

## Chapter 1. Optimization Overview

**Page 22.** About the gradient in Eq. (1.23),

$$\nabla f = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_n \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}$$

where the second entry should be  $\partial f / \partial x_2$ .

## Chapter 2. Gradient Based Optimization

**Page 48.** About the gradient descent formula in Eq. (2.16a),

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma(\mathbf{Ax}_k - \mathbf{b})$$

should be

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \gamma(\mathbf{Ax}_k + \mathbf{b})$$

## Chapter 2. Gradient Based Optimization

**Page 50.** About the gradient descent update in Eq. (2.19),

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k - \gamma \nabla f(\mathbf{x}_k) + \beta_k \mathbf{v}_k \\ \mathbf{v}_{k+1} &= \beta \mathbf{v}_k - \gamma \nabla f(\mathbf{x}_k)\end{aligned}$$

where the hyper-parameter  $\beta_k$  should be  $\beta$ .

In my mind, this equation could be simplified as follows,

$$\begin{aligned}\mathbf{v}_{k+1} &= \beta \mathbf{v}_k - \gamma \nabla f(\mathbf{x}_k) \\ \mathbf{x}_{k+1} &= \mathbf{x}_k + \mathbf{v}_{k+1}\end{aligned}$$

## Chapter 2. Gradient Based Optimization

**Page 51.** Multiplying both sides by  $\Delta t^2$  and grouping terms, this simplifies as

$$m(\mathbf{x}_{k+1} - 2\mathbf{x}_k + \mathbf{x}_{k+1}) = -\Delta t^2 \nabla f(\mathbf{x}_k) - d\Delta t(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

should be

$$m(\mathbf{x}_{k+1} - 2\mathbf{x}_k + \mathbf{x}_{k+1}) = -\Delta t^2 \nabla f(\mathbf{x}_k) - \delta\Delta t(\mathbf{x}_{k+1} - \mathbf{x}_k)$$

This typo also appears in the left-hand side:

$$(m + d\Delta t)\mathbf{x}_{k+1} = \dots$$

which should be

$$(m + \delta\Delta t)\mathbf{x}_{k+1} = \dots$$

## Chapter 2. Gradient Based Optimization

**Page 72.** For the matrix-vector product in Eq. (2.66a), it should be

$$\mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{nj}x_j \end{bmatrix}$$

Correspondingly, the gradient in Eq. (2.67) should be

$$\nabla(\mathbf{Ax}) = \begin{bmatrix} \frac{\partial}{\partial x_1} \left( \sum_{j=1}^n a_{1j}x_j \right) & \frac{\partial}{\partial x_2} \left( \sum_{j=1}^n a_{1j}x_j \right) & \cdots \\ \frac{\partial}{\partial x_1} \left( \sum_{j=1}^n a_{2j}x_j \right) & \frac{\partial}{\partial x_2} \left( \sum_{j=1}^n a_{2j}x_j \right) & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \mathbf{A}$$

## Chapter 2. Gradient Based Optimization

**Page 73.** The formula in Eq. (2.71a) should be

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} \sum_{j=1}^n a_{1j} x_j \\ \sum_{j=1}^n a_{2j} x_j \\ \vdots \\ \sum_{j=1}^n a_{nj} x_j \end{bmatrix}$$

## Chapter 4. Least-Squares Regression

**Page 115.** In the caption of Figure 4.2, “...or under-determined (left)...” should be “...or under-determined (right)...”.