Applications of Lindeberg Principle in Communications and Statistical Learning

Satish Babu Korada* and Andrea Montanari*,†

Abstract

We use a generalization of the Lindeberg principle developed by Sourav Chatterjee to prove universality properties for various problems in communications, statistical learning and random matrix theory. We also show that these systems can be viewed as the limiting case of a properly defined sparse system. The latter result is useful when the sparse systems are easier to analyze than their dense counterparts. The list of problems we consider is by no means exhaustive. We believe that the ideas can be used in many other problems relevant for information theory.

1 Introduction

The phenomenon of universality is common to many disciplines of science and engineering. A well known example is the central limit theorem which, in a simple version, says the following. Let $\{X_i\}_{i\geq 1}$ be a collection of i.i.d. random variables with mean zero and variance $\mathbb{E}[X_i^2] = 1$. Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{\mathrm{d}} \mathsf{N}(0,1),$$

where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution as $n \to \infty$, and $\mathsf{N}(0,1)$ is a Gaussian random variable with mean zero and variance one. In particular, the central limit theorem implies that the distribution of $n^{-1/2}\sum_{i=1}^{n} X_i$ is asymptotically independent of the details of the distribution of the summands X_i. In other words, its limit is "universal" for a large class of summands' distributions. Other examples include the limiting spectrum of random matrices [2], and various properties of statistical mechanics models [3].

Examples in communications theory where universal properties have been established include the MIMO communications problem [4]. In these problems it was shown the capacity of the system is independent of the distribution of the fading coefficients and the spreading sequences respectively.

A different research area in which universality ideas appear ubiquitous is compressed sensing. Donoho and Tanner [5] carried out a systematic empirical investigation of universality in this context. In particular they showed that the phase transition boundary in the sparsity-undersampling tradeoff is universal for a large class of sensing matrices. The precise location of this phase transition was determined earlier on in the case of Gaussian sensing matrices [6].

*Department of Electrical Engineering, Stanford University

†Department of Statistics, Stanford University

A related phenomenon which we study here is the sparse-dense equivalence. As an example consider a uniformly random regular graph G_n of degree d over n vertices. Let $A_n \in \mathbb{R}^{n\times n}$ be the symmetric matrix whose non-vanishing entries correspond to edges in G_n and take values in $\{+1/\sqrt{d}, -1/\sqrt{d}\}$ independently and uniformly at random. As $n \to \infty$ the spectral measure of such a matrix converges almost surely [7, 8] to a well defined limit $\rho_d(\mathrm{d}\lambda)$ supported on $[-2\sqrt{1-1/d}, 2\sqrt{1-1/d}]$, where:

$$\rho_d(\mathrm{d}\lambda) = \frac{1}{2\pi}\,\frac{\sqrt{4(1-1/d)-\lambda^2}}{1-\lambda^2/d}\,\mathrm{d}\lambda\,. \tag{1}$$

If we now consider the $d \to \infty$ limit, this distribution converges weakly to the celebrated semi-circle law

$$\rho_\infty(\mathrm{d}\lambda) = \frac{1}{2\pi}\,\sqrt{4-\lambda^2}\,\mathrm{d}\lambda\,. \tag{2}$$

This is the limiting spectrum of the standard (dense) Wigner matrices. We refer to this type of property as to a sparse-dense equivalence. Showing such a relationship can be particularly useful when the analysis of the sparse system is easier than its dense counterpart. Specific examples will be provided below.

Universality and sparse-dense equivalence can have far reaching consequences in communications and information theory. In this paper, we demonstrate this by studying both phenomena within a common framework, and obtaining new results in each of the above mentioned problems. The main tool that we use is the following generalization of Lindeberg's principle that was proved in [1].

1.1 Lindeberg Principle

Given $f : \mathbb{R}^n \to \mathbb{R}$, the generalized Lindeberg principle provides conditions under which the distribution of $f(X_1, \dots, X_n)$ is approximately insensitive to the distribution of its arguments $X_1, \dots, X_n$ which are assumed to be independent. This generalizes the classical Lindeberg proof of the central limit theorem, that focused on $f(x_1, \dots, x_n) = (x_1 + \cdots + x_n)/\sqrt{n}$.

Let us restate here the main result of [1].

Theorem 1 (Generalized Lindeberg Principle, [1])**.** *Let $U = (U_1, \dots, U_n)$ and $V = (V_1, \dots, V_n)$ be two random vectors with mutually independent components. For $1 \le i \le n$, define*

$$a_i \equiv |\mathbb{E}[U_i] - \mathbb{E}[V_i]|,$$
$$b_i = |\mathbb{E}[U_i^2] - \mathbb{E}[V_i^2]|.$$

and further assume $\max_i(\mathbb{E}\{|U_i|^3\} + \mathbb{E}\{|V_i|^3\}) \le M_3$. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a thrice continuously differentiable function, and for $r = 1, 2, 3$, let $L_r(f)$ be a finite constant such that $|\partial_i^r f(u)| \le L_r(f)$ for each i and $u \in \mathbb{R}^n$, where ∂_i^r denotes the r-fold derivative in the ith coordinate. Then

$$|\mathbb{E}[f(U)] - \mathbb{E}[f(V)]| \le \sum_{i=1}^{n}(a_i L_1(f) + \frac{1}{2}b_i L_2(f)) + \frac{1}{6}nL_3(f)M_3.$$

Notice that, while this theorem explicitly bounds the change in expectation $f(\,\cdot\,)$, it gives control on its distribution as well, by applying it to $g(f(\,\cdot\,))$, for $g : \mathbb{R} \to \mathbb{R}$ belonging to a suitable class of test functions.

In many problems of interest for this paper, the bound on the derivatives of f required by the last theorem does not hold, and a more careful analysis is needed. For that purpose we use the following theorem. The proof is analogous to the one of Theorem 1, and is provided in Section 3.

Theorem 2. *Let $U = (U_1, \ldots, U_n)$ and $V = (V_1, \ldots, V_n)$ be two random vectors with mutually independent components. Let $\{a_i\}_{1\le i\le n}$ and $\{b_i\}_{1\le i\le n}$ be as defined in Theorem 1. Then*

$$\begin{aligned}|\mathbb{E}[f(U)] - \mathbb{E}[f(V)]| \le \sum_{i=1}^{n} \Big\{ & a_i \mathbb{E}[|\partial_i f(U_1^{i-1}, 0, V_{i+1}^n)|] + \frac{1}{2} b_i \mathbb{E}[|\partial_i^2 f(U_1^{i-1}, 0, V_{i+1}^n)|] \\ & + \frac{1}{2}\mathbb{E}\int_0^{U_i} \big[|\partial_i^3 f(U_1^{i-1}, s, V_{i+1}^n)|\big](U_i - s)^2 \, \mathrm{d}s \\ & + \frac{1}{2}\mathbb{E}\int_0^{V_i} \big[|\partial_i^3 f(U_1^{i-1}, s, V_{i+1}^n)|\big](V_i - s)^2 \, \mathrm{d}s \Big\}.\end{aligned}$$

2 Applications

In this section we discuss the application of Theorem 2 to a problem from communications theory (code division multiple access channels), and one from statistical learning theory (estimation via LASSO). We also revisit a standard model from statistical mechanics (the Sherrington-Kirkpatrick model), and the spectrum of Wishart matrices, which is related to capacity of MIMO channels. In each of these cases, Theorem 2 implies both universality and sparse-dense equivalence results. We will not try to be exhaustive, but rather to point out some selected conclusion. This Section contains definitions and statements, while proofs are deferred to section 4.

In the following, we use uppercase letters, e.g, X, Y, to denote random variables and their lowercase counterparts, e.g. x, y, to denote realizations of such random variables. We also use boldface characters to denote random matrices, with the subscript to indicate their dimension, e.g. $\mathbf{A}_n$, $\mathbf{B}_n$.

Most of our results concern random matrices with i.i.d. entries and apply under some simple centering and normalization conditions, provided the entries have finite sixth moment. Rather than repeating these conditions at each of the results below, we introduce them once and for all.

Definition 1 (Random Matrices of Standard Type)**.** *Let $\mathbf{A}_n = \{A_{ij}\}_{1\le i\le m, 1\le j\le n}$ be a sequence of random matrices indexed by their dimensions m and n (with $m = m_n$ an appropriate sequence of integers). We say that $\mathbf{A}_n$ is a random matrix of* standard type *if the entries $\{A_{ij}\}_{i,j\ge 1}$ form an array of independent and identically distributed random variables with $\mathbb{E}[A_{ij}] = 0$, $\mathbb{E}[A_{ij}^2] = 1$ and $\mathbb{E}[A_{ij}^6] \le K < \infty$, for some K independent of m, n.*

2.1 Capacity of a CDMA System

Code Division Multiple Access (CDMA) is a widely used communication system between multiple users and a common receiver [9]. The scheme consists of n users modulating their information sequence by a signature sequence (spreading sequence) of length m and transmitting the resulting signal. The number m is sometimes referred to as the spreading gain or the number of chips per sequence. The receiver obtains the sum of all transmitted signals and the noise which is often assumed to be white and Gaussian (AWGN).

