ECE421: Introduction to Machine Learning — Fall 2024

Worksheet 1 Solution: Pocket Algorithm and Linear Regression

Notation

(a) We use a **underline** to represents **column vectors**, *e.g.*, $\underline{p} \in \mathbb{R}^k$ represents a column vector with k elements. We adopt the following notations to list the elements of a **column vector**

$$\underline{p} = (p_1, p_2, \ldots, p_k) = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{bmatrix}.$$

Note the usage of parentheses and brackets. The notation with parentheses provides a more compact representation of vectors and optimizes space usage.
Additionally, **row vectors** can be represented by $\underline{q}^\top = [p_1, p_2, \ldots, p_k]$. Note the use of transpose and brackets.
Finally, the context and notation should make it clear whether a vector is a column vector or a row vector.

(b) For all questions we denote the weight vector by $\underline{w} = (b, w_1, \ldots, w_d) \in \mathbb{R}^{d+1}$, where $b \in \mathbb{R}$ is the bias term, and we denote the example vectors by $\underline{x} = (1, x_1, x_2, \ldots, x_d) \in \mathbb{R}^{d+1}$.

(c) In the following, LFD refers to the textbook "Learning from Data."

Q0 Linear Algebra Review

0.a (The ℓ_p-norm) For a real number $p \geq 1$, define the ℓ_p-norm of a vector $\underline{x} \in \mathbb{R}^n$.

Answer. To answer this question, please review your notes for the linear algebra course or any introductory linear algebra textbook.

0.b (The ℓ_1, ℓ_2, and ℓ_∞-norm) Consider the vector $\underline{x} = (5, 2, -3)$. Find the ℓ_1, ℓ_2, and ℓ_∞-norm of $\underline{x}$.

Answer.

- ℓ_1-norm: $\mid 5 \mid + \mid 2 \mid + \mid -3 \mid = 10$
- ℓ_2-norm: $\sqrt{5^2 + 2^2 + (-3)^2} = \sqrt{38}$
- ℓ_∞-norm: $\max(\mid 5 \mid, \mid 2 \mid, \mid -3 \mid) = 5$

0.c (Matrix Multiplication) Let $\underline{w} = (w_0, w_1, \ldots, w_d)$ and $\underline{x}_i = (x_{i0}, x_{i1}, \ldots, x_{id})$ for $i \in \{1, 2, \ldots, N\}$. Let

$$X = \begin{bmatrix} x_{10} & x_{11} & \ldots & x_{1d} \\ x_{20} & x_{21} & \ldots & x_{2d} \\ \vdots & \vdots & \ldots & \vdots \\ x_{N0} & x_{N1} & \ldots & x_{Nd} \end{bmatrix} = \begin{bmatrix} \underline{x}_1^\top \\ \underline{x}_2^\top \\ \vdots \\ \underline{x}_N^\top \end{bmatrix},$$

$$\underline{\hat{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_N \end{bmatrix} = \begin{bmatrix} \underline{w}^\top \underline{x}_1 \\ \underline{w}^\top \underline{x}_2 \\ \vdots \\ \underline{w}^\top \underline{x}_N \end{bmatrix}.$$

Show that $\underline{\hat{y}} = X\underline{w}$.

Answer. To answer this question, please review your notes for the linear algebra course or any introductory linear algebra textbook.

Q1 Gradient and Optimization Fundamentals

1.a (Gradient) Prove that $\nabla_{\underline{x}}(\underline{a}^\top \underline{x}) = \underline{a}$, and $\nabla_{\underline{x}}(\underline{x}^\top \underline{a}) = \underline{a}$ and $\nabla_{\underline{x}}(\underline{x}^\top A\underline{x}) = 2A\underline{x}$, where $\underline{a}$ and $\underline{x}$ are vectors with k entries and A is a symmetric squared matrix.

