From Linear Algebra to Optimization

https://xinychen.github.io/tutorial/alg_opt.pdf

Xinyu Chen

November 5, 2025

Contents

1 Natural Gradient Descent

For any function $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ with respect to the vector $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^\top \in \mathbb{R}^n$ of length n, gradient is actually the first-order derivative such that

$$\nabla f(\boldsymbol{x}) = \frac{\mathrm{d}f(\boldsymbol{x})}{\mathrm{d}\boldsymbol{x}} = \begin{bmatrix} \frac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \frac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix} \in \mathbb{R}^n, \tag{1.1}$$

where $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$ is the partial derivative of $f(\boldsymbol{x})$ with respect to the i-th entry of $\boldsymbol{x}$, denoted by $x_i, \forall i \in \{1, 2, \ldots, n\}$.

2 Procrustes Problems

References