For the sake of simplicity, we will assume antipodal signals: each user wishes to communicate a symbol $X_k \in \{+1, -1\}$, to the common receiver. User k uses a signature sequence $(A_{1k}, \ldots, A_{mk})$, with $A_{ik} \in \mathbb{R}$. The received signal Y_i in the i-th time interval is given by

$$Y_i = \sum_{k=1}^{n} A_{ik} X_k + \sigma Z_i \,,$$

where Z_i are i.i.d. copies of $\mathsf{N}(0,1)$ and therefore the noise power is σ^2.

We use $x^{\mathsf{in}} = (x_1, \dots, x_n)$ to denote any specific realization of the transmitted symbols, and will assume that a realization of such symbols is used uniformly at random. The corresponding random vector is $X^{\mathsf{in}} = (X_1, \dots, X_n)$ while $Y = (Y_1, \dots, Y_m)$ is the received signal. Typically X^{in} is chosen to be uniformly distributed over $\{+1, -1\}^n$. In this paper we restrict to this case. However it is possible to generalize the results below to a large class of distributions for the symbol X_i.

We write A_n for the $m \times n$ matrix $\{A_{ik}\}_{1 \le i \le m, 1 \le j \le n}$. Let $C_n(\mathbf{A}_n)$ denote the capacity of such system, i.e. the number of bits per user that can be reliably transmitted to the common receiver under the above constraints. Explicitly we have

$$C_n(A_n) = \log 2 - \frac{1}{2}\alpha - \frac{1}{n}\mathbb{E}_Y \log \Big\{ \sum_{x \in \{+1,-1\}^n} e^{-\frac{1}{2\sigma^2}\|Y - A_n x\|_2^2} \Big\} . \tag{3}$$

Here expectation $\mathbb{E}_Y$ is taken over the received signal.

Random spreading sequences were initially considered in [10]. Here, the signature sequences are modeled as random vectors with i.i.d. components $\{A_{ik}\}_{1 \le i \le m, 1 \le k \le n}$. Without loss of generality we can assume $\mathbb{E}\{A_{ik}\} = 0$ and $\mathbb{E}\{A_{ik}^2\} = 1$. We will be interested in the large system limit $m, n \to \infty$ with $\alpha = m/n$ fixed.

In order to keep the average power (per symbol) equal to 1, we will rescale the signature matrix by a factor $1/n$. For a random signature matrix $\mathbf{A}_n$, we consider therefore the capacity $C_n(m^{-1/2}\mathbf{A}_n)$, which is itself random. As proved in [11], $C_n(m^{-1/2}\mathbf{A}_n)$ does in fact concentrate exponentially around its expectation. This motivates us to focus on its expectation.

Theorem 3 (Universality of the Capacity of random CDMA sytems)**.** *Let $\mathbf{A}_n = \{A_{ij}\}_{1 \le i \le m, 1 \le j \le n}$ and $\mathbf{B}_n = \{B_{ij}\}_{1 \le i \le m, 1 \le j \le n}$ denote two $m \times n$ dimensional random spreading matrices of standard type. Then*

$$\lim_{n \to \infty, m = n\alpha} \big\{ \mathbb{E}[C_n(m^{-1/2}\mathbf{A}_n)] - \mathbb{E}[C_n(m^{-1/2}\mathbf{B}_n)] \big\} = 0 .$$

The above theorem establishes that the per-user capacity of a CDMA channel is asymptotically independent of the distribution of the spreading sequences. The conditions required to be satisfied by the distributions are milder than the ones imposed in [11].

Our next result concerns the sparse-dense equivalence. Sparse signature schemes were proposed in [12] both as a tool for simplifying mathematical analysis and as a design option with potential practical advantages. Given a signature matrix $\mathbf{A} = \{A_{ij}\}_{1 \le i \le m, 1 \le j \le n}$ defined as above, its sparsification $\mathbf{A}_n^\gamma$ is given by

$$A_{ij}^\gamma = \begin{cases} A_{ij} & \text{with probability } \gamma/n, \\ 0 & \text{with probability } 1 - \gamma/n, \end{cases} \tag{4}$$

with $\gamma > 0$ a design parameter that is kept fixed in the large system limit. Under a sparse signature scheme, the power per symbol is normalized to 1 if we rescale the signatures by a factor $1/\sqrt{\gamma}$. The channel output is therefore

$$Y_i = \frac{1}{\sqrt{\gamma}} \sum_{k=1}^{n} A_{ik}^\gamma X_k + \sigma Z_i .$$

We can then prove the following sparse-dense equivalence result.

Theorem 4 (Sparse-Dense Equivalence for CDMA channels)**.** *Let* $\mathbf{A}_n = \{A_{ij}\}_{1\le i\le m, 1\le j\le n}$ *and* $\mathbf{B}_n = \{B_{ij}\}_{1\le i\le m, 1\le j\le n}$ *denote two* $m \times n$ *dimensional random spreading matrices of standard type. For* $\gamma > 0$*, let* $\mathbf{A}_n^\gamma$ *be the sparsification of* $\mathbf{A}_n$*. Then*

$$\lim_{\gamma\to\infty} \lim_{n\to\infty, m=n\alpha} \left\{ \mathbb{E}[C_n(\gamma^{-1/2}\mathbf{A}_n^\gamma)] - \mathbb{E}[C_n(n^{-1/2}\mathbf{B}_n)] \right\} = 0.$$

As already mentioned, establishing sparse-dense equivalence is particularly useful when the analysis of a sparse system is simpler than for its dense counterpart. In [12] it was shown that there exists $\alpha_{\rm s} > 0$ such that, for all $\alpha \le \alpha_s$,

$$\lim_{\gamma\to\infty} \lim_{n\to\infty, m=n\alpha} \mathbb{E}[C_n(\gamma^{-1/2}\mathbf{A}_n^\gamma)] = \min_{m\in[0,1]} C_{\rm RS}(q)\,, \tag{5}$$

where

$$C_{\rm RS}(q) = \frac{\lambda}{2}(1+q) - \frac{1}{2\alpha}\log\lambda\sigma^2 - \mathsf{E}_z\{\log(2\cosh(\sqrt{\lambda}Z+\lambda))\}\,, \tag{6}$$

$$\lambda \equiv \frac{1}{\sigma^2+\alpha(1-q)}\,, \tag{7}$$

where E_z denotes expectation with respect to $Z \sim \mathsf{N}(0,1)$. The parameter $\alpha_{\rm s}$ is defined as the largest α such that the maximizer in (5) is unique. Numerically $\alpha_{\rm s} \approx 1.49$. The same formula was derived earlier by Tanaka [13] using the non-rigorous replica method from statistical physics.

Combining this with Theorem 4 we can conclude the following result for the capacity of a random CDMA system.

Corollary 1 (Capacity of random CDMA systems)**.** *Let* $\mathbf{A}_n$ *denote an* $m \times n$ *dimensional random spreading matrix with i.i.d. entries. Assume* $\mathbb{E}[A_{ij}] = 0$*,* $\mathbb{E}[A_{ij}^2] = 1$ *and* $\mathbb{E}[A_{ij}^6] \le K < \infty$*. Then for* $\alpha \le \alpha_{\rm s}$

$$\lim_{n\to\infty, m=n\alpha} \mathbb{E}[C_n(m^{-1/2}\mathbf{A}_n)] = \min_{q\in[0,1]} C_{\rm RS}(q).$$

2.2 Estimation via LASSO

The LASSO (also known as basis pursuit de-noising) is a popular strategy in statistical learning, used for reconstructing high-dimensional parameter vectors from noisy measurements [14, 15]. It is particularly well suited when the underlying parameters vector is sparse in an appropriate basis. For this very reason, it is object of intense study within the compressed sensing literature.

We assume here that a signal $x_0 \in \mathbb{R}^n$ is observed through the sensing matrix A_n which has dimensions $m \times n$. The measurements $y \in \mathbb{R}^m$ are modeled as a noisy linear functions

$$y = A_n\, x_0 + z\,, \tag{8}$$

with $z \in \mathbb{R}^m$ a noise vector. Let the noise vector z be i.i.d. Gaussian vector. The recovery of x_0 from y is done using the following convex optimization problem

$$\widehat{x}(\lambda) = \mathrm{argmin}_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\|y - A_n x\|_2^2 + \lambda\,\|x\|_1 \right\}. \tag{9}$$

For some applications the sensing matrix A_n is not far from random or pseudo-random. It is important to ask to which degree results obtained for a specific distribution of A_n generalize to other

distributions [6, 5]. We consider the case in which the entries $x_{0,i}$ of x_0 are uniformly bounded, i.e., $|x_{0i}| \le x_{\max}$ for some constant $x_{\max} > 0$ independent of n, m. We further assume that the noise vector z has i.i.d. entries $z_i \sim \mathsf{N}(0, \sigma^2)$ and focus on the limit $m, n \to \infty$ with $m/n = \alpha$ fixed.

