最优化方法/实用优化算法
第一章 最优化问题概述

教师 强静

中国矿业大学 数学学院

April 26, 2023

目录

推荐书目

- 【教材】解可新, 韩健, 林友联, 最优化方法, 天津大学出版社.
- 孙文瑜, 徐成贤, 朱德通, 最优化方法, 高等教育出版社.
- 高立, 数值最优化方法, 北京大学出版社.
- E.K.P. Chong, S.H. Żak, 最优化导论, 电子工业出版社.
- A.D. Belegundu, T.R. Chandrupatla, 最优化技术导论与工程应用, 电子工业出版社.

问题的基本形式

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \ & \boldsymbol{x} \in \Omega \end{aligned}$$

其中，

问题的基本形式

$$\begin{aligned} &\min \quad f(\boldsymbol{x}) \\ &\text{s.t. } \boldsymbol{x} \in \Omega \end{aligned}$$

其中,

- $f : R^n \to f$ 称为目标函数;

问题的基本形式

$$\min \quad f(\boldsymbol{x})$$
$$\text{s.t. } \boldsymbol{x} \in \Omega$$

其中,

- $f: R^n \to f$ 称为目标函数;
- $\boldsymbol{x} = (x_1, \cdots, x_n)^T \in R^n$ 为决策变量;

问题的基本形式

$$\begin{aligned} &\min \quad f(\boldsymbol{x}) \\ &\text{s.t.}\ \boldsymbol{x} \in \Omega \end{aligned}$$

其中,

- $f: R^n \to f$ 称为目标函数;
- $\boldsymbol{x} = (x_1, \cdots, x_n)^T \in R^n$ 为决策变量;
- $\Omega \subset R^n$ 称为可行域.

数学模型

- $\Omega = R^n$：无约束;

数学模型

- $\Omega = R^n$: 无约束;
- $\Omega = [\boldsymbol{l}, \boldsymbol{u}], \boldsymbol{l}, \boldsymbol{u} \in R^n$: 简单界（盒子）约束;

数学模型

- $\Omega = R^n$：无约束;
- $\Omega = [\boldsymbol{l}, \boldsymbol{u}], \boldsymbol{l}, \boldsymbol{u} \in R^n$：简单界（盒子）约束;
- $\Omega = \{\boldsymbol{x} \mid h(\boldsymbol{x}) = 0\}$, $h(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \cdots, h_m(\boldsymbol{x}))^T$：等式约束;

数学模型

- $\Omega = R^n$: 无约束;
- $\Omega = [\boldsymbol{l}, \boldsymbol{u}], \boldsymbol{l}, \boldsymbol{u} \in R^n$: 简单界（盒子）约束;
- $\Omega = \{\boldsymbol{x} \mid h(\boldsymbol{x}) = 0\}, h(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \cdots, h_m(\boldsymbol{x}))^T$: 等式约束;
- $\Omega = \{\boldsymbol{x} \mid g(\boldsymbol{x}) \geq 0\}, g(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \cdots, g_p(\boldsymbol{x}))^T$: 不等式约束;

数学模型

- $\Omega = R^n$: 无约束;
- $\Omega = [\boldsymbol{l}, \boldsymbol{u}]$, $\boldsymbol{l}, \boldsymbol{u} \in R^n$: 简单界（盒子）约束;
- $\Omega = \{\boldsymbol{x} \mid h(\boldsymbol{x}) = 0\}$, $h(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \cdots, h_m(\boldsymbol{x}))^T$: 等式约束;
- $\Omega = \{\boldsymbol{x} \mid g(\boldsymbol{x}) \geq 0\}$, $g(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \cdots, g_p(\boldsymbol{x}))^T$: 不等式约束;
- $\Omega = \{\boldsymbol{x} \mid h(\boldsymbol{x}) = 0,\ g(\boldsymbol{x}) \geq 0\}$ 或

数学模型

- $\Omega = R^n$: 无约束;
- $\Omega = [\boldsymbol{l}, \boldsymbol{u}]$, $\boldsymbol{l}, \boldsymbol{u} \in R^n$: 简单界（盒子）约束;
- $\Omega = \{\boldsymbol{x} \mid h(\boldsymbol{x}) = 0\}$, $h(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \cdots, h_m(\boldsymbol{x}))^T$: 等式约束;
- $\Omega = \{\boldsymbol{x} \mid g(\boldsymbol{x}) \geq 0\}$, $g(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \cdots, g_p(\boldsymbol{x}))^T$: 不等式约束;
- $\Omega = \{\boldsymbol{x} \mid h(\boldsymbol{x}) = 0,\ g(\boldsymbol{x}) \geq 0\}$ 或
- $\Omega = \{\boldsymbol{x} \mid h(\boldsymbol{x}) = 0,\ g(\boldsymbol{x}) \geq 0, l \leq \boldsymbol{x} \leq u\}$: 一般约束.

最优化的历史

- 阿基米德证明：给定周长，圆所包围的面积为最大.

最优化的历史

- 阿基米德证明：给定周长，圆所包围的面积为最大.
- 最优化方法真正形成为科学方法则在17世纪以后。17世纪，I.牛顿和G.W.莱布尼茨在他们所创建的微积分中，提出求解具有多个自变量的实值函数的最大值和最小值的方法.

最优化的历史

- 阿基米德证明：给定周长，圆所包围的面积为最大.
- 最优化方法真正形成为科学方法则在17世纪以后。17世纪，I.牛顿和G.W.莱布尼茨在他们所创建的微积分中，提出求解具有多个自变量的实值函数的最大值和最小值的方法.
- 第二次世界大战前后，由于军事上的需要和科学技术和生产的迅速发展，许多实际的最优化问题已经无法用古典方法来解决，这就促进了近代最优化方法的产生.

关于最优化的几件事情

- 以苏联康托罗维奇和美国丹齐格为代表的线性规划.

关于最优化的几件事情

- 以苏联康托罗维奇和美国丹齐格为代表的线性规划.
- 以美国库恩和塔克尔为代表的非线性规划.

关于最优化的几件事情

- 以苏联康托罗维奇和美国丹齐格为代表的线性规划.
- 以美国库恩和塔克尔为代表的非线性规划.
- 以美国贝尔曼为代表的动态规划.

微积分中的极值问题

例2.1 (一元无约束问题)

求 $f(x) = (x^2 - 1)^3 + 1,\ x \in R$ 的最小值.

微积分中的极值问题

例2.1 (一元无约束问题)

求 $f(x) = (x^2 - 1)^3 + 1,\ x \in R$ 的最小值.

经常把这个问题写成如下形式

$$\min_{x \in R} f(x) = (x^2 - 1)^3 + 1.$$

微积分中的极值问题

例2.1 (一元无约束问题)

求 $f(x) = (x^2 - 1)^3 + 1$, $x \in R$ 的最小值.
经常把这个问题写成如下形式

$$\min_{x \in R} f(x) = (x^2 - 1)^3 + 1.$$

【解析：】令 $f'(x) = 0$ 并讨论即可.

微积分中的极值问题

例2.2 (二元无约束问题)

求

$$\min f(x,y) = x^3 + y^3 + \frac{3}{2}y^2 - 3x^2 + 1.$$

微积分中的极值问题

例2.2 (二元无约束问题)

求

$$\min f(x,y) = x^3 + y^3 + \frac{3}{2}y^2 - 3x^2 + 1.$$

【解析：】只需考虑

$$\frac{\partial f}{\partial x} = 3x^2 - 6x = 0, \quad \frac{\partial f}{\partial y} = 3y^2 + 3y = 0.$$

并进行讨论.

微积分中的极值问题

例2.3 (条件极值问题)

求$f(x,y)=x^2+y^2$ 在条件$x+y=1$ 下的极小值.

微积分中的极值问题

例2.3 (条件极值问题)

求$f(x,y)=x^2+y^2$ 在条件$x+y=1$ 下的极小值.

常将此问题写成如下形式

$$\min\ f(x,y)=x^2+y^2,$$
$$s.t.\ x+y=1.$$

微积分中的极值问题

例2.3 (条件极值问题)

求$f(x,y)=x^2+y^2$ 在条件$x+y=1$ 下的极小值.

常将此问题写成如下形式

$$\min\ f(x,y)=x^2+y^2,$$
$$s.t.\ x+y=1.$$

【解析：】写出Lagrange 函数

$$L(x,y,\lambda)=x^2+y^2-\lambda(x+y-1).$$

令Lagrange函数的导数等于0,

令Lagrange函数的导数等于0，得

$$\begin{cases} \frac{\partial L}{\partial x} = 2x - \lambda = 0, \\ \frac{\partial L}{\partial y} = 2y - \lambda = 0, \\ \frac{\partial L}{\partial \lambda} = x + y - 1 = 0. \end{cases}$$

令Lagrange函数的导数等于0，得

$$\begin{cases} \dfrac{\partial L}{\partial x} = 2x - \lambda = 0, \\ \dfrac{\partial L}{\partial y} = 2y - \lambda = 0, \\ \dfrac{\partial L}{\partial \lambda} = x + y - 1 = 0. \end{cases}$$

解得

$$x = y = \frac{1}{2}, \ \lambda = 1.$$

微积分中的Newton 法

问题：如果无法或很难求出前面三个例题中的方程（组）的精确解（解析解）,该怎么办？

微积分中的Newton 法

- Newton 法也叫切线法，其基本思想是迭代地求方程的近似解.

微积分中的Newton 法

- Newton 法也叫切线法，其基本思想是迭代地求方程的近似解.
- 设求解$f(\boldsymbol{x}) = 0$.

微积分中的Newton 法

- Newton 法也叫切线法，其基本思想是迭代地求方程的近似解.
- 设求解$f(\boldsymbol{x}) = 0$.
- 迭代公式

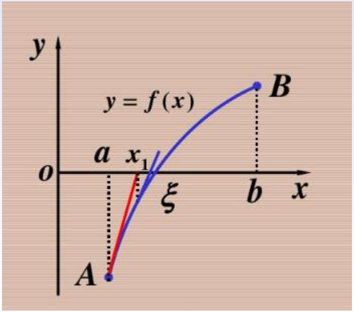
$$\boldsymbol{x}_k = \boldsymbol{x}_{k-1} - \frac{f(x_{k-1})}{f'(x_{k-1})}.$$

微积分中的Newton 法

- Newton 法也叫切线法, 其基本思想是迭代地求方程的近似解.
- 设求解$f(\boldsymbol{x})=0$.
- 迭代公式

$$\boldsymbol{x}_k=\boldsymbol{x}_{k-1}-\frac{f(x_{k-1})}{f'(x_{k-1})}.$$

- 如图



例2.4

用$Newton$法（切线法）求方程

$$f(x) = x^3 + 1.1x^2 + 0.9x - 1.4 = 0$$

的实根的近似值，使误差不超过10^{-3}.

【解:】取初始值为$x_0 = 1$. 连续应用迭代公式，得

【解:】取初始值为$x_0 = 1$. 连续应用迭代公式，得

$$x_1 = 1 - \frac{f(1)}{f'(1)} \approx 0.738,$$

【解:】取初始值为$x_0=1$. 连续应用迭代公式，得

$$x_1 = 1 - \frac{f(1)}{f'(1)} \approx 0.738,$$

$$x_2 = 0.738 - \frac{f(0.738)}{f'(0.738)} \approx 0.674,$$

【解:】取初始值为$x_0 = 1$. 连续应用迭代公式，得

$$x_1 = 1 - \frac{f(1)}{f'(1)} \approx 0.738,$$

$$x_2 = 0.738 - \frac{f(0.738)}{f'(0.738)} \approx 0.674,$$

$$x_3 = 0.674 - \frac{f(0.674)}{f'(0.674)} \approx 0.671,$$

【解:】取初始值为$x_0 = 1$. 连续应用迭代公式，得

$$x_1 = 1 - \frac{f(1)}{f'(1)} \approx 0.738,$$

$$x_2 = 0.738 - \frac{f(0.738)}{f'(0.738)} \approx 0.674,$$

$$x_3 = 0.674 - \frac{f(0.674)}{f'(0.674)} \approx 0.671,$$

$$x_4 = 0.671 - \frac{f(0.671)}{f'(0.671)} \approx 0.671,$$

算法终止.

二维最优化问题的目标函数

$$z = f(x_1, x_2)$$

表示三维空间R^3 中的曲面.

二维最优化问题的目标函数

$$z = f(x_1, x_2)$$

表示三维空间R^3 中的曲面.

等高线或等值线

称投影曲线

$$\begin{cases} f(x_1, x_2) = c, \\ z = 0 \end{cases}$$

为目标函数的等高线, 或等值线.

二维最优化问题的目标函数

$$z = f(x_1, x_2)$$

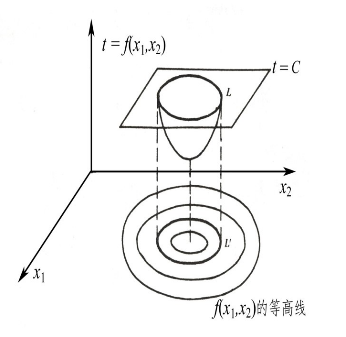
表示三维空间R^3 中的曲面.

等高线或等值线

称投影曲线

$$\begin{cases} f(x_1, x_2) = c, \\ z = 0 \end{cases}$$

为目标函数的等高线，或等值线.



例题

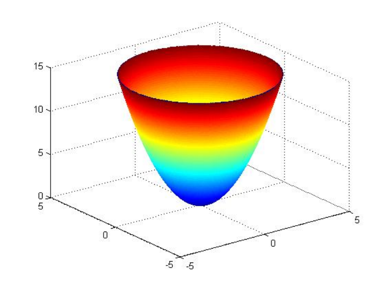
例2.5

画出曲面$f(x_1,x_2)=x_1^2+x_2^2$ 的图像和等高线.

