Pure Mathematics 理论数学, 2024, 14(8), 20-24
Published Online August 2024 in Hans. https://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2024.148299

Hans汉斯

多元函数梯度及其应用

刘　帅[1]，赵　迪[2]

[1]上海理工大学理学院，上海
[2]上海理工大学管理学院，上海

收稿日期：2024年7月1日；录用日期：2024年8月2日；发布日期：2024年8月14日

摘　要

多元函数的梯度是微积分中的一个重要概念，在分析学中占有举足轻重的地位，它允许我们在多维空间中对函数进行深入的理解和操作。梯度不仅揭示了函数在特定点的局部行为，还为优化问题提供了方向性指导。在数学、物理学、工程学以及其他科学领域，梯度的概念和应用都极为广泛。

关键词

多元函数梯度，方向导数，高等数学，梯度下降法

Multivariate Function Gradient and Its Application

Shuai Liu[1], Di Zhao[2]

[1]College of Science, University of Shanghai for Science and Technology, Shanghai
[2]Business School of University of Shanghai for Science and Technology, Shanghai

Received: Jul. 1st, 2024; accepted: Aug. 2nd, 2024; published: Aug. 14th, 2024

Abstract

The gradient of multivariate functions is an important concept in calculus and occupies a pivotal position in the field of analysis. It allows us to deeply understand and manipulate functions within multidimensional spaces. The gradient not only reveals the local behavior of a function at specific points but also provides directional guidance for optimization problems. The concept and application of the gradient are extremely broad in mathematics, physics, engineering, and other scientific fields.

文章引用: 刘帅, 赵迪. 多元函数梯度及其应用[J]. 理论数学, 2024, 14(8): 20-24. DOI: 10.12677/pm.2024.148299

Keywords

Gradient of Multiple Functions, Directional Derivative, Advanced Mathematics, Gradient Descent Method

1. 引言

17 世纪，牛顿和莱布尼茨独立发展了微积分的基本概念，包括导数，这是微积分学的基础。随着微积分理论的发展，数学家们开始研究多元函数的导数问题，即如何量化一个函数在多维空间中的变化率。在多元函数的情况下，导数的概念被推广为偏导数，它描述了函数沿某一特定方向的变化率。梯度是偏导数的自然扩展，它不仅告诉我们函数在哪个方向上增长最快，还提供了这个方向的量化表示[1]。19 世纪，数学家们引入了 Nabla 算子，这是一个向量微分算子，用于表示梯度、散度和旋度等概念。随着数学分析的深入，梯度的概念被进一步发展和完善，成为多元函数分析中的核心概念之一[2]。梯度的应用从最初的理论探索，扩展到物理学、工程学、经济学等多个领域，成为解决实际问题的重要工具。在现代数学中，梯度的概念与其他数学分支，如泛函分析、流形上的微积分等，相互融合，形成了更为丰富的数学理论体系。

梯度下降法是一种一阶迭代优化算法，用于求解目标函数的最小值。其核心思想是利用负梯度方向作为搜索方向，因为在多元函数的某一点处，函数值沿着负梯度方向下降最快[3]。在机器学习中，梯度下降法是训练神经网络和其他机器学习模型的基本算法，用于调整模型参数，降低损失函数的值；在深度学习中，梯度下降法用于优化模型的参数，通过不断迭代更新参数，使神经网络的损失函数最小；在经济学中，梯度下降法可以用于求解某些优化问题，如资源分配问题；在工程设计和控制理论中，梯度下降法可以用于系统参数的优化；在信号处理领域，梯度下降法可以用于滤波器设计和信号估计。梯度下降法的变体包括批量梯度下降、随机梯度下降[4]和小批量梯度下降。这些变体在不同的应用场景下有各自的优势，例如随机梯度下降适用于大规模数据集，而批量梯度下降在数据量较小时计算更为精确。在实际应用中，梯度下降法需要考虑一些实用技巧，如选择合适的学习率、检查梯度流问题以及使用自适应技术等，以确保算法的效率和准确性。此外，梯度下降法也存在一些问题，例如可能会陷入局部最小值或鞍点，这需要通过更高级的优化算法来解决[5]。

本文将从梯度的基本概念出发，进一步扩展梯度在瞎子爬山以及机器学习中的应用。

2. 梯度

2.1. 问题的引入

一个人被困在山上，需要从山上下来。但此时山上的浓雾很大，导致可视度很低。因此，下山的路径就无法确定，他必须利用自己周围的信息去找到下山的路径。这个时候，他就可以利用梯度下降算法来帮助自己下山。具体来说就是，以他当前的所处的位置为基准，寻找这个位置最陡峭的地方，然后朝着山的高度下降的地方走，然后每走一段距离，都反复采用同一个方法，最后就能成功的抵达山谷。