Answer. We prove that $\nabla_{\underline{x}}(\underline{x}^\top A\underline{x}) = (A + A^\top)\underline{x}$, for any square matrix A. Observe that $\underline{x}^\top A\underline{x} = \sum_{i=1}^{k} x_i \sum_{j=1}^{k} A_{i,j}x_j$. Hence, for any $t \in \{1, 2, \ldots, k\}$,

$$\begin{aligned}
\frac{\partial}{\partial x_t}(\underline{x}^\top A\underline{x}) &= \frac{\partial}{\partial x_t}(\sum_{i=1}^{k} x_i \sum_{j=1}^{k} A_{i,j}x_j) = \sum_{i=1}^{k} \frac{\partial}{\partial x_t}(x_i \sum_{j=1}^{k} A_{i,j}x_j) \\
&= \sum_{i=1}^{k} \left((\frac{\partial}{\partial x_t}x_i)(\sum_{j=1}^{k} A_{i,j}x_j) + x_i \frac{\partial}{\partial x_t}(\sum_{j=1}^{k} A_{i,j}x_j) \right) \\
&= \sum_{i=1}^{k} (\frac{\partial}{\partial x_t}x_i)(\sum_{j=1}^{k} A_{i,j}x_j) + \sum_{i=1}^{k} x_i \frac{\partial}{\partial x_t}(\sum_{j=1}^{k} A_{i,j}x_j) \\
&= \sum_{j=1}^{k} A_{t,j}x_j + \sum_{i=1}^{k} x_i A_{i,t}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\nabla_{\underline{x}}(\underline{x}^\top A\underline{x}) &= \begin{bmatrix} \frac{\partial}{\partial x_1}(\underline{x}^\top A\underline{x}) \\ \frac{\partial}{\partial x_2}(\underline{x}^\top A\underline{x}) \\ \vdots \\ \frac{\partial}{\partial x_k}(\underline{x}^\top A\underline{x}) \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{k} A_{1,j}x_j + \sum_{i=1}^{k} x_i A_{i,1} \\ \sum_{j=1}^{k} A_{2,j}x_j + \sum_{i=1}^{k} x_i A_{i,2} \\ \vdots \\ \sum_{j=1}^{k} A_{k,j}x_j + \sum_{i=1}^{k} x_i A_{i,k} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{k} A_{1,j}x_j \\ \sum_{j=1}^{k} A_{2,j}x_j \\ \vdots \\ \sum_{j=1}^{k} A_{k,j}x_j \end{bmatrix} + \begin{bmatrix} \sum_{i=1}^{k} x_i A_{i,1} \\ \sum_{i=1}^{k} x_i A_{i,2} \\ \vdots \\ \sum_{i=1}^{k} x_i A_{i,k} \end{bmatrix} \\
&= A\underline{x} + A^\top \underline{x} = (A + A^\top)\underline{x}.
\end{aligned}$$

Thus, $\nabla_{\underline{x}}(\underline{x}^\top A\underline{x}) = (A + A^\top)\underline{x}$. When A is symmetric, $\nabla_{\underline{x}}(\underline{x}^\top A\underline{x}) = 2A\underline{x}$.

1.b (Exercise 3.17 (a),(b) in LFD) Recall that for a scalar-valued function $f : \mathbb{R}^n \to \mathbb{R}$ and a vector $\underline{p} \in \mathbb{R}^n$, the first-order Taylor series approximation of $f(\underline{x}+\underline{p})$ is $f(\underline{x}+\underline{p}) \approx f(\underline{x}) + \nabla f(\underline{x})^\top \underline{p}$. Consider the function $E(u, v) = e^u + e^{2v} + e^{uv} + u^2 - 3uv + 4v^2 - 3u - 5v$, where u and v are scalars.

1.b.i Denote by $\hat{E}_1(\Delta u, \Delta v)$ the first-order Taylor series approximation of E at $(u, v) = (0, 0)$. We know that $\hat{E}_1(\Delta u, \Delta v)$ is of the form $\hat{E}_1(\Delta u, \Delta v) = a_u \Delta u + a_v \Delta v + a$. What are the values of a_u, a_v, and a?

Answer. Observe that $\nabla E(u, v) = \begin{bmatrix} \frac{\partial}{\partial u}E(u, v) \\ \frac{\partial}{\partial v}E(u, v) \end{bmatrix} = \begin{bmatrix} e^u + ve^{uv} + 2u - 3v - 3 \\ 2e^{2v} + ue^{uv} - 3v + 8v - 5 \end{bmatrix}$. Hence, by the first-order Taylor series approximation of E at $(0, 0)$,

$$E\left(\underline{0} + (\Delta u, \Delta v)\right) \approx E(\underline{0}) + \nabla E(\underline{0})^\top (\Delta u, \Delta v) = 3 + (-2, -3)^\top (\Delta u, \Delta v) = 3 - 2\Delta u - 3\Delta v,$$

Thus, $a_u = -2, a_v = -3$, and $a = 3$.

