Permutation Recovery by Linear Decoding

Optimality and Asymptotics

Minoh Jeong*, Alex Dytso†, Martina Cardone*, and H. Vincent Poor‡

* University of Minnesota, Electrical and Computer Engineering,

† New Jersey Institute of Technology, Department of Electrical and Computer Engineering,

‡ Princeton University, Department of Electrical Engineering.

Contact Information:

Email: jeong316@umn.edu, alex.dytso@njit.edu, mcardone@umn.edu, poor@princeton.edu

Abstract

The problem of recovering data permutations from noisy observations is becoming a common task of modern communication and computing systems. We investigate the following question on noisy data permutation recovery: Given a noisy observation of a permuted data set, according to which permutation was the original data sorted? Our focus is on scenarios where data is generated according to a given distribution, and the noise is additive Gaussian with zero-mean and a given covariance matrix. We pose this problem within a hypothesis testing framework, and our objective is two-fold. First, we characterize sufficient conditions on the noise covariance matrix that ensure that the optimal decision criterion (that is, the criterion that minimizes the probability of error) has a complexity that is at most polynomial in the data size. Towards this end, we focus on the linear regime, that is, the optimal decision criterion consists of computing a permutation-independent linear function of the noisy observation followed by a sorting operation. We find necessary and sufficient conditions for the optimality of such a linear regime. Second, we characterize a general expression for the probability of error and study its rate of convergence in the linear regime for some practically relevant asymptotic regimes, such as when the data size grows to infinity, and in the high and low noise regimes.

Problem Setting

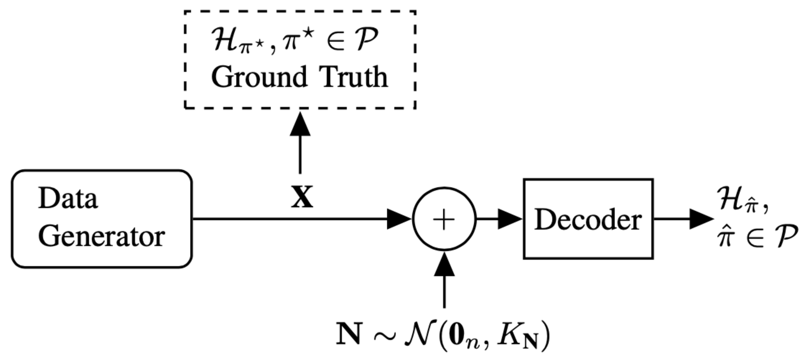

Figure 1: Graphical representation of the considered framework.

$$\mathbf{X} + \mathbf{N} = \mathbf{Y}, \tag{1}$$

- $\mathbf{X}$ is the unknown n-dimensional exchangeable random vector;
- $\mathbf{N} \sim \mathcal{N}(\mathbf{0}_n, K_\mathbf{N})$ is Gaussian noise;
- $\mathbf{Y}$ is the observation.

Question:

Given the observation $\mathbf{y}$, according to which permutation was $\mathbf{x}$ sorted?

Hypothesis:

$$\mathcal{H}_\pi = \{\mathbf{x} \in \mathbb{R}^n : x_{\pi_1} \le x_{\pi_2} \le \cdots \le x_{\pi_n}\},\ \pi \in \mathcal{P}, \tag{2}$$

where π indicates the permutation, and $\mathcal{P}$ is the set of all permutations.

Example: Let $n = 3$, then we have $\mathcal{H}_\pi, \pi \in \mathcal{P}$ defined as

$$\mathcal{H}_{\{1,2,3\}} : X_1 \le X_2 \le X_3, \quad \mathcal{H}_{\{1,3,2\}} : X_1 \le X_3 \le X_2,$$
$$\mathcal{H}_{\{2,1,3\}} : X_2 \le X_1 \le X_3, \quad \mathcal{H}_{\{2,3,1\}} : X_2 \le X_3 \le X_1,$$
$$\mathcal{H}_{\{3,1,2\}} : X_3 \le X_1 \le X_2, \quad \mathcal{H}_{\{3,2,1\}} : X_3 \le X_2 \le X_1.$$

Optimal decision rule (Optimal decoder):

$$\mathcal{H}_{\hat\pi} : \hat\pi = \operatorname*{argmin}_{\pi \in \mathcal{P}} \{\Pr(\mathcal{H}_\pi \ne \mathcal{H}_{\pi^\star})\}, \tag{3}$$

where $\pi^\star$ denotes the true permutation of $\mathbf{x}$.

