AFRL-AFOSR-VA-TR-2019-0266

Solving inference and inverse problems using soft data

Mark Davenport
GEORGIA TECH RESEARCH CORPORATION

09/12/2019
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
02/19/2018	Final Performance	9/15/2014 - 9/14/2017

4. TITLE AND SUBTITLE
(YIP) Solving Inference and Inverse Problems Using Soft Data

5a. CONTRACT NUMBER
FA9550-14-1-0342

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Davenport, Mark A.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Georgia Tech Research Corporation
505 10th St
Atlanta, GA 30318

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
801 N Randolph St, Rm 732
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The research funded by this grant involved four main thrusts: (1) Deepening our understanding of how to learn embeddings from similarity observations, (2) Developing techniques for adaptive sampling for estimation and learning, (3) Developing methods for exploiting dynamics in the context of low-rank matrix recovery, and (4) Improving our understanding of how to learn from point process data..

15. SUBJECT TERMS
Low-rank matrix recovery, dynamic subspace models, adaptive sampling, personalized learning

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	SAR		

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATE COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33315-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report. e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Final Performance Report

(YIP) Solving Inference and Inverse Problems Using Soft Data

AFOSR Grant FA9550-14-1-0342

Reporting Period: September 15, 2014 – September 14, 2017

Principal Investigator: Prof. Mark A. Davenport
School of Elec. & Comp. Engineering
Georgia Institute of Technology
777 Atlantic Drive NW, Atlanta, GA 30332
(404) 894-2881
mdav@gatech.edu

Submitted to: Dr. Doug Riecken
Air Force Office of Scientific Research
875 North Randolph Street, Arlington, VA 22203
(703) 696-7796
richard.riecken@us.af.mil

Summary

In recent years we have witnessed an explosion in the amounts of (often extremely high-dimensional) data being acquired. As a result, data-driven techniques are increasingly applied, not only in traditional quantitative settings, but also in a variety of non-traditional scenarios that challenge many common assumptions. For example, in contexts such as social network analysis, collaborative filtering, personalized medicine, and personalized learning, we face a variety of challenges due largely to the fact that an important – often the only – source of data is *people*. In such applications, we want to learn about *people* using data that *people* supply. This presents several difficulties, including the fact that such data is often "soft" in the sense that it consists of nonmetric data like categories or comparisons rather than precise numerical values. Such data is also often constantly changing, necessitating approaches that can handle dynamic models. Moreover, in many cases it is impossible to fully sample, and our observations are highly incomplete.

It might seem impossible to extract much information from such data. Fortunately, in the examples considered above, as well as in many other applications, the data often has a great deal of structure. For example, the data might exhibit low-rank structure, or it may be possible to embed the data in a low-dimensional subspace or union of subspaces. Such structure, when it exists, allows us to potentially overcome the limitations of missing data. The goal of the research funded by this grant was to develop algorithms that can tackle such inverse problems when confronted with "soft" data. While this is indeed possible, the approaches necessarily differ substantially from the approach taken when given numerical measurements. As a part of this research, we addressed a number of fundamental theoretical and algorithmic questions in the context of low-rank matrix recovery, multidimensional scaling, unfolding, and low-dimensional dynamic models.

1 Accomplishments

In this section I briefly summarize the research efforts supported by this grant. Please refer to the publications listed at the end of this report for a more detailed description of these efforts.

1.1 Overview

The act of *measurement* is the cornerstone of quantitative research in nearly every scientific and engineering discipline. Measurements provide the raw data we need to learn about the world around us, help limit the impact of our subjective biases, and make the replication of experiments far easier. The success of this "data-driven" approach to science can hardly be overstated, and as measurements have become increasingly easier and less expensive to obtain, we have witnessed an explosion in the amounts of (often extremely high-dimensional) data that is being acquired. Moreover, the success of this paradigm now extends far beyond the traditional quantitative "hard sciences." Data-driven techniques are increasingly being applied in a variety of non-traditional scenarios that challenge the assumptions taken for granted in more traditional settings.