2.2. 梯度的概念

2.2.1. 偏导数的定义

设二元函数 $f(x,y)$ 在点 $P(x_0,y_0)$ 处偏导存在，根据偏导数的定义，有

$$f_x(P)=\lim_{\Delta x\to 0}\frac{f(x_0+\Delta x,y_0)-f(x_0,y_0)}{\Delta x}$$

$f_x(P)$ 数值上表示函数 $f(x,y)$ 在点 P 沿 x 轴正方向的变化率。

2.2.2. 方向导数的定义

设二元函数 $f(x,y)$ 在点 $P(x_0,y_0)$ 处沿 l 的方向导数 $\frac{\partial f}{\partial l}$ 存在，l 方向单位向量 $\boldsymbol{e}_l=(\cos\alpha,\cos\beta)$，则方向导数可以表示为

$$\left.\frac{\partial f}{\partial l}\right|_P=\lim_{\rho\to 0^+}\frac{f(x_0+\rho\cos\alpha,y_0+\rho\cos\beta)-f(x_0,y_0)}{\rho}$$

例：求 $f(x,y,z)=xy+yz+zx$ 在点 $P(1,2,2)$ 沿方向 l 的方向导数，其中 l 的方向角分别为 $60^\circ,45^\circ,60^\circ$。

解：与 l 同向的单位向量为

$$\boldsymbol{e}_l=(\cos 60^\circ,\cos 45^\circ,\cos 60^\circ)=\left(\frac{1}{2},\frac{\sqrt{2}}{2},\frac{1}{2}\right)$$

计算可得方向导数为

$$\left.\frac{\partial f}{\partial l}\right|_P=\lim_{\rho\to 0^+}\frac{f\left(1+\frac{1}{2}\rho,2+\frac{\sqrt{2}}{2}\rho,2+\frac{1}{2}\rho\right)-f(1,2,2)}{\rho}=\frac{1}{2}\left(7+3\sqrt{2}\right)$$

2.2.3. 二元函数的梯度

二元函数 $f(x,y)$ 在点 P 的所有方向导数中，能否找到最大值？分析表明，$G=(f_x(P),f_y(P))$ 很特别，该方向上能取得方向导数的最大值，从而引出梯度的定义。

设函数 $f(x,y)$ 在平面区域 D 内具有一阶连续偏导数，则对于区域 D 内的每一点 $P(x_0,y_0)$，都可以确定出一个向量 $f_x(x_0,y_0)\boldsymbol{i}+f_y(x_0,y_0)\boldsymbol{j}$，称该向量为函数 $f(x,y)$ 在点 $P(x_0,y_0)$ 处的梯度，记作 $\boldsymbol{grad}\,f(x_0,y_0)$ 或 $\boldsymbol{grad}\,f(P)$，即

$$\boldsymbol{grad}\,f(x_0,y_0)=f_x(x_0,y_0)\boldsymbol{i}+f_y(x_0,y_0)\boldsymbol{j}$$

说明 1：二元函数的梯度是平面上的一个向量，是该函数关于 x 的偏导数向量与关于 y 的偏导数向量的和向量。

说明 2：函数在点 P 处的梯度方向即能够使得方向导数取得最大值的方向。

说明 3：由于方向导数表明了在点 P 处沿某方向函数值的变化率，方向导数取得最大值意味着函数值增加得最快，因此梯度方向是函数值增加最快的方向。

例：设函数 $f(x,y)=\ln(x+y^2)$，求其在点 $(1,1)$ 处的梯度。

解：函数在点 $(1,1)$ 处的偏导数为

$$f_x(1,1)=\left.\frac{1}{x+y^2}\right|_{(1,1)}=\frac{1}{2}$$

$$f_y(1,1)=\frac{2y}{x+y^2}\bigg|_{(1,1)}=1$$

可得该函数在$(1,1)$处的梯度为

$$\boldsymbol{grad}\, f(1,1)=\left(f_x(1,2)+f_y(1,1)\right)=\left(\frac{1}{2},1\right)$$

2.2.4. 三元函数的梯度

设函数$f(x,y,z)$在空间区域D内具有一阶连续偏导数，则对于区域D内的每一点$P(x_0,y_0,z_0)$，都可以确定出一个向量$f_x(x_0,y_0)\boldsymbol{i}+f_y(x_0,y_0)\boldsymbol{j}+f_z(x_0,y_0)\boldsymbol{k}$，称该向量为函数$f(x,y,z)$在点$P(x_0,y_0,z_0)$处的梯度，记作$\boldsymbol{grad}\, f(x_0,y_0,z_0)$或$\boldsymbol{grad}\, f(P)$，即

$$\boldsymbol{grad}\, f(x_0,y_0,z_0)=f_x(x_0,y_0)\boldsymbol{i}+f_y(x_0,y_0)\boldsymbol{j}+f_z(x_0,y_0)\boldsymbol{k}$$