1.b.ii Minimize $\hat{E}_1$ over all possible $(\Delta u, \Delta v)$ such that $\|(\Delta u, \Delta v)\|_2 = 0.5$, *i.e.*,

$$\begin{aligned}
\min_{\Delta u, \Delta v} \quad & \hat{E}_1(\Delta u, \Delta v) \\
\text{s.t.} \quad & \|(\Delta u, \Delta v)\|_2 = 0.5.
\end{aligned}$$

Recall that the column vector $(\Delta u^*, \Delta v^*)$ that minimizes $\hat{E}_1$ is in the direction of $-\nabla E(u, v)$, *i.e.*, the negative gradient direction. Compute $(\Delta u^*, \Delta v^*)$ that minimizes $\hat{E}_1$, and the resulting $\hat{E}_1(\Delta u^*, \Delta v^*)$.

Answer. Note that

$$\hat{E}_1(\Delta u, \Delta v) = 3 + (-2, -3)^\top (\Delta u, \Delta v) = 3 + \|(-2, -3)\|_2 \|(\Delta u, \Delta v)\|_2 \cos\theta,$$

where θ is the angle between the two vectors $(-2, -3)$ and $(\Delta u, \Delta v)$. Hence, the vector $(\Delta u, \Delta v)$ with $\|(\Delta u, \Delta v)\|_2 = 0.5$ that minimizes $\hat{E}_1$ must be in the opposite direction of $(-2, -3)$, *i.e.*, $\cos\theta = -1$. Therefore, $(\Delta u^*, \Delta v^*) = -0.5 \frac{(-2,-3)}{\|(-2,-3)\|_2} = \frac{(2,3)}{2\sqrt{13}}$.[1]

Q2 (Perceptron Learning Algorithm) Given a dataset $\mathcal{D} = \{(\underline{x}_n, y_n)\}_{n=1}^N$, where $\underline{x}_n \in \mathbb{R}^d$ and $y_n \in \{+1, -1\}$, we wish to train a Perceptron model

$$h(\underline{x}) = \text{sign}\left(b + \sum_{i=1}^d w_i x_i\right) = \text{sign}(\underline{w}^\top \underline{x})$$

that correctly classifies *all* examples in $\mathcal{D}$. Consider the perceptron weight update rule

$$\underline{w}(t+1) = \underline{w}(t) + y_n \underline{x}_n,$$

where $(\underline{x}_n, y_n)$ is the misclassified datapoint after iteration t. This weight update rule moves the weights in the direction of classifying examples correctly. To see this, show the following.

2.a If $\underline{x}(t)$ is misclassified by $\underline{w}(t)$, show that $y_n \underline{w}^\top(t) \underline{x}_n \leq 0$.

Answer. When $\underline{x}(t)$ is misclassified, we must have one of the following cases.

- $\underline{w}^\top(t) \underline{x}_n = 0$.[2]
- $\underline{w}^\top(t) \underline{x}_n > 0$ but $y_n = -1$.
- $\underline{w}^\top(t) \underline{x}_n < 0$ but $y_n = +1$.

Observe that in the above cases, $y_n \underline{w}^\top(t) \underline{x}_n \leq 0$.

2.b Use the equation for $\underline{w}(t+1)$ to show that $y_n \underline{w}^\top(t+1) \underline{x}_n > y_n \underline{w}^\top(t) \underline{x}_n$.

Answer.

$$y_n \underline{w}^\top(t+1) \underline{x}_n = y_n(\underline{w}(t) + y_n \underline{x}_n) \underline{x}_n = y_n \underline{w}(t) \underline{x}_n + (y_n)^2 \underline{x}_n \underline{x}_n = y_n \underline{w}(t) \underline{x}_n + \|\underline{x}_n\|_2^2 > y_n \underline{w}(t) \underline{x}_n,$$

where the last inequality is due to the fact that $\|\underline{x}_n\|_2^2 > 0$ since $x_{n,0} = 1$.

