J. Appl. Math. & Informatics Vol. 42(2024), No. 1, pp. 149 - 158
https://doi.org/10.14317/jami.2024.149

# FINDING EXPLICIT SOLUTIONS FOR LINEAR REGRESSION WITHOUT CORRESPONDENCES BASED ON REARRANGEMENT INEQUALITY†

MIJIN KIM, HYUNGU LEE, HAYOUNG CHOI*

Abstract. A least squares problem without correspondences is expressed as the following optimization:

$$\min_{\mathbf{\Pi}\in\mathcal{P}_m,\ \boldsymbol{x}\in\mathbb{R}^n} \|\mathbf{A}\boldsymbol{x}-\mathbf{\Pi}\mathbf{y}\|,$$

where $\mathbf{A}\in\mathbb{R}^{m\times n}$ and $\mathbf{y}\in\mathbb{R}^m$ are given. In general, solving such an optimization problem is highly challenging. In this paper we use the rearrangement inequalities to find the closed form of solutions for certain cases. Moreover, despite the stringent constraints, we successfully tackle the nonlinear least squares problem without correspondences by leveraging rearrangement inequalities.

AMS Mathematics Subject Classification : 62J05, 90C26, 49N30.
*Key words and phrases* : Permuted linear model, shuffled regression, rearrangement inequality, least squares, optimizatoin.

## 1. Introduction

Linear regression is a statistical modeling technique used to understand the relationship between a dependent variable and one or more independent variables. Linear regression has various applications in diverse fields such as economics, finance, social sciences, and engineering [8], [29], [3], [5], [18], [15], [14], [28], [20]. It is commonly used for prediction [7] and forecasting [10]. The main objective of linear regression is to find the best-fitting line that minimizes the overall difference between the observed data points and the predicted values. The line is

Received October 21, 2023. Revised December 12, 2023. Accepted January 5, 2024.
*Corresponding author.
†This research was supported by Kyungpook National University Development Project Research Fund, 2020

determined by estimating the coefficients or weights assigned to each independent variable, which represent the magnitude and direction of their impact on the dependent variables. See reference [16] and references therein.

To achieve this, linear regression employs the ordinary least squares problem which is written as the optimization problem [2]:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \|\mathbf{A}\boldsymbol{x} - \mathbf{y}\|, \tag{1.1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{y} \in \mathbb{R}^m$ are given. Additionally, it serves as a fundamental building block for more complex machine learning algorithms and techniques [13].

The least squares problem in (1.1) is solved by considering the correspondence between input and output. However, in many cases output of datasets are massed up, especailly in the absence of correspondence of output. *A least squares problem without correspondences* refers to the method of solving the least squares problem for given data without any correspondence relationship. Specifically, it is expressed as

$$\min_{\boldsymbol{\Pi}\in\mathcal{P}_m,\ \boldsymbol{x}\in\mathbb{R}^n} \|\mathbf{A}\boldsymbol{x} - \boldsymbol{\Pi}\mathbf{y}\|. \tag{1.2}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{y} \in \mathbb{R}^m$ are given. Here, $\mathcal{P}_m$ is the set of $m \times m$ permutation matrices. The problem in (1.2) is called a *linear regression without correspondences* [26], *shuffled linear regression* [26], or *permuted linear model* [24].

The problem (1.2) appears in many different fields. For example, it can be used in simultaneous pose-and-correspondence determination [4], signaling with identical tokens [22], relative dating from archaeological samples [21], pose-correspondence estimation cell tracking [23], genome-assembly [12], header-free communication [19], e.g., for data de-anonymization and low-latency communications in Internet-Of-Things networks [24]. Literature reviews on this topic can be found in [17], [27].

The main focus of previous research has been on developing a theoretical understanding of the conditions necessary for the unique recovery of $\boldsymbol{x}$ or $\boldsymbol{\Pi}$. [27] showed that if $\mathbf{A} \in \mathbb{R}^{m\times n}$ in (1.2) is randomly drawn from any continuous probability distribution, then an optimal solution $\boldsymbol{x}$ can be uniquely recovered with probability of 1, provided that $m \geq 2n$. However, finding an optimal solution involves the verification of whether the linear system $\boldsymbol{\Pi}\mathbf{y} = \mathbf{A}\boldsymbol{x}$ remains consistent for every permutation $\boldsymbol{\Pi}$. This verification process leads to a computational complexity of $\mathcal{O}((m)!mn^2)$. The algorithm in [11] can efficiently reduce computational complexity in situations where the measured values are noiseless. The prevailing approach for practical deployment appears to be the one that solves (1.2) through alternating minimization [1]. In this method, an estimate for $\boldsymbol{x}$ is computed based on a given estimation for $\boldsymbol{\Pi}$ through sorting, and vice versa. However, this approach generally works effectively only when data is partially shuffled. If we consider a more general scenario where the signal is unrestricted, it has been demonstrated that having $m \geq n+1$ measurements is adequate [26]. These findings were subsequently extended to encompass various linear transformations, going beyond permutations and down-sampling techniques, as discussed in [25].