The next result provides rigorous evidence towards the broader universality picture, by proving universality for the normalized cost

$$L(A_n) = \frac{1}{n} \min_{x \in [-x_{\max}, x_{\max}]^n} \left\{ \frac{1}{2} \|y - A_n x\|_2^2 + \lambda \|x\|_1 \right\}. \tag{10}$$

Theorem 5 (Universality for LASSO)**.** *Let $\mathbf{A}_n = \{A_{ij}\}_{1 \le i \le m, 1 \le j \le n}$ and $\mathbf{B}_n = \{B_{ij}\}_{1 \le i \le m, 1 \le j \le n}$ denote two $m \times n$ dimensional random sensing matrices of standard type. Then*

$$\lim_{n \to \infty, m = n\alpha} \left\{ \mathbb{E}[L(n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[L(n^{-1/2}\mathbf{B}_n)] \right\} = 0 .$$

2.3 Spectrum of Wishart matrices and capacity of MIMO channels

Given an $n \times n$ symmetric matrix W_n, let $\{\lambda_i(W_n)\}_{1 \le i \le n}$ denote its eigenvalues. The spectral measure of W_n is the probability measure

$$\mu_n \equiv \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(W_n)} . \tag{11}$$

The study of the limit of μ_n as $n \to \infty$, for a sequence of random matrices $\mathbf{W}_n$ is a central topic in random matrix theory, with important applications in multi-antenna communications. A well-studied example is the family of Wishart matrices. Here, $\mathbf{W}_m = \frac{1}{n}\mathbf{A}_n^\top \mathbf{A}_n$, where $\mathbf{A}_n$ is an $m \times n$ matrix, whose entries are i.i.d. realizations of a zero mean random variable with variance 1.

A standard approach to characterizing the spectral measure is through its Stieltjes transform [16] which is defined as

$$S_n(\mathbf{W}_n, z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{z + \lambda_i(\mathbf{W}_n)} = \frac{1}{n} \mathrm{Tr}\big((\mathbf{W}_n + z I_n)^{-1}\big),$$

where $z \in \mathbb{C} \backslash \mathbb{R}$ and I_n is the n-dimensional identity matrix. The limiting spectrum of the family $\{\mathbf{W}_n\}_{n \ge 1}$ can be obtained by computing $\lim_{n \to \infty} S_n(\mathbf{W}_n, z)$. The universality of Wishart matrices is a well known result [2]. The following is a sparse-dense equivalence result for this class of matrices.

Theorem 6 (Sparse-Dense Equivalence for Wishart Matrices)**.** *Let $\mathbf{A}_n = \{A_{ij}\}_{1 \le i \le m, 1 \le j \le n}$ and $\mathbf{B}_n = \{B_{ij}\}_{1 \le i \le m, 1 \le j \le n}$ denote two $m \times n$ dimensional random matrices of standard type. For $\gamma > 0$, let $\mathbf{A}_n^\gamma$ be the sparsification of $\mathbf{A}_n$. Let $\mathbf{W}_{\mathbf{A},n}^\gamma \equiv \gamma^{-1}(\mathbf{A}_n^\gamma)^\top \mathbf{A}_n^\gamma$ and $\mathbf{W}_{\mathbf{B},n} \equiv n^{-1}(\mathbf{B}_n)^\top \mathbf{B}_n$. Then for all $z \in \mathbb{C} \backslash \mathbb{R}$*

$$\lim_{\gamma \to \infty} \lim_{n \to \infty, m = n\alpha} \left\{ \mathbb{E}[S_n(\mathbf{W}_{\mathbf{A},n}^\gamma, z)] - \mathbb{E}[S_n(\mathbf{W}_{\mathbf{B},n}, z)] \right\} = 0 .$$

Under appropriate tightness conditions, convergence of Stieltjes transforms implies weak convergence of the spectrum μ_n, which further implies the convergence of the empirical average $\frac{1}{n}\sum_i f(\lambda_i)$ for any continuous bounded function f. As a particular application of this remark, we consider the capacity of multi-input multi-output (MIMO) communication systems. The channel model is very

similar to the CDMA system discussed in Section 2.1. For a channel input $X = (X_1, \ldots, X_n)$, the channel output is a vector $Y = (Y_1, \ldots, Y_m)$ in $\mathbb{R}^m$, with components

$$Y_i = \sum_{k=1}^{n} H_{ik} X_k + \sigma\, Z_i$$

where Z_i are i.i.d. realizations of $\mathsf{N}(0,1)$. However, in this case it is customary to not restrict the inputs to be $\{+1,-1\}$, but rather to impose a power constraint $n^{-1}\sum_{i=1}^{n}\mathbb{E}\{X_i^2\} \leq 1$. Given a channel gains matrix $H_n = \{H_{ij}\}_{1\leq i\leq m, 1\leq j\leq n}$, the average capacity per input antenna [4] is then given by

$$C_n(H_n) = \max_{\{Q\succeq 0:\frac{1}{n}\sum_{i=1}^{n} Q_{ii}=1\}} \frac{1}{2n}\mathbb{E}\Big\{\log \mathrm{Det}\Big(\mathbf{I}_m + \frac{1}{\sigma^2} H_n Q H_n^{\top}\Big)\Big\}\,.$$

when the input covariance is Q For the case of H_{ij} being i.i.d. symmetric Gaussian random variables it was shown in [4] that the above maximum is achieved for $Q = \mathbf{I}_n$. Here, we assume that little is known about the channel gains and therefore this covariance matrix is used for other matrices H_n as well. Under this assumption, the achievable average rate is given by

$$C_n(H_n) = \frac{1}{2n}\sum_{i=1}^{m}\log\Big\{1+\frac{1}{\sigma^2}\lambda_i(H_n H_n^{\top})\Big\} = \frac{1}{2n}\sum_{i=1}^{n}\log\Big\{1+\frac{1}{\sigma^2}\lambda_i(H_n^{\top} H_n)\Big\}.$$

Under the above theorem implies the following result for the MIMO channels.

Corollary 2 (Sparse-Dense Equivalence for the MIMO Capacity)**.** *Let $\mathbf{A}_n = \{A_{ij}\}_{1\leq i\leq m,1\leq j\leq n}$ and $\mathbf{B}_n = \{B_{ij}\}_{1\leq i\leq m,1\leq j\leq n}$ denote two $m\times n$ dimensional random matrices of standard type. For $\gamma > 0$, let $\mathbf{A}_n^{\gamma}$ be the sparsification of $\mathbf{A}_n$. Then*

$$\lim_{\gamma\to\infty}\lim_{n\to\infty, m=n\alpha}\big\{\mathbb{E}[C_n(\gamma^{-1/2}\mathbf{A}_n^{\gamma})] - \mathbb{E}[C_n(n^{-1/2}\mathbf{B}_n)]\big\} = 0\,.$$

2.4 Spin glass models

Spin glass models have been object of intense interest within statistical mechanics, mathematical physics and probability theory. Both rigorous and heuristic techniques from this domain have been applied with success in information theory [17].

A number of universality and sparse-dense equivalence results have been proved in this context [1, 18, 19]. We re-derive two of these results here because they provide a very simple and instructive illustration of the proof technique that is used in the more intricate examples listed in the previous sections.

We focus in particular on the Sherrington-Kirkpatrick (SK) model. The model is defined by the Hamiltonian function $\mathcal{H} : \{+1,-1\}^n \times \mathbb{R}^{n\times n} \to \mathbb{R}$ given by

$$\mathcal{H}(x, A_n) = -\frac{1}{\sqrt{2}}\sum_{i,j=1}^{n} A_{ij}x_i x_j = -\frac{1}{\sqrt{2}}x^{\top}A_n x\,,$$

for an $n\times n$ dimensional matrix A_n and $x = (x_1,\ldots,x_n) \in \{+1,-1\}^n$. An important object of interest in this context is the free entropy density at inverse temperature β, which is defined by

$$f(\beta, A_n) \equiv \frac{1}{n}\log\Big\{\sum_{x\in\{+1,-1\}^n} e^{-\beta\mathcal{H}(x,A_n)}\Big\}.$$

Universality of the free energy for the SK model was established in [20] and was later extended to general distributions in [21]. As shown in [1] the current approach gives a stronger result.

Theorem 7 (Universality for the SK model [1])**.** *Let $\mathbf{A}_n = \{A_{ij}\}_{1\le i,j\le n}$ and $\mathbf{B}_n = \{B_{ij}\}_{1\le i,j\le n}$ be two $n \times n$ dimensional random matrices. Assume that both $\{A_{ij}\}$ and $\{B_{ij}\}$ are collections of i.i.d. random variables with $\mathbb{E}[A_{ij}] = \mathbb{E}[B_{ij}] = 0$, $\mathbb{E}[A_{ij}^2] = \mathbb{E}[B_{ij}^2] = 1$, and $\mathbb{E}[|A_{ij}|^3], \mathbb{E}[|B_{ij}|^3] \le K < \infty$. Then*

$$\lim_{n\to\infty} \left\{\mathbb{E}[f(\beta, n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[f(\beta, n^{-1/2}\mathbf{B}_n)]\right\} = 0\,.$$

The sparse-dense equivalence was proved in [19] under the slightly stronger assumption of uniformly bounded entries $|A_{ij}| \le 1$ with even distribution.