Decision region:

$$\mathcal{R}_{\pi,K_\mathbf{N}} = \{\mathbf{y} \in \mathbb{R}^n : \hat\pi = \pi\},\ \forall \pi \in \mathcal{P}. \tag{4}$$

Linear decision region (Linear decoder):

$$\mathcal{R}_{\pi,K_\mathbf{N}} = A\mathcal{H}_\pi + \mathbf{b},\ \forall \pi \in \mathcal{P}, \tag{5}$$

where $A \in \mathbb{R}^{n\times n}$ and $\mathbf{b} \in \mathbb{R}^n$ are fixed for all $\pi \in \mathcal{P}$.

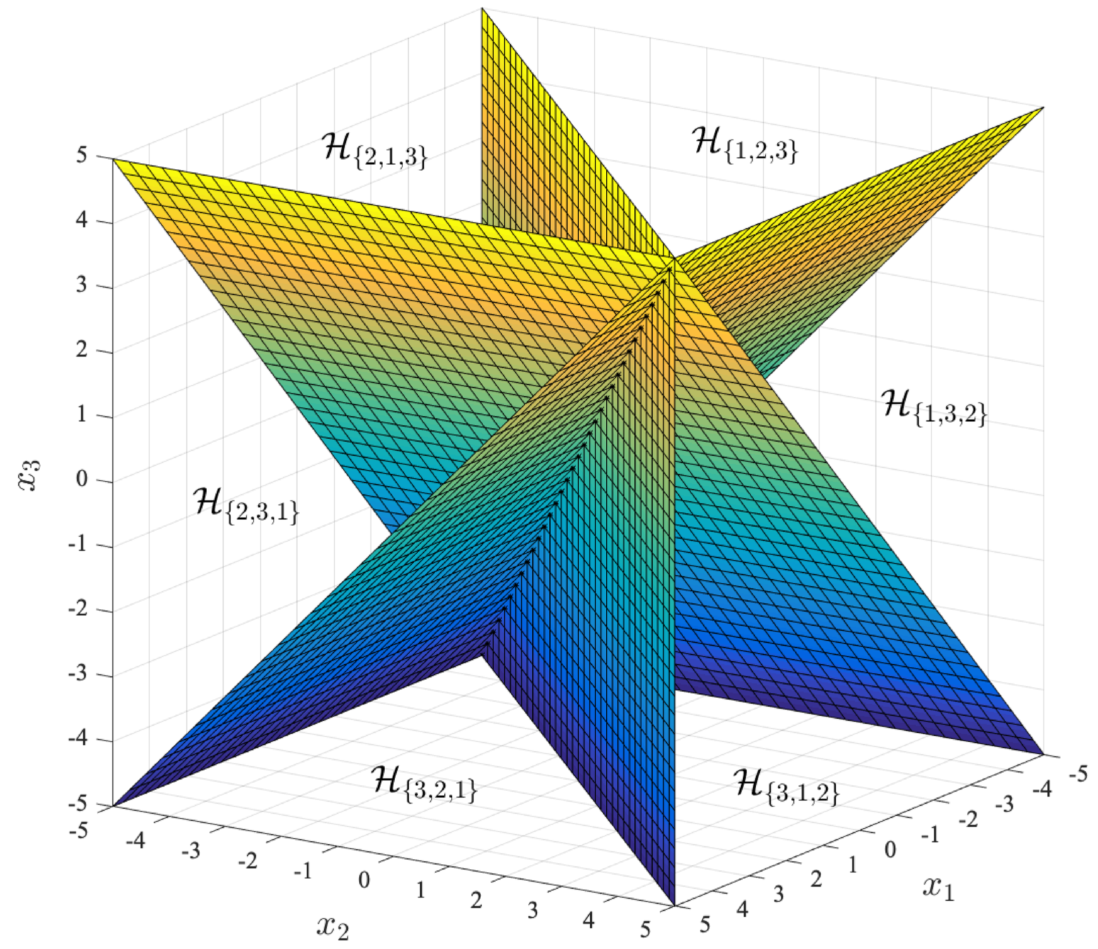

Figure 2: $\mathcal{H}_\pi$'s defined in (2) when $n = 3$.