However, to the best of our knowledge, there have not been studied to find an exact solution. In this paper, we use the rearrangement inequalities to find an exact solution for a least squares problem without correspondences.

## 2. Main results

Denote the set of real numbers as $\mathbb{R}$ and the set of nonnegative real numbers as $\mathbb{R}_+$ and the set of positive real numbers as $\mathbb{R}_{++}$. We let $\mathcal{P}_m$ denote the set of permutations on $\{1,\dots,m\}$ as well as the corresponding set of permutation matrices. For a given vector $\mathbf{v} \in \mathbb{R}^m$, we denote the vector with entries of $\mathbf{v}$ sorted in non-decreasing (resp. non-increasing) order as $\mathbf{v}^\uparrow$ (resp. $\mathbf{v}^\downarrow$). The usual inner product is denoted by $\langle \cdot, \cdot \rangle$.

We first give an interesting result which is a vital tool in this paper. The rearrangement inequality was introduced by Hardy, Littlewood, Polya as follows [9].

**Lemma 2.1** (Rearrangement inequality)**.** For any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$,

$$\langle \mathbf{u}^\uparrow, \mathbf{v}^\downarrow \rangle \leq \langle \mathbf{u}, \mathbf{v} \rangle \leq \langle \mathbf{u}^\uparrow, \mathbf{v}^\uparrow \rangle.$$

Using the rearrangement inequality, we solve the least squares problem without correspondences. The first result is obtained for the simplest setting.

**Theorem 2.2.** Let $\mathbf{A} \in \mathbb{R}^{m\times 1}$ with $\mathbf{A} = \mathbf{A}^\uparrow$ and $\mathbf{y} \in \mathbb{R}^m$ be given. Then an optimal solution of

$$\min_{\mathbf{\Pi}\in\mathcal{P}_m,\ x\in\mathbb{R}} \|\mathbf{A}x - \mathbf{\Pi}\mathbf{y}\| \tag{2.1}$$

is explicitly written as follows:

(i) if $\langle \mathbf{A}^\uparrow, \mathbf{y}^\uparrow \rangle^2 \geq \langle \mathbf{A}^\uparrow, \mathbf{y}^\downarrow \rangle^2$, then $(\mathbf{\Pi}, x) = \left(\widetilde{\mathbf{\Pi}}, \frac{\langle \mathbf{A}, \widetilde{\mathbf{\Pi}}\mathbf{y}\rangle}{\|\mathbf{A}\|^2}\right)$ such that $\widetilde{\mathbf{\Pi}}\mathbf{y} = \mathbf{y}^\uparrow$. In such case, the minimum value is $\|\mathbf{y}\| \sin\theta^+$,

(ii) if $\langle \mathbf{A}^\uparrow, \mathbf{y}^\uparrow \rangle^2 \leq \langle \mathbf{A}^\uparrow, \mathbf{y}^\downarrow \rangle^2$, then $(\mathbf{\Pi}, x) = \left(\widetilde{\mathbf{\Pi}}, \frac{\langle \mathbf{A}, \widetilde{\mathbf{\Pi}}\mathbf{y}\rangle}{\|\mathbf{A}\|^2}\right)$ such that $\widetilde{\mathbf{\Pi}}\mathbf{y} = \mathbf{y}^\downarrow$. In such case, the minimum value is $\|\mathbf{y}\| \sin\theta^-$,

where $\theta^+$ is the angle between $\mathbf{A}^\uparrow$ and $\mathbf{y}^\uparrow$ and $\theta^-$ is the angle between $\mathbf{A}^\uparrow$ and $\mathbf{y}^\downarrow$.