2.2.5. 三元函数梯度与方向导数的关系

设三元函数$f(x,y,z)$在$P(x_0,y_0,z_0)$处可微，$\boldsymbol{e}_l=(\cos\alpha,\cos\beta,\cos\gamma)$是与方向$l$同向的单位向量，则

$$\frac{\partial f}{\partial l}\bigg|_P=f_x(P)\cos\alpha+f_y(P)\cos\beta+f_z(P)\cos\gamma=\boldsymbol{grad}\, f(P)\cdot\boldsymbol{e}_l$$

3. 多元函数梯度的应用

3.1. 瞎子爬山问题的解决方法

“瞎子爬山”是一种形象的比喻，用来描述一种优化算法，其中“瞎子”代表了算法在没有全局信息的情况下寻找最优解的过程。在这种比喻中，梯度下降法可以视为一种“瞎子爬山”的策略，其核心思想如下：

1) 局部最优：瞎子(即优化算法)只能感知到局部的梯度信息，就像一个盲人只能通过脚来感知地面的坡度。

2) 随机起点：瞎子可能从山的任意位置开始，这对应于优化算法中的随机初始化。

3) 梯度指导：瞎子通过感知地面的坡度(即梯度)来决定下一步的移动方向。在优化问题中，这意味着算法根据目标函数的梯度来更新参数。

4) 局部搜索：由于只能感知局部信息，瞎子可能会陷入局部最高点，而不是全局最高点。同样，梯度下降法可能会找到局部最小值而非全局最小值。

5) 迭代过程：瞎子通过不断移动来逐渐接近山顶，这对应于梯度下降法中的迭代过程。

6) 学习率：瞎子每次移动的步伐大小可以类比于梯度下降法中的学习率。如果步伐太大，可能会跳过山顶；步伐太小，则可能收敛缓慢。

7) 收敛条件：瞎子爬山的停止条件可以是达到一定的高度或不再有更高的点可移动，这对应于梯度下降法中的收敛条件，如梯度足够小或达到预设的迭代次数。

3.2. 梯度在机器学习中的应用

梯度下降法是机器学习中最基本的优化技术之一，广泛应用于线性回归、逻辑回归、神经网络等模型的训练过程中。梯度下降是一种优化算法，用于最小化损失函数，即模型预测值与实际值之间的差异。梯度下降法的核心思想是通过迭代调整模型参数，以减少预测误差。以下是梯度下降法在机器学习中的应用和关键点：

1) 损失函数：首先定义一个损失函数(也称为目标函数或代价函数)，该函数衡量模型预测值与实际值之间的差距。常见的损失函数包括均方误差(MSE)、交叉熵损失等。

2) 参数初始化：在训练开始之前，需要初始化模型的参数。这些参数可以随机初始化，也可以根据问题的特性进行特定的初始化。

3) 梯度计算：在每次迭代中，计算损失函数对每个参数的偏导数，即梯度。梯度指向损失函数增长最快的方向。

4) 参数更新：根据计算得到的梯度和预设的学习率，更新模型参数。更新规则通常是：$\theta_{\text{new}} = \theta_{\text{old}} - \mu\nabla L(\theta)$，其中 θ 表示模型参数，μ 是学习率，$\nabla L(\theta)$ 是损失函数 L 对参数 θ 的梯度。

5) 迭代过程：重复步骤 3)和 4)，直到模型的损失函数达到一个可接受的最小值，或者达到预设的迭代次数。

6) 学习率调整：学习率是梯度下降算法中的关键超参数。太小的学习率会导致收敛速度慢，而太大的学习率可能导致跳过最小值或不稳定的收敛。

7) 收敛条件：梯度下降算法的停止条件可以是梯度的大小小于某个阈值，或者迭代次数达到预设的上限。

8) 局部最小值和鞍点：梯度下降可能会陷入局部最小值或鞍点，特别是当损失函数是非凸函数时。为了解决这个问题，可以使用更高级的优化算法，如牛顿方法、共轭梯度法等。

9) 正则化：为了防止过拟合，可以在损失函数中添加正则化项，如 L1 或 L2 正则化。

4. 结语

本文首先介绍了梯度的起源和发展，然后给出了梯度的概念以及与方向导数的关系，最后重点探讨了梯度在瞎子爬山以及机器学习中的应用。

参考文献

[1] 同济大学数学科学学院. 高等数学[M]. 北京: 高等教育出版社, 2021.

[2] 陈纪修, 於崇华, 金路. 数学分析[M]. 北京: 高等教育出版社, 2019.

[3] 史加荣, 王丹, 尚凡华, 张鹤. 随机梯度下降算法研究进展[J]. 自动化学报, 2021, 47(9): 2103-2119.

[4] 邱松强. 特征值与梯度下降算法[J]. 高等数学研究, 2023, 26(3): 36-39.

[5] Ruder, S. (2016) An Overview of Gradient Descent Optimization Algorithms. arXiv.1609.04747.