For example, as the dimensionality of the data we would like to work with grows, it can become challenging or impossible to fully sample the data. This occurs in a variety of sensing applications (e.g., high-resolution medical and scientific imaging) but also when working with the kind of data that arises in many common machine learning problems (e.g., recommender systems, personalized medicine, and intelligent tutoring systems), where *incomplete data* poses a significant challenge. Moreover, while centuries of experience has trained us to instinctively think of a measurement as a numerical quantity that can be read off a scale (e.g., mass, length, time, voltage, etc.), in many modern applications, the data is often of a more "soft" nature. For example, many machine learning systems often involve an attempt to measure some quality of a person – such as a preference, opinion, feeling, ability, etc. – for which there is no natural numerical scale. Moreover, in these contexts our main (or only) source of data often consists of responses which are typically heavily quantized. More generally, due to various practical limitations, it is increasingly common to encounter *nonlinear* measurement models in a wide array of applications.

In recent years, there has been a tremendous amount of progress in developing techniques for dealing with incomplete data and nonlinearities in the observation process when the data exhibits certain forms of low-dimensional structure. To name some particularly prominent examples, the fields of *compressive sensing*, *matrix completion*, and *phase retrieval* all aim to address such problems by exploiting sparsity and/or low-rank structure. While these approaches are extremely successful when these models are good matches to the structure truly present in the data, there can often be significant gaps between these models and the structure actually present in the real-world. For example, our measurements may be generated according to some underlying continuous phenomenon, as opposed to the discrete models (such as sparsity in a fixed dictionary) that are more commonly considered. Similar challenges can arise whenever the underlying models are dynamically evolving over time. In such cases, novel theoretical and algorithmic tools are required to develop computationally friendly extensions of existing models.

The focus of the research supported by this grant was to further develop theory and algorithms that can help us tackle the kinds of inverse problems described above when confronted with incomplete and/or nonlinear measurements, even extending to the case where the underlying signal of interest is dynamically changing with time. While performing inference in these contexts is indeed possible, the approaches necessarily differ substantially from the existing methods, and required the development of novel low-dimensional signal models as well as new theoretical tools and algorithms. Below I summarize the research accomplishments supported by this grant in four broad areas: (1) learning from paired comparisons, (2) adaptive sampling for estimation and learning, (3) dynamic low-rank matrix recovery, and (4) learning from point processes.

1.2 Learning from paired comparisons

1.2.1 Problem statement

In the context of learning from paired comparisons, our goal is to determine the location of a point in Euclidean space based on distance comparisons to a set of known points, where our observations are nonmetric. In particular, let $x \in \mathbb{R}^n$ be the true vector that we are trying to estimate, and let $(p_1, q_1), \ldots, (p_m, q_m)$ be pairs of "landmark" points in $\mathbb{R}^n$ which we initially assume to be known *a priori*. Rather than directly observing the raw distances from x, i.e., $\|x - p_i\|$ and $\|x - q_i\|$, we instead obtain only paired comparisons of the form $\|x - p_i\| < \|x - q_i\|$. Our goal is to estimate x from a set of such inequalities. Nonmetric observations of this type arise in numerous applications and have

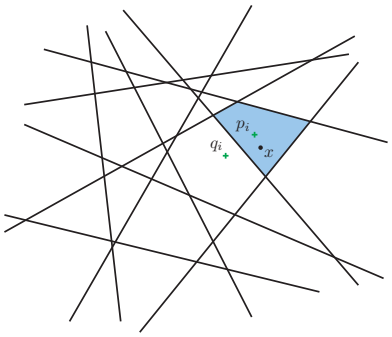

Figure 1: An illustration of the localization problem from paired comparisons. The information that x is closer to p_i than q_i tells us which side of a hyperplane x lies. Through many such comparisons we can hope to localize x to a high degree of accuracy.

seen considerable interest in recent literature. These methods are often applied in situations where we have a collection of items and hypothesize that it is possible to embed the items in $\mathbb{R}^n$ in such a way that the Euclidean distance between points corresponds to their "dissimilarity," with small distances corresponding to similar items. Here, we focus on the sub-problem of adding a new point to a known (or previously learned) configuration of landmark points.