Summary of Main Results

Main Results	
· **Theorem 1**	· **Theorem 2**
Optimality condition for linear decoder when $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$ [1]	Characterization of P_e when $n \to \infty$ when $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$ [2]
· **Theorem 3**	· **Theorem 4**
Characterization of P_e when $\sigma \to 0$ [3]	Characterization of P_e when $\sigma \to \infty$ [3]

Table 1: Summary of main results

Theorem 1. *[1] Assume that $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$. Then, the following conditions are equivalent:*

1. *$\mathcal{R}_{\pi,K_\mathbf{N}}$ is a permutation-independent linear transformation of $\mathcal{H}_\pi$;*
2. *$\mathbf{0}_n \in \bigcap_{\pi\in\mathcal{P}} \mathcal{R}_{\pi,K_\mathbf{N}}$;*
3. *The ellipsoid $\left(K_\mathbf{N}^{-1} + I_n\right)^{-\frac{1}{2}} \mathcal{B}^n(\mathbf{0}_n, 1)$ projected onto the hyperplane $\mathcal{W} = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{1}_n^T\mathbf{x} = 0\}$ is an $(n-1)$-dimensional ball of radius γ for some constant $\gamma \in (0,1)$;*
4. *Let $\mathcal{Q} = \left\{Q \in \mathcal{SO}(n) : \mathbf{q}_n = \frac{1}{\sqrt{n}}\mathbf{1}_n\right\}$, where $\mathcal{SO}(n)$ is the set of $n \times n$ orthonormal matrices, and $\mathbf{q}_n$ is the n-th column of Q. Then,*
$$\left(K_\mathbf{N}^{-1} + I_n\right)^{-1} = Q \begin{bmatrix} \gamma I_{n-2} & 0_{n-2\times 2} \\ 0_{2\times n-2} & S \end{bmatrix} Q^T, \tag{6}$$
where $Q \in \mathcal{Q}$, $S = \left[\begin{smallmatrix} \gamma & v \\ v & a \end{smallmatrix}\right]$ and $\gamma \in (0,1)$, $a \in (0,1)$, $v \in \mathbb{R}$ such that $v^2 < \min\{a\gamma, (1-a)(1-\gamma)\}$; and
5. *$\mathcal{R}_{\pi,K_\mathbf{N}} = (K_\mathbf{N} + I_n)\,\mathcal{H}_\pi$, for all $\pi \in \mathcal{P}$.*

