EE381V Course Project

Signal Recovery from Permuted Observations

Shanshan Wu (sw33323)

May 8th, 2015

1 Problem

We start with the following problem: let $s \in \mathbb{R}^n$ be an unknown n-dimensional real-valued signal, we observe all the values of s with an unknown order, the goal is to recover the original signal s. More formally, given the scrambled signal $g \in \mathbb{R}^n := Ps$, where P is an unknown permutation matrix with $\{0,1\}$ values satisfying that $P_{ij} = 1$ if and only if $g_i = s_j$, can we reconstruct s (exactly or approximately)?

The answer is "No" if no further assumptions are made for the signal s. In this case, the only condition we can use to recover s is that s is a sequence of samples coming from a continuous-time signal. However, according to the theory of interpolation [1], for any sequence of real numbers, we can always construct a continuous-time signal by computing the convolution with a sinc function (or by passing the samples through an ideal low-pass filter). As a result, for a general signal s, recovery from the permuted signal g is impossible.

To make the above recovery problem tractable, we add one restriction to the original signal s and assume that s is k-sparse in the frequency domain. In other words, let F be the n-by-n matrix for computing the Discrete Fourier Transform (DFT) of s, i.e., the entry of F in the i-th row and m-th column is defined as

$$F_{lm} = e^{-i2\pi(l-1)(m-1)/n}, \quad \forall l \in [n], \ \ m \in [n],$$

we assume that the frequency-domain representation of the original signal $S \in \mathbb{C}^n := Fs$ has at most k nonzero values. More formally, let $||v||_0$ denote the number of non-zero terms of vector v, then our problem can be formulated as follows:

Given: $||Fs||_0 \leq k$, and $g = Ps$, where P is an unknown permutation matrix,
Goal: recover $s \in \mathbb{R}^n$ (exactly or approximately).

Note that our problem of recovering from its permutations is related to but is different from the conventional *sparse signal recovery* problem [2, 3, 4, 5]. In the setting of a sparse signal recovery problem, the goal is to reconstruct a k-sparse signal x from a limited number of linear measurements $y := Ax$, where A is a known m-by-n measurement matrix.

Ideally, we want to solve the following optimization problem

$$\begin{aligned} &\text{minimize: } ||F\hat{P}g||_0 \\ &\text{subject to: } \hat{P} \text{ is a permutation matrix} \end{aligned} \tag{1}$$

We have encountered two issues regarding solving (1), as explained follows.

Complexity issue. Directly solving (1) is computationally costly as it would involve exhaustive search over the region of all permutation matrices (which has a size of $n!$). One computationally efficient approach is to relax (1) into a convex optimization problem: first relax the objective function to its convex envelope, which is l_1-norm; then relax the constraint to the convex hull of permutation matrices, which is the set of all doubly stochastic matrices (the Birkhoff polytope [7, 8]). Accordingly, (1) is relaxed to the following convex problem:

$$\begin{aligned} \text{minimize: } & ||F\hat{P}g||_1 \\ \text{subject to: } & \sum_j \hat{P}_{ij} = 1, \quad \forall i \in [n] \\ & \sum_i \hat{P}_{ij} = 1, \quad \forall j \in [n] \\ & \hat{P}_{ij} \geq 0, \quad \forall i, j \end{aligned} \tag{2}$$

The relaxed convex optimization problem (2) can be solved in polynomial time, however, the following theorem says that the optimal solution to (2) provides no information of what the original signal s looks like.

Claim 1. $\hat{P}^* = \frac{1}{n}\mathbf{1} \cdot \mathbf{1}^T$ *is the optimal solution to (2).*

Remark. Proof of this claim can be found in Appendix A. Claim 1 indicates that the optimal solution to (2) does not depend on s. In other words, there is no way to recover or estimate the original signal s by solving the convex relaxation problem (2).