As a motivating example, we consider the problem of estimating a user's preferences from limited response data. This is useful, for instance, in recommender systems, information retrieval, targeted advertising, and psychological studies. A common and intuitively appealing way to model preferences is via the *ideal point model*, which supposes preference for a particular item varies inversely with Euclidean distance in a feature space. We assume that the items to be rated are represented by points p_i and q_i in an n-dimensional Euclidean space. A user's preference is modeled as an additional point x in this space (called the individual's "ideal point"). This represents a hypothetical "perfect" item satisfying all of the user's criteria for evaluating items.

Using response data consisting of paired comparisons between items (e.g., "user x prefers item p_i to item q_i") is a natural approach when dealing with human subjects since it avoids requiring people to assign precise numerical scores to different items (which is generally a quite difficult task, especially when preferences may depend on multiple factors). In contrast, human subjects often find pairwise judgements much easier to make. Data consisting of paired comparisons is often generated implicitly in contexts where the user has the option to act on two (or more) alternatives; for instance they may choose to watch a particular movie, or click a particular advertisement, out of those displayed to them. In such contexts, the "true distances" in the ideal point model's preference space are generally inaccessible directly, but it is nevertheless still possible to obtain an estimate of a user's ideal point.

1.2.2 Main results

The fundamental question which interests us here is how many comparisons we need (and how should we choose them) to estimate x to a desired degree of accuracy. Thus, we consider the case where we are given an existing embedding of the items (as in a mature recommender system) and focus on the on-line problem of locating a single new user from their feedback (consisting of binary data generated from paired comparisons). The item embedding could be generated using various methods, such as multidimensional scaling applied to a set of item features, or even using the results of previous paired comparisons via an approach we describe below. Given such an embedding of ℓ items, there are a total of $\binom{\ell}{2} = \Theta(\ell^2)$ possible paired comparisons. Clearly, in a system with thousands (or more) items, it will be prohibitive to acquire this many comparisons as a typical user will likely only provide comparisons for a handful of items. Fortunately, in general we can expect that many, if not most, of the possible comparisons are actually redundant. For example, of the comparisons illustrated in Figure 1, all but four are redundant and – at least in the absence of noise – add no additional information.

Any precise answer to this question would depend on the underlying geometry of the item embedding. Each comparison essentially divides $\mathbb{R}^n$ in two, indicating on which side of a hyperplane x lies, and some arrangements of hyperplanes will yield better tessellations of the preference space than others. Thus, to gain some intuition on this problem without reference to the geometry of a particular embedding, we will instead consider a probabilistic model where the items are generated at random from a particular distribution. In this case we show in [6] (and the preliminary work of [5]) that under certain natural assumptions on the distribution, it is possible to estimate the

location of any x to within an error of ϵ using a number of comparisons which, up to log factors, is proportional to n/ϵ. This is essentially optimal, so that no set of comparisons can provide a uniform guarantee with significantly fewer comparisons. We also demonstrate several stability and robustness guarantees for various settings in which the comparisons are subject to noise or errors in [6] (and the preliminary work of [4]). We also develop a simple extension to an *adaptive* scheme where we adaptively select the comparisons (manifested here in adaptively altering the mean and variance of the distribution generating the items) to substantially reduce the required number of comparisons, described in [6]. Finally, in [11] we propose practical methods for learning the embedding of the items using only the comparisons themselves by considering the setting where we have many users whose ideal points (the x's) are known and studying how to estimate the item embeddings. Combining these approaches, we can iteratively learn the full embedding using only paired comparisons.

1.3 Adaptive sampling for estimation and learning

1.3.1 Problem statement

Suppose that we wish to estimate a sparse vector from a small number of noisy linear measurements. In the setting where the measurements are selected in advance (independently of the signal) we now have a rich understanding of both practical algorithms and the theoretical limits on the performance of these algorithms. A typical result from this literature states that for a suitable measurement design, one can estimate a sparse vector with an accuracy that matches the minimax lower bound up to a constant factor. Such results have had a tremendous impact in a variety of practical settings. In particular, they provide the mathematical foundation for compressive sensing, a paradigm for efficient sampling that has inspired a range of new sensor designs over the last decade.