例题

例2.5

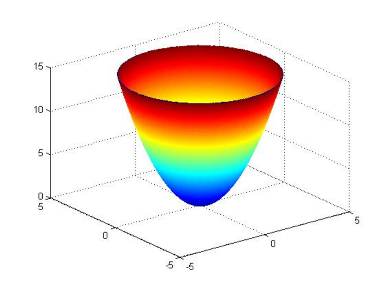
画出曲面$f(x_1,x_2)=x_1^2+x_2^2$的图像和等高线.

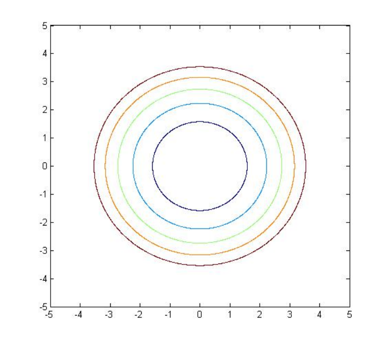


例题

例2.5

画出曲面$f(x_1,x_2)=x_1^2+x_2^2$ 的图像和等高线.





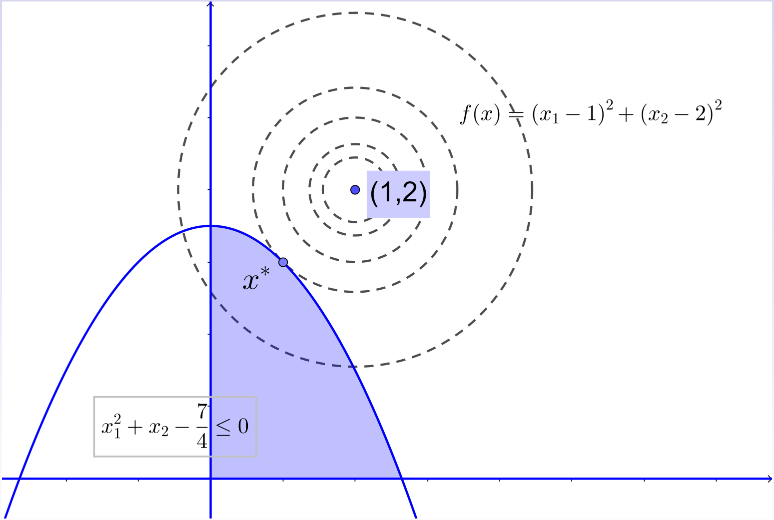
例题

例2.6

$$
\begin{aligned}
\min\ & (x_1-1)^2+(x_2-2)^2 \\
s.t.\ & x_1^2+x_2-\frac{7}{4}\le 0, \\
& x_1,x_2\ge 0.
\end{aligned}
$$

例题

如图



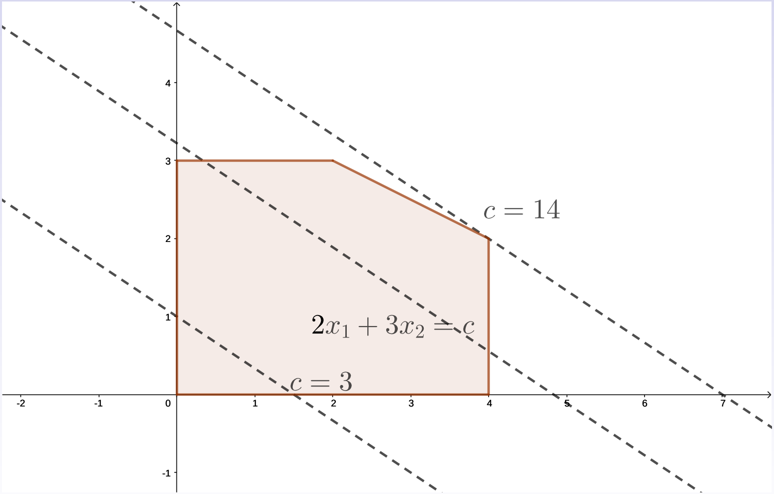
例题

例2.7

$$
\begin{aligned}
\max\ & 2x_1 + 3x_2 \\
s.t.\ & x_1 + 2x_2 \leq 8, \\
& x_1 \leq 4, \\
& x_2 \leq 3, \\
& x_1, x_2 \geq 0.
\end{aligned}
$$

例题

如图



微积分基础

给定函数$f: R^n \to R$可微, 则称

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \dfrac{\partial f}{\partial x_2}(\boldsymbol{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}$$

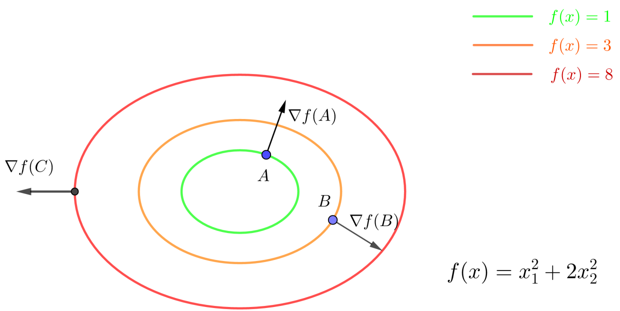
微积分基础

给定函数$f : R^n \to R$ 可微, 则称

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \dfrac{\partial f}{\partial x_2}(\boldsymbol{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{bmatrix}$$

为f 的**梯度**. 常将梯度表示为$g(\boldsymbol{x}) = \nabla f(\boldsymbol{x})$.

梯度的几何意义



梯度的几何意义：指向函数值增长最快的方向
（上述三个向量的长度均经标准化变为 1.）

微积分基础

若 f 是二次可微的, 则称

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_1} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n} & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

微积分基础

若 f 是二次可微的，则称

$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_1} \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f}{\partial x_1 \partial x_n} & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

为 f 在点 $\boldsymbol{x}$ 处的**Hesse 矩阵.**

微积分基础

给定函数$c: R^n \to R^m$ 可微, 即

$$c(\boldsymbol{x}) = \begin{bmatrix} c_1(\boldsymbol{x}) \\ \vdots \\ c_m(\boldsymbol{x}) \end{bmatrix}, \text{ 其中} c_i(\boldsymbol{x}): R^n \to R.$$

则$c(\boldsymbol{x})$ 的梯度可表示为

微积分基础

$$\nabla c(\boldsymbol{x}) = \begin{bmatrix} \nabla c_1(\boldsymbol{x}) & \nabla c_2(\boldsymbol{x}) & \cdots & \nabla c_m(\boldsymbol{x}) \end{bmatrix}$$

微积分基础

$$
\begin{aligned}
\nabla c(\boldsymbol{x}) &= \left[\nabla c_1(\boldsymbol{x}) \quad \nabla c_2(\boldsymbol{x}) \quad \cdots \quad \nabla c_m(\boldsymbol{x})\right] \\
&= \begin{bmatrix}
\frac{\partial c_1}{\partial x_1}(\boldsymbol{x}) & \frac{\partial c_2}{\partial x_1}(\boldsymbol{x}) & \cdots & \frac{\partial c_m}{\partial x_1}(\boldsymbol{x}) \\
\frac{\partial c_1}{\partial x_2}(\boldsymbol{x}) & \frac{\partial c_2}{\partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial c_m}{\partial x_2}(\boldsymbol{x}) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\partial c_1}{\partial x_n}(\boldsymbol{x}) & \frac{\partial c_2}{\partial x_n}(\boldsymbol{x}) & \cdots & \frac{\partial c_m}{\partial x_n}(\boldsymbol{x})
\end{bmatrix}
\end{aligned}
$$

微积分基础

泰勒展开

泰勒展开

- 线性展开

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) + o(\|\boldsymbol{x} - \boldsymbol{x}_0\|).$$

微积分基础

泰勒展开

- 线性展开

$$f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) + o(\|\boldsymbol{x} - \boldsymbol{x}_0\|).$$

- 二阶展开

$$\begin{aligned} f(\boldsymbol{x}) = {} & f(\boldsymbol{x}_0) + \nabla f(\boldsymbol{x}_0)^T(\boldsymbol{x} - \boldsymbol{x}_0) \\ & + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T \nabla^2 f(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + o(\|\boldsymbol{x} - \boldsymbol{x}_0\|^2). \end{aligned}$$

微积分基础

方向导数

称

$$\frac{\partial f(\boldsymbol{x}_0)}{\partial p}=\lim_{t\to 0^+}\frac{f(\boldsymbol{x}_0+tp)-f(\boldsymbol{x}_0)}{t\|p\|}$$

为函数$f(\boldsymbol{x})$ 在$\boldsymbol{x}_0$ 处沿方向p 的方向导数.

微积分基础

定理1 (链式法则)

如果$g: \mathcal{D} \to R$ 在开集$\mathcal{D} \subset R^n$ 上是可微的，且$f: (a,\ b) \to \mathcal{D}$ 在$(a,\ b)$ 可微. 那么它们的复合函数$h:\ (a,\ b) \to R,\ h(t) = g(f(t))$ 在$(a,\ b)$ 上是可微的，且导数

$$h'(t) = \nabla g(f(t))^T \begin{bmatrix} f_1'(t) \\ \vdots \\ f_n'(t) \end{bmatrix}.$$

微积分基础

乘法法则

设$f: R^n \to R^m$和$g: R^n \to R^m$为两个可微函数，定义$h(\boldsymbol{x}) = f(\boldsymbol{x})^T g(\boldsymbol{x})$. 则$h(\boldsymbol{x})$可微，且

微积分基础

乘法法则

设$f: R^n \to R^m$和$g: R^n \to R^m$为两个可微函数，定义$h(\boldsymbol{x}) = f(\boldsymbol{x})^T g(\boldsymbol{x})$. 则$h(\boldsymbol{x})$可微，且

$$\nabla h(\boldsymbol{x}) = \nabla g(\boldsymbol{x}) f(\boldsymbol{x}) + \nabla f(\boldsymbol{x}) g(\boldsymbol{x}).$$

几个常用的关于x 的梯度公式

微积分基础

几个常用的关于$\boldsymbol{x}$ 的梯度公式

- 给定矩阵$A \in R^{m\times n}$ 和向量$\boldsymbol{y} \in R^m$, 则

$$\nabla(\boldsymbol{y}^T A\boldsymbol{x}) = A^T\boldsymbol{y}.$$

微积分基础

几个常用的关于$\boldsymbol{x}$ 的梯度公式

- 给定矩阵$A \in R^{m\times n}$ 和向量$\boldsymbol{y} \in R^m$, 则

$$\nabla(\boldsymbol{y}^T A\boldsymbol{x}) = A^T\boldsymbol{y}.$$

- 当$m = n$, 有

$$\nabla(\boldsymbol{x}^T A\boldsymbol{x}) = (A + A^T)\boldsymbol{x}.$$

微积分基础

几个常用的关于$\boldsymbol{x}$ 的梯度公式

- 给定矩阵$A \in R^{m\times n}$ 和向量$\boldsymbol{y} \in R^m$, 则

$$\nabla(\boldsymbol{y}^T A\boldsymbol{x}) = A^T\boldsymbol{y}.$$

- 当$m = n$, 有

$$\nabla(\boldsymbol{x}^T A\boldsymbol{x}) = (A + A^T)\boldsymbol{x}.$$

- 若$\boldsymbol{y} \in R^n$, 则$\nabla(\boldsymbol{y}^T\boldsymbol{x}) = \boldsymbol{y}$.

微积分基础

几个常用的关于$\boldsymbol{x}$ 的梯度公式

- 给定矩阵$A \in R^{m\times n}$ 和向量$\boldsymbol{y} \in R^m$, 则

$$\nabla(\boldsymbol{y}^T A\boldsymbol{x}) = A^T\boldsymbol{y}.$$

- 当$m = n$, 有

$$\nabla(\boldsymbol{x}^T A\boldsymbol{x}) = (A + A^T)\boldsymbol{x}.$$

- 若$\boldsymbol{y} \in R^n$, 则$\nabla(\boldsymbol{y}^T\boldsymbol{x}) = \boldsymbol{y}$.
- 若Q 是对称矩阵，则

$$\nabla(\boldsymbol{x}^T Q\boldsymbol{x}) = 2Q\boldsymbol{x}.$$

微积分基础

- $$\frac{\partial f(\boldsymbol{x}_0)}{\partial p} = \nabla f(\boldsymbol{x}_0)^T \frac{p}{\|p\|}.$$

微积分基础

- $\dfrac{\partial f(\boldsymbol{x}_0)}{\partial p} = \nabla f(\boldsymbol{x}_0)^T \dfrac{p}{\|p\|}.$
- $\nabla(A^T x) = A.$

微积分基础

- $\dfrac{\partial f(\boldsymbol{x}_0)}{\partial p} = \nabla f(\boldsymbol{x}_0)^T \dfrac{p}{\|p\|}$.
- $\nabla(A^T x) = A$.
- $\nabla \boldsymbol{x} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$.

微积分基础

- $\dfrac{\partial f(\boldsymbol{x}_0)}{\partial p} = \nabla f(\boldsymbol{x}_0)^T \dfrac{p}{\|p\|}$.
- $\nabla(A^T x) = A$.
- $\nabla \boldsymbol{x} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}$.
- 若$\varphi(t) = f(\boldsymbol{x} + t\boldsymbol{d})$, 则

$$\varphi'(t) = \nabla f(\boldsymbol{x} + t\boldsymbol{d})^T \boldsymbol{d}, \quad \varphi''(t) = \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x} + t\boldsymbol{d})\boldsymbol{d}.$$

线性代数基础

定义1 (内积)

对于$\boldsymbol{x}, \boldsymbol{y} \in R^n$，定义欧式内积为

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x}^T \boldsymbol{y}.$$

内积的性质

(1) 非负性: $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, $\boldsymbol{x} = 0 \Leftrightarrow \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$;

内积的性质

(1) 非负性: $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, $\boldsymbol{x} = 0 \Leftrightarrow \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$;

(2) 对称性: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$;

内积的性质

(1) 非负性: $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, $\boldsymbol{x} = 0 \Leftrightarrow \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$;

(2) 对称性: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$;

(3) 可加性: $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$;

内积的性质

(1) 非负性: $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, $\boldsymbol{x} = 0 \Leftrightarrow \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$;

(2) 对称性: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$;

(3) 可加性: $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$;

(4) 齐次性: 对于任意$r \in R$, 总有$\langle r\boldsymbol{x}, \boldsymbol{y} \rangle = r\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ 成立.