*Proof.* By Lemma 2.1, it follows that for any $\mathbf{\Pi} \in \mathcal{P}_m$

$$\begin{aligned}\|\mathbf{A}x - \mathbf{\Pi}\mathbf{y}\|^2 &= \|\mathbf{A}\|^2 x^2 - 2\langle \mathbf{A}, \mathbf{\Pi}\mathbf{y}\rangle x + \|\mathbf{y}\|^2 \\ &= \|\mathbf{A}\|^2 \left(x - \frac{1}{\|\mathbf{A}\|^2}\langle \mathbf{A}, \mathbf{\Pi}\mathbf{y}\rangle\right)^2 + \frac{1}{\|\mathbf{A}\|^2}\left(\|\mathbf{A}\|^2\|\mathbf{y}\|^2 - \langle \mathbf{A}, \mathbf{\Pi}\mathbf{y}\rangle^2\right)\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{1}{\|\mathbf{A}\|^2}\left(\|\mathbf{A}\|^2\|\mathbf{y}\|^2 - \langle \mathbf{A}, \mathbf{\Pi y}\rangle^2\right) \\
&\geq \frac{1}{\|\mathbf{A}\|^2}\left(\|\mathbf{A}\|^2\|\mathbf{y}\|^2 - \max\left\{\langle \mathbf{A}, \mathbf{y}^{\uparrow}\rangle^2, \langle \mathbf{A}, \mathbf{y}^{\downarrow}\rangle^2\right\}\right) \\
&= \|\mathbf{y}\|^2 \min\left\{\sin\theta^{+}, \sin\theta^{-}\right\}.
\end{aligned}
$$

□

Note that since $\|\widetilde{\mathbf{\Pi}}(\mathbf{A}x - \mathbf{\Pi y})\| = \|\mathbf{A}x - \mathbf{\Pi y}\|$ for all $\widetilde{\mathbf{\Pi}} \in \mathcal{P}_m$, Theorem 2.2 holds for $\mathbf{A} \neq \mathbf{A}^{\uparrow}$.

**Example 2.3.** Consider

$$
\mathbf{A} = \begin{bmatrix} 3 \\ 2 \\ 5 \\ 1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 1 \end{bmatrix}. \tag{2.2}
$$

It is easy to check that optimal solutions of

$$
\min_{\mathbf{\Pi} \in \mathcal{P}_m,\ x \in \mathbb{R}} \|\mathbf{A}x - \mathbf{\Pi y}\| \tag{2.3}
$$

are

$$
(\mathbf{\Pi}, x) = \left(\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, 0.1795\right), \quad \left(\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, 0.1795\right).
$$

And the minimum value is 1.3205. Note that although the optimal solution $x$ for the problem (1.2) is unique, the corresponding permutation matrices may possibly not be unique.

**Corollary 2.4.** Let $\mathbf{A} \in \mathbb{R}^{m \times 1}$ and $\mathbf{y} \in \mathbb{R}^m$ be given. Then the following holds:

(i) if $\langle \mathbf{A}^{\uparrow}, \mathbf{y}^{\downarrow}\rangle \geq 0$, then the optimal solution of (2.1) is $(\mathbf{\Pi}, x) = (\widetilde{\mathbf{\Pi}}, \|\mathbf{y}\| \sin\theta^{+})$ such that $\widetilde{\mathbf{\Pi}}\mathbf{y} = \mathbf{y}^{\uparrow}$.

(ii) if $\langle \mathbf{A}^{\uparrow}, \mathbf{y}^{\uparrow}\rangle \leq 0$, then the optimal solution of (2.1) is $(\mathbf{\Pi}, x) = (\widetilde{\mathbf{\Pi}}, \|\mathbf{y}\| \sin\theta^{-})$ such that $\widetilde{\mathbf{\Pi}}\mathbf{y} = \mathbf{y}^{\downarrow}$.

*Proof.* If $0 \leq \langle \mathbf{A}^{\uparrow}, \mathbf{y}^{\downarrow}\rangle \leq \langle \mathbf{A}^{\uparrow}, \mathbf{y}^{\uparrow}\rangle$, it holds that $\langle \mathbf{A}^{\uparrow}, \mathbf{y}^{\downarrow}\rangle^2 \leq \langle \mathbf{A}^{\uparrow}, \mathbf{y}^{\uparrow}\rangle^2$. Then (i) can be derived by Theorem 2.2. In a similar way, one can verify (ii) easily. □

Next we consider when a given matrix $\mathbf{A}$ is $m \times 2$. Denote $m \times 1$ column vector whose entries all are 1 as $\mathbf{1}$.

**Corollary 2.5.** Let $\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ c & c & \cdots & c \end{bmatrix}^T \in \mathbb{R}^{m \times 2}$, where $\mathbf{a} = [a_1\ a_2\ \ldots\ a_m]^T$ with $\mathbf{a} = \mathbf{a}^{\uparrow}$ and $\mathbf{y} \in \mathbb{R}^m$ be given. Then the optimal solution of

$$
\min_{\mathbf{\Pi} \in \mathcal{P}_m,\ \boldsymbol{x} \in \mathbb{R}^2} \|\mathbf{A}\boldsymbol{x} - \mathbf{\Pi y}\| \tag{2.4}
$$

is explicitly written as follows:

(i) if $\langle \mathbf{a}^{\uparrow}, \mathbf{y}^{\uparrow}\rangle^2 \geq \langle \mathbf{a}^{\uparrow}, \mathbf{y}^{\downarrow}\rangle^2$, then $(\mathbf{\Pi}, \boldsymbol{x}) = \left(\widetilde{\mathbf{\Pi}}, (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\widetilde{\mathbf{\Pi}}\mathbf{y}\right)$ such that $\widetilde{\mathbf{\Pi}}\mathbf{y} = \mathbf{y}^{\uparrow}$.