A distinguishing feature of the standard compressive sensing paradigm is that the measurements are *nonadaptive*, meaning that a fixed set of measurements are designed and acquired without allowing for any possibility of adapting as the measurements begin to reveal the structure of the signal. While this can be attractive in the sense that it enables simpler hardware design, in the context of sparse estimation this also leads to some clear drawbacks. In particular, this would mean that even once the acquired measurements show us that portions of the signal are very likely to be zero, we may still expend significant effort in "measuring" these zeros! In such a case, by *adaptively* choosing the measurements, dramatic improvements may be possible.

Inspired by this potential, recent investigations have shown that we can often acquire a sparse (or compressible) signal via far fewer measurements or far more accurately if we choose them adaptively. This body of work demonstrates that adaptive sensing indeed offers the potential for dramatic improvements over nonadaptive sensing in many settings. However, the existing approaches to adaptive sensing, which rely on being able to acquire *arbitrary* linear measurements, cannot be applied in most real-world applications where the measurements must respect certain physical *constraints*. Our focus in this thrust was on *constrained adaptive sensing or sampling*, where our measurements are restricted to be chosen from a particular set of allowable measurements/samples. To address such problems within this more restrictive setting, new algorithms and theoretical analysis were required.

1.3.2 Main results

Our main results on this thrust consist of a detailed study of two main problem settings. The first consists of adaptively selecting measurements from a fixed set of candidates, with the goal being to estimate a sparse vector as accurately as possible (when these measurements are corrupted by noise). In [2] (and the preliminary work of [1]) we provide a theoretical analysis of this problem for some sample cases. Our analysis demonstrates that there is a strong dependence between the underlying structure of the measurement ensemble and the sparsity model. In some cases, such as choosing Fourier measurements to estimate a sparse signal, we demonstrate that the benefits of adaptivity are *much* less significant than is the case for more general measurements, where dramatic improvements are possible. However, we also show that in other contexts, significant improvement remains possible. We demonstrate this empirically in [2] by proposing a convex algorithm for selecting measurements and then demonstrating its benefits in a stylized MRI application. (Note that an analysis of this algorithm was subsequently provided by applying novel results related to graph sparsification which demonstrated that this algorithm is near optimal.)

Out second thrust consists of the problem where we would like to solve any kind of supervised learning problem, but where the data lives on a low-dimensional manifold as opposed to being sparse. In [12] (and in the preliminary work of [13]), we developed several novel algorithms for doing this by studying various eigenvalue properties of the

alignment matrix that underlies many common manifold learning techniques. Via extensive empirical comparisons, we showed that our proposed algorithms achieve state-of-the-art performance on a wide variety of real-world data sets.

1.4 Dynamic low-rank matrix recovery

1.4.1 Problem statement

While a rich literature now exists that studies the problem of when and how we can recover a low-rank matrix from a small number of samples or measurements (summarized in [3]), these approaches are often limited in practice by the fact that the underlying matrix may be dynamically evolving over time. For instance, in a recommendation system a persons's preferences may gradually change, and in a personalized learning system a student's skills should hopefully be (perhaps not so gradually) improving over time. To address such scenarios, we consider a simple first model for how a low-rank matrix might be changing over time during the measurement process. For simplicity we will model this through the following discrete dynamic process: at time t, we have a low-rank matrix $X^t \in \mathbb{R}^{n_1 \times n_2}$ with rank r, which we assume is related to the matrix at previous time-steps via

$$X^t = f(X^1, \ldots, X^{t-1}) + \epsilon^t,$$

where ϵ^t represents noise. Then we observe each X^t through a linear operator $\mathcal{A}^t : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{m^t}$,

$$y^t = \mathcal{A}^t(X^t) + z^t, \quad y^t, z^t \in \mathbb{R}^{m^t},$$

where z^t is measurement noise. In our problem we will suppose that we observe up to d time steps, and our goal is to recover $\{X^t\}_{t=1}^d$ jointly from $\{y^t\}_{t=1}^d$.

The above model is sufficiently flexible to incorporate a wide variety of dynamics, but we will make several simplifications. First, we note that we can impose the low-rank constraint explicitly by factorizing X^t as $X^t = U^t \left(V^t\right)^T, U^t \in \mathbb{R}^{n_1 \times r}, V^t \in \mathbb{R}^{n_2 \times r}$. In general, both U^t and V^t may be changing over time. However, in some applications, it is reasonable to assume that only one set of factors is changing. For example, in a recommendation system where our matrix represents user preferences, if the rows correspond to items and the columns correspond to users, then U^t contains the latent properties of the items and V^t models the latent preferences of the users. In this context it is reasonable to assume that only V^t changes over time, and that there is a fixed matrix U (which we may assume to be orthonormal) such that we can write $X^t = UV^t$ for all t. Similar arguments can be made in a variety of other applications, including personalized learning systems, blind signal separation, and more.

