Lecture 36 — November 13

Instructor: Stas Minsker *Scribe: Mose Wintner*

36.1 Matrix recover problems and sparse linear regression

Let $A_1, A_2 \in \mathbb{R}^{m_1 \times m_2}$. Define the inner product

$$\langle A_1, A_2 \rangle := \operatorname{tr}(A_1^T A_2).$$

The norm corresponding to this inner product is the Frobenius norm $||A||_F^2 = \sum a_{i,j}^2$.

Model: $Y = \langle X, A_0 \rangle + \xi$ where $A, X \in \mathbb{R}^{m_1 \times m_2}$ are possibly random, and ξ is independent of X with mean 0.

Goal: Recover A_0 based on a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$.

Example 1. Diagonal matrices: $X = \operatorname{diag}(x_1, \ldots, x_n)$, $A_0 = \operatorname{diag}(a_1^0, \ldots, a_m^0)$.

The natural definition of "sparsity" for a matrix shouldn't be the number of 0s–this is not invariant wrt change of basis. In other words, it just treats the matrix as a table of numbers. Instead we'll use the rank of the matrix.

Example 2. $X_{i,j}$ are iid $\mathcal{N}(0,1)$. Solution attempt: minimize $\operatorname{rank}(A)$ subject to

$$|\langle X_j, A \rangle - Y_j| \le \varepsilon_j \qquad j = 1, \ldots, n.$$

Not a convex problem. We'll take the convex hull.

Let K be the set of all rank 1 matrices of unit Frobenius norm. Any rank 1 matrix can be written as uv^T with $u \in \mathbb{R}^{m_1}$ and $v \in \mathbb{R}^{m_2}$ both of unit F norm. Consider th set $K = \{uv^T\}$. What is the convex hull of K?

Definition 1. The nuclear norm, or Schatten-1 norm, is defined as

$$||A||_* = \sum_{i=1}^{\operatorname{rank}(A)} \sigma_j(A),$$

where $\sigma_1(A) \ge \sigma_2(A) \ge \cdots \ge \sigma_r(A)$ are the singular values of A and r is the rank of A. Recall the (reduced) singular value decomposition of A is

$$A = \sum \sigma_i u_i v_i^T$$

where $\sigma_i(A) > 0$.

Lemma 1. The convex hull of K is $\{M \in \mathbb{R}^{m_1 \times m_2} | \ ||M||_* \le 1\}$.
Proof:

$$M = \sum_j \frac{\sigma_j(M)}{\sum \sigma_j(M)} u_j v_j^T \sigma_j(M) \in co(K).$$

Asusme that $M \in co(K)$. $A \mapsto \sigma_1(A) + \cdots + \sigma_k(A)$ defines a norm (Ky-Fam norm) for any $k = 1, \ldots, \min(m_1, m_2)$, where $\sigma_i(A)$ are ordered in descending order.
Idea of the proof: $\sum_{j=1}^k \sigma_j(A) = \sup_{u_i, v_i} \sum_{j=1}^k u_i^T A v_i$. Let $M - \sum \alpha_j u_j v_j^T$, $\sum \alpha_j = 1$, $\alpha_j \ge 0$.
Then

$$\sum \sigma_j \left(\sum \alpha_j u_j v_j^T \right) \le \sum_j \alpha_j \sum_i \sum \sigma_i (u_j v_i^T) \le \sum \alpha_j = 1.$$