quantization precision and sampling rate in systems that operate under memory constraints. In my works [KGE15] and [KEG15], this trade-off was studied using a sampling system and a quantizer restricted to operate under a fixed number of bits per source sample. The main result of this research implies that the optimal sampling rate that minimizes the overall error is achieved at a sub-Nyquist sampling rate. This result implies that some sampling distortion is preferred in order to achieve an optimal digital representation of the source signal, and that lifting memory restrictions such as a strict bit-budget challenges the well-established convention that sampling should be performed above the Nyquist rate. Since these memory restrictions are often used in practice, this part of my work has immediate implications to the design of analog-to-digital sensors [KGE16].

Distributed Estimation under Communication Constraints

Estimating signals from their partial or noisy observations is one of the most fundamental tasks in signal processing. Bandwidth and power restrictions in distributed sensor technology impose further limitation on communication rates from the sensors to some central estimation unit. Therefore, each sensor must compress its observations at a rate not exceeding the capacity of the communication link, and the overall quality of estimation is affected by inaccuracies due to both observation noise and lossy compression.

My publications [KRG15], [KRG16b], [KRG16a], [SKRG16] have characterized the minimal distortion that can be attained as a function of the allowed communication rates under various assumptions on the amount of cooperation possible among the different sensors. In particular, we showed that when the quality of observations is low to begin with, the estimation cannot be improved by allowing the different sensors to devise a single compression strategy. In contrast, the system can greatly benefit from a joint estimation strategy when quality of observations is high. Another important contribution of our research is the development and analysis of a particular compression and estimation strategy that does not require the exchange of offline information among the sensors. This strategy allows the design of ad-hoc distributed sensor network, where each sensor is designed independently of the end estimation task.

Other Works

This section briefly explains a few of my other results in signal processing and communication. While these works do not fall under the categories above, they all describe novel results in signal processing and statistical estimation with restricted signal information that is not necessarily the result of data compression.

In [RKJ +16], we provided convergence guarantees for estimating the covariance matrix of vector autoregressive processes, under the assumption that observation of the process are only partially given. In [CKG15], we have shown that using multiple receiving antennas, reliable communication is possible even if the bit-resolution of an analog-to-digital converter at each antenna is very coarse compared to the received signal strength. Finally, in my Masters work [Kip12], I have proposed an extension of the Itô stochastic integral to a class of processes with possibly dependent differences. The Itô integral is widely used in financial mathematics and dynamical systems, and our extension allows the formulation of similar dynamical models driven by random perturbations correlated across long time lags. In continuation of this work, my Masters adviser and I have developed a framework for solving classical prediction problems with respect to this class of processes [AK15].

Future Research

In this section I describe my research vision to develop a unified theory for signal processing and statistical estimation under compression and communication constraints. The long-term goal of my
Research Statement

Alon Kipnis

research is to incorporate these restrictions into existing statistical theories and analysis. In addition to fundamental performance guarantees, this would also provide optimal compression and communication strategies that can be implemented in designing novel sensing devices and more efficient distributed computing schemes. Rapid changes in cloud computing, optimization and sensor fusion technology make the pursuit of this goal both exciting and challenging.

In the short term, under the umbrella of a unified characterization for the performance of various statistical theories under strict memory or communication requirements, I plan to develop tools to analyze hypothesis testing, pattern learning and parametric estimation under these constraint. Along with this theoretical work, I will investigate applications of these tools in optimization, classification and sensing problems, where memory or communication rates are limited. Below I have listed three specific research projects that I have begun to explore in this area of research. These projects are good starting points for understanding limitations of existing techniques due to memory and communication constraints, which would provide the foundation to develop a unified theory.

Combined compressed sensing and lossy compression - This research investigation asks if and how the phase transition curve in compressed sensing, which describes the ratio between sparsity and sampling rate allowing perfect reconstruction, changes as a result of compressing the measurements. Another question I intend to explore is to determine the optimal trade-off between bit-precision and number of samples in this setting. In addition to the theoretical insight, this research also bridges the gap between compressed sensing theories and their implementation in practical bit-constrained systems. It would therefore assist in implementing novel systems based on compressed sensing ideas, such as radar, cognitive radio and ultrasound.

Parametric estimation from few-bits measurements - The goal in this research is to derive statistical guarantees and strategies in estimating parameters under severe limitation on the number of bits available on each sample from the distribution. An important question this research would answer is determining the rate of convergence of a mean estimator from one-bit measurements. This question highlights the fact that even the most basic statistical estimation tasks lack simple solutions under severe bit constraints.

Optimization under communication constraint - This problem investigates the performance of optimization techniques where information on the curvature of the objective function is limited. For instance, we may ask what are the theoretical convergence guarantees in optimizing a simple convex function where only 1-bit gradient measurements are available.

Summary

Today’s signal processing demands call for a unified framework of statistical inference and communication. Such a framework should include all stages of data processing – from acquisition and analog-to-digital conversion through compression and retrieval. As memory and communication constraints become the limiting factor in current sensors and distributed computing technologies, availability of my unified analysis framework would be crucial for optimization and scaling of existing techniques and developing new applications based on these technologies. These applications include low-power devices employing low communication rates strategies, and distributed algorithms for classification and optimization in machine learning that can process only small portions of available data.

References