Second, we assume a Markov property on f, so that X^t (or equivalently, V^t) only depends on the previous X^{t-1} (or V^{t-1}). Furthermore, although other dynamic models could be accommodated, for the sake of simplicity in our analysis we consider the simple model on V^t where

$$V^t = V^{t-1} + \epsilon^t, \quad t = 2, \ldots, d.$$

We will also assume that both ϵ^t and the measurement noise z^t are i.i.d. zero-mean Gaussian noise. The main thrust of this effort is to develop algorithms that can provably exploit this dynamic structure to achieve improved guarantees than what would be possible by, for instance, simply recovering each V^t independently.

1.4.2 Main results

To simplify our discussion, we will assume that our goal is to recover the matrix at the most recent time-step, i.e., we wish to estimate X^d from $\{y^t\}_{t=1}^d$. Our general approach can be stated as follows. In [14] we proposed a novel estimator (denoted LOWEMS for *locally weighted matrix smoothing*) given by the following optimization program:

$$\hat{X}^d = \underset{X \in \mathbb{C}(r)}{\arg\min} \ \frac{1}{2} \sum_{t=1}^{d} w_t \left\| \mathcal{A}^t(X) - y^t \right\|_2^2,$$

where $\mathbb{C}(r) = \{X \in \mathbb{R}^{n_1 \times n_2} : \text{rank}(X) \leq r\}$, and $\{w_t\}_{t=1}^d$ are non-negative weights. We further assume $\sum_{t=1}^d w_t = 1$ to avoid ambiguity.

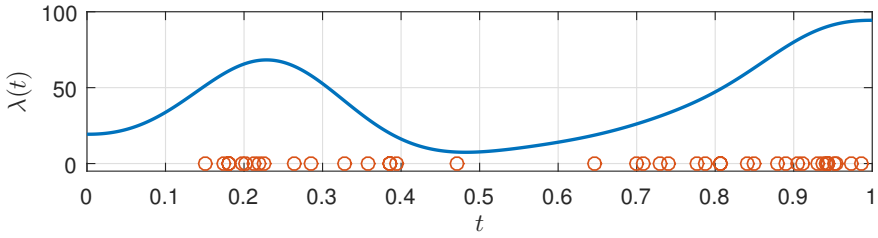

Figure 2: Example realization of a point process. We observe event locations (circles) but their empirical density provides only a sketch of the underlying intensity (line).

In [14] we provide theoretical guarantees on the performance of this estimator for two common choices for the measurement operators $\mathcal{A}^t$: the Gaussian measurement ensemble and the matrix sampling ensemble (which corresponds to the matrix completion setting). In both cases, we provide the first theoretical guarantees for the dynamic setting described above. We further demonstrated practical benefits in the context of a (Netflix) recommendation system. The upshot of these results is that we provide guarantees that provide clear practical guidance in how to select the weights w_t in the LOWEMS estimator so as to optimally exploit observations from previous time samples. This generalizes previous results for matrix completion/recovery in the static setting.

Additional empirical observations for this problem are described in [16] and [15]. In [16] we study a range of different estimators aside from the LOWEMS estimator described above. Some lead to improved performance, but at severe computational costs, with a compromise approach dubbed *simultaneous LOWEMS* achieving most of the improved performance benefits at an only slightly increased computational cost (albeit, without the theoretical support available for LOWEMS). In [15] we study an extension of the LOWEMS estimator to the case where our observations are binary quantized outcomes. Our proposed algorithm leads to moderately improved performance in a practical application to personalized learning systems (a matrix completion problem where the data consists of incomplete student responses, graded with correct/incorrect outcomes, on a set of homework problems). In the future we plan to continue to study this setting. We hypothesize that the random walk model studied here is not particularly well-suited to this application, and plan to explore the use of methods that aim to learn more accurate models of the underlying dynamics from the data.