Theorem 8 (Sparse-Dense Equivalence)**.** *Let $\mathbf{A}_n = \{A_{ij}\}_{1\le i,j\le n}$ and $\mathbf{B}_n = \{B_{ij}\}_{1\le i,j\le n}$ be two $n\times n$ dimensional random matrices. Assume that both $\{A_{ij}\}$ and $\{B_{ij}\}$ are collections of i.i.d. random variables with $\mathbb{E}[A_{ij}] = \mathbb{E}[B_{ij}] = 0$, $\mathbb{E}[A_{ij}^2] = \mathbb{E}[B_{ij}^2] = 1$, and $\mathbb{E}[|A_{ij}|^3], \mathbb{E}[|B_{ij}|^3] \le K < \infty$. For $\gamma > 0$, let $\mathbf{A}_n^\gamma$ be the sparsification of $\mathbf{A}_n$. Then*

$$\lim_{\gamma\to\infty}\lim_{n\to\infty} \left\{\mathbb{E}[f(\beta, \gamma^{-1/2}\mathbf{A}_n^\gamma)] - \mathbb{E}[f(\beta, n^{-1/2}\mathbf{B}_n)]\right\} = 0\,.$$

3 Proof of Theorem 2

Proof of Theorem 2. Let $\partial_i^r f$ denote $\frac{\partial^r f}{\partial x_i^r}$. Let

$$\overline{W}_i = (U_1, \dots, U_i, V_{i+1}, \dots, V_n),$$
$$\overline{W}_i^0 = (U_1, \dots, U_{i-1}, 0, V_{i+1}, \dots, V_n).$$

Then

$$\mathbb{E}[f(U)] - \mathbb{E}[f(V)] = \sum_{i=1}^n (\mathbb{E}[f(\overline{W}_i)] - \mathbb{E}[f(\overline{W}_{i-1})])\,. \tag{12}$$

From the third-order Taylor expansion, we have

$$f(\overline{W}_i) = f(\overline{W}_i^0) + U_i\partial_i f(\overline{W}_i^0) + \frac{U_i^2}{2}\partial_i^2 f(\overline{W}_i^0) + \frac{1}{2}\int_0^{U_i} \partial_i^3 f(U_1^{i-1}, s, V_{i+1}^n)(U_i - s)^2 \mathrm{d}s\,. \tag{13}$$

Similarly, we get

$$f(\overline{W}_{i-1}) = f(\overline{W}_i^0) + V_i\partial_i f(\overline{W}_i^0) + \frac{V_i^2}{2}\partial_i^2 f(\overline{W}_i^0) + \frac{1}{2}\int_0^{V_i} \partial_i^3 f(U_1^{i-1}, s, V_{i+1}^n)(V_i - s)^2 \mathrm{d}s. \tag{14}$$

From Eq. (12), using (13) and (14), we get

$$\begin{aligned}\mathbb{E}[f(U)] - \mathbb{E}[f(V)] = \sum_{i=1}^n \Big\{&\mathbb{E}[(U_i - V_i)\partial_i f(\overline{W}_i^0)] + \frac{1}{2}\mathbb{E}[(U_i^2 - V_i^2)\partial_i^2 f(\overline{W}_i^0)] \\ &+ \mathbb{E}\big[\frac{1}{2}\int_0^{U_i} \partial_i^3 f(U_1^{i-1}, s, V_{i+1}^n)(U_i - s)^2 \mathrm{d}s\big] + \mathbb{E}\big[\frac{1}{2}\int_0^{V_i} \partial_i^3 f(U_1^{i-1}, s, V_{i+1}^n)(V_i - s)^2 \mathrm{d}s\big]\Big\}.\end{aligned}$$

The result follows by noting that $f(\overline{W}_i^0)$ is independent of $\{U_i, V_i\}$. □

4 Proofs of statements from Section 2

We will present the proofs starting from the last example, i.e. the Sherrington-Kirkpatrick model in Section 2.4. As mentioned, this is a particularly simple example of the general proof strategy.

4.1 SK Model

As mentioned in Section 2.4, the Hamiltonian for this model is given by

$$\mathcal{H}(x, A_n) = -\frac{1}{\sqrt{2}}\, x^\top A_n x,$$

where A_n is an $n \times n$ dimensional matrix. For a function $(x, A_n) \mapsto g(x, A_n)$, we denote by $\langle g(x, A_n)\rangle$ its expectation with respect to the probability distribution $p_{A_n}(x) \propto \exp\{-\beta\mathcal{H}(x, A_n)\}$ on $\{+1, -1\}^n$. Explicitly:

$$\langle g(x, A_n)\rangle = \frac{\sum_{x\in\mathcal{X}^n} g(x, A_n)\, e^{-\mathcal{H}(x,Z,A_n)}}{\sum_{x\in\mathcal{X}^n} e^{-\mathcal{H}(x,Z,A_n)}}\,. \tag{15}$$

Denote by ∂^k_{rc} the k-th partial derivative with respect to A_{rc} (row r, column c). A straightforward calculation shows that third derivative $\partial^3_{rc} f(\beta, A_n)$ is given by

$$\partial^3_{rc} f(\beta, A_n) = \frac{\beta^3}{\sqrt{2}n}\langle x_r x_r\rangle(1 - \langle x_r x_c\rangle^2)\ ,$$

which implies $L_3(f) \le \beta^3/(\sqrt{2}n)$ (with L_3 defined as in Theorem 1).

Proof of Theorem 7. From the definition of the random matrices $\mathbf{A}_n$ and $\mathbf{B}_n$, we have we have $\mathbb{E}[A_{ij}] = \mathbb{E}[B_{ij}]$, $\mathbb{E}[A^2_{ij}] = \mathbb{E}[B^2_{ij}]$ and $\mathbb{E}[|A_{ij}|^3] \le (1 + K)$, $\mathbb{E}[|B_{ij}|^3] \le (1 + K)$. Using Theorem 1 we get

$$|\mathbb{E}[f(n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[f(n^{-1/2}\mathbf{B}_n)]| \le \frac{1}{6}n^2\frac{\beta^3}{\sqrt{2}\,n}\max_{r,c\in[n]}\Big\{\mathbb{E}\Big[\frac{|A_{rc}|^3}{n^{3/2}}\Big], \mathbb{E}\Big[\frac{|B_{rc}|^3}{n^{3/2}}\Big]\Big\} = O\Big(\frac{1}{\sqrt{n}}\Big).$$

□

Proof of Theorem 8. From the definition of the random matrices $\mathbf{A}^\gamma_n$ and $\mathbf{B}_n$, we have we have $\mathbb{E}[A_{ij}] = \mathbb{E}[B_{ij}]$, $\mathbb{E}[A^2_{ij}] = \mathbb{E}[B^2_{ij}]$ and $\mathbb{E}[|A^\gamma_{ij}|^3] \le (1+K)\gamma/n$, $\mathbb{E}[|B_{ij}|^3] \le (1+K)$ (with K independent of γ and n). Therefore using the estimate on $L_3(f)$ fro the previous proof, together with Theorem 1 we have

$$|\mathbb{E}[f(\gamma^{-1/2}\mathbf{A}^\gamma_n)] - \mathbb{E}[f(n^{-1/2}\mathbf{B}_n)]| \le \frac{1}{6}n^2\frac{\beta^3}{\sqrt{2}\,n}\max_{r,c\in[n]}\Big\{\mathbb{E}\Big[\frac{|A^\gamma_{rc}|^3}{\gamma^{3/2}}\Big], \mathbb{E}\Big[\frac{|B_{rc}|^3}{n^{3/2}}\Big]\Big\} \le K'\beta^3\max\Big\{\frac{1}{\sqrt{\gamma}}, \frac{1}{\sqrt{n}}\Big\}.$$

Therefore, $\lim_{\gamma\to\infty}\lim_{n\to\infty}\big\{\mathbb{E}[f(\gamma^{-1/2}\mathbf{A}^\gamma_n)] - \mathbb{E}[f(n^{-1/2}\mathbf{B}_n)]\big\} = 0$. □

4.2 CDMA

For any $m \times n$ matrix A_n, the capacity (3) can be expressed as

$$C_n(A_n) = \log 2 - \frac{1}{2}\alpha - \frac{1}{n} \sum_{x^{\mathsf{in}} \in \{+1,-1\}^n} \frac{1}{2^n} \mathbb{E}_Z \log \Big\{ \sum_{x \in \{+1,-1\}^n} e^{-\frac{1}{2\sigma^2}\|Z + A_n(x^{\mathsf{in}} - x)\|_2^2} \Big\},$$

where Z is an m-dimensional random vector, whose entries are i.i.d. $\mathsf{N}(0, \sigma^2)$. By a simple change of variables in the sum over x, we get

$$C_n(A_n) = \log 2 - \frac{1}{2}\alpha - \frac{1}{n} \sum_{x^{\mathsf{in}} \in \{+1,-1\}^n} \frac{1}{2^n} \mathbb{E}_Z \log \Big\{ \sum_{x \in \{0,2\}^n} e^{-\frac{1}{2\sigma^2}\|Z + A_n x^{\mathsf{in}} x\|_2^2} \Big\}.$$