(ii) if $\langle \mathbf{a}^{\uparrow}, \mathbf{y}^{\uparrow}\rangle^2 \leq \langle \mathbf{a}^{\uparrow}, \mathbf{y}^{\downarrow}\rangle^2$, then $(\mathbf{\Pi}, \boldsymbol{x}) = \left(\widetilde{\mathbf{\Pi}}, (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\widetilde{\mathbf{\Pi}}\mathbf{y}\right)$ such that $\widetilde{\mathbf{\Pi}}\mathbf{y} = \mathbf{y}^{\downarrow}$.

*Proof.* Let $\boldsymbol{x} = (x_1, x_2)$. Then one can check that for any $\mathbf{\Pi} \in \mathcal{P}_m$ it holds that

$$\begin{aligned}\|\mathbf{A}\boldsymbol{x} - \mathbf{\Pi}\mathbf{y}\|^2 &= \|x_1\mathbf{a} + cx_2\mathbf{1} - \mathbf{\Pi}\mathbf{y}\|^2 \\ &= \|x_1\mathbf{a} - \mathbf{\Pi}(\mathbf{y} - cx_2\mathbf{1})\|^2.\end{aligned}$$

Note that, for any $x_1$ and $x_2 \in \mathbb{R}$,

$$\arg\min_{\mathbf{\Pi}\in\mathcal{P}_m} \|\mathbf{a}x_1 - \mathbf{\Pi}(\mathbf{y} - cx_2\mathbf{1})\|^2 = \arg\min_{\mathbf{\Pi}\in\mathcal{P}_m} \|\mathbf{a}x_1 - \mathbf{\Pi}\mathbf{y}\|^2$$

since $x_2$ has no effect on the order of $\mathbf{y}$. Thus it is enough to solve the optimization problem

$$\min_{\mathbf{\Pi}\in\mathcal{P}_m,\ x_1\in\mathbb{R}} \|\mathbf{a}x_1 - \mathbf{\Pi}\mathbf{y}\|^2,$$

which can be solved by Theorem 2.2.

$\square$

**Corollary 2.6.** Let $\mathbf{A} = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \end{bmatrix}^T \in \mathbb{R}^{m\times 2}$, where $\mathbf{a} = [a_1\ a_2\ \ldots\ a_m]^T$, $\mathbf{b} = [b_1\ b_2\ \ldots\ b_m]^T$ with $\mathbf{a} = \mathbf{a}^{\uparrow}$, $\mathbf{b} = \mathbf{b}^{\uparrow} = -\mathbf{b}^{\downarrow}$ and $\mathbf{y} \in \mathbb{R}^m$ be given. Then the optimal solution of optimization problem

$$\min_{\mathbf{\Pi}\in\mathcal{P}_m,\ \boldsymbol{x}\in\mathbb{R}^2} \|\mathbf{A}\boldsymbol{x} - \mathbf{\Pi}\mathbf{y}\| \tag{2.5}$$

is written as $(\mathbf{\Pi}, \boldsymbol{x}) = \left(\widetilde{\mathbf{\Pi}}, (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\widetilde{\mathbf{\Pi}}\mathbf{y}\right)$ such that $\widetilde{\mathbf{\Pi}}\mathbf{y} = \mathbf{y}^{\uparrow}$ or $\mathbf{y}^{\downarrow}$.

*Proof.* Note that

$$\|\mathbf{A}\boldsymbol{x} - \mathbf{\Pi}\mathbf{y}\|^2 = \|(x_1\mathbf{a} + x_2\mathbf{b}) - \mathbf{\Pi}\mathbf{y}\|^2$$

Since $\mathbf{a} = \mathbf{a}^{\uparrow}$ and $\mathbf{b} = \mathbf{b}^{\uparrow} = -\mathbf{b}^{\downarrow}$, $x_1\mathbf{a} + x_2\mathbf{b} = (x_1\mathbf{a} + x_2\mathbf{b})^{\uparrow}$ or $(x_1\mathbf{a} + x_2\mathbf{b})^{\downarrow}$ for all $\boldsymbol{x} \in \mathbb{R}^2$. Then, by Theorem 2.2, $\mathbf{\Pi} = \widetilde{\mathbf{\Pi}}$ such that $\widetilde{\mathbf{\Pi}}\mathbf{y} = \mathbf{y}^{\uparrow}$ or $\mathbf{y}^{\downarrow}$. And $\boldsymbol{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\widetilde{\mathbf{\Pi}}\mathbf{y}$. $\square$