2.c Argue that the weight update from $\underline{w}(t)$ to $\underline{w}(t+1)$ is a move "in the right direction."

Answer. Note that we eventually want a weight vector that can successfully classify $\underline{x}_n$, *i.e.*, we want $y_n \underline{w}\ \underline{x}_n$ to be positive. By updating $\underline{w}$ according to the Perceptron weight update rule, we make $y_n \underline{w}\ \underline{x}_n$ more positive. Therefore, this is a move "in the right direction."

[**REMARK:** Problem 1.3 in LFD, page 33, shows steps towards a rigorous proof of convergence of the Perceptron algorithm. Feel free to attempt solving this problem on your own. This is an optional exercise.]

Q3 (Linear Regression) Given a dataset $\mathcal{D} = \{(\underline{x}_n, y_n)\}_{n=1}^N$, where $\underline{x}_n \in \mathbb{R}^d$ and $y_n \in \mathbb{R}$, we wish to train a linear regression model

$$h(x) = b + \sum_{i=1}^d w_i x_i = w^\top x.$$

[1] Note that $(-2, -3)$ is in fact $\nabla E(\underline{0})$ and $(\Delta u^*, \Delta v^*)$ must be in the opposite direction of $\nabla E(\underline{0})$.

[2] The Perceptron model identifies any point on the decision boundary as a misclassification.

The in-sample error associated with the linear regression model is

$$E_{\text{in}}(w) = \frac{1}{2N}\sum_{n=1}^{N}(\underline{w}^\top \underline{x}_n - y_n)^2. \tag{1}$$

Define the data matrix X and target vector $\underline{y}$ as:

$$X = \begin{bmatrix} x_{10} & x_{11} & \dots & x_{1d} \\ \vdots & \vdots & \dots & \vdots \\ x_{N0} & x_{N1} & \dots & x_{Nd} \end{bmatrix} = \begin{bmatrix} \underline{x}_1^\top \\ \vdots \\ \underline{x}_N^\top \end{bmatrix} \in \mathbb{R}^{N\times(d+1)},$$

$$\underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \in \mathbb{R}^N.$$

where $\underline{x}_i = (x_{i0}, x_{i1}, \dots, x_{id})$ and $x_{i0} = 1$ for all $i \in \{1, 2, \dots, N\}$.

3.a Show that the in-sample error can be written as:

$$E_{\text{in}}(\underline{w}) = \frac{1}{2N}\|X\underline{w} - \underline{y}\|_2^2 = \frac{1}{2N}\left(\underline{w}^\top X^\top X\underline{w} - 2\underline{w}^\top X^\top \underline{y} + \|\underline{y}\|_2^2\right). \tag{2}$$

Answer.

$$\begin{aligned} E_{\text{in}}(\underline{w}) &= \frac{1}{2N}\|X\underline{w} - \underline{y}\|_2^2 = \frac{1}{2N}(X\underline{w} - \underline{y})^\top(X\underline{w} - \underline{y}) = \frac{1}{2N}(\underline{w}^\top X^\top - \underline{y}^\top)(X\underline{w} - \underline{y}) \\ &= \frac{1}{2N}(\underline{w}^\top X^\top X\underline{w} - \underline{w}^\top X^\top \underline{y} - \underline{y}^\top X\underline{w} + \underline{y}^\top \underline{y}) \\ &= \frac{1}{2N}\left(\underline{w}^\top X^\top X\underline{w} - \underline{w}^\top X^\top \underline{y} - \underline{y}^\top X\underline{w} + \|\underline{y}\|_2^2\right) \\ &= \frac{1}{2N}\left(\underline{w}^\top X^\top X\underline{w} - 2\underline{w}^\top X^\top \underline{y} + \|\underline{y}\|_2^2\right), \end{aligned}$$

where the last equality follows from the fact that the transpose of a scalar is equal to itself.

3.b Find the expressions for the gradient of (1) and (2) with respect to $\underline{w}$. Verify that the gradients of the two forms are equivalent.

