# *Linear Algebra, Data Science, and Machine Learning*

by Jeff Calder and Peter J. Olver

Springer Undergraduate Texts in Mathematics and Technology, 2025

## Errata

Last updated November 17, 2025

- Page 141: Equation with $R \circ S$ on left hand side:

  Change $\mathbf{u}_j \mathbf{u}_j^T$ to $\mathbf{u}_i \mathbf{u}_i^T$.

- Page 188: Proof of Theorem 6.9:

  Change $b = \mathbf{y}^T \mathbf{f}$ to $b = -\mathbf{y}^T \mathbf{f}$.

- Page 205, Exercise 4.8:

  Change $A$ to $H$ twice:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k (H \mathbf{x}_k - \mathbf{b}) \qquad \text{and} \qquad \mathbf{r}_k = H \mathbf{x}_k - \mathbf{b}.$$

- Page 213:

  Change $x$ to $\mathbf{x}$ on the line immediately below Eq. (6.79).

- Page 214: Definition 6.32:

  Change the definition to read "is the $n \times n$ self-adjoint matrix defined by the inequality".

- Page 225: Equation 6.109:

  Change $\mu^2$ to $\mu$.

- Page 275: Equation 7.62:

  Change the equation to read

$$\min_{\substack{\mathbf{w} \in \mathbb{R}^n \\ b \in \mathbb{R}}} \left\{ \| \mathbf{w} \|^2 \;\middle|\; y_i(\mathbf{x}_i \cdot \mathbf{w} - b) \geq 1, \quad i = 1, \ldots, m \right\}.$$

- Page 275: Equation 7.63:

  Change the equation to read

$$\min_{\mathbf{w}, b} \left\{ \| \mathbf{w} \|^2 \;\middle|\; \mathbf{z} \cdot \mathbf{w} - b \geq 1, \quad \mathbf{z} \cdot \mathbf{w} + b \geq 1 \right\}.$$

- Page 306: Theorem 7.22:

  The assumptions on convexity/concavity of $F$ should be switched, and $D$ must be nonempty. Also, while the result in Fan [72] holds for nonconvex setes $D$, Fan has a more general notion of convex/concave functions than we have given in the book, so to be more clear we also assume $D$ is convex. The correct statement of the theorem should read:

  **Theorem 7.22:** Let $D \subset \mathbb{R}^m$ be convex, compact, i.e., closed and bounded, and nonempty. Let $F\colon D \times \mathbb{R}^n \to \mathbb{R}$ be continuous, and suppose that $F(\mathbf{x}, \mathbf{y})$ is convex as a function of $\mathbf{y}$ for each fixed $\mathbf{x} \in D$, while $F(\mathbf{x}, \mathbf{y})$ is concave, i.e., $-F(\mathbf{x}, \mathbf{y})$ is convex, as a function of $\mathbf{x}$ for each fixed $\mathbf{y} \in \mathbb{R}^n$. Then,

  $$\min_{\mathbf{y} \in \mathbb{R}^n} \max_{\mathbf{x} \in D} F(\mathbf{x}, \mathbf{y}) = \max_{\mathbf{x} \in D} \min_{\mathbf{y} \in \mathbb{R}^n} F(\mathbf{x}, \mathbf{y}).$$

  As a side note, Theorem 7.22 is one place in the book where the minimum is really an infimum.

- Page 307: Equation (7.112):

  Change $\sum_{i=1}^{n}$ to $\sum_{i=1}^{m}$.

- Page 307: Remark 7.23:

  Replace “exactly those” with “a subset of the”.

- Page 309:

  Change $\sum_{i=1}^{n}$ to $\sum_{i=1}^{m}$.

- Page 319: Section 8.8.1. Kernel Principal Component Analysis

  Throughout this section there should be a standing assumption that $d \geq m$ (i.e., the feature dimension is at least as large as the numbrer of data points).

- Page 319: Section 8.8.1. Kernel Principal Component Analysis

  We can only define the principal components $\mathbf{q}_i$ for $\lambda_i > 0$. Thus, lines -10 through -5 should be replaced with:
  Notice that
  $$\| \mathbf{v}_i \|^2 = \| \underline{Z}^T \mathbf{p}_i \|^2 = \mathbf{p}_i^T \underline{Z}\, \underline{Z}^T \mathbf{p}_i = \lambda_i \| \mathbf{p}_i \|^2 = \lambda_i.$$
  Thus, if $\lambda_1, \ldots, \lambda_k > 0$, then the top $k$ principal components of the feature vector data $\mathbf{z}_1, \ldots, \mathbf{z}_m$ are given by
  $$\mathbf{q}_i = \frac{\mathbf{v}_i}{\| \mathbf{v}_i \|} = \lambda_i^{-1/2} \underline{Z}^T \mathbf{p}_i \qquad \text{for} \qquad i = 1, \ldots, k.$$

- Page 320: Theorem 8.4:

  The formula for $\mathbf{q}_i$ should read $\mathbf{q}_i = \lambda_i^{-1/2} Z^T \mathbf{p}_i$.

- Page 333: Equation (8.33):

  Change $Q$ to $Q^T$.

- Page 333: Equation (8.34):

  Change $Q_k$ to $Q_k^T$.

- Page 346, Exercise 4.1:

  Delete "How does the accuracy change with the number of principal components used?"

- Page 541, Eq. (10.94):

  Remove extra left parenthesis, so equation should read

$$\Phi(\mathbf{x}) = \prod_{i=1}^{n} \theta(x_i) = \prod_{i=1}^{n} \big( 1 - |\, x_i \,| \big)_+ .$$

- Page 543, line -6:

  Remove first "the" in "...it is the only the lower intrinsic...".

*Acknowledgements:*

Many thanks to Gilad Lerman and Markus Schweighofer for sending us their corrections to the printed text.