Consider

$$\min_{\mathbf{\Pi}\in\mathcal{P}_m,\ \boldsymbol{x}\in\mathbb{R}^{m-1}} \|\mathbf{A}\boldsymbol{x} - \mathbf{\Pi}\mathbf{y}\|, \tag{2.6}$$

where $\mathbf{A} \in \mathbb{R}^{m\times(m-1)}$, $\mathbf{y} \in \mathbb{R}^{m\times 1}$. By singular value decomposition, there exist orthogonal matrices $\mathbf{U}$, $\mathbf{V}$ such that $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Then it follows that

$$\begin{aligned}\mathbf{H} :&= \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T - \mathbf{I} \\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T(\mathbf{V}\mathbf{\Sigma}^T\mathbf{U}^T\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^{-1}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T - \mathbf{I}\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{U}\left(\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^T - \mathbf{I}\right)\mathbf{U}^T \\
&= \mathbf{U}\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & -1 \end{pmatrix}\mathbf{U}^T \\
&= -\mathbf{u}_m\mathbf{u}_m^T
\end{aligned}
$$

where $\mathbf{u}_m$ is the $m$th column vector of $\mathbf{U}$. Using the fact that $\boldsymbol{x} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Pi}\mathbf{y}$,

$$
\begin{aligned}
\|\mathbf{A}\boldsymbol{x} - \boldsymbol{\Pi}\mathbf{y}\|^2 &= \|\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\boldsymbol{\Pi}\mathbf{y} - \boldsymbol{\Pi}\mathbf{y}\|^2 \\
&= \|(\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T - \mathbf{I})\boldsymbol{\Pi}\mathbf{y}\|^2 \\
&= \|\mathbf{H}\boldsymbol{\Pi}\mathbf{y}\|^2 \\
&= (\boldsymbol{\Pi}\mathbf{y})^T(\mathbf{H}^T\mathbf{H})\boldsymbol{\Pi}\mathbf{y} \\
&= (\boldsymbol{\Pi}\mathbf{y})^T\mathbf{u}_m\mathbf{u}_m^T\boldsymbol{\Pi}\mathbf{y} \\
&= \langle \mathbf{u}_m, \boldsymbol{\Pi}\mathbf{y}\rangle^2.
\end{aligned}
$$

If $\langle \mathbf{u}_m, \boldsymbol{\Pi}\mathbf{y}\rangle \geq 0$ (resp, $\langle \mathbf{u}_m, \boldsymbol{\Pi}\mathbf{y}\rangle \leq 0$) for all $\boldsymbol{\Pi} \in \mathcal{P}_m$, then by rearrangement inequality an optimal solution $\boldsymbol{\Pi}$ for (2.6) can be found. However, $\langle \mathbf{u}_m, \boldsymbol{\Pi}\mathbf{y}\rangle$ has both positive and negative signs for some $\boldsymbol{\Pi} \in \mathcal{P}_m$, it may not possibly be easy to use rearrangement inequality.

Now we consider nonlinear regression without correspondences. Let $f(x;\boldsymbol{\beta}) : \mathbb{R} \to \mathbb{R}$ be a real-valued function where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_n) \in \mathbb{R}^n$ is a parameter of $f$. Denote $\boldsymbol{x} = (x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$ and denote

$$
\mathbf{f}(\boldsymbol{x};\boldsymbol{\beta}) := [f(x_1;\boldsymbol{\beta})\ f(x_2;\boldsymbol{\beta})\ \ldots\ f(x_m;\boldsymbol{\beta})]^T.
$$

A nonlinear least squares problem without correspondences is written as

$$
\min_{\boldsymbol{\Pi}\in\mathcal{P}_m,\ \boldsymbol{\beta}\in\mathbb{R}^n} \|\mathbf{f}(\boldsymbol{x};\boldsymbol{\beta}) - \boldsymbol{\Pi}\mathbf{y}\|^2, \tag{2.7}
$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)$, $\mathbf{y} = (y_1, y_2, \ldots, y_m) \in \mathbb{R}^m$ are given.

Let $\mathcal{I}_1, \mathcal{I}_2$ be closed bounded intervals and let $\phi : \mathcal{I}_1 \times \mathcal{I}_2 \to \mathbb{R}$ be a real-valued function and let $\Phi : \mathcal{I}_1^m \times \mathcal{I}_2^m \to \mathbb{R}$ be a real-valued function defined by

$$
\Phi(\mathbf{u}, \mathbf{v}) := \sum_{j=1}^{m} \phi(u_j, v_j)
$$

where $\mathbf{u} = (u_1, u_2, \ldots, u_m) \in \mathcal{I}_1^m$ and $\mathbf{v} = (v_1, v_2, \ldots, v_m) \in \mathcal{I}_2^m$.