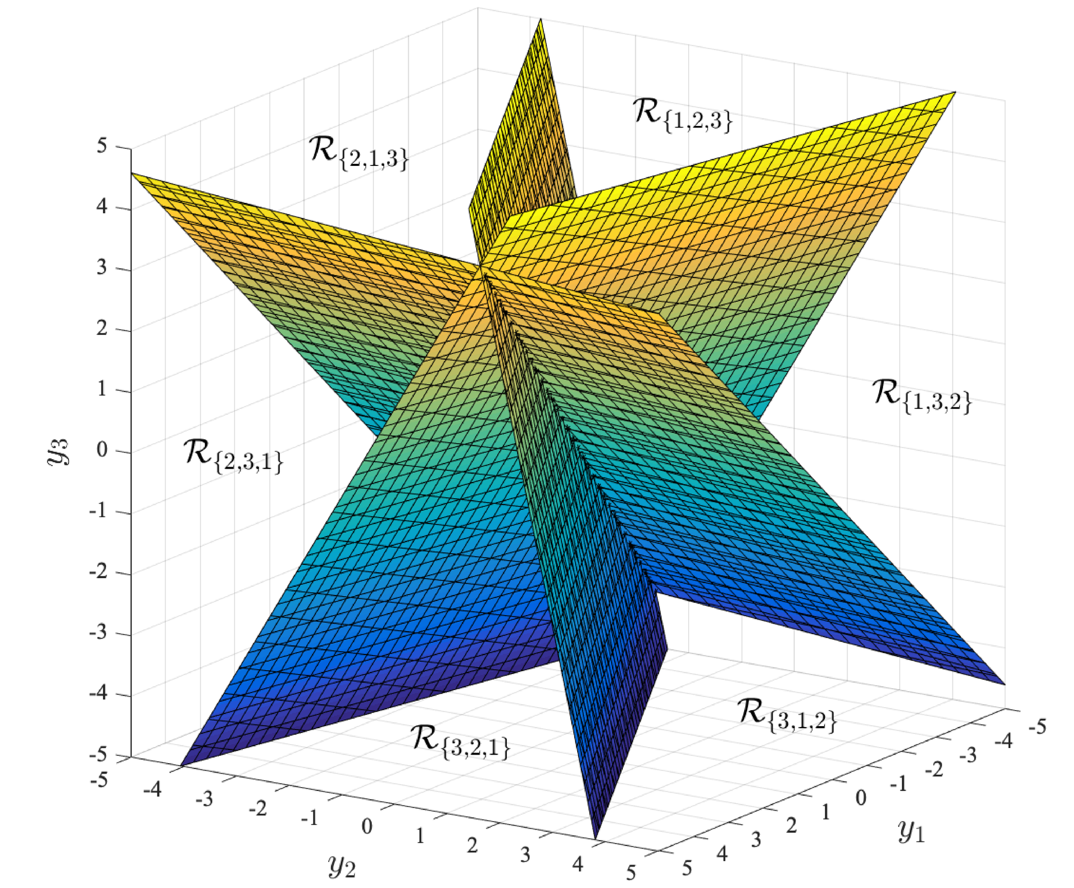

Figure 3: An example of $K_\mathbf{N}$ that induces the linear regime can be obtained by considering $n = 3$ and $(\gamma, a, v) = (0.5, 0.5, 0.2)$ in (6) in Theorem 1.

Theorem 2. *[2] Assume that $\mathbf{X} \sim \mathcal{N}(\mathbf{0}_n, I_n)$ and $\mathbf{N} \sim \mathcal{N}(0_n, \sigma^2 I_n)$. Then, the probability of correctness can be upper and lower bounded as*

$$\frac{1}{n!} \le P_c \le \frac{1}{n!}\frac{\|A\|^{2n}}{\sigma^n}, \tag{7a}$$

where $A = \begin{bmatrix} I_n & 0_{n\times n} \\ I_n & \sigma I_n \end{bmatrix} \in \mathbb{R}^{2n}$ and

$$\|A\| = \left(\frac{(\sigma^4+4)^{\frac{1}{2}}}{2} + \frac{\sigma^2}{2} + 1\right)^{\frac{1}{2}}. \tag{7b}$$

Consequently,

$$\lim_{n\to\infty} \frac{\log\frac{1}{P_c}}{\log(n!)} = 1. \tag{7c}$$

Theorem 3. *[3] Let $\mathbf{X}$ consist of n i.i.d. random variables generated according to X. Let X' be an independent copy of X and assume that*

$$f_{X-X'}(x) < \infty,\ \forall x \in \mathbb{R}. \tag{8}$$

Then,

$$\lim_{\sigma\to 0} \frac{P_e(\sigma)}{\sigma} = \sum_{i=1}^{n-1} \frac{f_{W_i}(0^+)}{\sqrt{\pi}}, \tag{9}$$

where $W_i = X_{i+1:n} - X_{i:n},\ i \in [1:n-1]$.