1.5 Learning from point processes

1.5.1 Problem statement

The traditional sensing paradigm in most of signal processing and machine learning is based upon the assumption that we can begin by obtaining regularly-spaced (or potentially irregularly-spaced) samples of the phenomenon of interest. Ideally, samples are taken sufficiently densely so that no detail falls below the observation resolution. However, for many phenomenon, when viewed at an appropriate (and often the most appropriate) scale, it is much more natural to model the data as a set of discrete events occuring at specific times. For example, when studying brain activity at a gross scale (as in functional MRI) we can think of regularly sampling the "activity level" in each "voxel". However, at finer resolutions, a more natural way to model the data is as a spatio-temporal process of neurons located at points in space and firing at specific moments in time. Similar considerations arise in many other physical phenomena (e.g., modeling of earthquakes, lightning strikes, etc.), but these are perhaps even more present in various social systems/phenomena. For example, the most natural way of modeling a social network consist of individual users participating in discrete interactions at specific times. Analyzing such data requires a different set of tools. We model these event-based interactions using point processes, focusing primarily on inference using Poisson processes and Hawkes processes.

Poisson processes work on the principle of independence between events and forms a useful model for a large range of event-based applications including nuclear imaging, fluoroscopy, mass spectrometry, low-light imaging, and medical image processing, just to name a few. We have specifically explored their application in the context of time-of-flight mass spectrometry in [10]. The simplicity of Poisson models makes them the preferred choice over other point processes when their independence assumption can be (adequately) met.

Unlike Poisson processes, Hawkes processes are autoregressive in that actions can create reactions. This allows for feedback within the system. Hawkes processes can, for example, be used to model the relationship between earthquakes and aftershocks. Other applications of Hawkes processes include modeling social networks, communication

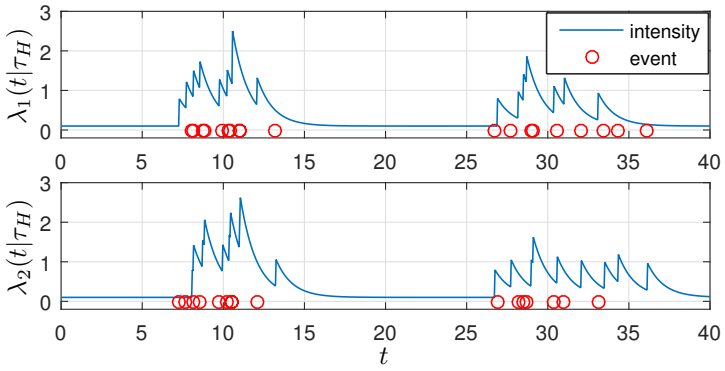

Figure 3: Intensity functions and events for a multivariate Hawkes process with two subprocesses.

networks, (biological) neural networks, financial transactions, ecological systems, and sociological patterns.

In the cases of both Poisson processes and Hawkes processes, there is surprisingly little that is known regarding the accuracy with which one can estimate the underlying parameters under certain natural linear modelling assumptions. Below we describe our results in these two contexts.

1.5.2 Main results

A Poisson process is defined by an intensity function λ that maps the observation domain $\mathbb{T}$ to a density in $\mathbb{R}_+$. Higher density regions are likely to have more events, as seen in Figure 2. When λ lies in the linear span of some dictionary, we established guarantees regarding the accuracy with which the dictionary coefficients $x \in \mathbb{R}^n$ are estimated. Our results analyze the case where $\mathbb{T}$ is some uncountable set, while existing results were largely limited to countable $\mathbb{T}$. Under mild assumptions and using (possibly unnecessary) regularization, we showed in [8, 9] that the estimated maximum-likelihood basis weights $\widehat{x}$ for a dictionary with Gram matrix Γ approximate the true weights x according to the ℓ_2 norm bound

$$\|\widehat{x} - x\|_2 \leq c\|\Gamma^{-1}\|_2\sqrt{\|\Gamma\|_*\|\lambda\|_\infty}$$

with high probability. Operator $\|\Gamma^{-1}\|_2$ represent the spectral norm of Γ^{-1} and $\|\Gamma\|_*$ represents the nuclear norm of Γ, while $\|\lambda\|_\infty = \sup_{t\in\mathbb{T}} \lambda(t)$. We also showed that improved results are possible when the weights are known to be sparse [8].