**Theorem 2.7.** [6] Let $\mathbf{u} \in \mathcal{I}_1^m$, $\mathbf{v} \in \mathcal{I}_2^m$. Assume that the first order partial derivatives of $\phi$ exist everywhere, and second order partial derivative $\partial^2\phi(x,y)/\partial x\partial y$ exists for all $(x,y) \in \mathcal{I}_1 \times \mathcal{I}_2$. Then

(i) if $\frac{\partial^2\phi(x,y)}{\partial x\partial y} \geq 0$ for all $(x,y) \in \mathcal{I}_1 \times \mathcal{I}_2$, it holds

$$\Phi(\mathbf{u}^{\uparrow}, \mathbf{v}^{\downarrow}) \leq \Phi(\mathbf{u}, \mathbf{v}) \leq \Phi(\mathbf{u}^{\uparrow}, \mathbf{v}^{\uparrow}),$$

(ii) if $\frac{\partial^2\phi(x,y)}{\partial x\partial y} \leq 0$ for all $(x,y) \in \mathcal{I}_1 \times \mathcal{I}_2$, it holds

$$\Phi(\mathbf{u}^{\uparrow}, \mathbf{v}^{\downarrow}) \geq \Phi(\mathbf{u}, \mathbf{v}) \geq \Phi(\mathbf{u}^{\uparrow}, \mathbf{v}^{\uparrow}).$$

For a given vector $\mathbf{v} = (v_1, v_2, \ldots, v_m)$ denote $\mathbf{v}_{\min} = \min_{1\leq i\leq m} v_i$ and $\mathbf{v}_{\max} = \max_{1\leq i\leq m} v_i$. Denote the set of functions with $k$ continuous derivatives as $\mathcal{C}^k$.

**Theorem 2.8.** Let $\boldsymbol{x}, \mathbf{y} \in \mathbb{R}^m$ be given. Assume that $f \in \mathcal{C}^2(a,b)$ such that $[\boldsymbol{x}_{\min}, \boldsymbol{x}_{\max}] \subset (a,b)$. Then it holds that

(i) if $f'(x;\boldsymbol{\beta}) \leq 0$ for all $a \leq x \leq b$ and $\boldsymbol{\beta} \in \mathbb{R}^n$, then

$$\|\mathbf{f}(\boldsymbol{x}^{\uparrow};\boldsymbol{\beta}) - \mathbf{y}^{\downarrow}\| \leq \|\mathbf{f}(\boldsymbol{x};\boldsymbol{\beta}) - \mathbf{\Pi}\mathbf{y}\| \leq \|\mathbf{f}(\boldsymbol{x}^{\uparrow};\boldsymbol{\beta}) - \mathbf{y}^{\uparrow}\|,$$

(ii) if $f'(x;\boldsymbol{\beta}) \geq 0$ for all $a \leq x \leq b$ and $\boldsymbol{\beta} \in \mathbb{R}^n$, then

$$\|\mathbf{f}(\boldsymbol{x}^{\uparrow};\boldsymbol{\beta}) - \mathbf{y}^{\uparrow}\| \leq \|\mathbf{f}(\boldsymbol{x};\boldsymbol{\beta}) - \mathbf{\Pi}\mathbf{y}\| \leq \|\mathbf{f}(\boldsymbol{x}^{\uparrow};\boldsymbol{\beta}) - \mathbf{y}^{\downarrow}\|.$$

*Proof.* Set $\phi(x,y) = |f(x;\boldsymbol{\beta}) - y|^2$. If $f'(x;\boldsymbol{\beta}) \leq 0$ for all $x \in [\boldsymbol{x}_{\min}, \boldsymbol{x}_{\max}]$, then $\frac{\partial^2\phi(x,y)}{\partial x\partial y} = -2f'(x;\boldsymbol{\beta}) \geq 0$ for all $(x,y) \in [\boldsymbol{x}_{\min}, \boldsymbol{x}_{\max}] \times \mathbb{R}$. □

## 3. Computational results

In the real world, data often comes with noise. However, to facilitate more straightforward estimation of a parameter, we conducted experiments using noiseless data. Let $(x,y) \in \mathbb{R}^2$ be given shuffled data generated by a nonlinear function $y = f(x;\beta) = e^{\beta^2 x}$, where $\beta \in \mathbb{R}$ is unknown as follows.