For a matrix $A_n = \{A_{i,j}\}_{1 \le i \le m, 1 \le j \le n}$, and a vector $x^{\mathsf{in}} \in \{+1,-1\}^n$, define $A_n(x^{\mathsf{in}})$ by letting $[A_n(x^{\mathsf{in}})]_{ij} = A_{ij} x_j^{\mathsf{in}}$. Further, define the Hamiltonian function $\mathcal{H} : \{0,2\}^n \times \mathbb{R}^m \times \mathbb{R}^{m \times n} \to \mathbb{R}$ by

$$\mathcal{H}(x, Z, A_n) = \frac{1}{2\sigma^2}\|Z + A_n x\|_2^2 = \frac{1}{2\sigma^2} \sum_{i=1}^m \Big(Z_i + \sum_{j=1}^n A_{ij} x_j \Big)^2.$$

Then we have

$$\begin{aligned} C_n(A_n) &= \log 2 - \frac{1}{2}\alpha - \frac{1}{2^n} \sum_{x^{\mathsf{in}} \in \{+1,-1\}^n} \mathbb{E}_Z f(A_n(x^{\mathsf{in}}), Z), \\ f(A_n, Z) &\equiv \frac{1}{n} \log \Big\{ \sum_{x \in \{0,2\}^n} e^{-\mathcal{H}(x,Z,A_n)} \Big\}. \end{aligned}$$

If $\mathbf{A}_n$ is a random matrix of standard type, and $x^{\mathsf{in}} \in \{+1,-1\}^n$, then $\mathbf{A}_n(x^{\mathsf{in}})$ is also a random matrix of standard type. In order to prove the universality results, theorems 3 and 4, it is therefore sufficient to fix –say– $x^{\mathsf{in}} = (+1, \ldots, +1)$, and prove universality of $\mathbb{E}_Z f(A_n, Z)$.

Analogously to the proof in the previous section, for a function $(x, Z, A_n) \mapsto g(x, Z, A_n)$, we let

$$\langle g(x, Z, A_n) \rangle \equiv \frac{\sum_{x \in \{0,2\}^n} g(x, Z, A_n) e^{-\mathcal{H}(x,Z,A_n)}}{\sum_{x \in \mathcal{X}^n} e^{-\mathcal{H}(x,Z,A_n)}}. \tag{16}$$

In order use Theorem 2 we need to estimate the third derivatives of f. Again, $\partial_{rc}^k f$ denote the k-th derivative of f with respect to the $A_{r,c}$. The third derivative is then given by

$$\partial_{rc}^3 f(A_n, Z) = \frac{1}{n(2\sigma^2)^3} \Big(- \langle (\partial_{rc}\mathcal{H}(x, Z, A_n))^3 \rangle + 3 \langle \partial_{rc}\mathcal{H}(x, Z, A_n) \rangle \langle (\partial_{rc}\mathcal{H}(x, Z, A_n))^2 \rangle - 2 \langle \partial_{rc}\mathcal{H}(x, Z, A_n) \rangle^3 \Big). \tag{17}$$

Proof of Theorem 3. Let $\mathbf{A}_n$ and $\mathbf{B}_n$ be as defined in the theorem. Let $\mathbf{D}_n(r, c, s)$ denote the matrix with entries

$$D_{ij} = \begin{cases} \frac{1}{\sqrt{m}} A_{ij}, & \text{if } i < r \text{ or } i = r \text{ and } j < c, \\ s, & \text{if } i = r, \text{ and } j = c, \\ \frac{1}{\sqrt{m}} B_{ij}, & \text{otherwise.} \end{cases}$$

From now onwards we use $\mathcal{H}(x)$ to denote $\mathcal{H}(x, Z, \mathbf{D}_n(r,c,s))$ and let $\langle\cdot\rangle$ denote the corresponding average, as per Eq. (16). Further, for $r \in [m]$, let $\Theta_r(x) \equiv (Z_r + \sum_{j=1}^n D_{rj}x_j)/(\sqrt{2}\sigma)$ and $\mathcal{H}_{\sim r}(x) = \mathcal{H}(x) - \Theta_r(x)^2$. Notice that

$$\mathcal{H}(x) = \sum_{i\in[m]} \Theta_i(x)^2\,, \qquad \mathcal{H}_{\sim r}(x) = \sum_{i\in[m]\setminus r} \Theta_i(x)^2\,.$$

Accordingly, we let $\langle\cdot\rangle_{\sim r}$ denote the average as defined in (16) with the Hamiltonian $\mathcal{H}_{\sim r}(x)$.

The derivative of $\mathcal{H}(x)$ with respect to A_{rc} is

$$\partial_{rc}\mathcal{H}(x) = \frac{1}{2\sigma^2}\Big(Z_r + \sum_{j=1}^n D_{rj}x_j\Big)2x_c = \frac{1}{\sqrt{2}\sigma}2x_c\Theta_r(x).$$

Its fourth moment can then be bounded as

$$\begin{aligned}\mathbb{E}\langle(\partial_{rc}\mathcal{H})^4\rangle_s &= \mathbb{E}\Big\{\frac{\sum_x e^{-\mathcal{H}(x)}(\partial_{rc}\mathcal{H})^4}{\sum_x e^{-\mathcal{H}(x)}}\Big\}\\ &\le \mathbb{E}\Big\{\frac{\sum_x e^{-\mathcal{H}_{\sim r}(x)-\Theta_r(x)^2}(64/\sigma^4)\Theta_r(x)^4}{\sum_x e^{-\mathcal{H}_{\sim r}(x)-\Theta_r(x)^2}}\Big\}\,.\end{aligned}$$

Since the random variables $e^{-\Theta_r(x)^2}$ and $\Theta_r(x)^4$ are negatively correlated, we have

$$\langle e^{-\Theta_r(x)^2}\Theta_r(x)^4\rangle_{\sim r} \le \langle e^{-\Theta_r(x)^2}\rangle_{\sim r}\langle\Theta_r(x)^4\rangle_{\sim r},$$

which implies

$$\mathbb{E}\langle(\partial_{rc}\mathcal{H})^4\rangle_s \le \frac{64}{\sigma^4}\mathbb{E}\langle\Theta_r(x)^4\rangle_{\sim r}. \tag{18}$$

Using the inequality $(a+b+c)^4 \le 27(a^4+b^4+c^4)$ and the definition of $\{A_{ij}\}$ and $\{B_{ij}\}$ in Theorem 3, we get

$$\begin{aligned}\mathbb{E}\langle(\partial_{rc}\mathcal{H})^4\rangle &\le \frac{27\cdot 64}{4\sigma^4}\Big\{\mathbb{E}[Z_r^4] + \mathbb{E}[\langle(D_{rc}x_c)^4\rangle_{\sim r}] + \mathbb{E}[\langle(\sum_{i\in[n]\setminus c} D_{ri}x_i)^4\rangle_{\sim r}]\Big\}\\ &\le K_1 + K_1\,s^4 + K_1\,\mathbb{E}[\langle(\sum_{i\in[n]\setminus c} D_{ri}x_i)^4\rangle_{\sim r}\,,\end{aligned}$$

where $K = K(\sigma)$ is a constant independent of m, n. If we use the subscript $i \ne j \ne k \ne \dots$ to denote all the tuples of distinct indices and we expand the power, we get

$$\begin{aligned}\mathbb{E}[\langle(\sum_{i\in[n]\setminus c} D_{ri}x_i)^4\rangle_{\sim r}] &= \sum_{i,j,k,l\in[n]\setminus c} \mathbb{E}[D_{ri}D_{rj}D_{rk}D_{rl}\langle x_ix_jx_kx_l\rangle_{\sim r}]\\ &= \sum_{i,j,k,l\in[n]\setminus c} \mathbb{E}[D_{ri}D_{rj}D_{rk}D_{rl}]\mathbb{E}[\langle x_ix_jx_kx_l\rangle_{\sim r}] =\\ &= \sum_{i\in[n]\setminus c} \mathbb{E}[D_{ri}^4]\mathbb{E}[\langle x_i^4\rangle_{\sim r}] + +3\sum_{i\ne j\in[n]\setminus c} \mathbb{E}[D_{ri}^2D_{rj}^2]\mathbb{E}[\langle x_i^2x_j^2\rangle_{\sim r}]\,,\end{aligned}$$

Here we used the fact that $\{D_{ri}\}_{1\le i\le n}$ are independent of $\mathcal{H}_{\sim r}(x)$, and therefore of $\langle x_i x_j x_k x_l\rangle_{\sim r}$ Further all the terms with one of the indices i,j,k,l distinct from the all others vanish because $\mathbb{E}\{D_{ri}\}=0$ for all $i\neq c$ by our assumption on $\mathbf{A}_n$, $\mathbf{B}_n$. Using $x_i\in\{0,2\}$, we then get

$$\mathbb{E}[\langle(\sum_{i\in[n]\setminus c} D_{ri}x_i)^4\rangle_{\sim r}] \le \sum_{i\in[n]\setminus c}\frac{(1+K)^2}{m^2}\cdot 16 + 3\sum_{i\neq j\in[n]\setminus c}\frac{1}{m^2}\cdot 16 \le K_2$$

where $K_2=K_2(\alpha)$ is another constant. Putting everything together, we get

$$\mathbb{E}\langle(\partial_{rc}\mathcal{H})^4\rangle \le K_3\,(1+s^4)\,.$$

and therefore, by Jensen inequality, we get $\mathbb{E}\langle|\partial_{rc}\mathcal{H}|^3\rangle \le K_3(1+|s|^3)$ (by eventually enlarging the constant K_3. Using Eq. (17), this finally implies that

$$\mathbb{E}[|\partial^3_{rc} f(\mathbf{D}_n(r,c,s),Z)|] \le \frac{K_4}{n}(1+|s|^3)\,.$$