数学基础

定义2 (范数)

在n 维线性空间R^n 中，若函数$\|\boldsymbol{x}\|$ 满足：

数学基础

定义2 (范数)

在n 维线性空间R^n 中，若函数$\|\boldsymbol{x}\|$ 满足：

(i) 对任意$\boldsymbol{x} \in R^n$，有$\|\boldsymbol{x}\| \geq 0$，且$\boldsymbol{x} = 0 \Leftrightarrow \|\boldsymbol{x}\| = 0$;

数学基础

定义2 (范数)

在n 维线性空间R^n 中，若函数$\|\boldsymbol{x}\|$ 满足：

(i) 对任意$\boldsymbol{x} \in R^n$，有$\|\boldsymbol{x}\| \geq 0$，且$\boldsymbol{x} = 0 \Leftrightarrow \|\boldsymbol{x}\| = 0$;

(ii) 对任意$\boldsymbol{x} \in R^n$ 及$\alpha \in R$，有$\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$;

数学基础

定义2 (范数)

在n 维线性空间R^n 中，若函数$\|\boldsymbol{x}\|$ 满足：

(i) 对任意$\boldsymbol{x} \in R^n$，有$\|\boldsymbol{x}\| \geq 0$，且$\boldsymbol{x} = 0 \Leftrightarrow \|\boldsymbol{x}\| = 0$;

(ii) 对任意$\boldsymbol{x} \in R^n$ 及$\alpha \in R$，有$\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$;

(iii) 对任意$\boldsymbol{x}, \boldsymbol{y} \in R^n$，有$\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$;

数学基础

定义2 (范数)

在n 维线性空间R^n 中，若函数$\|\boldsymbol{x}\|$ 满足：

(i) 对任意$\boldsymbol{x} \in R^n$，有$\|\boldsymbol{x}\| \geq 0$，且$\boldsymbol{x} = 0 \Leftrightarrow \|\boldsymbol{x}\| = 0$;

(ii) 对任意$\boldsymbol{x} \in R^n$ 及$\alpha \in R$，有$\|\alpha\boldsymbol{x}\| = |\alpha|\|\boldsymbol{x}\|$;

(iii) 对任意$\boldsymbol{x}, \boldsymbol{y} \in R^n$，有$\|\boldsymbol{x}+\boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$;

则称$\|\boldsymbol{x}\|$ 为R^n 上的向量范数.

数学基础

定义3 (p-范数)

对$1 \le p < +\infty$，称

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

为向量$\boldsymbol{x}$ 的p-范数.

数学基础

常用的范数

数学基础

常用的范数

- 2-范数(ℓ_2 范数): $\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$;

数学基础

常用的范数

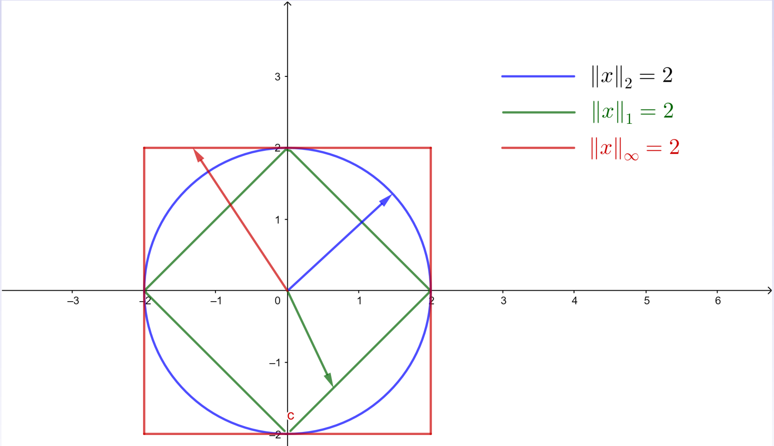
- 2-范数(ℓ_2 范数): $\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$;
- 1-范数(ℓ_1 范数): $\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$;

数学基础

常用的范数

- 2-范数(ℓ_2 范数): $\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$;
- 1-范数(ℓ_1 范数): $\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$;
- ∞-范数(ℓ_∞ 范数): $\|\boldsymbol{x}\|_\infty = \max_{1\le i\le n} |x_i|$.

常用的范数



三种常见范数的几何意义

极小点

定义4 (极小点)

设$f(\boldsymbol{x})$ 为一个n 元实值函数，定义域为$\Omega \subset R^n$. 对于定义域中的一个点$\boldsymbol{x}^*$,

极小点

定义4 (极小点)

设$f(\boldsymbol{x})$ 为一个n 元实值函数，定义域为$\Omega \subset R^n$. 对于定义域中的一个点$\boldsymbol{x}^*$,

(1) 若存在$\epsilon > 0$, 使得

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*),\ \forall \boldsymbol{x} \in B(\boldsymbol{x}^*, \epsilon) \cap \Omega,$$

其中，$B(\boldsymbol{x}^*, \epsilon) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon\}$, 则称$\boldsymbol{x}^*$ 为f 在Ω 中的(局部)极小点.

极小点

定义4 (极小点)

设$f(\boldsymbol{x})$ 为一个n 元实值函数，定义域为$\Omega \subset R^n$. 对于定义域中的一个点$\boldsymbol{x}^*$,

(1) 若存在$\epsilon > 0$, 使得

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*),\ \forall \boldsymbol{x} \in B(\boldsymbol{x}^*, \epsilon) \cap \Omega,$$

其中，$B(\boldsymbol{x}^*, \epsilon) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon\}$, 则称$\boldsymbol{x}^*$ 为f 在Ω 中的(局部)极小点.

(2) 若

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*),\ \forall \boldsymbol{x} \in \Omega,$$

则称$\boldsymbol{x}^*$ 为f 在Ω 中的全局极小点（或最小点）.

极小点

定义5 (严格极小点)

设$f(\boldsymbol{x})$ 为一个n 元实值函数，定义域为$\Omega \subset R^n$. 对于定义域中的一个点$\boldsymbol{x}^*$,

极小点

定义5 (严格极小点)

设$f(\boldsymbol{x})$ 为一个n 元实值函数，定义域为$\Omega \subset R^n$. 对于定义域中的一个点$\boldsymbol{x}^*$,

(1) 若存在$\epsilon > 0$, 使得

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*), \ \forall \boldsymbol{x} \in B(\boldsymbol{x}^*, \epsilon) \cap (\Omega \setminus \{\boldsymbol{x}^*\}),$$

其中，$B(\boldsymbol{x}^*, \epsilon) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon\}$, 则称$\boldsymbol{x}^*$ 为f 在Ω 中的严格(局部)极小点.

极小点

定义5 (严格极小点)

设$f(\boldsymbol{x})$ 为一个n 元实值函数，定义域为$\Omega \subset R^n$. 对于定义域中的一个点$\boldsymbol{x}^*$,

(1) 若存在$\epsilon > 0$, 使得

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*),\ \forall \boldsymbol{x} \in B(\boldsymbol{x}^*, \epsilon) \cap (\Omega \setminus \{\boldsymbol{x}^*\}),$$

其中，$B(\boldsymbol{x}^*, \epsilon) = \{\boldsymbol{x} \mid \|\boldsymbol{x} - \boldsymbol{x}^*\| < \epsilon\}$, 则称$\boldsymbol{x}^*$ 为f 在Ω 中的严格(局部)极小点.

(2) 若

$$f(\boldsymbol{x}) > f(\boldsymbol{x}^*),\ \forall \boldsymbol{x} \in \Omega \setminus \{\boldsymbol{x}^*\},$$

则称$\boldsymbol{x}^*$ 为f 在Ω 中的严格全局极小点（或最小点）.

数学基础

如图

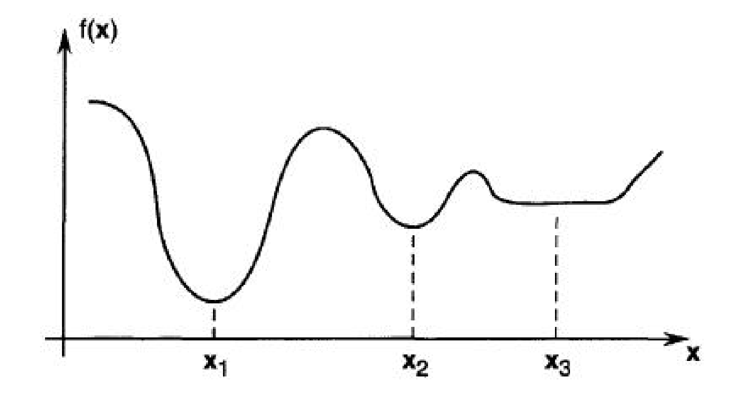


图: 极小点

数值最优化方法的基本思想

通过求解一系列简单问题来求解一个复杂问题.

数值最优化方法的基本思想

通过求解一系列简单问题来求解一个复杂问题.

小和尚下山

- 用线性函数代替非线性函数;

- 用线性函数代替非线性函数;
- 用二次函数来代替非线性非二次的函数;

- 用线性函数代替非线性函数;
- 用二次函数来代替非线性非二次的函数;
- 用没有约束的问题来代替有约束的问题;

- 用线性函数代替非线性函数;
- 用二次函数来代替非线性非二次的函数;
- 用没有约束的问题来代替有约束的问题;
- …….

最优化问题的一般解法

最优化问题的一般迭代格式(线性搜索)

【步 0】 给定初始点$\boldsymbol{x}_0$, 令$k=0$;

最优化问题的一般解法

最优化问题的一般迭代格式(线性搜索)

【步 0】 给定初始点$\boldsymbol{x}_0$, 令$k=0$;

【步 1】 确定点$\boldsymbol{x}_k$ 处的搜索方向d_k;

最优化问题的一般解法

最优化问题的一般迭代格式(线性搜索)

【步 0】 给定初始点$\boldsymbol{x}_0$, 令$k=0$;

【步 1】 确定点$\boldsymbol{x}_k$ 处的搜索方向d_k;

【步 2】 确定步长α_k;

最优化问题的一般解法

最优化问题的一般迭代格式(线性搜索)

【步 0】 给定初始点$\boldsymbol{x}_0$, 令$k = 0$;

【步 1】 确定点$\boldsymbol{x}_k$ 处的搜索方向d_k;

【步 2】 确定步长α_k;

【步 3】 令$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k$;

最优化问题的一般解法

最优化问题的一般迭代格式(线性搜索)

【步 0】 给定初始点$\boldsymbol{x}_0$, 令$k=0$;

【步 1】 确定点$\boldsymbol{x}_k$ 处的搜索方向d_k;

【步 2】 确定步长α_k;

【步 3】 令$\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\alpha_k d_k$;

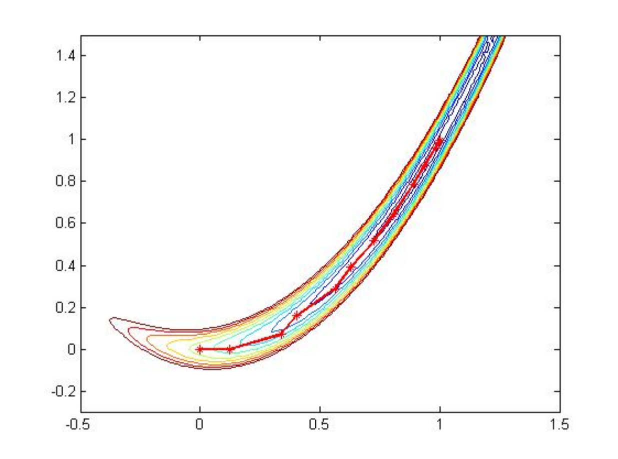
【步 4】 若$\boldsymbol{x}_{k+1}$ 满足某种终止准则, 则停止迭代, 以$\boldsymbol{x}_{k+1}$ 为近似最优解. 否则令$k=k+1$, 转步1.

一个求解的例子

用Newton法求Rosenbrock 函数

$$f(x, y) = 100(y - x^2)^2 + (1 - x)^2$$

的最小值.



基本概念

定义6 (下降方向1)

对于向量$0 \neq d \in R^n$, 和某个点$\boldsymbol{x}$ 处. 如果存在一个实数$\alpha_0 > 0$, 使得对于所有$\alpha \in [0, \alpha_0]$, 有

$$f(\boldsymbol{x} + \alpha d) < f(\boldsymbol{x}),$$

则称d 为$\boldsymbol{x}$ 处的一个下降方向.

基本概念

定义6 (下降方向1)

对于向量$0 \neq d \in R^n$, 和某个点$\boldsymbol{x}$ 处. 如果存在一个实数$\alpha_0 > 0$, 使得对于所有$\alpha \in [0, \alpha_0]$, 有

$$f(\boldsymbol{x} + \alpha d) < f(\boldsymbol{x}),$$

则称d 为$\boldsymbol{x}$ 处的一个下降方向.

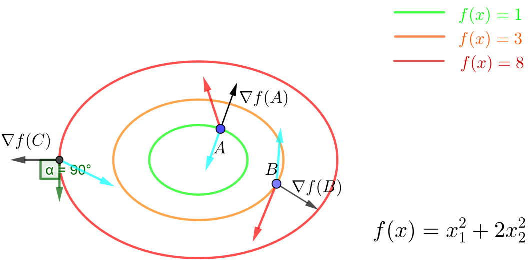
定义7 (下降方向2(常用))

对于向量$0 \neq d \in R^n$, 和某个点$\boldsymbol{x}$ 处. 若d 满足

$$\nabla f(\boldsymbol{x})^T d < 0,$$

则称d 为$\boldsymbol{x}$ 处的一个下降方向.