Uniqueness issue. One natural question about the optimization problem (1) would be: is the original signal s the unique solution to (1)? Equivalently, we want to know that among all the possible permuted versions of s, whether s is the unique signal that has a k-sparse frequency-domain representation. The answer to this question is crucial to our problem, because if v is a rearrangement of s, and $||Fv||_0 \leq k$, then we cannot recover s by exactly solving (1). Unfortunately, the following claim says that we can always find such a signal v.

Claim 2. *Let x be a n-dimensional signal and Fx be its Fourier spectrum. Let σ be any integer that is invertible mod n, and for any integer $\tau \in [n]$, we can define a permutation matrix P such that $(Px)_i = x_{\sigma i+\tau}$, for $i \in [n]$. Then in the frequency domain, we have $|(FPx)_{\sigma i}| = |(Fx)_i|$.*

Remark. Proof of this claim can be found in [6]. It provides a way to permute a signal x in the time-domain, but at the same time retains the level of sparsity in the frequency domain. Therefore, there is no way to recover s from its permuted version by only knowing that s is k-sparse in the frequency domain. This indicates that the problem that we consider is not a valid problem. To continue with this project, we would like to consider a related problem, as defined in the next section.

2 Problem Redefined

We now redefine the problem as follows: given a vector $g \in \mathbb{R}^n$, and a matrix $M \in \mathbb{R}^{n\times n}$, where each entry $m_{ij} \sim N(0, 1/n)$ is i.i.d. Gaussian distributed, the goal is to find an optimal permutation matrix P such that $||MPg||_1$ is minimized. This corresponds to the following optimization program:

$$\begin{aligned} \text{OPT} = \text{minimize: } & ||MPg||_1 \\ \text{subject to: } & \sum_j P_{ij} = 1, \quad \forall i \in [n] \\ & \sum_i P_{ij} = 1, \quad \forall j \in [n] \\ & P_{ij} \in \{0,1\} \quad \forall i,j. \end{aligned} \tag{3}$$

Exhaustive search over all the permutation matrices would need $n!$ time. Our goal here is to find a polynomial-time α-approximation algorithm. The following claim indicates that a effective approximation algorithm should provide an α less than or equal to $\sqrt{n}$.

Claim 3. *Randomly choosing a permutation matrix gives a $\sqrt{n}$-approximation algorithm for $n \to \infty$.*

Proof. Since M is a random Gaussian matrix, the Norm preservation lemma (which is a key step to prove the Johnson-Linderstrauss lemma [9]) shows that $||MPg||_2 \to ||g||_2$ as $n \to \infty$. Furthermore, for any permutation matrix P and a sufficiently large n, we have

$$||g|| \approx ||MPg||_2 \leq ||MPg||_1 \leq \sqrt{n}||MPg||_2 \approx \sqrt{n}||g||_2.$$

This indicates that OPT$\geq ||g||_2$. And for a randomly chosen permutation matrix $\hat{P}$, we have $||M\hat{P}g||_1 \leq \sqrt{n}||g||_2 \leq \sqrt{n} \cdot \text{OPT}$. □

Similar to the optimization problem stated in Section 1, program (3) has a convex objective function, but an non-convex feasible set because of the integer constraint $P_{ij} \in \{0,1\}$. By relaxing it to $P_{ij} \geq 0$, we can obtain a convex optimization program relaxation, which looks exactly as in (2), except that the DFT matrix F is now changed to a random Gaussian matrix M. Suppose P^* is the optimal solution of the relaxed convex program, we then get a lower bound on OPT: OPT$\geq ||MP^*g||_1$. Note that P^* is a doubly stochastic matrix, but P^* does not necessarily equal $\frac{1}{n}\mathbf{1} \cdot \mathbf{1}^T$ since Claim 1 depends on the special structure of F. Then a key problem is: how to get a permutation matrix from P^* and bound the corresponding approximation error?

Idea 1: add a regularization term. This technique is commonly used in machine learning problems (e.g., Lasso and Ridge regression) to prevent overfitting. The idea is to add a term to the objective function so as to penalize solutions with undesired properties. For instance, in the regularized regression problem, a term of $\lambda|| \cdot ||$ is added to the loss function so that the optimal solution does not have a large norm. The parameter λ can be tuned so as to balance between how much we penalize a large norm and how much we want to deviate from the original loss function.