Answer.
Gradient of (1)**:** Observe that $\frac{\partial}{\partial w_i}(\frac{1}{2N}\sum_{n=1}^{N}(\underline{w}^\top \underline{x}_n - y_n)^2) = \frac{1}{2N}\sum_{n=1}^{N}\frac{\partial}{\partial w_i}((\underline{w}^\top \underline{x}_n - y_n)^2)$. Thus,

$$\begin{aligned} \frac{\partial}{\partial w_i}(\frac{1}{2N}\sum_{n=1}^{N}(\underline{w}^\top \underline{x}_n - y_n)^2) &= \frac{1}{2N}\sum_{n=1}^{N}\frac{\partial}{\partial w_i}((\underline{w}^\top \underline{x}_n - y_n)^2) \\ &= \frac{1}{2N}\sum_{n=1}^{N}2(\underline{w}^\top \underline{x}_n - y_n)\frac{\partial}{\partial w_i}(\underline{w}^\top \underline{x}_n - y_n) \\ &= \frac{1}{2N}\sum_{n=1}^{N}2(\underline{w}^\top \underline{x}_n - y_n)\underline{x}_{n,i} \\ &= \frac{1}{2N}\left(2\sum_{n=1}^{N}\underline{x}_{n,i}\underline{x}_n^\top \underline{w} - 2\sum_{n=1}^{N}\underline{x}_{n,i}y_n\right). \end{aligned}$$

Thus,

$$\nabla_{\underline{w}} E_{\text{in}}(\underline{w}) = \frac{1}{N}\begin{bmatrix} \sum_{n=1}^{N} \underline{x}_{n,0}\underline{x}_n^\top \\ \sum_{n=1}^{N} \underline{x}_{n,1}\underline{x}_n^\top \\ \vdots \\ \sum_{n=1}^{N} \underline{x}_{n,d}\underline{x}_n^\top \end{bmatrix} \underline{w} - \frac{1}{N}\begin{bmatrix} \sum_{n=1}^{N} \underline{x}_{n,0} y_n \\ \sum_{n=1}^{N} \underline{x}_{n,0} y_n \\ \vdots \\ \sum_{n=1}^{N} \underline{x}_{n,d} y_n \end{bmatrix} = \frac{1}{N} X^\top X \underline{w} - \frac{1}{N} X^\top \underline{y}.$$

[**NOTE:** You can derive the gradient by using the chain rule, as well. In the solution above, we implicitly applied the chain rule. See the next wroksheet and the lecture notes for more details on the chain rule.]
Gradient of (2)**:**

$$\begin{aligned} \nabla_{\underline{w}} E_{\text{in}}(\underline{w}) &= \nabla_{\underline{w}} \left(\frac{1}{2N} \left(\underline{w}^\top X^\top X \underline{w} - 2\underline{w}^\top X^\top \underline{y} + \|\underline{y}\|_2^2 \right) \right) \\ &= \frac{1}{2N} \nabla_{\underline{w}} \left(\underline{w}^\top X^\top X \underline{w} - 2\underline{w}^\top X^\top \underline{y} + \|\underline{y}\|_2^2 \right) \\ &= \frac{1}{2N} (2X^\top X \underline{w} - 2X^\top \underline{y}) \\ &= \frac{1}{N} X^\top X \underline{w} - \frac{1}{N} X^\top \underline{y}. \end{aligned}$$

3.c Suppose $X^\top X$ is invertible. Let $\underline{w}^\star = (X^\top X)^{-1} X^\top \underline{y}$. Show that $E_{\text{in}}(\underline{w})$ can be decomposed as:

$$E_{\text{in}}(\underline{w}) = \frac{1}{2N} \left(\|X\underline{w} - \underline{y}_{\text{ls}}\|_2^2 + \|\underline{y} - \underline{y}_{\text{ls}}\|_2^2 \right),$$

where $\underline{y}_{\text{ls}} = X\underline{w}^\star$.
Answer. Note that

$$\begin{aligned} E_{\text{in}}(\underline{w}) &= \frac{1}{2N} \|X\underline{w} - \underline{y}\|_2^2 = \frac{1}{2N} \|X\underline{w} - \underline{y}_{\text{ls}} + \underline{y}_{\text{ls}} - \underline{y}\|_2^2 = \\ &= \frac{1}{2N} \left(\|X\underline{w} - \underline{y}_{\text{ls}}\|_2^2 + \|\underline{y}_{\text{ls}} - \underline{y}\|_2^2 + 2(X\underline{w} - \underline{y}_{\text{ls}})^\top (\underline{y}_{\text{ls}} - \underline{y}) \right). \end{aligned}$$

