COMS E6998-9: Algorithms for Massive Data (Fall'23) Oct. 2, 2023

Lecture 8: Compressed Sensing

Instructor: *Alex Andoni* Scribe: *Krish Singal*

1 Numerical Linear Algebra

1.1 Closing Remarks

Numerical linear algebra is concerned with designing faster algorithms for

(i) Least Square Regression: $\min_{x\in\mathbb{R}^n} \|Ax-b\|_2$. Last lecture, we used a matrix S to be an Oblivious Space Embedding (OSE) to get an approximation $\approx \min_{x\in\mathbb{R}^n} \|S(Ax-b)\|_2$. Methods seen from last lecture can be extended to achieve run-time $O(\text{nnz}(A) + (\frac{d}{\epsilon})^{O(1)})$ where $\text{nnz}(A)$ denotes the number of non-zero entries in A. In some cases, dimension reduction matrix S can be chosen not to be OSE, but rather a specific construction dependent on A. This can achieve run-time $O(\log \frac{1}{\epsilon} \cdot (\text{nnz}(A) + d^{O(1)}))$.

(ii) Regression under different norms: $\min_{x\in R^n} \|Ax-b\|_l$ for $l \neq 2$. The 1-norm corresponds to the lasso-regression which promotes sparsity. An analogous construction of an OSE S for ℓ_1 with approximation factor $\alpha = d^{O(1)}$ can be used to achieve run-time $O(\frac{1}{\epsilon^{O(1)}} \cdot (\text{nnz}(A) + d^{O(1)}))$.

(iii) Rank-k Approximation or Matrix multiplication approximation: This problem can be approached similarly by applying a dimension reduction and solving the problem in lower dimensions. In some cases, one may consider applying a dimension reduction $S \in \mathbb{R}^{k\times n}$ to $A \in \mathbb{R}^{n\times n}$ from both directions (i.e. SAS^T).

2 Compressed Sensing

2.1 Problem Introduction

Compressed sensing is a problem originating from digital signal processing. Given a vector $x \in \mathbb{R}^n$, we design a "sensing matrix" $A \in \mathbb{R}^{m\times n}$ to make m linear measurements on x. In particular, $y = Ax$ and our goal is to recover x from y. We generally assume $m \ll n$, so A is not necessarily invertible. Therefore, x cannot be fully recovered, so we make the basic assumption that x is k-sparse in some basis. Suppose that x is k-sparse in some basis apart from the standard basis. Then, $x = \varphi \cdot z$ where z is the k-sparse representation of x in the new basis and φ is the change of basis linear transformation. It follows that

$$y = A \cdot \varphi \cdot z = A' \cdot z$$

where $A' = A \cdot \varphi$ is our new sensing matrix acting on k-sparse vector z.

There is a natural trade-off between the number of measurements made, m, and how well we can recover x. Note that the constants in our setting of m are an important and active area of research.

2.2 Formalization

In more precise terms, we assume the original signal x is *well-approximated* by a k-sparse vector. This motivates the following problem

$$L_0(y) = \operatorname*{arg\,min}_{\substack{x^* \in \mathbb{R}^n \\ Ax^* = y}} \|x^*\|_0$$

Where A is designed such that $L_0(y)$ approximately recovers x. For a fixed A, computing $L_0(y)$ is known to be NP-Hard. Work done in [CT05] motivates a relaxation of this problem to the following ℓ_1 minimization problem

$$L_1(y) = \operatorname*{arg\,min}_{\substack{x^* \in \mathbb{R}^n \\ Ax^* = y}} \|x^*\|_1$$

Observe that $L_1(y)$ is a linear programming problem that can be expressed as

$$\begin{aligned} \text{Minimize} \quad & \sum_{i=1}^{n} l_i \\ \text{Subject to} \quad & Ax^* = y \\ & -l_i \leq x_i^* \leq l_i \quad \forall i \in [n] \end{aligned}$$

which can be solved in polynomial time. Because x is *well-approximated* by a k-sparse vector, we would ideally like to find

$$x^* = \operatorname*{arg\,min}_{\substack{x' \in \mathbb{R}^n \\ \|x'\|_0 \leq k}} \|x - x'\|_1$$