The same technique may be applied here. Since permutation matrices have a larger Frobenius norm than any doubly stochastic matrices (note that their l_1 norms are the same), we want to penalize solutions with a small Frobenius norm. In other words, we want to change the objective function of (3) to $||MPg||_1 - \lambda||P||_F$. However, the main problem is that the regularized objective function is now non-convex in P. Actually, we can argue that any regularization method would resulted in a non-convex objective function: since permutation matrices are the extreme points of a Birkhoff polytope [7, 8], any regularized term would give a small value at the extreme points and a large value inside the polytope, which corresponds to a concave function.

Idea 2: round P^* to its nearest permutation matrix. This idea is quite intuitive. Finding a permutation matrix P that is closest to P^* can be formulated as the following optimization problem:

$$\begin{aligned} &\text{maximize: } \sum_{i,j} P_{ij} P^*_{ij} \\ &\text{subject to: } P \text{ is a permutation matrix.} \end{aligned} \tag{4}$$

Note that (4) is indeed the *assignment problem*: suppose there are n agents and n tasks, the benefit of assigning agent i to task j is given by P^*_{ij}, the goal is to assign exactly one agent (with no overlapping) to each task such that the total benefits are maximized. The assignment problem can be solved exactly in polynomial time by solving the relaxed LP or by applying other heuristic algorithms such as the Hungarian algorithm [10].

Although (4) can be exactly solved in polynomial time, we find it hard to derive a good bound for $||MPg||_1$ using $||MP^*g||_1$, given that P is the optimal solution to (4). This is because $||MPg||_1$ is not a linear function of P; and its value depends on the interaction between M, P, and g. Therefore, finding a permutation matrix by choosing as many as possible large-valued entries of P^* does not guarantee the closeness between $||MPg||_1$ and $||MP^*g||_1$.

Idea 3: sort g in the same order as P^*g. The intuition is that since $||MP^*g||_1$ is the optimal value to the relaxed convex program, we may want to rearrange g in such a way that Pg is most similar to P^*g. This algorithm gives us an $O(\sqrt{n})$-approximation factor. Before we prove this approximation factor, let us fist introduce two lemmas.

Lemma 1. *Rearranging g in the same order as P^*g provides the optimal solution to the following problem:*

$$\begin{aligned} &\textit{maximize: } (Pg)^T(P^*g) \\ &\textit{subject to: } P \textit{ is a permutation matrix.} \end{aligned} \tag{5}$$

Proof. This lemma follows directly from the rearrangement inequality, which says that if two vectors $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$ satisfy $x_1 \leq ... \leq x_n$ and $y_1 \leq ... \leq y_n$, then $(Px)^Ty \leq x^Ty$, for any permutation matrix P. Therefore, rearranging g in the same order as P^*g will give the maximum value of $(Pg)^T(P^*g)$ over all permutation matrices. □

Lemma 2. *For any vector $g \in \mathbb{R}^n$ and any doubly stochastic matrix P_{ds}, $||P_{ds}g||_2 \leq ||g||_2$.*

Proof. According to the Birkhoff-von Neumann theorem [7, 8], any doubly stochastic matrix can be expressed as a convex combination of permutation matrices, i.e., $P_{ds} = \sum_i \alpha_i P_i$, where $\alpha_i > 0$, $\sum_i \alpha_i = 1$, and P_i's are permutation matrices. Since $||\cdot||_2$ is a convex function, applying Jensen's inequality gives

$$||P_{ds}g||_2 \leq \sum_i \alpha_i ||P_i g||_2 = ||g||_2.$$

$\square$

Claim 4. *Suppose P is the optimal solution to (5), and let $\beta = (Pg)^T(P^*g)$, then it gives a $\left(1 + \sqrt{n} \cdot \sqrt{\frac{4}{\pi}(1 - \frac{\beta}{||g||_2})}\right)$-approximation algorithm when n approaches infinity.*