下降方向



黑色箭头为梯度方向；蓝色箭头均为下降方向；
红色箭头为上升方向；
绿色箭头在定义6的意义下是上升方向，
在定义7的意义下既非上升方向也非下降方向。

基本概念

定义8 (可行方向)

对于向量$0 \neq d \in R^n$, 和可行域Ω 中的某个点$\boldsymbol{x} \in \Omega$. 如果存在一个实数$\alpha_0 > 0$, 使得对于所有$\alpha \in [0, \alpha_0]$, $\boldsymbol{x} + \alpha d$ 仍然属于Ω, 即$\boldsymbol{x} + \alpha d \in \Omega$, 则称$d$ 为$\boldsymbol{x}$ 处的一个可行方向.

基本概念

如图

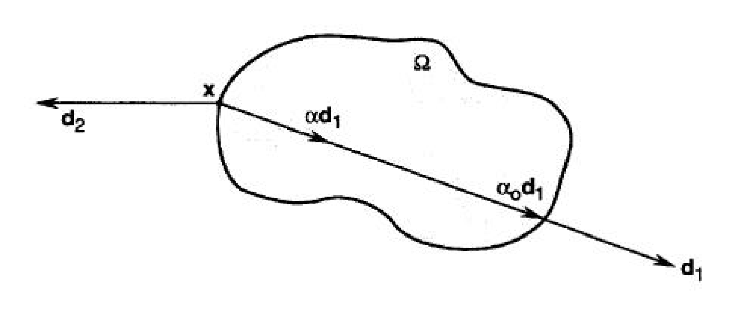


图: 可行方向

基本概念

一维搜索

由$\boldsymbol{x}_k$ 出发沿方向d_k 求步长α_k 的过程叫做一维搜索或线性搜索.

基本概念

一维搜索

由$\boldsymbol{x}_k$ 出发沿方向d_k 求步长α_k 的过程叫做一维搜索或线性搜索.

算法收敛

如果某算法构造出的点列$\{\boldsymbol{x}_k\}$ 能够在有限步之内得到最优化问题的最优解$\boldsymbol{x}^*$, 或者点列$\{\boldsymbol{x}_k\}$ 有聚点, 且聚点是最优解$\boldsymbol{x}^*$, 则称这种算法是收敛的.

基本概念

定义9 (线性、超线性收敛)

设序列$\{\boldsymbol{x}_k\}$ 收敛于$\boldsymbol{x}^*$，而且

$$\lim_{k\to\infty}\frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|}=\beta,$$

基本概念

定义9 (线性、超线性收敛)

设序列$\{\boldsymbol{x}_k\}$ 收敛于$\boldsymbol{x}^*$, 而且

$$\lim_{k\to\infty}\frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|}=\beta,$$

- 若$0<\beta<1$, 则称序列$\{\boldsymbol{x}_k\}$ 为**线性收敛的**; 称β 为**收敛比**;

基本概念

定义9 (线性、超线性收敛)

设序列$\{\boldsymbol{x}_k\}$ 收敛于$\boldsymbol{x}^*$, 而且

$$\lim_{k\to\infty}\frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|}=\beta,$$

- 若$0<\beta<1$, 则称序列$\{\boldsymbol{x}_k\}$ 为**线性收敛的**; 称β 为**收敛比**;
- 若$\beta=0$, 则称序列$\{\boldsymbol{x}_k\}$ 为**超线性收敛的**;

基本概念

定义10 (收敛阶数)

设序列$\{\boldsymbol{x}_k\}$ 收敛于$\boldsymbol{x}^*$, 若对于某个实数$p \geq 1$, 有

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^*\|}{\|\boldsymbol{x}_k-\boldsymbol{x}^*\|^p} = \beta,\ 0<\beta<\infty$$

则称序列$\{\boldsymbol{x}_k\}$ 为p 阶收敛的.

凸组合

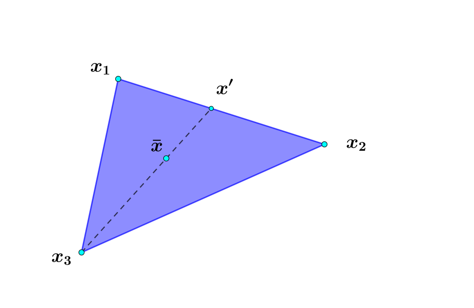
定义11

已知$X \subset R^n$, $\bar{\boldsymbol{x}} \in X$. 如果存在$k$ 个点$\boldsymbol{x}_i \in X$ 和常数$\alpha_i \geq 0$, $i = 1, 2, \cdots, k$ 且$\sum_{i=1}^{n} \alpha_i = 1$, 使得

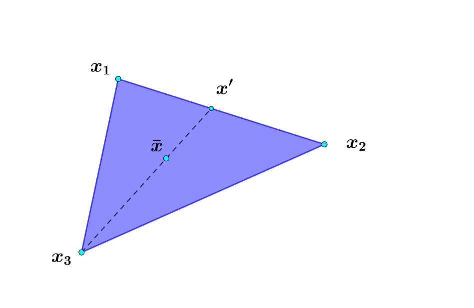
$$\bar{\boldsymbol{x}} = \sum_{i=1}^{k} \alpha_i \boldsymbol{x}_i.$$

则称$\bar{\boldsymbol{x}}$ 为$\boldsymbol{x}_i \in X$, $i = 1, 2, \cdots, k$ 的一个凸组合.

凸组合的几何意义



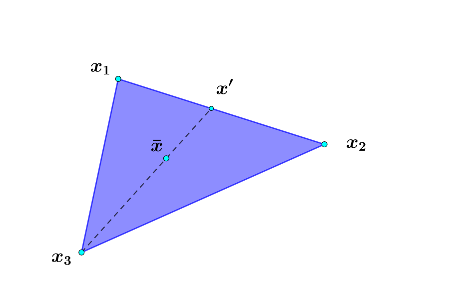
凸组合的几何意义



上图中，存在$\lambda, \eta \in [0,1]$ 使得

$$\boldsymbol{x}' = \lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2,$$
$$\bar{\boldsymbol{x}} = \eta \boldsymbol{x}_3 + (1-\eta)\boldsymbol{x}'.$$

凸组合的几何意义



上图中，存在$\lambda, \eta \in [0,1]$ 使得

$$\begin{aligned} x' &= \lambda x_1 + (1-\lambda)x_2, \\ \bar{x} &= \eta x_3 + (1-\eta)x'. \end{aligned}$$

则

$$\bar{x} = (\lambda - \lambda\eta)x_1 + (1 - \lambda - \eta + \lambda\eta)x_2 + \eta x_3.$$

凸集

定义12

凸集 设$X \subset R^n$. 若集合X 中的任意两点的凸组合仍然属于X, 则称X 为凸集.

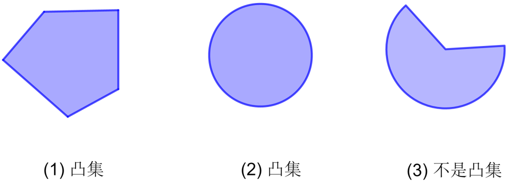
凸集

定义12

凸集 设$X \subset R^n$. 若集合X 中的任意两点的凸组合仍然属于X, 则称X 为凸集. 即

$$X \text{ 是凸集} \Leftrightarrow \begin{aligned} &\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in X, \\ &\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in X, \lambda \in [0,1]. \end{aligned}$$

如图所示



凸函数

定义13

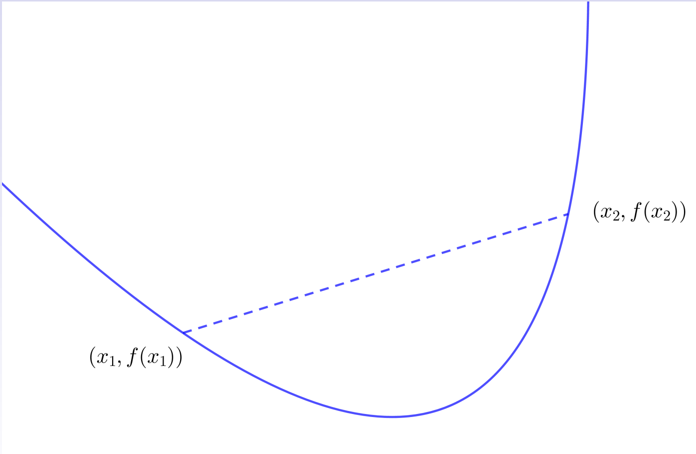
凸函数 设$X \subset R^n$ 是一个凸集. 若对任意的$x_1, x_2 \in X$, 有

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2), \ \forall \lambda \in [0,1],$$

则称f 是定义在X 上的凸函数.

凸函数

如图所示



凸集和凸函数的例子

- 凸集
  - 单位球: $\{\boldsymbol{x} \in R^n | \|\boldsymbol{x}\| \leq 1\}$;
  - 多面体: $\{\boldsymbol{x} \in R^n | A\boldsymbol{x} = b, C\boldsymbol{x} \leq d\}$, 其中$A$, C 为矩阵, b, d 是向量.
- 凸函数
  - 线性函数: $f(\boldsymbol{x}) = c^T\boldsymbol{x} + \alpha$, 其中$c \in R^n$, $\alpha \in R$;
  - 凸二次函数: $f(\boldsymbol{x}) = \boldsymbol{x}^T H\boldsymbol{x}$, 其中$H$ 是对称半正定矩阵.

凸集和凸函数的两个性质

定理2

集合$X \in R^n$ 是凸集的充要条件是X 中任意m 个元素的凸组合仍然属于X，即对任意$\boldsymbol{x}_i \in X$，$i = 1, 2, \cdots, m$，有

$$\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i \in X,$$

其中，$\lambda_i \geq 0$ 且$\sum_{i=1}^{m} \lambda_i = 1$.

凸集和凸函数的两个性质

定理2

集合$X \in R^n$ 是凸集的充要条件是X 中任意m 个元素的凸组合仍然属于X, 即对任意$\boldsymbol{x}_i \in X$, $i=1,2,\cdots,m$, 有

$$\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i \in X,$$

其中, $\lambda_i \geq 0$ 且$\sum_{i=1}^{m} \lambda_i = 1$.

这个性质可以看作是凸集定义的推广.

凸集和凸函数的两个性质

定理3 (Jensen 不等式)

设f 是定义在凸集$X \in R^n$ 上的凸函数. 则对任意$\boldsymbol{x}_i \in X$, $\lambda_i \geq 0,\ i = 1, 2, \cdots, m$ 且$\sum_{i=1}^{m} \lambda_i = 1$, 有

$$f(\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i) \leq \sum_{i=1}^{m} \lambda_i f(\boldsymbol{x}_i).$$

凸集和凸函数的两个性质

定理3 (Jensen 不等式)

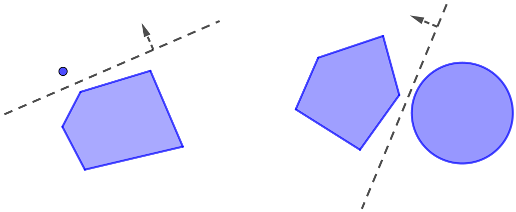
设f 是定义在凸集$X \in R^n$ 上的凸函数. 则对任意$\boldsymbol{x}_i \in X$, $\lambda_i \geq 0$, $i = 1, 2, \cdots, m$ 且$\sum_{i=1}^{m} \lambda_i = 1$, 有

$$f(\sum_{i=1}^{m} \lambda_i \boldsymbol{x}_i) \leq \sum_{i=1}^{m} \lambda_i f(\boldsymbol{x}_i).$$

这个定理可以看作是凸函数定义的推广.

凸集分离定理

几何直观



(1) 点和凸集分离　　(2) 凸集和凸集分离

凸集分离的定义

定义14

设$D_1, D_2 \subset R^n$ 为两个非空凸集，若存在非零向量$a \in R^n$ 和实数β 使得

$$D_1 \subset H^+ = \{\boldsymbol{x} \in R^n \mid a^T\boldsymbol{x} \geq \beta\}$$
$$D_2 \subset H^- = \{\boldsymbol{x} \in R^n \mid a^T\boldsymbol{x} \leq \beta\}$$

则称超平面

$$H = \{\boldsymbol{x} \in R^n \mid a^T\boldsymbol{x} = \beta\}$$

分离了D_1, D_2. 如果更有

$$D_1 \subset H_0^+ = \{\boldsymbol{x} \in R^n \mid a^T\boldsymbol{x} > \beta\}$$
$$D_2 \subset H_0^- = \{\boldsymbol{x} \in R^n \mid a^T\boldsymbol{x} < \beta\}$$

则称超平面H 严格分离了D_1, D_2.

点到凸集的投影

设$D \subset R^n$ 是非空闭凸集, $\boldsymbol{y} \in R^n$ 但是$\boldsymbol{y} \notin D$, 则

(1) 存在唯一的点$\bar{\boldsymbol{x}} \in D$, 使得集合$D$ 到点$\boldsymbol{y}$ 的距离最小, 即

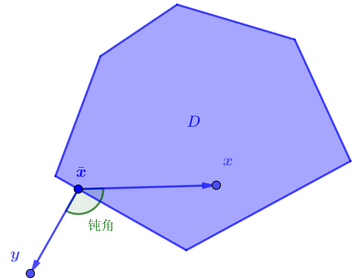
$$\|\bar{\boldsymbol{x}} - \boldsymbol{y}\| = \inf\{\|\boldsymbol{x} - \boldsymbol{y}\|, \boldsymbol{x} \in D\}.$$

(2) $\bar{\boldsymbol{x}} \in D$ 是$\boldsymbol{y}$ 到集合D 的最短距离点的充分必要条件是

$$(\boldsymbol{x} - \bar{\boldsymbol{x}})^T(\boldsymbol{y} - \bar{\boldsymbol{x}}) \leq 0, \ \forall \boldsymbol{x} \in D.$$

点到凸集的投影的几何意义

如图所示:



证明思路

函数$\varphi(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ 是一个连续函数. 按照数学分析的知识, 连续函数在有界闭集上一定能取到最小值.