We are now in position to apply Theorem 2. Since the means and variances of the entries of $\mathbf{A}_n$ and $\mathbf{B}_n$ are equal, we have $a_i=b_i=0$. We get therefore

$$\begin{aligned}|\mathbb{E}[f(m^{-1/2}\mathbf{A}_n,Z)]-\mathbb{E}[f(m^{-1/2}\mathbf{B}_n,Z)]| &\le \frac{K_4}{n}\sum_{r=1}^m\sum_{c=1}^n\big(\mathbb{E}_{A_{rc}}\int_0^{A_{rc}/\sqrt{m}}(1+|s|^3)(\frac{A_{rc}}{\sqrt{m}}-s)^2\mathrm{d}s\\ &\qquad + \mathbb{E}_{B_{rc}}\int_0^{B_{rc}/\sqrt{m}}(1+|s|^3)(\frac{B_{rc}}{\sqrt{m}}-s)^2\mathrm{d}s\big)\\ &\le mK'\sum_{i=3}^6\Big\{\mathbb{E}\Big[\Big(\frac{A_{rc}}{\sqrt{m}}\Big)^i\Big]+\mathbb{E}\Big[\Big(\frac{B_{rc}}{\sqrt{m}}\Big)^i\Big]\Big\}=O\Big(\frac{1}{\sqrt{n}}\Big).\end{aligned}$$

□

The proof of Theorem 4 is very similar to the one above. We only stress the differences below.

Proof of Theorem 4. Let $\mathbf{A}^\gamma_n$ and $\mathbf{B}_n$ be as defined in the statement. We modify the definition of $\mathbf{D}_n(r,c,s)$ used in the last proof, as follows

$$D_{ij}=\begin{cases}\frac{1}{\sqrt{\gamma}}A^\gamma_{ij}, & \text{if } i<r \text{ or } i=r \text{ and } j<c,\\ s, & \text{if } i=r, \text{ and } j=c,\\ \frac{1}{\sqrt{m}}B_{ij}, & \text{otherwise.}\end{cases}$$

Now following the proof of Theorem 3, and assuming without loss of generality $\gamma\ge 1$, we get again

$$\mathbb{E}\langle(\partial_{rc}\mathcal{H})^4\rangle \le K_1\Big(1+s^4\Big)\,.$$

(The final step consists as in the previous proof, in bounding the sums $\sum_{i\in[n]\setminus c}\mathbb{E}[D^4_{ri}]$ and $\sum_{i\neq j\in[n]\setminus c}\mathbb{E}[D^2_{ri}D^2_{rj}]\mathbb{E}[\langle x_i^2x_j^2\rangle_{\sim r}]$.) This in turn implies $\mathbb{E}[|\partial^3_{rc}f(\mathbf{D}_n(r,c,s),Z)|]\le (K_1'/n)(1+|s|^3)$. Since the means and variances of the entries of $\mathbf{A}^\gamma_n$ and $\mathbf{B}_n$ are equal, we have $a_i=b_i=0$. Applying Theorem 2, we get

$$|\mathbb{E}[f(\gamma^{-1/2}\mathbf{A}^\gamma_n,Z)]-\mathbb{E}[f(m^{-1/2}\mathbf{B}_n,Z)]| \le \frac{K_1'}{n}\sum_{r=1}^m\sum_{c=1}^n\Big\{\mathbb{E}_{A^\gamma_{rc}}\int_0^{A^\gamma_{rc}/\sqrt{\gamma}}(1+|s|^3)(\frac{A^\gamma_{rc}}{\sqrt{\gamma}}-s)^2\mathrm{d}s$$

$$+\mathbb{E}_{B_{rc}}\int_0^{B_{rc}/\sqrt{m}}(1+|s|^3)(\frac{B_{rc}}{\sqrt{m}}-s)^2\mathrm{d}s\Big\}$$
$$\le mK_2\sum_{i=3}^{6}\Big\{\mathbb{E}\Big[\Big(\frac{A^\gamma_{rc}}{\sqrt{\gamma}}\Big)^i\Big]+\mathbb{E}\Big[\Big(\frac{B_{rc}}{\sqrt{m}}\Big)^i\Big]\Big\}$$
$$\le K_3\Big(\frac{1}{\sqrt{\gamma}}+\frac{1}{\sqrt{n}}\Big).$$

Now taking the limit $n\to\infty$ first and then the limit $\gamma\to\infty$ gives the result. □

4.3 LASSO

The proof of Theorem 5 repeats some arguments already present in the proof of Theorem 3 presented in the previous section. We shall omit such repetitions and instead focus on the new ideas required.

Proof of Theorem 5. Without loss of generality, we will assume $x_{\max}=1$. Define $\mathcal{X}=[-1,1]$ and, for $\delta>0$, define $\mathcal{X}_\delta=\{k\delta\,:\,k\in\mathbb{Z},\,|k\delta|\le 1\}$. In words $\mathcal{X}_\delta$ is a grid of points in the interval $[-1,1]$ with spacing δ. Recall that x_0 is a fixed deterministic signal with $||x_0||_\infty\le 1$, and the resulting measurements read $Y=A_nx_0+Z$, where Z is noise vector with i.i.d. Gaussian component. Define the Hamiltonian function $\mathcal{H}:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^{m\times n}\to\mathbb{R}$ by letting

$$\begin{aligned}\mathcal{H}(x,z,A_n) &= \lambda\|x\|_1+\frac{1}{2}\|y-A_nx\|_2^2\\ &= \lambda\|x\|_1+\frac{1}{2}\|z-A_n(x-x_0)\|_2^2\,.\end{aligned}$$

With this definition, $L(A_n)=\frac{1}{n}\min_{x\in\mathcal{X}^n}\{\mathcal{H}(x,z,A_n)\}$. Let $L_\delta(A_n)=\frac{1}{n}\min_{x\in\mathcal{X}^n_\delta}\{\mathcal{H}(x,z,A_n)\}$. Our proof follows by first showing that there exists a constant C such that

$$\big|\mathbb{E}[L_\delta(n^{-1/2}\mathbf{A}_n)]-\mathbb{E}[L(n^{-1/2}\mathbf{A}_n)]\big|\le C\,\delta,\quad \big|\mathbb{E}[L_\delta(n^{-1/2}\mathbf{B}_n)]-\mathbb{E}[L(n^{-1/2}\mathbf{B}_n)]\big|\le C\,\delta\,.$$

Obviously $L_\delta(A_n)\ge L(A_n)$. In order to prove the converse bound, let $\widehat{x}$ be a minimizer of $\mathcal{H}(x,z,A_n)$ in $\mathcal{X}^n$, and denote by x_δ its closest approximation in $\mathcal{X}^n_\delta$. Obviously $|x_{\delta,i}-\widehat{x}_i|\le\delta$ for all $i\in[n]$. We then have

$$\begin{aligned}\frac{1}{n}|\mathcal{H}(\widehat{x},z,n^{-1/2}\mathbf{A}_n)-\mathcal{H}(x_\delta,z,n^{-1/2}\mathbf{A}_n)| &\le \lambda\delta+\frac{1}{2n}\Big|\|y-n^{-1/2}\mathbf{A}_n\widehat{x}\|_2^2-\|y-n^{-1/2}\mathbf{A}_nx_\delta\|_2^2\Big|\\ &= \lambda\delta+\frac{1}{2n}\Big|\big(\frac{1}{\sqrt{n}}\mathbf{A}_n(x_\delta-\widehat{x})\big)^\top 2z+\big(\frac{1}{\sqrt{n}}\mathbf{A}_n(\widehat{x}-x_\delta)\big)^\top\frac{1}{\sqrt{n}}\mathbf{A}_n(\widehat{x}+x_\delta-2x_0)\Big|\\ &\le \lambda\delta+\frac{1}{2n}\Big\|\frac{1}{\sqrt{n}}\mathbf{A}_n(\widehat{x}-x_\delta)\Big\|_2 2\|z\|_2+\frac{1}{2n^2}\,\sigma_{\max}(\mathbf{A}_n)^2\|\widehat{x}-x_\delta\|_2\|\widehat{x}+x_\delta-2x_0\|_2\\ &\le \lambda\delta+\sigma_{\max}(n^{-1/2}\mathbf{A}_n)\frac{1}{\sqrt{n}}\,\delta\,\|z\|_2+2\,\sigma_{\max}(n^{-1/2}\mathbf{A}_n)^2\delta\,,\end{aligned}\tag{19}$$

where we used $\|x_\delta\|_2$, $\|\widehat{x}\|_2$, $\|x_0\|_2\le\sqrt{n}$ and $\|\widehat{x}-x_\delta\|_2\le\delta\sqrt{n}$. Here $\sigma_{\max}(A_n)$ is the largest singular value of A_n. From [22] we know that $\mathbb{E}[\sigma_{\max}(n^{-1/2}\mathbf{A}_n)^2]<K$, for some constant $K<\infty$. Combining this with Eq. (19), and using the Cauchy-Schwartz inequality, we get

$$\big|\mathbb{E}[L_\delta(n^{-1/2}\mathbf{A}_n)]-\mathbb{E}[L(n^{-1/2}\mathbf{A}_n)]\big|\le C\delta\,.$$

A similar result obviously holds for the matrix ensemble $\mathbf{B}_n$ as well. By triangular inequality, we have

$$\lim_{n\to\infty} \left|\mathbb{E}[L(n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[L(n^{-1/2}\mathbf{B}_n)]\right| \le C\delta + \lim_{n\to\infty} \left|\mathbb{E}[L_\delta(n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[L_\delta(n^{-1/2}\mathbf{B}_n)]\right| .$$

Since this inequality holds for any $\delta > 0$, the proof of the theorem reduces to showing that $\lim_{n\to\infty} \left|\mathbb{E}[L_\delta(n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[L_\delta(n^{-1/2}\mathbf{B}_n)]\right| = 0$.