Remark. Proof of this claim can be found in Appendix B. The dependence on β can be explained as follows. Since β is defined as the optimal value of (5), using Cauchy-Schwarz inequality and Lemma 2, we have $\beta \leq ||Pg||_2||P^*g||_2 \leq ||g||_2^2$, where equality is achieved only when $Pg = P^*g$. Therefore, $\beta/||g||_2^2$ approaches 1 if P^* is close to a permutation matrix; $\beta/||g||_2^2$ approaches 0 if P^* is far from any permutation matrix (i.e., P^* is close to $\frac{1}{n}\mathbf{1} \cdot \mathbf{1}^T$). In other words, $\beta/||g||_2^2$ characterizes the deviation between the optimal solutions of the original non-convex program (4) and its convex relaxation.

3 Conclusion

In this project we started with a new and interesting problem: can we recover a signal (exactly or approximately) from its time-domain permuations, given that the original signal is sparse in the frequency domain? However, we found that the answer is no because for any signal there exits at least $O(n \cdot \phi(n))$ permutations that preserves the same level of sparsity in the frequency domain, where $\phi(n)$ is the Euler's totient function. Since the original problem is not a valid problem, we then considered another related problem: find an optimal permutation P of a given vector g such that $||MPg||_1$ is minimized, where m_{ij} is i.i.d. $N(0, 1/n)$ distributed. We tried three different methods to get an approximated solution. Although the first two do not work, we showed that the third one give an approximation factor of $O(\sqrt{n} \cdot \sqrt{\frac{4}{\pi}(1 - \frac{\beta}{||g||_2})})$, where the value of $\beta/||g||_2^2 \in [0, 1]$ depends on M and g, and is a measure of how close the optimal solution of the relaxed convex optimization program is to a permutation matrix.

References

[1] A. V. Oppenheim, R. W. Schafer, *Discrete-Time Signal Processing*, 3d. ed. Pearson, 2010.

[2] E. Cands, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[3] E. J. Cands and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, Sept. 2006.

[5] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[6] A. Gilbert, M. Muthukrishnan, and M. Strauss, "Improved time bounds for near-optimal space fourier representations," *SPIE Conference, Wavelets*, 2005.

[7] G. Birkhoff, "Three Observations on Linear Algebra," *Revi. Univ. Nac. Tucuman*, ser A, 5, pp. 147?151, 1946.

[8] J. von Neumann, "A certain zero-sum two-person game equivalent to the optimal assignment problem," in *Contributions to the theory of games*, vol. 2, ed. by H. W. Kuhn, and A. W. Tucker. Princeton University Press, Princeton, New Jersey, 1953.

[9] S. Kakade and G. Shakhnarovich, "Random Projections," lecture notes of *CMSC 35900 Large Scale Learning*, Spring 2009.

[10] H. W. Kuhn, "The Hungarian Method for the assignment problem", *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.

A Proof of Claim 1

Claim: $\hat{P}^* = \frac{1}{n}\mathbf{1} \cdot \mathbf{1}^T$ is the optimal solution to (2).

Proof. To prove the above claim, we need to make use of the special structure of the DFT matrix F. Specifically, two properties of F are used in our proof: 1) the first row of F is $\mathbf{1}^T$, i.e., an all-ones vector; 2) $\sum_j F_{ij} = 0$, for $i = 2, 3, ..., n$, i.e., the sum of elements in each row (except the first row) of F equals 0.