A (multivariate) Hawkes process is composed of N subprocesses (each a point process on its own) that can potentially influence each other. The set of events attributed to subprocess j is S_j and the conditional intensity function for subprocess i, given the history of event times τ_H, is

$$\lambda_i(t|\tau_H) = \mu_i(t) + \sum_{j=1}^{N} A_{ij} \sum_{k\in S_j} \phi(t - \tau_k).$$

The base intensity $\mu_i(t)$ describes a Poisson process while the excitation weights A_{ij} and kernel $\phi(t)$ are responsible for the autoregressive behavior of the process. An example of a two-subprocess Hawkes process with mutual excitation is shown in Figure 3.

It is common to estimate the weights A_{ij} of a Hawkes process to learn the connectivity of an event-driven network. We explored the use and effectiveness of Hawkes processes in this context via an empirical study in [7]. We have continued to explore this an in unpublished results have established theoretical lower bounds and asymptotic guarantees regarding the accuracy with which the A_{ij} can be estimated. Such results represent an important step in understanding the behavior of Hawkes processes and, in particular, placing the appropriate degree of confidence in Hawkes models learned from data. We expect these results to be formally published soon.

2 Personnel

The following individuals have performed research supported in whole or part by the Air Force Office of Scientific Research under grant FA9550-14-0342.

1. Prof. Mark A. Davenport, Associate Professor of Electrical & Computer Engineering, Georgia Tech
2. Hongteng Xu, Ph.D. student (Graduated 2017, now Postdoctoral Researcher at Duke University)
3. Andrew Massimino, Ph.D. student (Ph.D. in progress, expected 2018)
4. Michael Moore, Ph.D. student (Ph.D. in progress, expected 2018)
5. Liangbei Xu, Ph.D. student (Ph.D. in progress, expected 2019)
6. Matthew O'Shaughnessy, Undergraduate student (Graduated 2016, now Ph.D. student at Georgia Tech)

3 Publications

The following publications were prepared in whole or part with the support of the Air Force Office of Scientific Research under grant FA9550-14-0342.

[1] M. A. Davenport, A. K. Massimino, D. Needell, and T. Woolf. Constrained adaptive sensing. In *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, Cambridge, United Kingdom, July 2015.

[2] M. A. Davenport, A. K. Massimino, D. Needell, and T. Woolf. Constrained adaptive sensing. *IEEE Trans. on Signal Processing*, 64(20):5437–5449, October 2016.

[3] M. A. Davenport and J. Romberg. An overview of low-rank matrix recovery from incomplete observations. *IEEE J. of Selected Topics in Signal Processing*, 10(4):608–622, June 2016.

[4] A. K. Massimino and M. A. Davenport. Binary stable embedding via paired comparisons. In *Proc. IEEE Work. on Statistical Signal Processing (SSP)*, Palma de Mallorca, Spain, June 2016.

[5] A. K. Massimino and M. A. Davenport. The geometry of random paired comparisons. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, New Orleans, Louisiana, March 2017.

[6] A. K. Massimino and M. A. Davenport. As you like it: Localization via paired comparisons. *To be submitted to J. Machine Learning Research*, 2018.

[7] M. G. Moore and M. A. Davenport. Learning network structure via Hawkes processes. In *Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, Cambridge, United Kingdom, July 2015.

[8] M. G. Moore and M. A. Davenport. Sparse parametric estimation of poisson processes. In *Proc. Work. on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, Lisbon, Portugal, June 2017.

[9] M. G. Moore and M. A. Davenport. Estimation of Poisson arrival processes under linear models. *To be submitted to IEEE Trans. on Information Theory*, 2018.

[10] M. G. Moore, A. K. Massimino, and M. A. Davenport. Randomized multi-pulse time-of-flight mass spectrometry. In *Proc. IEEE Int. Work. on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Cancun, Mexico, December 2015.

[11] M. R. O'Shaughnessy and M. A. Davenport. Localizing users and items from paired comparisons. In *Proc. IEEE Int. Work. on Machine Learning for Signal Processing (MLSP)*, Vietri sul Mare, Salerno, Italy, September 2016.