In order to prove this, define

$$f(\delta, \beta, z, A_n) = -\frac{1}{\beta n} \log \Big\{ \sum_{x\in\mathcal{X}_\delta^n} e^{-\beta\mathcal{H}(x,z,A_n)} \Big\} .$$

It is easy to see that

$$\lim_{\beta\to\infty} f(\delta, \beta, z, A_n) = \frac{1}{n} \min_{x\in\mathcal{X}_\delta^n} \mathcal{H}(x, z, A_n) . \tag{20}$$

Further, a straightforward calculation shows that

$$\beta^2 \frac{\partial f}{\partial \beta}(\delta, \beta, z, A_n) = H(p_{\beta,A_n}) ,$$

where $H(p)$ denotes Shannon's entropy of the probability distribution p and $p_{\beta,A_n}(x) \propto \exp\{-\beta\mathcal{H}(x, z, A_n)\}$. Of course $0 \le H(p_{\beta,A_n}) \le n \log|\mathcal{X}_\delta|$ whence

$$-\frac{1}{\beta^2} \log\Big(\frac{2}{\delta}\Big) \le \frac{\partial f}{\partial \beta}(\delta, \beta, z, A_n) \le 0 . \tag{21}$$

Therefore,

$$\begin{aligned}
&\lim_{n\to\infty} \left|\mathbb{E}[L_\delta(n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[L_\delta(n^{-1/2}\mathbf{B}_n)]\right| \\
&\quad = \lim_{n\to\infty} \Big| \lim_{\beta\to\infty} \mathbb{E}[f(\delta, \beta, Z, n^{-1/2}\mathbf{A}_n)] - \lim_{\beta\to\infty} \mathbb{E}[f(\delta, \beta, Z, n^{-1/2}\mathbf{B}_n)]\Big| \\
&\quad \le \lim_{n\to\infty} \left|\mathbb{E}[f(\delta, \beta, Z, n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[f(\delta, \beta, Z, n^{-1/2}\mathbf{B}_n)]\right| + \int_\beta^\infty \frac{1}{s^2} \log\Big(\frac{2}{\delta}\Big)\, \mathrm{d}s ,
\end{aligned} \tag{22}$$

where the first step follows from (20) and the second from (21). Notice the close resemblance between the function $f(\delta, \beta, Z, A_n)$ defined here and the one used in the previous section. Using the same arguments developed there for the proof of Theorem 3 it is immediate to show that

$$\left|\mathbb{E}[f(\delta, \beta, Z, n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[f(\delta, \beta, Z, n^{-1/2}\mathbf{B}_n)]\right| \le O\Big(\frac{1}{\sqrt{n}}\Big) .$$

Combining this with Eq. (22), we get

$$\lim_{n\to\infty} \left|\mathbb{E}[L_\delta(n^{-1/2}\mathbf{A}_n)] - \mathbb{E}[L_\delta(n^{-1/2}\mathbf{B}_n)]\right| \le \frac{1}{\beta} \log\Big(\frac{2}{\delta}\Big) .$$

The proof is completed by letting $\beta \to \infty$. □

4.4 Wishart Matrices

The proof is analogous the proof for universality of the Wigner's semi-circle law developed in [23].

Proof of Theorem 6. By the analiticity of the Stieltjes transform, it is sufficient to prove the claim for $\mathrm{Im}(z)$ large enough.

For an $m \times n$ matrix A_n and any $z \in \mathbb{C}\backslash\mathbb{R}$, let

$$f(A_n) \equiv \frac{1}{n}\mathrm{Tr}\big((A_n^\top A_n + zI_n)^{-1}\big)\,.$$

In order to simplify the notation we drop the subscript n and denote the partial derivative with respect to A_{ij} by ∂_{ij}. Define $R = (A^\top A + zI)^{-1}$. Therefore $(A^\top A + zI)R = I$, which implies $\partial_{ij}((A^\top A + zI)R) = 0$. This yields

$$\partial_{ij} R = -R\partial_{ij}(A^\top A)R\,.$$

Let 1_{ij} denote the matrix with (ij)-th entry equal to 1 and the remaining entries equal to 0. Then

$$\begin{aligned}
\partial_{ij}(A^\top A) &= 1_{ji}A + A^\top 1_{ij},\\
\partial_{ij}^2(A^\top A) &= 21_{ii}\,,\\
\partial_{ij}^3(A^\top A) &= 0\,.
\end{aligned}$$

Using the identity $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, we get

$$\begin{aligned}
\partial_{ij} f &= -\frac{1}{n}\mathrm{Tr}\Big(\partial_{ij}(A^\top A)R^2\Big),\\
\partial_{ij}^2 f &= \frac{2}{n}\mathrm{Tr}\Big(\partial_{ij}(A^\top A)R\partial_{ij}(A^\top A)R^2\Big) - \frac{1}{n}\mathrm{Tr}\Big(\partial_{ij}^2(A^\top A)R^2\Big),\\
\partial_{ij}^3 f &= -\frac{6}{n}\mathrm{Tr}\Big(\partial_{ij}(A^\top A)R\partial_{ij}(A^\top A)R\partial_{ij}(A^\top A)R^2\Big) \quad + \frac{3}{n}\mathrm{Tr}\Big(\partial_{ij}^2(A^\top A)R\partial_{ij}(A^\top A)R^2\Big)\\
&\quad + \frac{3}{n}\mathrm{Tr}\Big(\partial_{ij}(A^\top A)R\partial_{ij}^2(A^\top A)R^2\Big). \qquad (23)
\end{aligned}$$

Note that R is a symmetric matrix and therefore is diagonalizable. Moreover, note that the singular values of R^{-1} are bounded by $|v|^{-1}$, where $v = \mathrm{Im}(z)$. Let $\|A\|$ and $\|A\|_2$ denote the Frobenius norm and the spectral norm of A respectively. From Cauchy-Schwartz inequality we have $|\mathrm{Tr}(AB)| \le \|A\|\,\|B\|$. Therefore, we can bound the first term as

$$\begin{aligned}
|\mathrm{Tr}(\partial_{ij}(A^\top A)R\partial_{ij}(A^\top A)R\partial_{ij}(A^\top A)R^2)| &\le \|(\partial_{ij}(A^\top A)R)^2\|\|\partial_{ij}(A^\top A)R^2\|\\
&\overset{(a)}{\le} \frac{1}{|v|}\,\|\partial_{ij}(A^\top A)R\|^3\\
&\overset{(b)}{\le} \frac{1}{|v|^4}\,\|\partial_{ij}(A^\top A)\|^3\,, \qquad (24)
\end{aligned}$$

where we have used $\|AB\| \le \|A\|\|B\|$ in (a) and $\|AB\| \le \|A\|\|B\|_2$ in both (a) and (b).

Similarly one can bound the second and third terms of (23) as

$$|\mathrm{Tr}(\partial_{ij}^2(A^\top A)R\partial_{ij}(A^\top A)R^2)| \le \|\partial_{ij}^2(A^\top A)\|\|\partial_{ij}(A^\top A)\|\frac{1}{|v|^3} = \|\partial_{ij}(A^\top A)\|\frac{2}{|v|^3}. \qquad (25)$$

Finally, we can bound $\|\partial_{ij}(A^\top A)\|$ as follows

$$\|\partial_{ij}(A^\top A)\| \leq \|1_{ji}A\| + \|A^\top 1_{ij}\| = 2\|A^\top 1_{ij}\| = 2\Big(\sum_{k=1}^{m} A_{kj}^2\Big)^{1/2}. \tag{26}$$

Let us now consider the random matrices $\mathbf{A}_n^\gamma$ and $\mathbf{B}_n$ as defined in the theorem. Let $\mathbf{C}_n(r,c,s)$ denote the matrix as defined in Section 4.2, i.e.,

$$C_{ij} = \begin{cases} \frac{1}{\sqrt{\gamma}} A_{ij}^\gamma, & \text{if } i < r \text{ or } i = r \text{ and } j < c, \\ s, & \text{if } i = r, \text{ and } j = c, \\ \frac{1}{\sqrt{m}} B_{ij}, & \text{otherwise.} \end{cases}$$

Using the equations (23), (24), (25), and (26), we get

$$\begin{aligned} \mathbb{E}\big\{|\partial_{rc}^3 f(\mathbf{C}_n(r,c,s))|\big\} &\leq \frac{K_0}{n}\mathbb{E}\Big\{\Big(1+\sum_{k=1}^{m} C_{kc}^2\Big)^{3/2}\Big\} \\ &\leq \frac{K_1}{n}(1+s^3)\,. \end{aligned}$$

The proof is finished as for Theorem 4. □

4.5 Proof of Corollary 2

Throughout this proof we will assume $\sigma = 1$, for simplicity of notation (general $\sigma > 0$ follows exactly the same argument).