Let $(F\hat{P}g)_i$ be the i-th element of vector $F\hat{P}g$, then we can get a lower bound for $||F\hat{P}g||_1$:

$$||F\hat{P}g||_1 = \sum_i |(F\hat{P}g)_i| \geq |(F\hat{P}g)_1| = |\mathbf{1}^T\hat{P}g| = |\sum_i \sum_j \hat{P}_ijg_j| = |\sum_j g_j \sum_i \hat{P}_ij| = |\sum_j g_j|.$$

On the other hand, the above lower bound is achievable if we substitute a special doubly stochastic matrix $\hat{P}^*$ with $\hat{P}^*_{ij} = \frac{1}{n}$ for all i and j. In this case, $\hat{P}^*g = (\sum_k g_k/n) \cdot \mathbf{1}$. Since the row sum of F equals 0 for all rows except the first row, we then have

$$||F\hat{P}^*g||_1 = |(\sum_k g_k/n)| \cdot ||F \cdot \mathbf{1}||_1 = |(\sum_k g_k/n)| \cdot |(\sum_j F_{1j})| = |(\sum_k g_k/n)| \cdot n = |\sum_k g_k|.$$

Therefore, $\hat{P}^* = \frac{1}{n}\mathbf{1} \cdot \mathbf{1}^T$ is the optimal solution to (2), and the corresponding optimal value is $|\sum_j g_j|$, i.e., the DC term of signal g. □

B Proof of Claim 4

Claim: Suppose P is the optimal solution to (5), and let $\beta = (Pg)^T(P^*g)$, then it gives a $\left(1 + \sqrt{n} \cdot \sqrt{\frac{4}{\pi}(1 - \frac{\beta}{||g||_2})}\right)$-approximation algorithm when n approaches infinity.

Proof. For any real numbers x and y, $|x| - |y| \leq |x + y|$ holds, so we have

$$\frac{1}{n}(||MPg||_1 - ||MP^*g||_1) \leq \frac{1}{n}||MPg - MP^*g||_1 = \frac{1}{n}\sum_i |\sum_j m_{ij}(Pg - P^*g)|. \quad (6)$$

Because $m_{ij} \sim N(0, 1/n)$ is i.i.d. Gaussian distributed, $\sum_j m_{ij}(Pg - P^*g)$ is also Gaussian distributed with mean 0 and variance $\sigma^2 = ||Pg - P^*g||_2^2/n$. Then for any i, $|\sum_j m_{ij}(Pg - P^*g)|$ follows a half-normal distribution with mean $\sigma\sqrt{2/\pi}$ and variance $\sigma^2(1-2/\pi)$. Therefore, $\frac{1}{n}\sum_i |\sum_j m_{ij}(Pg - P^*g)|$ is the average of n i.i.d. half-normal distributed random variables. By the law of large numbers, we have for any $\epsilon > 0$, with n sufficient large,

$$P(\frac{1}{n}\sum_i |\sum_j m_{ij}(Pg - P^*g)| \leq \sigma\sqrt{2/\pi} + \epsilon) \geq 1 - \epsilon. \quad (7)$$

Suppose P is the optimal solution to (5), and let $\beta = (Pg)^T(P^*g)$, then we have

$$||Pg - P^*g||_2^2 = ||Pg||_2^2 + ||P^*g||_2^2 - 2 \cdot (Pg)^T(P^*g) \leq 2||g||_2^2 - 2 \cdot \beta, \quad (8)$$

where the inequality follows from Lemma 2. Substituting (7) and (8) into (6), we get that as n approaches infinity,

$$||MPg||_1 - ||MP^*g||_1 \leq n\sigma\sqrt{2/\pi} = \sqrt{n}||Pg - P^*g||_2\sqrt{2/\pi} \leq \sqrt{n}\sqrt{4/\pi(||g||_2^2 - \beta)}. \quad (9)$$

We have two lower bounds for OPT: 1) P^* is the optimal solution for the relaxed optimization program, so OPT $\geq ||MP^*g||_1$; 2) in the proof of Claim 3, we have shown that OPT $\geq ||g||_2$. Combining the two lower bounds with (9) gives us the desired approximation factor:

$$||MPg||_1 \leq ||MP^*g||_1 + \sqrt{n}\sqrt{\frac{4}{\pi}(1 - \frac{\beta}{||g||_2^2})} \cdot \text{OPT} \leq \left(1 + \sqrt{n} \cdot \sqrt{\frac{4}{\pi}(1 - \frac{\beta}{||g||_2})}\right) \cdot \text{OPT}. \quad (10)$$

$\square$