证明思路

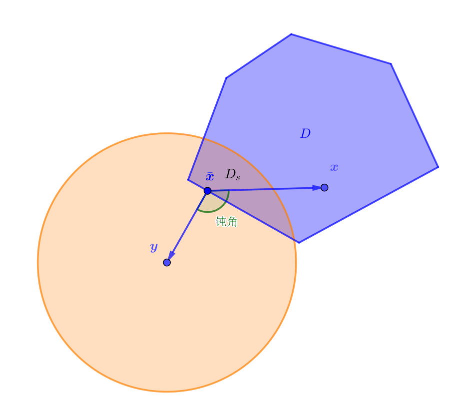
函数$\varphi(\boldsymbol{x}) = \|\boldsymbol{x}-\boldsymbol{y}\|$ 是一个连续函数. 按照数学分析的知识, 连续函数在有界闭集上一定能取到最小值.
但是凸集不一定是有界集.

证明思路

函数$\varphi(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ 是一个连续函数. 按照数学分析的知识, 连续函数在有界闭集上一定能取到最小值.

但是凸集不一定是有界集.

其实我们不一定要在整个凸集D 上考虑$\varphi(\boldsymbol{x})$ 的最小值, 如下图



证明： (1) 令

$$S = \{x \in R^n \,|\, \|x\| \le 1\}.$$

证明： (1) 令

$$S = \{x \in R^n \,|\, \|x\| \leq 1\}.$$

取充分大的$\mu > 0$ 使得

$$D_S = D \cap (y + \mu S) \neq \emptyset.$$

证明： (1) 令

$$S = \{x \in R^n \,|\, \|x\| \leq 1\}.$$

取充分大的$\mu > 0$ 使得

$$D_S = D \cap (y + \mu S) \neq \emptyset.$$

因此连续函数$\varphi(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ 在D_S 上必定可以取到最小点$\bar{\boldsymbol{x}}$.

证明： (1) 令

$$S = \{x \in R^n \,|\, \|x\| \leq 1\}.$$

取充分大的$\mu > 0$ 使得

$$D_S = D \cap (y + \mu S) \neq \emptyset.$$

因此连续函数$\varphi(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ 在D_S 上必定可以取到最小点$\bar{\boldsymbol{x}}$.

存在性证毕.

唯一性的证明为常用技巧: 假设存在两个不相等的投影, 证明矛盾即可.

唯一性的证明为常用技巧：假设存在两个不相等的投影，证明矛盾即可.

假设存在$\tilde{\boldsymbol{x}} \in D$, $\tilde{\boldsymbol{x}} \neq \bar{\boldsymbol{x}}$, 使得

$$\|\boldsymbol{y}-\bar{\boldsymbol{x}}\| = \|\boldsymbol{y}-\tilde{\boldsymbol{x}}\| = \gamma.$$

唯一性的证明为常用技巧: 假设存在两个不相等的投影, 证明矛盾即可.

假设存在$\tilde{\boldsymbol{x}} \in D$, $\tilde{\boldsymbol{x}} \neq \bar{\boldsymbol{x}}$, 使得

$$\|\boldsymbol{y}-\bar{\boldsymbol{x}}\|=\|\boldsymbol{y}-\tilde{\boldsymbol{x}}\|=\gamma.$$

则由范数的三角不等式由

$$\begin{aligned}&\left\|\boldsymbol{y}-\frac{\bar{\boldsymbol{x}}+\tilde{\boldsymbol{x}}}{2}\right\|=\left\|\frac{1}{2}(\boldsymbol{y}-\bar{\boldsymbol{x}})+\frac{1}{2}(\boldsymbol{y}-\tilde{\boldsymbol{x}})\right\|\\ \leq&\frac{1}{2}\|\boldsymbol{y}-\bar{\boldsymbol{x}}\|+\frac{1}{2}\|\boldsymbol{y}-\tilde{\boldsymbol{x}}\|=\gamma.\end{aligned}$$

由于D 是凸集，故有$\frac{\bar{\boldsymbol{x}}+\tilde{\boldsymbol{x}}}{2} \in D$，又因为$\gamma$ 是集合D 到$\boldsymbol{y}$ 的最短距离，所以上式的等于号成立.

由于D 是凸集, 故有$\frac{\bar{\boldsymbol{x}}+\tilde{\boldsymbol{x}}}{2} \in D$, 又因为$\gamma$ 是集合D 到$\boldsymbol{y}$ 的最短距离, 所以上式的等于号成立.

因此存在实数α 使得

$$\boldsymbol{y}-\bar{\boldsymbol{x}}=\alpha(\boldsymbol{y}-\tilde{\boldsymbol{x}}).$$

由于D 是凸集, 故有$\frac{\bar{\boldsymbol{x}}+\tilde{\boldsymbol{x}}}{2} \in D$, 又因为$\gamma$ 是集合D 到$\boldsymbol{y}$ 的最短距离, 所以上式的等于号成立.
因此存在实数α 使得

$$\boldsymbol{y}-\bar{\boldsymbol{x}}=\alpha(\boldsymbol{y}-\tilde{\boldsymbol{x}}).$$

又因为$\|\boldsymbol{y}-\bar{\boldsymbol{x}}\|=\|\boldsymbol{y}-\tilde{\boldsymbol{x}}\|$, 故而$|\alpha|=1$.

由于D 是凸集, 故有$\frac{\bar{\boldsymbol{x}}+\tilde{\boldsymbol{x}}}{2} \in D$, 又因为$\gamma$ 是集合D 到$\boldsymbol{y}$ 的最短距离, 所以上式的等于号成立.
因此存在实数α 使得

$$\boldsymbol{y}-\bar{\boldsymbol{x}}=\alpha(\boldsymbol{y}-\tilde{\boldsymbol{x}}).$$

又因为$\|\boldsymbol{y}-\bar{\boldsymbol{x}}\|=\|\boldsymbol{y}-\tilde{\boldsymbol{x}}\|$, 故而$|\alpha|=1$.
若$\alpha=1$, 则$\bar{\boldsymbol{x}}=\tilde{\boldsymbol{x}}$.

由于D 是凸集，故有$\frac{\bar{\boldsymbol{x}}+\tilde{\boldsymbol{x}}}{2} \in D$，又因为$\gamma$ 是集合D 到$\boldsymbol{y}$ 的最短距离，所以上式的等于号成立.
因此存在实数α 使得

$$\boldsymbol{y}-\bar{\boldsymbol{x}}=\alpha(\boldsymbol{y}-\tilde{\boldsymbol{x}}).$$

又因为$\|\boldsymbol{y}-\bar{\boldsymbol{x}}\|=\|\boldsymbol{y}-\tilde{\boldsymbol{x}}\|$，故而$|\alpha|=1$.
若$\alpha=1$，则$\bar{\boldsymbol{x}}=\tilde{\boldsymbol{x}}$.
若$\alpha=-1$，则$\boldsymbol{y}=\frac{\bar{\boldsymbol{x}}+\tilde{\boldsymbol{x}}}{2} \in D$，这与$\boldsymbol{y} \notin D$ 矛盾.

(2) 充分性较为容易，留作作业.下证必要性.

(2) 充分性较为容易，留作作业.下证必要性.

设$\bar{\boldsymbol{x}}$ 是D 到$\boldsymbol{y}$ 距离最短点, 即

$$\|\bar{\boldsymbol{x}}-\boldsymbol{y}\| \leq \|\boldsymbol{x}-\boldsymbol{y}\|,\ \forall x \in D.$$

(2) 充分性较为容易，留作作业.下证必要性.

设$\bar{\boldsymbol{x}}$ 是D 到$\boldsymbol{y}$ 距离最短点, 即

$$\|\bar{\boldsymbol{x}}-\boldsymbol{y}\| \leq \|\boldsymbol{x}-\boldsymbol{y}\|, \ \forall x \in D.$$

令$\lambda \in (0,1)$, 有

$$\bar{\boldsymbol{x}}+\lambda(\boldsymbol{x}-\bar{\boldsymbol{x}}) = \lambda\boldsymbol{x}+(1-\lambda)\bar{\boldsymbol{x}} \in D.$$

(2) 充分性较为容易，留作作业.下证必要性.

设$\bar{\boldsymbol{x}}$ 是D 到$\boldsymbol{y}$ 距离最短点, 即

$$\|\bar{\boldsymbol{x}}-\boldsymbol{y}\| \leq \|\boldsymbol{x}-\boldsymbol{y}\|, \ \forall x \in D.$$

令$\lambda \in (0,1)$, 有

$$\bar{\boldsymbol{x}}+\lambda(\boldsymbol{x}-\bar{\boldsymbol{x}}) = \lambda\boldsymbol{x}+(1-\lambda)\bar{\boldsymbol{x}} \in D.$$

于是有

$$\begin{aligned}\|\bar{\boldsymbol{x}}-\boldsymbol{y}\|^2 \leq & \|\boldsymbol{y}-\bar{\boldsymbol{x}}-\lambda(\boldsymbol{x}-\bar{\boldsymbol{x}})\|^2 \\ = & \|\bar{\boldsymbol{x}}-\boldsymbol{y}\|^2+\lambda^2\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2-2\lambda(\boldsymbol{y}-\bar{\boldsymbol{x}})^T(\boldsymbol{x}-\bar{\boldsymbol{x}}).\end{aligned}$$

故而，对任意$\lambda \in (0,1)$，均有

$$\lambda^2\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2 - 2\lambda(\boldsymbol{y}-\bar{\boldsymbol{x}})^T(\boldsymbol{x}-\bar{\boldsymbol{x}}) \geq 0.$$

故而，对任意$\lambda \in (0,1)$, 均有

$$\lambda^2\|\boldsymbol{x}-\bar{\boldsymbol{x}}\|^2-2\lambda(\boldsymbol{y}-\bar{\boldsymbol{x}})^T(\boldsymbol{x}-\bar{\boldsymbol{x}})\geq 0.$$

令$\lambda \to 0$, 便有

$$(\boldsymbol{y}-\bar{\boldsymbol{x}})^T(\boldsymbol{x}-\bar{\boldsymbol{x}})\leq 0.$$

凸规划

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } x \in D, \end{aligned}$$

其中$f(x)$ 是凸函数, D 是凸集.

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t. } g_i(x) \le 0,\ i = 1, 2, \cdots, l; \\ &\qquad h_j(x) = 0,\ j = 1, 2, \cdots, k. \end{aligned}$$

其中, $f(x)$, $g_i(x)$ 是凸函数, $h_j(x)$ 是线性函数.

水平集

设$f(x)$ 为凸函数, $X \subset R^n$ 是一个凸集, 则称

$$D_\alpha = \{x \mid f(x) \leq \alpha, x \in X\}$$

为f 的水平集.

性质1

水平集一定是凸集.

凸函数的性质

定理4

凸函数的局部极小点就是全局极小点.

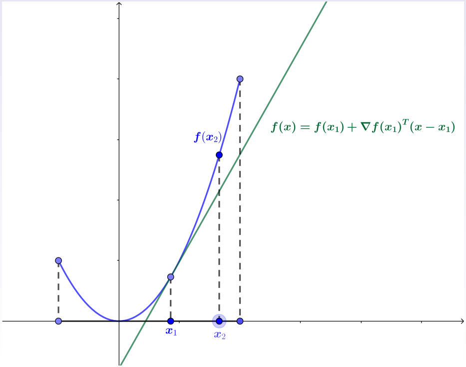
凸函数的判断条件

定理5

$f(\boldsymbol{x})$ 是凸集X 上的可微函数，$f(\boldsymbol{x})$是凸函数的充要条件是$\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in X$，有

$$f(\boldsymbol{x}_2) \geq f(\boldsymbol{x}_1) + \nabla f(\boldsymbol{x}_1)^T(\boldsymbol{x}_2 - \boldsymbol{x}_1).$$

如下图所示



凸函数的判断条件

定理6

设$f(\boldsymbol{x})$ 在凸集X 上二阶连续偏导数，则$f(\boldsymbol{x})$ 是凸函数的充要条件是$\forall \boldsymbol{x} \in X$，有$\nabla^2 f(\boldsymbol{x})$ 半正定.

例3.1

正定二次函数

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T A\boldsymbol{x} + b^T\boldsymbol{x} + c$$

是凸函数，其中A 是正定矩阵

- 一维最优化问题

$$\min_{x \in R} f(x) \text{ 或 } \min_{a \le x \le b} f(x)$$

- 一维最优化问题

$$\min_{x\in R} f(x) \text{ 或 } \min_{a\leq x\leq b} f(x)$$

- 一维搜索用于求解一维最优化问题.

- 一维最优化问题

$$\min_{x\in R} f(x) \text{ 或 } \min_{a\le x\le b} f(x)$$

- 一维搜索用于求解一维最优化问题.
  - 常被用于**求最优步长**.

- 一维最优化问题

$$\min_{x \in R} f(x) \text{ 或 } \min_{a \le x \le b} f(x)$$

- 一维搜索用于求解一维最优化问题.
  - 常被用于**求最优步长**.
    已知当前点$\boldsymbol{x}_k$, 搜索方向d_k, 求解

$$\min_{\alpha > 0} \varphi(\alpha) = f(\boldsymbol{x}_k + \alpha d_k).$$

例题：手动求最优步长

例4.1

设$f(\boldsymbol{x}) = x_1^2 + 2x_2^2$，$\boldsymbol{x}_k = (1,2)^T$，$d_k = (-2,-8)^T$. 求最优步长$\alpha_k$.