[12] H. Xu, L. Yu, M. A. Davenport, and H. Zha. A unified framework for manifold landmarking. *Submitted to IEEE Trans. on Signal Processing*, December 2017.

[13] H. Xu, H. Zha, and M. A. Davenport. Active manifold learning via Gershgorin circle guided sample selection. In *Proc. AAAI Conf. on Artificial Intelligence (AAAI-15)*, Austin, Texas, January 2015.

[14] L. Xu and M. A. Davenport. Dynamic matrix recovery from incomplete observations under an exact low-rank constraint. In *Proc. Advances in Neural Information Processing Systems (NIPS)*, Barcelona, Spain, December 2016.

[15] L. Xu and M. A. Davenport. Dynamic one-bit matrix completion. In *Proc. Work. on Signal Processing with Adaptive Sparse Structured Representations (SPARS)*, Lisbon, Portugal, June 2017.

[16] L. Xu and M. A. Davenport. Simultaneous recovery of a series of low-rank matrices by locally weighted matrix smoothing. In *Proc. IEEE Int. Work. on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Curaçao, Dutch Antilles, December 2017.

4 Interaction and Transitions

Prof. Mark A. Davenport presented research supported in whole or part by the Air Force Office of Scientific Research under grant FA9550-14-0342 at the following venues.

Invited presentations

1. "Low-rank matrix recovery from incomplete observations: An overview and some recent advances," Alan Turing Institute, London, United Kingdom, June 2017.
2. "Localization via paired comparisons and nonmetric multidimensional scaling," Department of Electrical and Computer Engineering, University of Washington, Seattle, Washington, August 2016.
3. "Localization and nonmetric multidimensional scaling via paired comparisons," Simons Institute Workshop on Real-Time Decision Making, Berkeley, California, June 2016.
4. "Localization from paired comparisons," Conference on Information Sciences and Systems (CISS), Princeton, New Jersey, March 2016.
5. "1-bit matrix completion," Information and Inference Best Paper Prize Meeting, Oxford, United Kingdom, August 2015.
6. "Constrained adaptive sensing," ISI World Statistics Congress, Rio de Janiero, Brazil, July 2015.
7. "Localization via paired comparisons," ShanghaiTech Symposium on Data Science, Shanghai, China, June 2015.
8. "Adaptive sensing for sparse images," Lorentz Center Workshop on Transformations in Optics, Leiden, The Netherlands, May 2015.
9. "Compressive sensing," Lorentz Center Workshop on Transformations in Optics, Leiden, The Netherlands, May 2015.
10. "Localization via paired comparisons," Department of Mathematics, University of Georgia, Athens, Georgia, March 2015.
11. "Localization via paired comparisons," Department of Statistics, University of Wisconsin, Madison, Wisconsin, February 2015.
12. "Learning from pairwise comparisons," Information Theory and Applications Workshop (ITA), San Diego, California, February 2015.
13. "Matrix recovery from coarse observations," Foundations of Computational Mathematics (FoCM): Workshop on Computational Harmonic Analysis, Image, and Signal Processing, Montevideo, Uruguay, December 2014.
14. "On the power and limits of adaptivity for sparse signal acquisition," Georgia Institute of Technology School of Industrial and Systems Engineering Statistics Seminar, Atlanta, Georgia, October 2014.

Short courses and symposia

1. Symposium on "Low-dimensional dynamical systems in signal processing and data analysis," at IEEE Int. Work. on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2017.
2. Special session on "Signal processing and adaptive systems," at Asilomar Conf. on Signals, Systems, and Computers, 2017.
3. "Low-rank matrix completion: An overview and some recent advances," Minitutorial on Compressed Sensing/Dimensionality Reduction, SIAM Annual Meeting, Pittsburgh, Pennsylvania, July 2017. (Part of a series of lectures co-organized with Deanna Needell and Jeff Blanchard)

4. Symposium on "Randomness and efficient computation in signal processing," at IEEE Int. Work. on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2015.

5. Mini-symposium on "Finding and exploiting structure in data," at SIAM Conf. on Applied Linear Algebra, 2015.