Here x^* is the vector that keeps the largest k coordinates of x and zeroes out the others. Because we only have linear measurements of x, we resort to a more modest approximation via error

$$\mathrm{Err}_1^k(x) := \min_{\substack{x \in \mathbb{R}^n \\ \|x\|_0 \leq k}} \|x - x'\|_1$$

We will show that $x^* = L_1(y)$ satisfies

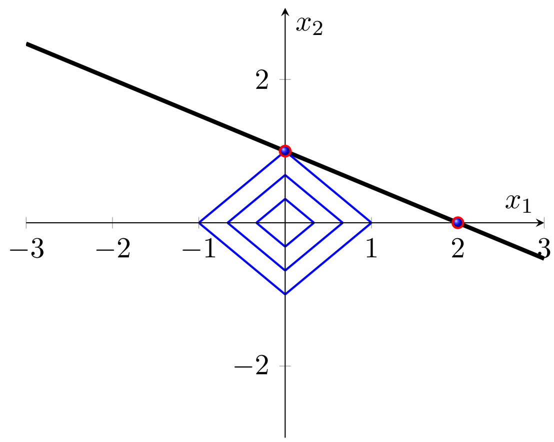
$$\|x - x^*\|_1 \leq c \cdot \mathrm{Err}_1^k(x) \tag{1}$$

where we typically take $c = 1 + \epsilon$ for $\epsilon > 0$, but sometimes c can also be a concrete constant.

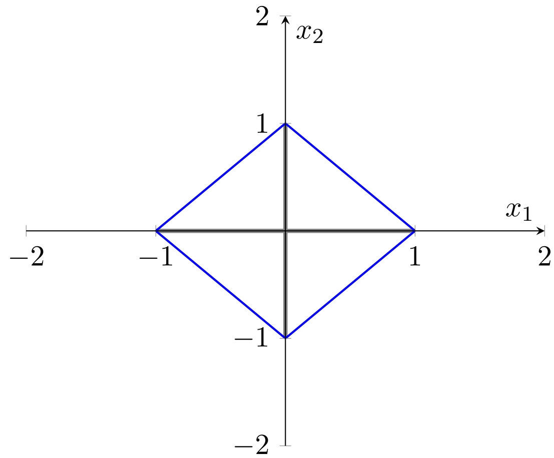
Theorem 1. *If A is i.i.d. $\mathcal{N}(0,1)$ with $m = O(k \log(\frac{n}{k}))$ then Eq.(1) holds for $c = O(1)$ with 90% probability.*

We can achieve $c = 1 + \epsilon$ for m a function of ϵ. Furthermore, $x^* = L_1(y)$ is not necessarily k-sparse. However, it is the case that if x is k-sparse, then $x^* = x$. A priori, this last point is not immediately clear in the setting of the ℓ_1 relaxation. At a high level, this is a result of careful choice of A. We will see in the following lectures that for A an RIP matrix, theorem 1 is true with probability 1. Then, it suffices to show that a random Gaussian matrix is RIP with high probability.

But why is it okay to relax $L_0(y)$ to $L_1(y)$? $L_1(y)$ is the "closest" convex relaxation of $L_0(y)$. Consider the following example, with $n = 2, m = 1, k = 1$



Here, the black line corresponds to $Ax^* = y$. The red points are the solution set of $L_0(y)$. The blue points correspond to those vectors x for which $\|x\|_1 = \epsilon$ for increasing values of ϵ. We note that the solution the $L_1(y)$ problem lies at the corner of ℓ_1 ball in this case. The following diagram illustrates the idea of "convex relaxation".



Here, the gray points represent those x with $\|x\|_0 = 1$ and $\|x\|_\infty \leq \epsilon$ and the blue represent the x with $\|x\|_1 \leq \epsilon$. Notice that the blue set is the smallest convex body containing the gray.