例题：手动求最优步长

例4.1

设 $f(\boldsymbol{x}) = x_1^2 + 2x_2^2$, $\boldsymbol{x}_k = (1,2)^T$, $d_k = (-2,-8)^T$. 求最优步长α_k.

【解：】计算

$$\boldsymbol{x}_k + \alpha d_k = \begin{pmatrix} 1-2\alpha \\ 2-8\alpha \end{pmatrix}.$$

例题：手动求最优步长

例4.1

设$f(\boldsymbol{x}) = x_1^2 + 2x_2^2$，$\boldsymbol{x}_k = (1, 2)^T$，$d_k = (-2, -8)^T$．求最优步长$\alpha_k$．

【解：】计算

$$\boldsymbol{x}_k + \alpha d_k = \begin{pmatrix} 1 - 2\alpha \\ 2 - 8\alpha \end{pmatrix}.$$

则

$$\begin{aligned} \varphi(\alpha) &= f(\boldsymbol{x}_k + \alpha d_k) \\ &= (1 - 2\alpha)^2 + 2(2 - 8\alpha)^2 \\ &= 132\alpha^2 - 68\alpha + 9. \end{aligned}$$

例题：手动求最优步长

$$\varphi'(\alpha) = 264\alpha - 68.$$

例题：手动求最优步长

$$\varphi'(\alpha) = 264\alpha - 68.$$

令 $\varphi'(\alpha) = 0$ 得最优步长

$$\alpha_k = \frac{17}{66}.$$

精确搜索

对精确一维搜索问题

$$\min_{\alpha>0} \varphi(\alpha) = f(\boldsymbol{x}_k + \alpha d_k),$$

精确搜索

对精确一维搜索问题

$$\min_{\alpha>0} \varphi(\alpha) = f(\boldsymbol{x}_k + \alpha d_k),$$

精确搜索

对精确一维搜索问题

$$\min_{\alpha>0} \varphi(\alpha) = f(\boldsymbol{x}_k + \alpha d_k),$$

则有

$$\varphi'(\alpha_k) = 0, \text{即 } \nabla f(\boldsymbol{x}_k + \alpha_k d_k)^T d_k = 0.$$

精确搜索

对精确一维搜索问题

$$\min_{\alpha>0} \varphi(\alpha) = f(\boldsymbol{x}_k + \alpha d_k),$$

则有

$$\varphi'(\alpha_k) = 0, \text{即 } \nabla f(\boldsymbol{x}_k + \alpha_k d_k)^T d_k = 0.$$

即：当α_k 为精确最优步长时

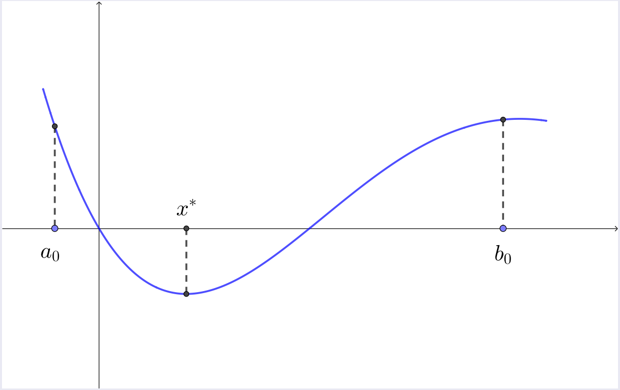
有用的性质： $g_{k+1}^T d_k = 0$, 其中 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k$.

（下）单峰函数

称目标函数在给定区间$[a,b]$ 上是（下）单峰的，如果它在该区间内有唯一的局部极小点. 如图

（下）单峰函数

称目标函数在给定区间$[a,b]$ 上是（下）单峰的，如果它在该区间内有唯一的局部极小点. 如图



性质

通过计算$[a,b]$内两个不同点的函数值，就可以确定一个包含极小点的子区间.

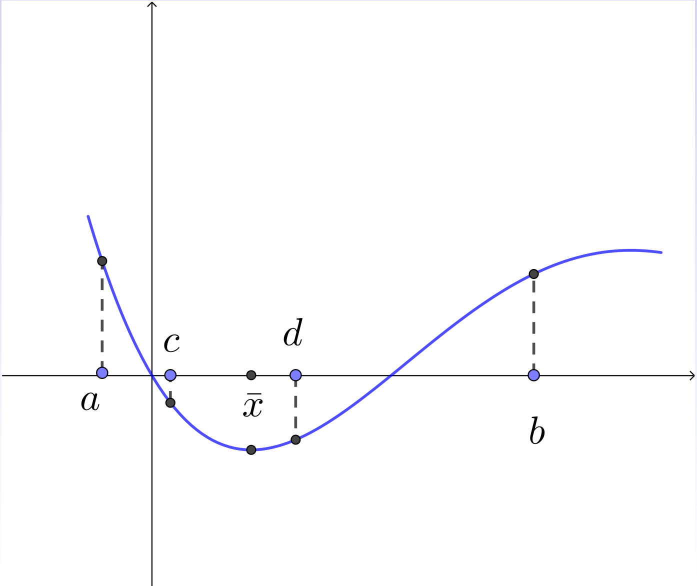
定理7

设$f(x)$是区间$[a,b]$上的一元下单峰函数，$\bar{x}$是$f(x)$在$[a,b]$上的极小点. 任取点$c<d\in[a,b]$，则有

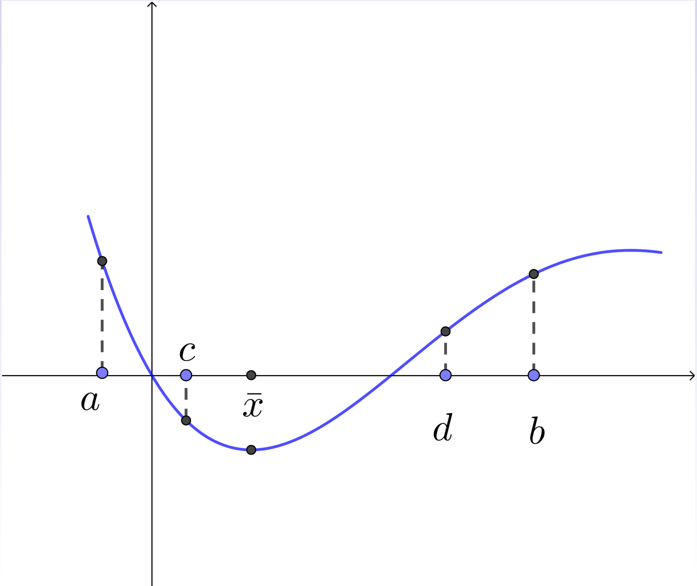
(1) 如果$f(c)>f(d)$，则$\bar{x}\in[c,b]$;

(2) 如果$f(c)\leq f(d)$，则$\bar{x}\in[a,d]$.

不需证明定理，只要看一下下图即可



不需证明定理，只要看一下下图即可



黄金分割法

思路

- 挑选区间$[a,b]$ 中的点，计算对应的目标函数值;

黄金分割法

思路

- 挑选区间$[a, b]$ 中的点, 计算对应的目标函数值;
- 通过比较不断缩小极小点所在的区间.

黄金分割法

思路

- 挑选区间$[a,b]$中的点，计算对应的目标函数值;
- 通过比较不断缩小极小点所在的区间.
- 其核心是：如何选择合适的分点.

黄金分割法

思路

- 挑选区间$[a,b]$ 中的点, 计算对应的目标函数值;
- 通过比较不断缩小极小点所在的区间.
- 其核心是：如何选择合适的分点.
  - 按对称压缩的方式缩小区间，即选择x_1, x_2,使得

$$x_1 - a = b - x_2 = \rho(b-a).$$

黄金分割法

思路

- 挑选区间$[a,b]$ 中的点, 计算对应的目标函数值;
- 通过比较不断缩小极小点所在的区间.
- 其核心是：如何选择合适的分点.
  - 按对称压缩的方式缩小区间，即选择x_1, x_2,使得
$$x_1 - a = b - x_2 = \rho(b-a).$$
  - 选择合适的ρ, 使得每次迭代只需计算一次函数值.

黄金分割法

思路

- 挑选区间$[a, b]$ 中的点, 计算对应的目标函数值;
- 通过比较不断缩小极小点所在的区间.
- 其核心是: 如何选择合适的分点.
  - 按对称压缩的方式缩小区间, 即选择x_1, x_2,使得

$$x_1 - a = b - x_2 = \rho(b - a).$$

  - 选择合适的ρ, 使得每次迭代只需计算一次函数值.
  - 为此, 容易算得$\rho = \frac{3-\sqrt{5}}{2} \approx 0.382$.

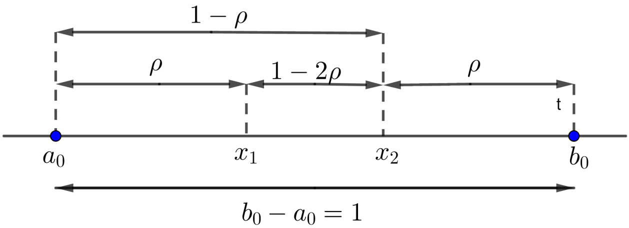
黄金分割法

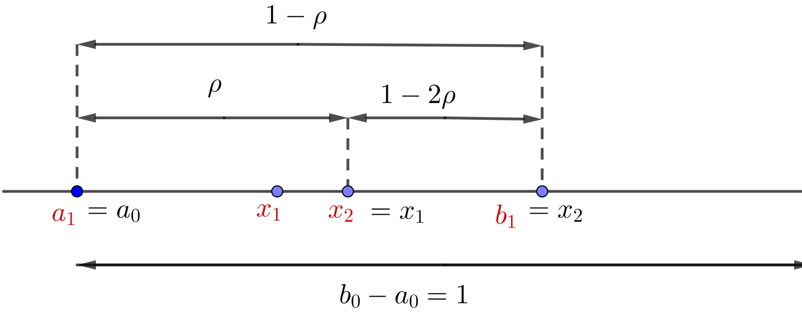
思路

- 挑选区间$[a,b]$ 中的点，计算对应的目标函数值;
- 通过比较不断缩小极小点所在的区间.
- 其核心是：如何选择合适的分点.
  - 按对称压缩的方式缩小区间，即选择x_1, x_2,使得

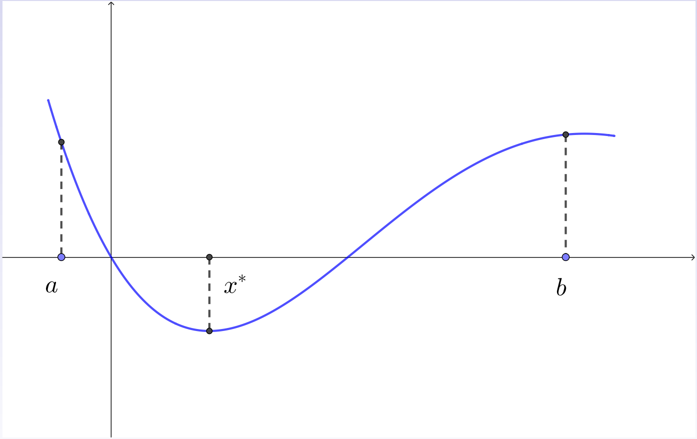
$$x_1 - a = b - x_2 = \rho(b-a).$$

  - 选择合适的ρ, 使得每次迭代只需计算一次函数值.
  - 为此，容易算得$\rho = \frac{3-\sqrt{5}}{2} \approx 0.382$.
- 每次去掉一个点缩短区间，确保剩下的三个点保持“高-低-高”的分布.

