Algorithmic Aspects of Machine Learning: Problem Set # 3

Instructor: Ankur Moitra

Due: December 4 at 11:59 p.m.

You can work with other students, but you must write-up your solutions by yourself and indicate at the top who you worked with!

Problem 1

Let $\widehat{x}$ be a k-sparse vector in n-dimensions. Let ω be the nth root of unity. Suppose we are given $v_\ell = \sum_{j=1}^{n} \widehat{x}_j \omega^{\ell j}$ for $\ell = 0, 1, ..., 2k-1$. Let $A, B \in \mathbb{R}^{k \times k}$ be defined so that $A_{i,j} = v_{i+j-2}$ and $B_{i,j} = v_{i+j-1}$.

(a) Express both A and B in the form $A = V D_A V^T$ and $B = V D_B V^T$ where V is a Vandermonde matrix, and D_A, D_B are diagonal.

(b) Prove that the solutions to the generalized eigenvalue problem $Ax = \lambda Bx$ can be used to recover the locations of the non-zeros in $\widehat{x}$.

(c) Given the locations of the non-zeros in $\widehat{x}$, and $v_0, v_1, ..., v_{k-1}$, given an algorithm to recover the values of the non-zero coefficients in $\widehat{x}$.

This is called the matrix pencil method. If you squint, it looks like a topic we didn't cover called *Prony's method* (Section 4.4) and has similar guarantees. Both are (somewhat) robust to noise if and only if the Vandermonde matrix is well-conditioned, and when exactly that happens is a longer story...

Problem 2

In the *angular synchronization* problem we observe

$$Y = \frac{\lambda}{n} x x^* + \frac{1}{\sqrt{n}} W.$$

Here the true signal $x \in \mathbb{C}^n$ has each entry drawn uniformly at random from the unit circle in $\mathbb{C}$ (so $|x_i| = 1$). The $n \times n$ noise matrix W is Hermitian with $W_{ii} \sim \mathcal{N}(0,1)$, $W_{ij} \sim \mathbb{C}\mathcal{N}(0,1) := \mathcal{N}(0,1/2) + \mathcal{N}(0,1/2)\mathbf{i}$ (with real and imaginary components independent) and $\{W_{ij}\}_{i \leq j}$ independent. The parameter $\lambda \geq 0$ is the signal-to-noise ratio. One can think of each x_i (which is a unit-norm complex number) as an angle and of Y_{ij} as a noisy measurement of the relative angle $x_i\overline{x_j}$. The goal is to (approximately) recover xx^*, i.e. recover x up to a global rotation.

Consider the iterative algorithm

$$v \leftarrow Yf(v)$$

where $f(v)$ projects each entry of $v \in \mathbb{C}^n$ onto the unit circle in $\mathbb{C}$ via $z \mapsto z/|z|$ (i.e. set the magnitude to 1 while leaving the phase unchanged). The initialization should be a random vector.

(a) Give a heuristic analysis of this algorithm by deriving the *state evolution* equations as follows. Suppose $v = \mu x + \sigma g$ with $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}_{\geq 0}$, and each entry of g distributed independently as $\mathbb{C}\mathcal{N}(0,1)$. Write down a recurrence for how μ and σ evolve after one iteration, in the limit $n \to \infty$. Assume that g and W are independent (i.e. assume that we're running the algorithm with an appropriate Onsager term).

(b) Predict the exact threshold for λ above which the algorithm achieves non-trivial correlation with the truth (in the $n \to \infty$ limit). You may either numerically (using a computer) or analytically analyze the recurrence from (a).

The above algorithm was first proposed by [Boumal '16]. For large values of λ it has near-optimal performance but it is sub-optimal for small λ. The optimal threshold for non-trivial recovery is $\lambda = 1$, which is achieved by PCA (top eigenvector) or AMP.

Problem 3

Consider the sparse coding model $y = Ax$ where A is a fixed $n \times n$ matrix with orthonormal columns a_i, and x has i.i.d. coordinates drawn from the distribution

$$x_i = \begin{cases} +1 & \text{with probability } \alpha/2, \\ -1 & \text{with probability } \alpha/2, \\ 0 & \text{with probability } 1-\alpha. \end{cases}$$

The goal is to recover the columns of A (up to sign and permutation) given many independent samples y. Construct the matrix

$$M = \mathbb{E}_y[\langle y^{(1)}, y\rangle\langle y^{(2)}, y\rangle yy^\top]$$

where $y^{(1)} = Ax^{(1)}$ and $y^{(2)} = Ax^{(2)}$ are two fixed samples from the sparse coding model, and the expectation is over a third sample y from the sparse coding model. Let $\hat{z}$ be the (unit-norm) eigenvector of M corresponding to the largest (in absolute value) eigenvalue.

(a) Write an expression for M in terms of $\alpha, x^{(1)}, x^{(2)}, \{a_i\}$.

(b) Assume for simplicity that $x^{(1)}$ and $x^{(2)}$ each have support size exactly αn and that their supports intersect at a single coordinate i^*. Show that $\langle \hat{z}, a_{i^*}\rangle^2 \geq 1 - O(\alpha^2 n)$ in the limit $\alpha \to 0$.

This method can be used to find a good starting point for alternating minimization.