| $(x,y)$ | | |
|---|---|---|
| $(-0.9495, 0.7754)$ | $(-0.3341, 0.5444)$ | $(0.2963, 0.9587)$ |
| $(-0.4089, 1.2089)$ | $(-0.0658, 1.0785)$ | $(0.6844, 0.8075)$ |
| $(-0.3973, 1.5734)$ | $(0.1180, 1.5496)$ | $(0.7082, 0.7697)$ |

Note that $f'(x;\beta) = \beta^2 e^{\beta^2 x} \geq 0$ for all $x \in \mathbb{R}$ and $\beta \in \mathbb{R}$. Then by Theorem 2.8

$$\|\mathbf{f}(\boldsymbol{x}^{\uparrow};\beta) - \mathbf{y}^{\uparrow}\| \leq \|\mathbf{f}(\boldsymbol{x};\beta) - \mathbf{\Pi}\mathbf{y}\| \leq \|\mathbf{f}(\boldsymbol{x}^{\uparrow};\beta) - \mathbf{y}^{\downarrow}\|, \tag{3.1}$$

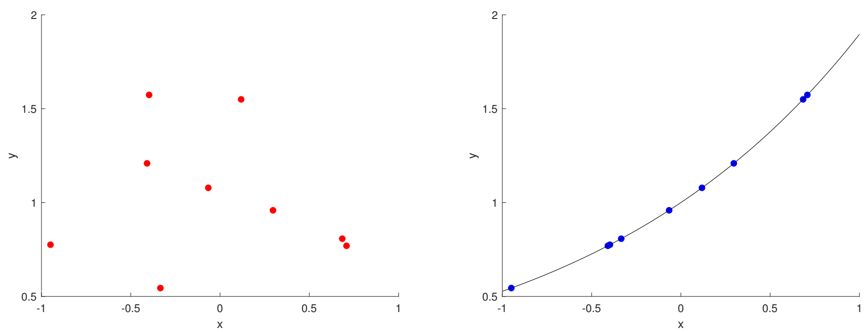
Consider a nonlinear least squares problem without correspondences

$$\min_{\mathbf{\Pi}\in\mathcal{P}_9,\ \beta\in\mathbb{R}} \|\mathbf{f}(\boldsymbol{x};\beta) - \mathbf{\Pi}\mathbf{y}\|, \tag{3.2}$$

where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^9$ is the vector form of given shuffled data $(x, y) \in \mathbb{R}^2$. Using Theorem 3.1, it is easy to find that the optimal solution of the optimization in (3.2) is $\beta = 0.8$ and

$$\boldsymbol{\Pi} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Hence, the nonlinear function can be theoretically found (see FIGURE 1).




(A) Given shuffled data (B) The optimal solution

FIGURE 1. For given shuffled data with unknown correspondence the nonlinear fitting curve can be found.

**Final remark** : we have established the closed forms of optimal solutions in (1.2) for certain cases. However, it is still questionable if there exists an explicit solution for other cases. When it comes to the nonlinear shuffle regression problem, our current results heavily rely on strong constraints, such as requiring the function to be non-increasing or non-decreasing. In order to address the nonlinear shuffle regression problem in more general circumstances, it appears necessary to reduce these constraints or propose a new methodology.

**Conflicts of interest** : The authors declare no conflict of interest.

**Data availability** : Not applicable

**Acknowledgments** : The authors are deeply grateful to the referee for suggestions that led to significant improvements in the presentation.