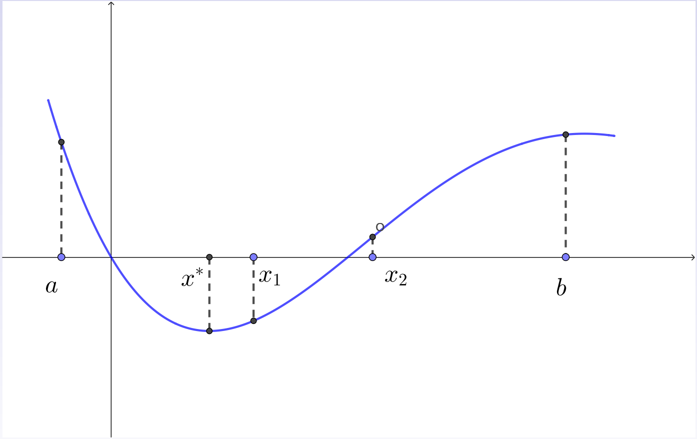

$1-\rho$
ρ
$1-2\rho$
ρ
t
a_0
x_1
x_2
b_0
$b_0 - a_0 = 1$


$1-\rho$
ρ
$1-2\rho$
$a_1 = a_0$
x_1
$x_2 = x_1$
$b_1 = x_2$
$b_0 - a_0 = 1$

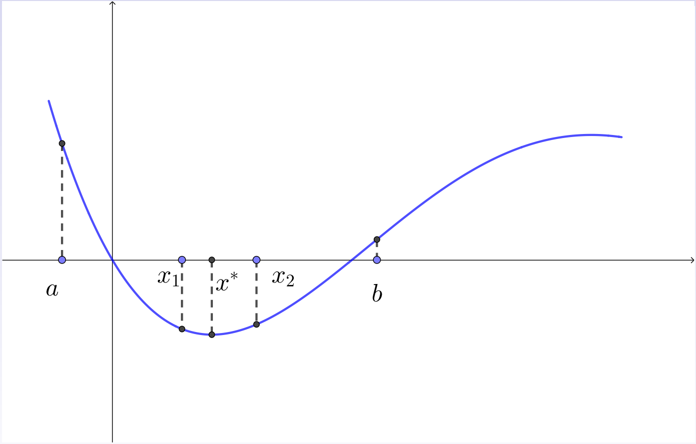
黄金分割法



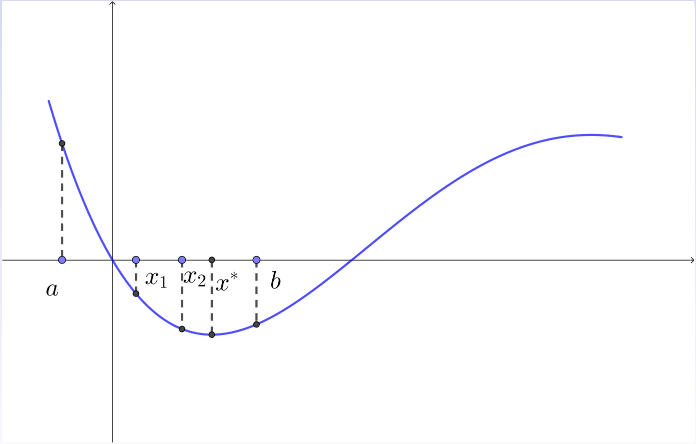
黄金分割法



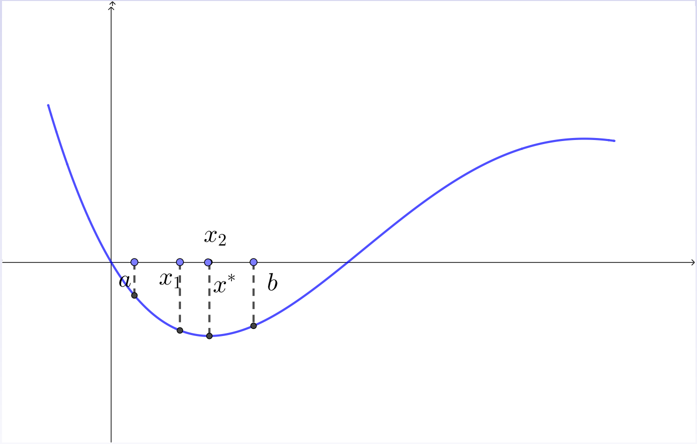
黄金分割法



黄金分割法



黄金分割法



黄金分割法

算法描述：黄金分割法

给定a, b $(a<b)$ 及$\epsilon>0$.

【步 1】 令$x_2=a+0.618(b-a)$, $f_2=f(x_2)$, 转步2;

【步 2】 令$x_1=a+0.382(b-a)$, $f_1=f(x_1)$, 转步3;

【步 3】 若$|b-a|\le\epsilon$, 则取$x^*=\frac{a+b}{2}$, 停; 否则,转步4;

【步 4】 若$f_1<f_2$, 则$b=x_2$, $x_2=x_1$, $f_2=f_1$, 转步2;
若$f_1=f_2$, 则$a=x_1$, $b=x_2$, 转步1;
若$f_1>f_2$, 则$a=x_1$, $x_1=x_2$, $f_1=f_2$, 转步5;

【步 5】 令$x_2=a+0.618(b-a)$, $f_2=f(x_2)$, 转步3.

例题

例4.2

用黄金分割法求函数 $f(x) = x^2 - x + 2$ 在区间 $[-1, 3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

例题

例4.2

用黄金分割法求函数$f(x) = x^2 - x + 2$ 在区间$[-1, 3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

取

$$x_1 = a + 0.382(b - a) = 0.528,$$
$$x_2 = a + 0.618(b - a) = 1.472.$$

例题

例4.2

用黄金分割法求函数$f(x)=x^2-x+2$ 在区间$[-1,3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

取

$$x_1 = a + 0.382(b-a) = 0.528,$$
$$x_2 = a + 0.618(b-a) = 1.472.$$

则

$$f_1 = f(x_1) = 1.751,\ f_2 = f(x_2) = 2.695.$$

例题

例4.2

用黄金分割法求函数$f(x) = x^2 - x + 2$ 在区间$[-1, 3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

取

$$x_1 = a + 0.382(b-a) = 0.528,$$
$$x_2 = a + 0.618(b-a) = 1.472.$$

则

$$f_1 = f(x_1) = 1.751,\ \ f_2 = f(x_2) = 2.695.$$

因为$f_1 < f_2$, 所以得到的新区间为$[-1, 1.472]$.

例题

仍把此区间记为$[a, b]$, 并令$x_2 = x_1$. 取

$$x_1 = a + 0.382(b - a).$$

计算过程如下表.

例题

仍把此区间记为$[a, b]$, 并令$x_2 = x_1$. 取

$$x_1 = a + 0.382(b - a).$$

计算过程如下表.

迭代	$[a, b]$	$x_1,\ x_2$	$f_1,\ f_2$	终止

例题

仍把此区间记为$[a,b]$, 并令$x_2 = x_1$. 取

$$x_1 = a + 0.382(b-a).$$

计算过程如下表.

迭代	$[a,b]$	$x_1,\ x_2$	$f_1,\ f_2$	终止
0	$[-1,3]$	0.528,1.472	1.751,2.695	否

例题

仍把此区间记为$[a, b]$, 并令$x_2 = x_1$. 取

$$x_1 = a + 0.382(b - a).$$

计算过程如下表.

迭代	$[a, b]$	$x_1,\ x_2$	$f_1,\ f_2$	终止
0	$[-1, 3]$	0.528,1.472	1.751,2.695	否
1	$[-1, 1.472]$	-0.056,0.528	2.059,1.751	否

例题

仍把此区间记为$[a,b]$, 并令$x_2 = x_1$. 取

$$x_1 = a + 0.382(b-a).$$

计算过程如下表.

迭代	$[a,b]$	x_1, x_2	f_1, f_2	终止
0	$[-1,3]$	0.528,1.472	1.751,2.695	否
1	$[-1,1.472]$	-0.056,0.528	2.059,1.751	否
2	$[-0.056,1.472]$	0.528,0.888	1.751,1.901	否

例题

仍把此区间记为$[a,b]$, 并令$x_2 = x_1$. 取

$$x_1 = a + 0.382(b-a).$$

计算过程如下表.

迭代	$[a,b]$	$x_1,\ x_2$	$f_1,\ f_2$	终止
0	$[-1,3]$	0.528,1.472	1.751,2.695	否
1	$[-1,1.472]$	-0.056,0.528	2.059,1.751	否
2	$[-0.056,1.472]$	0.528,0.888	1.751,1.901	否
3	$[-0.506,0.888]$	0.305,0.528	1.788,1.751	否
4	$[0.305,0.888]$	0.528,0.665	1.751,1.777	否
5	$[0.305,0.665]$	0.443,0.528	1.753,1.751	否
6	$[0.443,0.665]$	0.528,0.580	1.751,1.757	是
7	$[0.443,0.580]$			

例题

经过6 次迭代以满足精度要求，（近似）最优解与最优值为

$$x^* = \frac{0.443+0.665}{2} = 0.554,$$
$$f^* = f(x^*) = 1.753.$$

例题

例4.3

用黄金分割法求

$$f(x) = -2x^3 + 21x^2 - 60x + 50$$

在区间$[-1, 4]$ 内的最小值.

例题

例4.3

用黄金分割法求

$$f(x) = -2x^3 + 21x^2 - 60x + 50$$

在区间$[-1, 4]$ 内的最小值.

- 编写程序求解goldsec.m

例题

例4.3

用黄金分割法求

$$f(x) = -2x^3 + 21x^2 - 60x + 50$$

在区间$[-1, 4]$ 内的最小值.

- 编写程序求解goldsec.m
- 调用matlab 函数fminbnd

斐波那契法

思想：

斐波那契法

思想：

利用黄金分割法进行压缩时，参数ρ 始终保持不变.

斐波那契法

思想：

利用黄金分割法进行压缩时，参数ρ 始终保持不变.

如果在区间压缩过程中，允许参数ρ 不断调整，则可产生一种新的搜索方法.

斐波那契法

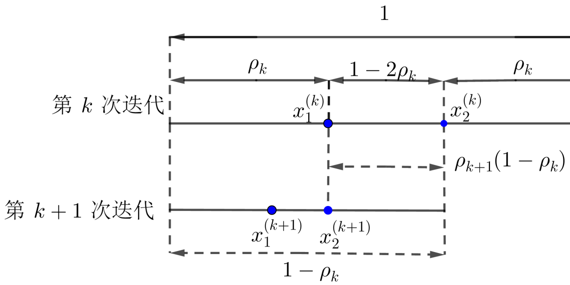
思想：

利用黄金分割法进行压缩时，参数ρ 始终保持不变.

如果在区间压缩过程中，允许参数ρ 不断调整，则可产生一种新的搜索方法.

设第k 次迭代的参数分别为ρ_k，为能够像黄金分割法一样,每次迭代中只需要计算一次目标函数，则需要

$$\rho_{k+1} = 1 - \frac{\rho_k}{1-\rho_k}, \rho_k \leq \frac{1}{2}.$$


1
ρ_k
$1-2\rho_k$
ρ_k
第 k 次迭代
$x_1^{(k)}$
$x_2^{(k)}$
$\rho_{k+1}(1-\rho_k)$
第 $k+1$ 次迭代
$x_1^{(k+1)}$
$x_2^{(k+1)}$
$1-\rho_k$

斐波那契法

思想 假设$\rho_1, \rho_2, \cdots$ 能够满足上述条件, 则经过N 次迭代后, 可将极小点所在区间的长度压缩为原始区间的

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N)$$

倍.

斐波那契法

思想 假设$\rho_1, \rho_2, \cdots$ 能够满足上述条件，则经过N 次迭代后，可将极小点所在区间的长度压缩为原始区间的

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N)$$

倍.

自然，我们希望总压缩比越小越好.

斐波那契法

思想 假设$\rho_1, \rho_2, \cdots$ 能够满足上述条件, 则经过N 次迭代后，可将极小点所在区间的长度压缩为原始区间的

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N)$$

倍.

自然，我们希望总压缩比越小越好. 斐波那契法可以做到总压缩比最小.

斐波那契法

斐波那契数列

$F_{-1}=0,\ F_0=1$; 对于$k\geq 0$, 有

$$F_{k+1}=F_k+F_{k-1}.$$

斐波那契法

斐波那契数列

$F_{-1}=0,\ F_0=1$; 对于$k\geq 0$, 有

$$F_{k+1}=F_k+F_{k-1}.$$

例4.4

斐波那契数列的前8项为

斐波那契法

斐波那契数列

$F_{-1}=0,\ F_0=1$; 对于$k\geq 0$, 有

$$F_{k+1}=F_k+F_{k-1}.$$

例4.4

斐波那契数列的前8项为

F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
1	2	3	5	8	13	21	34

斐波那契法

斐波那契法的参数

$$\rho_1 = 1 - \frac{F_N}{F_{N+1}} = \frac{F_{N-1}}{F_{N+1}}$$

$$\rho_2 = 1 - \frac{F_{N-1}}{F_N} = \frac{F_{N-2}}{F_N}$$

$$\vdots$$

$$\rho_k = 1 - \frac{F_{N-k+1}}{F_{N-k+2}} = \frac{F_{N-k}}{F_{N-k+2}}$$

$$\vdots$$

$$\rho_N = 1 - \frac{F_1}{F_2} = \frac{F_0}{F_2}$$

斐波那契法

确定迭代次数N

斐波那契法的总压缩比为

斐波那契法

确定迭代次数N

斐波那契法的总压缩比为

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N)=\frac{1}{F_{N+1}}$$

斐波那契法

确定迭代次数N

斐波那契法的总压缩比为

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N)=\frac{1}{F_{N+1}}$$

若初始区间为$[a,b]$, 要求精度为ϵ,

斐波那契法

确定迭代次数N

斐波那契法的总压缩比为

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N)=\frac{1}{F_{N+1}}$$

若初始区间为$[a,b]$，要求精度为ϵ,(即终止条件为区间长度小于等于ϵ). 则由

$$\frac{b-a}{F_{N+1}}<\epsilon$$

斐波那契法

确定迭代次数N

斐波那契法的总压缩比为

$$(1-\rho_1)(1-\rho_2)\cdots(1-\rho_N)=\frac{1}{F_{N+1}}$$

若初始区间为$[a,b]$, 要求精度为ϵ,(即终止条件为区间长度小于等于ϵ). 则由

$$\frac{b-a}{F_{N+1}}<\epsilon$$

可以确定N.

斐波那契法

算法描述

给定区间$[a,b]$ 及$\epsilon > 0$.

斐波那契法

算法描述

给定区间$[a,b]$ 及$\epsilon > 0$.

步 1　令$c = \frac{b-a}{\epsilon}$, $n = 1$, $F_0 = 1$, $F_1 = 1$, 转步2;

斐波那契法

算法描述

给定区间$[a, b]$ 及$\epsilon > 0$.

步 1 令$c = \frac{b-a}{\epsilon}$, $n = 1$, $F_0 = 1$, $F_1 = 1$, 转步2;

步 2 $n = n + 1$, $F_n = F_{n-2} + F_{n-1}$, 转步3;

斐波那契法

算法描述

给定区间$[a,b]$ 及$\epsilon > 0$.

步 1　令$c = \frac{b-a}{\epsilon}$, $n = 1$, $F_0 = 1$, $F_1 = 1$, 转步2;

步 2　$n = n+1$, $F_n = F_{n-2} + F_{n-1}$, 转步3;

步 3　若$F_n < c$, 则转步2; 否则转步4;

斐波那契法

算法描述

给定区间$[a,b]$ 及$\epsilon > 0$.

步 1 令$c = \frac{b-a}{\epsilon}$, $n = 1$, $F_0 = 1$, $F_1 = 1$, 转步2;

步 2 $n = n+1$, $F_n = F_{n-2} + F_{n-1}$, 转步3;

步 3 若$F_n < c$, 则转步2; 否则转步4;

步 4 令$k = 1$,

$$x_1 = a + \frac{F_{n-2}}{F_n}(b-a),\ x_2 = a + \frac{F_{n-1}}{F_n}(b-a),$$

令$f_1 = f(x_1)$, $f_2 = f(x_2)$, 转步5.