Example 1. Consider $X \sim \text{Unif}(a,b),\ 0 \le a < b < \infty$. Then,

$$\lim_{\sigma\to 0} \frac{P_e(\sigma)}{\sigma} = \frac{n(n-1)}{(b-a)\sqrt{\pi}}. \tag{10}$$

Example 2. Consider $X \sim \text{Exp}(\lambda),\ \lambda > 0$. Then,

$$\lim_{\sigma\to 0} \frac{P_e(\sigma)}{\sigma} = \frac{\lambda n(n-1)}{2\sqrt{\pi}}. \tag{11}$$

Example 3. Consider $X \sim \mathcal{N}(0,1)$. Then,

$$\frac{\sqrt{2}n(n-1)}{6\pi} \le \lim_{\sigma\to 0} \frac{P_e(\sigma)}{\sigma} \le \frac{n(n-1)}{\sqrt{2\pi}}. \tag{12}$$

Theorem 4. *[3] Let $\mathbf{X}$ be an exchangeable random vector such that $\mathbb{E}[\|\mathbf{X}\|] < \infty$. Then,*

$$\lim_{\sigma\to\infty} \frac{P_e(\infty) - P_e(\sigma)}{\frac{1}{\sigma}} = \frac{1}{\sqrt{2\pi}} \sum_{i=1}^{n-1} \alpha_i \mathbb{E}[W_i], \tag{13}$$

where $W_i = X_{i+1:n} - X_{i:n},\ i \in [1:n-1]$ and

$$\alpha_i = \frac{\text{Vol}\left(\mathcal{E}(\mathbf{0}_{n-1}, i) \cap \mathcal{H}_{[1:n-1]}\right)}{\text{Vol}\left(\mathcal{B}^{n-1}(\mathbf{0}_{n-1}, 1)\right)}, \tag{14}$$

where $\mathcal{B}(\mathbf{0}_{n-1}, 1)$ is the $(n-1)$-dimensional unit ball centered at the origin, and $\mathcal{E}(\mathbf{0}_{n-1}, i)$ is the $(n-1)$-dimensional ellipsoid centered at the origin with unit radii along standard axes except a $\frac{1}{\sqrt{2}}$ radius along the i-th axis.

Proposition 1. *[3] In the high noise regime, the convergence rate of the probability of correctness can be bounded as*

$$\frac{\mathbb{E}[R_n]}{\sqrt{\pi}(n-1)!2^{\frac{n}{2}}} \le \lim_{\sigma\to\infty} \frac{P_e(\infty) - P_e(\sigma)}{\frac{1}{\sigma}} \le \frac{\mathbb{E}[R_n]}{\sqrt{2\pi}(n-1)!},$$

where $R_n = X_{n:n} - X_{1:n}$.

Example 1. Consider $X \sim \text{Unif}(a,b),\ 0 \le a < b < \infty$. Then,

$$\mathbb{E}[R_n] = (b-a)(n-1)(n+1)^{-1}. \tag{15}$$

Example 2. Consider $X \sim \text{Exp}(\lambda),\ \lambda > 0$. Then,

$$\mathbb{E}[R_n] = \frac{1}{\lambda}\sum_{k=1}^{n-1}\frac{1}{k} = \Theta\left(\frac{1}{\lambda}\log(n)\right). \tag{16}$$

Example 3. Let X be γ^2-sub-Gaussian. Then,

$$\mathbb{E}[R_n] \le 2\sqrt{2\gamma^2\log(n)}. \tag{17}$$

References

[1] M. Jeong, A. Dytso, M. Cardone, and H. V. Poor, "Recovering data permutations from noisy observations: The linear regime," *IEEE Journal on Selected Areas in Information Theory*, vol. 1, no. 3, pp. 854–869, 2020.

[2] ——, "Recovering structure of noisy data through hypothesis testing," in *Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT)*, June 2020, pp. 1307–1312.

[3] M. Jeong, A. Dytso, and M. Cardone, "Retrieving data permutations from noisy observations: High and low noise asymptotics," *arXiv:2105.03015*, 2021.

[4] K. Nomakuchi and T. Sakata, "Characterizations of the forms of covariance matrix of an elliptically contoured distribution," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 205–210, 1988.

[5] ——, "Characterization of conditional covariance and unified theory in the problem of ordering random variables," *Annals of the Institute of Statistical Mathematics*, vol. 40, no. 1, pp. 93–99, 1988.

Acknowledgements

The work of M. Jeong and M. Cardone was supported in part by the U.S. National Science Foundation under Grant CCF-1849757. The work of A. Dytso and H. V. Poor was supported in part by the U.S. National Science Foundation under Grant CCF-1908308.