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 19, 2025

- Page 102: Remark:

  Change "two case" to "two cases".

- 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 191: Remark after Corollary 6.13:

  Insert "in Chapter 4" after "Exercise 2.6".

- Page 195:

  Change "Proposition 6.17" to "Theorem 6.17".

- 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 206: Motivation for the Conjugate Gradient Method:

  The formulas for the conjugate gradient method are all correct, but the derivation and motivation has an error. The method should minimize the corresponding quadratic function, and not the residual vector. The text starting right after Equation (6.65) and ending before "We begin with an initial guess" should be replaced with the text below.

  where $\mathbf{x}_0$ is some initial approximation to the solution. In view of the orthogonality condition, the coordinates of the solution vector are

  $$t_k = \frac{\langle\, \mathbf{x}^\star - \mathbf{x}_0 \,, \mathbf{v}_k \,\rangle_H}{\|\, \mathbf{v}_k \,\|_H^2}.$$

The conjugate gradient algorithm, to be derived below, computes the t_k and $\mathbf{v}_k$ iteratively, so that the k-th approximation to the solution is

$$\mathbf{x}_k = \mathbf{x}_0 + t_1 \mathbf{v}_1 + \ \cdots \ + t_k \mathbf{v}_k, \qquad \text{or, equivalently} \qquad \mathbf{x}_k = \mathbf{x}_{k-1} + t_k \mathbf{v}_k.$$

where, once $\mathbf{v}_k$ is specified, t_k is chosen to minimize the quadratic function (6.44). The secret is not to try to specify the conjugate basis vectors in advance, but rather to successively construct them during the course of the algorithm.

- 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 μ^2 to μ.

- 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 300: Line directly above Theorem 7.17:

  Remove the text ", as well as soft-margin SVM (7.64)". Technically soft-margin SVM does not fit into this loss function due to the extra parameter b.

- Page 302: Definition 7.18:

  The definition of Mercer kernel should be amended to include a requirement that $\mathcal{K}$ is continuous.

- 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 number of data points).

- Page 319: Section 8.8.1. Kernel Principal Component Analysis

  We can of course only define the principal components $\mathbf{q}_i$ when $\lambda_i > 0$, and we should also check that $\mathbf{v}_1, \ldots, v_k$ are orthogonal. Thus, lines -10 through -5 should be replaced with:

Notice that
$$\mathbf{v}_i \cdot \mathbf{v}_j = (\underline{Z}^T \mathbf{p}_i)^T \underline{Z}^T \mathbf{p}_j = \mathbf{p}_i^T \underline{Z}\,\underline{Z}^T \mathbf{p}_j = \lambda_j \, \mathbf{p}_i \cdot \mathbf{p}_j.$$
Thus, if $\lambda_1, \ldots, \lambda_k > 0$, then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are orthogonal and the top k principal components of the feature vector data $\mathbf{z}_1, \ldots, \mathbf{z}_m$ are given by the corresponding unit vectors
$$\mathbf{q}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} = \lambda_i^{-1/2}\, \underline{Z}^T \mathbf{p}_i \qquad \text{for} \qquad i = 1, \ldots, k.$$

- 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 Larry Baker, Gilad Lerman and Markus Schweighofer for sending us their corrections to the printed text.