It suffices to show that $(X\underline{w} - \underline{y}_{\text{ls}})^\top (\underline{y}_{\text{ls}} - \underline{y}) = 0$. Observe that

$$\begin{aligned} (X\underline{w} - \underline{y}_{\text{ls}})^\top (\underline{y}_{\text{ls}} - \underline{y}) &= \underline{w}^\top X^\top \underline{y} - \underline{w}^\top X^\top \underline{y}_{\text{ls}} - \underline{y}_{\text{ls}}^\top \underline{y}_{\text{ls}} + \underline{y}_{\text{ls}}^\top \underline{y} \\ &= \underline{w}^\top X^\top \underline{y} - \underline{w}^\top X^\top X (X^\top X)^{-1} X^\top \underline{y} \\ &\quad - (X(X^\top X)^{-1} X^\top \underline{y})^\top X (X^\top X)^{-1} X^\top \underline{y} \\ &\quad + (X(X^\top X)^{-1} X^\top \underline{y})^\top \underline{y} \\ &= \underline{w}^\top X^\top \underline{y} - \underline{w}^\top X^\top \underline{y} \\ &\quad - \underline{y}^\top X (XX^\top)^{-1} X^\top X (X^\top X)^{-1} X^\top \underline{y} \\ &\quad + \underline{y}^\top X (XX^\top)^{-1} X^\top \underline{y} \\ &= 0 - \underline{y}^\top X (XX^\top)^{-1} X^\top \underline{y} + \underline{y}^\top X (XX^\top)^{-1} X^\top \underline{y} = 0. \end{aligned}$$

3.d Use the result in **3.c** to show that the least-squares solution is $\underline{w}^\star = (X^\top X)^{-1} X^\top \underline{y}$.

Answer. by the result in **3.c**, $E_{\text{in}}(\underline{w}) = \frac{1}{2N} \left(\|X\underline{w} - X(X^\top X)^{-1} X^\top \underline{y}\|_2^2 + \|X(X^\top X)^{-1} X^\top \underline{y} - \underline{y}\|_2^2 \right)$. Thus, $\underline{w}^\star$, the minimizer of $E_{\text{in}}(\underline{w})$, is the minimizer of $\|X\underline{w} - X(X^\top X)^{-1} X^\top \underline{y}\|_2^2$. Observe that $\|X\underline{w} - X(X^\top X)^{-1} X^\top \underline{y}\|_2^2$ is minimized when $X\underline{w} - X(X^\top X)^{-1} X^\top \underline{y} = 0$. Hence, at the optimal solution, $\underline{w}^\star$, we have

$$X\underline{w}^\star - X(X^\top X)^{-1} X^\top \underline{y} = 0 \Rightarrow X\underline{w}^\star = X(X^\top X)^{-1} X^\top \underline{y} \Rightarrow \underline{w}^\star = (X^\top X)^{-1} X^\top \underline{y}.$$

3.e Explain geometrically why for any $\underline{w}$, $(X\underline{w} - \underline{y}_{\text{ls}})^\top (\underline{y} - \underline{y}_{\text{ls}}) = 0$.

Answer. Note that $\underline{y}_{\text{ls}}$ is the projection of $\underline{y}$ onto the hyperplane col-span(X). Thus, $(\underline{y} - \underline{y}_{\text{ls}})$ is orthogonal to any line segment on this hyperplane. Assume an arbitrary $\underline{w}$. Observe that $\underline{y}_{\text{ls}}$ and $X\underline{w}$ are two points on col-span(X). Thus, $(X\underline{w} - \underline{y}_{\text{ls}})$ is a line segment on this hyperplane. Therefore, it is orthogonal to $(\underline{y} - \underline{y}_{\text{ls}})$. Hence, $(X\underline{w} - \underline{y}_{\text{ls}})^\top (\underline{y} - \underline{y}_{\text{ls}}) = 0$. See the figure below for a 3D illustration.