## References

1. A. Abid and J. Zou, *Stochastic EM for shuffled linear regression*, arXiv preprint arXiv:1804.00681 (2018).
2. Å. Bjö rck, *Least squares methods*, Handbook of numerical analysis **1** (1990), 465-652.
3. E. Borensztein, J. De Gregorio and J.W. Lee, *How does foreign direct investment affect economic growth?*, J. Int. Econ. **45** (1998), 115-135.
4. P. David, D. Dementhon, R. Duraiswami and H. Samet, *Softposit: Simultaneous pose and correspondence determination*, Int. J. Comput. Vis. **59** (2004), 259-284.
5. C. Flammer, *Does corporate social responsibility lead to superior financial performance*? *A regression discontinuity approach*, Manage Sci. **61** (2015), 2549-2568.
6. R. Geretschläger and W. Janous, *Generalized Rearrangement Inequalities*, Am. Math. Mon. **108** (2001), 158-165.
7. F.S. Gharehchopogh, T.H. Bonab and S.R. Khaze, *A linear regression approach to prediction of stock market trading volume: a case study*, Int. J. Supply Chain Manag. **4** (2013), 25.
8. P.C. Hansen, V. Pereyra and G. Scherer, *Least squares data fitting with applications*, JHU Press, 2013.
9. G.H. Hardy, J.E. Littlewood and G. Pólya, *Inequalities*, Cambridge: Cambridge University Press, 1952.
10. B. Heshmaty and A. Kandel, *Fuzzy linear regression and its applications to forecasting in uncertain environment*, Fuzzy Sets Syst. **15** (1985), 159-191.
11. D.J. Hsu, K. Shi and X. Sun, *Linear regression without correspondence*, Adv. Neural Inf. Process. Syst. **30** (2017).
12. X. Huang and A. Madan, *CAP3: A DNA sequence assembly program*, Genome Res. **9** (1999), 868-877.
13. M.G. Lagoudakis and R. Parr, *Least-squares policy iteration*, J. Mach. Learn. Res. **4** (2003), 1107-1149.
14. P. Meer, D. Mintz, A. Rosenfeld and D.Y. Kim, *Robust regression methods for computer vision: A review*, Int. J. Comput. Vis. **6** (1991), 59-70.
15. J. Merlo, B. Chaix, H. Ohlsson, A. Beckman, K. Johnell, P. Hjerpe and K. Larsen, *A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena*, J. Epidemiology Community Health **60** (2006), 290-297.
16. D.C. Montgomery, E.A. Peck and G.G. Vining, *Introduction to Linear Regression Analysis*, John Wiley and Sons, 2021.
17. A. Pananjady, M.J. Wainwright and T.A. Courtade, *Linear regression with shuffled data: Statistical and computational limits of permutation recovery*, IEEE Trans. Inf. Theory **64** (2017), 3286-3300.
18. R.W. Parks, *Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated*, J. Am. Stat. Assoc. **62** (1967), 500-509.
19. L. Peng, X. Song, M.C. Tsakiris, H. Choi, L. Kneip and Y. Shi, *Algebraically-initialized expectation maximization for header-free communication*, In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5182-5186), IEEE, 2019.
20. K. Rose, *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems*, Proceedings of the IEEE **86** (1998), 2210-2239.
21. W.S. Robinson, *A method for chronologically ordering archaeological deposits*, Am. Antiq. **16.4** (1951), 293-301.

22. C. Rose and I.S. Mian, *Signaling with identical tokens: Upper bounds with energy constraints*, 2014 IEEE International Symposium on Information Theory, 2014, 1817-1821.
23. C. Rose, I.S. Mian and R. Song, *Timing channels with multiple identical quanta*, arXiv preprint arXiv:1208.1070 (2012).
24. X. Song, H. Choi and Y. Shi, *Permuted linear model for header-free communication via symmetric polynomials*, 2018 IEEE International Symposium on Information Theory (ISIT) (pp. 661-665), IEEE, 2018.
25. L. Peng and M.C. Tsakiris, *Homomorphic sensing of subspace arrangements*, Appl. Comput. Harmon. Anal. **55** (2021), 466-485.
26. M.C. Tsakiris, L. Peng, A. Conca, L. Kneip, Y. Shi and H. Choi, *An algebraic-geometric approach for linear regression without correspondences*, IEEE Trans. Inf. Theory **66** (2020), 5130-5144.
27. J. Unnikrishnan, S. Haghighatshoar and M. Vetterli, *Unlabeled sensing with random linear measurements*, IEEE Trans. Inf. Theory **64** (2018), 3237-3253.
28. V. Vapnik, S. Golowich and A. Smola, *Support vector method for function approximation, regression estimation and signal processing*, Adv. Neural Inf. Process. Syst. **9** (1996).
29. D.J. Zimmerman, *Regression toward mediocrity in economic stature*, Am. Econ. Rev. (1992), 409-429.

**Mijin Kim** received M.Sc. in the department of mathematics from Kyungpook National University in 2022. She is currently a researcher at MDA which is an artificial intelligence company and is also a Ph.D. candidate in the department of computer science at Kyungpook National University.

Department of Mathematics, Kyungpook National University, Daegu 41566, Republic of Korea.
e-mail: `rlamj0922@knu.ac.kr`

**Hyungu Lee** is currently a undergraduate student and will start the master program in the department of mathematics at Kyungpook National University. He is interested in graph neural networks and data analysis and so on.

Department of Mathematics, Kyungpook National University, Daegu 41566, Republic of Korea.
e-mail: `lewis9910@knu.ac.kr`

**Hayoung Choi** received M.Sc. from Kyungpook National University, and Ph.D. from University of Wyoming. He was previously a postdoctoral researcher at School of Information Science and Technology, ShanghaiTech University in China. He is currently an assistant professor in the Department of Mathematics at Kyungpook National University since 2020. His research concerns matrix analysis, numerical linear algebra, large-scale data analysis, and applications of these topics to problems in machine learning, data mining, and communication.

Department of Mathematics, Kyungpook National University, Daegu 41566, Republic of Korea.
e-mail: `hayoung.choi@knu.ac.kr`