Convergence of Stieltjes transform implies weak convergence of the expected distribution of eigenvalues [16, Theorem 2.4.4]. This means that for any continuous bounded function f[1]

$$\lim_{\gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[f(\lambda_i(\gamma^{-1}(\mathbf{A}_n^\gamma)^\top\mathbf{A}_n^\gamma))] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[f(\lambda_i(n^{-1}\mathbf{B}_n^\top\mathbf{B}_n))]\,. \tag{27}$$

The limit on the right hand side exists because the expected distribution of eigenvalues of Wishart matrices converges [2, 24]. Moreover, the limiting distribution function is continuous. Therefore, the convergence of the distributions implies the convergence of expectations for any bounded measurable function, not necessarily continuous (by the bounded convergence theorem). We are interested in estabilishing a result of the form (27) for the function $f(x) = \log(1+x)$, which is not bounded. However, note that only the behavior of f in the region $x \geq 0$ is relevant, because $\lambda_i \geq 0$. In the domain of interest the function f is bounded from below. In order to tackle the issue of boundedness from above, we use a standard truncation trick. We define $g_M(x) = f(x)\mathbb{1}_{\{x\leq M\}}$, for some $0 < M < \infty$. Note that the function g_M is bounded on $\mathbb{R}_+$. Therefore

$$\lim_{\gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[g_M(\lambda_i(\gamma^{-1}(\mathbf{A}_n^\gamma)^\top\mathbf{A}_n^\gamma))] = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[g_M(\lambda_i(n^{-1}\mathbf{B}_n^\top\mathbf{B}_n))]\,. \tag{28}$$

[1]Note that we have two limits on the left hand side. This can be taken care of by noticing that $\lim_{\gamma\to\infty}\lim_{n\to\infty} f(\gamma,n,x) = f(x)$ is equivalent to saying that $\lim_{n\to\infty} f(\gamma_n,n,x) = f(x)$ along any sequence of $\{\gamma_n\}$s satisfying $\lim_{n\to\infty}\gamma_n = \infty$.

Note that

$$\begin{aligned}
&\lim_{\gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big\{\big|g_M(\lambda_i(\gamma^{-1}(\mathbf{A}_n^\gamma)^\top\mathbf{A}_n^\gamma))-f(\lambda_i(\gamma^{-1}(\mathbf{A}_n^\gamma)^\top\mathbf{A}_n^\gamma))\big|\Big\}\\
&\quad=\lim_{\gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big\{\log(1+\lambda_i(\gamma^{-1}(\mathbf{A}_n^\gamma)^\top\mathbf{A}_n^\gamma))\mathbb{1}_{\{\lambda_i>M\}}\Big\}\\
&\quad\le\lim_{\gamma\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\Big\{\lambda_i^2(\gamma^{-1}(\mathbf{A}_n^\gamma)^\top\mathbf{A}_n^\gamma))/M\Big\}\\
&\quad=\lim_{\gamma\to\infty}\lim_{n\to\infty}\frac{1}{Mn\gamma^2}\mathbb{E}\,\mathrm{Tr}\Big\{\Big((A_n^\gamma)^\top A_n^\gamma\Big)^2\Big\}\\
&\quad=\lim_{\gamma\to\infty}\lim_{n\to\infty}\frac{1}{Mn\gamma^2}\mathbb{E}\Big\{\sum_{i,j}\Big(\sum_{k}A_{ki}A_{kj}\Big)^2\Big\}\\
&\quad=\lim_{\gamma\to\infty}\lim_{n\to\infty}\frac{1}{Mn\gamma^2}\Big\{\sum_{i\neq j}\sum_{k}\mathbb{E}\big\{A_{ki}^2\big\}\mathbb{E}\big\{A_{kj}^2\big\}+\sum_{i}\sum_{k_1\neq k_2}\mathbb{E}\big\{A_{k_1i}^2\big\}\mathbb{E}\big\{A_{k_2i}^2\big\}+\sum_{i,k}\mathbb{E}\big\{A_{ki}^4\big\}\Big\}\\
&\quad\le\frac{K}{M}\,,
\end{aligned}\tag{29}$$

for a constant K independent of M, γ as long as $\gamma\ge 1$. Using a similar argument we can show that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\big\{|g_M(\lambda_i(n^{-1}\mathbf{B}_n^\top\mathbf{B}_n))]-f(\lambda_i(n^{-1}\mathbf{B}_n^\top\mathbf{B}_n))|\big\}\ \le\frac{K'}{M}\tag{30}$$

for a constant K' independent of M. From (28), (29), (30) we get

$$\lim_{\gamma\to\infty}\lim_{n\to\infty}\big|\mathbb{E}[C_n(\gamma^{-1/2}\mathbf{A}_n^\gamma)]-\mathbb{E}[C_n(n^{-1/2}\mathbf{B}_n)]\big|\le\frac{K+K'}{M}\,.$$

Now taking the $\lim_{M\to\infty}$ gives the desired result.

References

[1] S. Chatterjee, "A generalization of the Lindeberg principle," *The Annals of Probability.*, vol. 34, no. 6, pp. 2061–2076, 2006.

[2] V. A. Marčenko and L.A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices", Math. USSR Sb. 1 (1967) 457-483

[3] J. Cardy, *Scaling and Renormalization in Statistical Physics*, Cambridge University Press, Cambridge, 1996

[4] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, 1999.

[5] D. L. Donoho and J. Tanner, "Observed Universality of Phase Transitions in High-Dimensional Geometry, with Implications for Modern Data Analysis and Signal Processing," Phil. Trans. R. Soc. A 13, November 2009, 4273-4293

[6] D. L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *Journal of the AMS*, vol. 22, no. 1, pp. 1–53, 2009.

[7] H. Kesten, "Symmetric random walks on groups," *Trans. Amer. Math. Soc.* 92 (1959), 336-354

[8] B. D. McKay, "The expected eigenvalue distribution of a large regular graph" *Linear Algebra Appl.* 40 (1981) 203-216

[9] S. Verdu, *Multiuser Detection*, Cambridge University Press, Cambridge, 1998

[10] A. J. Grant and P. D. Alexander, "Randomly selected spreading sequences for coded CDMA," in *4th Int. Spread Spectrum Techniques and Applications*, Mainz, Germany, Sept. 1996, pp. 54–57.

[11] S. B. Korada and N. Macris, "Tight bounds on the capacity of binary input random CDMA systems," *accepted in IEEE Trans. Inform. Theory* arXiv:0803.1454 (2008)

[12] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or : How to prove Tanaka's formula)," in *Proc. of the IEEE Inform. Theory Workshop*, Punta del Este, Uruguay, Mar 13–Mar 17 2006.

[13] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," IEEE Trans. on Inform. Theory, 48 (2002), 2888-2910

[14] R. Tibshirani, "Regression shrinkage and selection via the LASSO," J. Roy. Statist. Soc. B 58 (1996), 267–288

[15] S.S. Chen, D.L. Donoho, and M.A. Saunders, "Atomic decomposition by basis pursuit," SIAM Review 43 (2001), 129–159

[16] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge University Press, Cambridge, 2009

[17] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford University Press, Oxford 2009

[18] M. Talagrand, *Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models*, Springer, New York, 2003

[19] F. Guerra and F. L. Toninelli, "The High Temperature Region of the Viana-Bray Diluted Spin Glass Model," J. Stat. Phys. 115 (2004) 531-555

[20] M. Talagrand, "Gaussian averages, Bernoulli averages, and Gibbs' measures," *Random Structures and Algorithms*, vol. 21, no. 3-4, pp. 197–204, 2002.

[21] F. L. Toninelli, "Rigorous results for mean field spin glasses: thermodynamic limit and sum rules for the free energy," Ph.D. dissertation, Scuola Normale Superiore, Pisa, Italy, 2002.

[22] Y. Seginer, "The Expected Norm of Random Matrices", *Combinatorics, Probability and Computing*, vol. 9 pp. 149-166.

[23] S. Chatterjee, "A simple invariance theorem," *unpublished.*, vol. http://arxiv.org/abs/math/0508213v1, 2005.

[24] Z. Bai and J. W. Silverstein *Spectral Analysis of Large Dimensional Random Matrices*, Springer, New York, 2009