斐波那契法

算法描述

斐波那契法

算法描述

步 5 若$f_1 < f_2$, 则令$b = x_2$, $x_2 = x_1$, $f_2 = f_1$,

$$x_1 = a + \frac{F_{n-k-2}}{F_{n-k}}(b-a), f_1 = f(x_1),$$

转步6.

若$f_1 \geq f_2$, 则令$a = x_1$, $x_1 = x_2$, $f_1 = f_2$,

$$x_2 = a + \frac{F_{n-k-1}}{F_{n-k}}(b-a), f_2 = f(x_2),$$

转步6.

斐波那契法

算法描述

步 5　若$f_1 < f_2$, 则令$b = x_2$, $x_2 = x_1$, $f_2 = f_1$,

$$x_1 = a + \frac{F_{n-k-2}}{F_{n-k}}(b-a), f_1 = f(x_1),$$

转步6.

若$f_1 \geq f_2$, 则令$a = x_1$, $x_1 = x_2$, $f_1 = f_2$,

$$x_2 = a + \frac{F_{n-k-1}}{F_{n-k}}(b-a), f_2 = f(x_2),$$

转步6.

步 6　令$k = k+1$, 若$k < n-2$, 则转步5; 若$k = n-2$, 则转步7.

斐波那契法

算法描述

斐波那契法

算法描述

步 7　若$f_1 < f_2$, 则令$b = x_2$, $x_2 = x_1$, $f_2 = f_1$, 转步8;
若$f_1 \geq f_2$, 则令$a = x_1$, 转步8;

斐波那契法

算法描述

步 7 若$f_1 < f_2$, 则令$b = x_2$, $x_2 = x_1$, $f_2 = f_1$, 转步8;
若$f_1 \geq f_2$, 则令$a = x_1$, 转步8;

步 8 令$x_1 = x_2 - 0.1(b-a)$, $f_1 = f(x_1)$,
若$f_1 < f_2$, 则$x^* = \frac{1}{2}(a + x_2)$;
若$f_1 = f_2$, 则$x^* = \frac{1}{2}(x_1 + x_2)$;
若$f_1 > f_2$, 则$x^* = \frac{1}{2}(x_1 + b)$.

例题

例4.5

用斐波那契法求函数$f(x) = x^2 - x + 2$ 在区间$[-1, 3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

例题

例4.5

用斐波那契法求函数$f(x)=x^2-x+2$ 在区间$[-1,3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

【解：】函数$f(x)$ 在区间$[-1,3]$ 上为单峰函数，$\epsilon \le (3+1)\times 0.08 = 0.32$.

例题

例4.5

用斐波那契法求函数$f(x)=x^2-x+2$ 在区间$[-1,3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

【解：】函数$f(x)$ 在区间$[-1,3]$ 上为单峰函数，$\epsilon\le(3+1)\times 0.08=0.32$. 由$F_n\ge\frac{b-a}{\epsilon}=12.5$ 可知，应取的试点个数$n=6$，共5 次迭代.

例题

例4.5

用斐波那契法求函数$f(x) = x^2 - x + 2$ 在区间$[-1, 3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

【解：】函数$f(x)$ 在区间$[-1, 3]$ 上为单峰函数，$\epsilon \le (3+1) \times 0.08 = 0.32$. 由$F_n \ge \frac{b-a}{\epsilon} = 12.5$ 可知，应取的试点个数$n = 6$，共5 次迭代.

第一次迭代

例题

例4.5

用斐波那契法求函数$f(x)=x^2-x+2$ 在区间$[-1,3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

【解：】函数$f(x)$ 在区间$[-1,3]$ 上为单峰函数，$\epsilon \le (3+1)\times 0.08 = 0.32$. 由$F_n \ge \frac{b-a}{\epsilon} = 12.5$ 可知，应取的试点个数$n=6$, 共5 次迭代.

第一次迭代

最初的两个试点分别为

$$x_1 = a + \frac{F_4}{F_6}(b-a) = -1 + \frac{5}{13}\times 4 = 0.538;$$

$$x_2 = a + \frac{F_5}{F_6}(b-a) = 1.462.$$

例题

例4.5

用斐波那契法求函数$f(x) = x^2 - x + 2$ 在区间$[-1,3]$ 上的极小点，要求最终区间不大于原始区间的*0.08* 倍.

【解：】函数$f(x)$ 在区间$[-1,3]$ 上为单峰函数，$\epsilon \le (3+1) \times 0.08 = 0.32$. 由$F_n \ge \frac{b-a}{\epsilon} = 12.5$ 可知，应取的试点个数$n = 6$, 共5 次迭代.

第一次迭代

最初的两个试点分别为

$$x_1 = a + \frac{F_4}{F_6}(b-a) = -1 + \frac{5}{13} \times 4 = 0.538;$$
$$x_2 = a + \frac{F_5}{F_6}(b-a) = 1.462.$$

且$f_1 = 1.751$, $f_2 = 2.675$. 因为$f_1 < f_2$, 所以缩短后的新区间为$[-1, 1.462]$.

例题

第二次迭代

例题

第二次迭代

令$x_2 = 0.538,\ f_2 = f_1 = 1.751,\ b = 1.462.$ 取

$$x_1 = a + \frac{F_3}{F_5}(b-a) = -1 + \frac{3}{8} \times (1.462+1) = -0.077.$$

例题

第二次迭代

令$x_2 = 0.538,\ f_2 = f_1 = 1.751,\ b = 1.462.$ 取

$$x_1 = a + \frac{F_3}{F_5}(b - a) = -1 + \frac{3}{8} \times (1.462 + 1) = -0.077.$$

则$f_1 = 2.083.$ 因为$f_1 > f_2$, 所以得新区间为$[-0.077, 1.462]$.

例题

第二次迭代

令$x_2 = 0.538$, $f_2 = f_1 = 1.751$, $b = 1.462$. 取

$$x_1 = a + \frac{F_3}{F_5}(b - a) = -1 + \frac{3}{8} \times (1.462 + 1) = -0.077.$$

则$f_1 = 2.083$. 因为$f_1 > f_2$, 所以得新区间为$[-0.077, 1.462]$.

第三次迭代

例题

第二次迭代

令 $x_2 = 0.538,\ f_2 = f_1 = 1.751,\ b = 1.462.$ 取

$$x_1 = a + \frac{F_3}{F_5}(b-a) = -1 + \frac{3}{8} \times (1.462 + 1) = -0.077.$$

则 $f_1 = 2.083.$ 因为 $f_1 > f_2$, 所以得新区间为 $[-0.077, 1.462]$.

第三次迭代

令 $x_1 = 0.538,\ f_2 = f_1 = 1.751,\ a = -0.077.$ 取

$$x_2 = a + \frac{F_3}{F_4}(b-a) = -1 + \frac{3}{5} \times (1.462 + 0.077) = 0.846.$$

例题

第二次迭代

令$x_2 = 0.538$, $f_2 = f_1 = 1.751$, $b = 1.462$. 取

$$x_1 = a + \frac{F_3}{F_5}(b - a) = -1 + \frac{3}{8} \times (1.462 + 1) = -0.077.$$

则$f_1 = 2.083$. 因为$f_1 > f_2$, 所以得新区间为$[-0.077, 1.462]$.

第三次迭代

令$x_1 = 0.538$, $f_2 = f_1 = 1.751$, $a = -0.077$. 取

$$x_2 = a + \frac{F_3}{F_4}(b - a) = -1 + \frac{3}{5} \times (1.462 + 0.077) = 0.846.$$

则$f_2 = 1.870$. 因为$f_1 < f_2$, 所以得新区间为$[-0.077, 0.846]$.

例题

第四次迭代

例题

第四次迭代

令$x_2 = 0.538$, $f_2 = f_1 = 1.751$, $a = -0.077$. 取

$$x_1 = a + \frac{F_1}{F_3}(b-a) = -0.077 + \frac{1}{3} \times (0.846 + 0.077) = 0.231.$$

例题

第四次迭代

令$x_2 = 0.538$, $f_2 = f_1 = 1.751$, $a = -0.077$. 取

$$x_1 = a + \frac{F_1}{F_3}(b-a) = -0.077 + \frac{1}{3} \times (0.846 + 0.077) = 0.231.$$

则$f_1 = 1.822$. 因为$f_1 > f_2$, 所以得新区间为$[0.231, 0.846]$.

例题

第四次迭代

令$x_2 = 0.538$, $f_2 = f_1 = 1.751$, $a = -0.077$. 取

$$x_1 = a + \frac{F_1}{F_3}(b-a) = -0.077 + \frac{1}{3} \times (0.846 + 0.077) = 0.231.$$

则$f_1 = 1.822$. 因为$f_1 > f_2$, 所以得新区间为$[0.231, 0.846]$.

第五次迭代

例题

第四次迭代

令$x_2 = 0.538,\ f_2 = f_1 = 1.751,\ a = -0.077$. 取

$$x_1 = a + \frac{F_1}{F_3}(b-a) = -0.077 + \frac{1}{3} \times (0.846 + 0.077) = 0.231.$$

则$f_1 = 1.822$. 因为$f_1 > f_2$, 所以得新区间为$[0.231, 0.846]$.

第五次迭代

令$x_2 = 0.538,\ f_2 = f_1 = 1.751,\ a = 0.231$. 取

$$x_1 = x_2 - 0.1(b-a) = 0.477,\ f_1 = 1.751.$$

例题

第四次迭代

令$x_2 = 0.538$, $f_2 = f_1 = 1.751$, $a = -0.077$. 取

$$x_1 = a + \frac{F_1}{F_3}(b-a) = -0.077 + \frac{1}{3} \times (0.846 + 0.077) = 0.231.$$

则$f_1 = 1.822$. 因为$f_1 > f_2$, 所以得新区间为$[0.231, 0.846]$.

第五次迭代

令$x_2 = 0.538$, $f_2 = f_1 = 1.751$, $a = 0.231$. 取

$$x_1 = x_2 - 0.1(b-a) = 0.477, f_1 = 1.751.$$

因为$f_1 = f_2$, 所以最优解可取为$x^* = \frac{1}{2}(x_1 + x_2) = 0.508$, $f^* = 1.750$

收敛性

黄金分割法的迭代效果:

第k 次迭代后所得区间长度为初始区间长度的$\left(\frac{\sqrt{5}-1}{2}\right)^k$ 倍.

收敛性

黄金分割法的迭代效果:

第k 次迭代后所得区间长度为初始区间长度的$\left(\frac{\sqrt{5}-1}{2}\right)^k$ 倍.

例4.6

证明: *Fibonacci* 数列$\{F_n\}$ 具有如下表达式:

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right], n = 0, 1, 2, \cdots$$

定理8

设F_n 表示*Fibonacci* 数，则

$$\lim_{n \to +\infty} \frac{F_{n-1}}{F_n} = \frac{\sqrt{5}-1}{2} \approx 0.618.$$

定理8

设F_n 表示$Fibonacci$ 数，则

$$\lim_{n\to+\infty}\frac{F_{n-1}}{F_n}=\frac{\sqrt{5}-1}{2}\approx 0.618.$$

证明:

定理8

设F_n 表示*Fibonacci* 数，则

$$\lim_{n\to+\infty}\frac{F_{n-1}}{F_n}=\frac{\sqrt{5}-1}{2}\approx 0.618.$$

证明:

$$\lim_{n\to+\infty}\frac{F_{n-1}}{F_n}=\lim_{n\to+\infty}\frac{\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n}{\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}-\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}}$$

定理8

设F_n 表示*Fibonacci* 数，则

$$\lim_{n\to+\infty}\frac{F_{n-1}}{F_n}=\frac{\sqrt{5}-1}{2}\approx 0.618.$$

证明:

$$\lim_{n\to+\infty}\frac{F_{n-1}}{F_n}=\lim_{n\to+\infty}\frac{\left(\frac{1+\sqrt{5}}{2}\right)^n-\left(\frac{1-\sqrt{5}}{2}\right)^n}{\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}-\left(\frac{1-\sqrt{5}}{2}\right)^{n+1}}$$

$$=\lim_{n\to+\infty}\frac{1-\left(\frac{1-\sqrt{5}}{1+\sqrt{5}}\right)^n}{\frac{1+\sqrt{5}}{2}-\left(\frac{1-\sqrt{5}}{1+\sqrt{5}}\right)^n\frac{1-\sqrt{5}}{2}}$$

定理8

设F_n 表示*Fibonacci* 数，则

$$\lim_{n\to+\infty}\frac{F_{n-1}}{F_n}=\frac{\sqrt{5}-1}{2}\approx 0.618.$$

证明:

$$\lim_{n\to+\infty}\frac{F_{n-1}}{F_n}=\lim_{n\to+\infty}\frac{\left(\frac{1+\sqrt5}{2}\right)^n-\left(\frac{1-\sqrt5}{2}\right)^n}{\left(\frac{1+\sqrt5}{2}\right)^{n+1}-\left(\frac{1-\sqrt5}{2}\right)^{n+1}}$$

$$=\lim_{n\to+\infty}\frac{1-\left(\frac{1-\sqrt5}{1+\sqrt5}\right)^n}{\frac{1+\sqrt5}{2}-\left(\frac{1-\sqrt5}{1+\sqrt5}\right)^n\frac{1-\sqrt5}{2}}$$

$$=\frac{2}{1+\sqrt5}=\frac{\sqrt5-1}{2}.$$

线性收敛性

可以证明黄金分割法与Fibonacci 法都是线性收敛的，收敛比为 $\frac{\sqrt{5}-1}{2}$.

平分法

若函数$f(x)$ 在区间$[a,b]$ 上为下单峰函数且具有连续的一阶导数.

算法描述

已知$a < b$, $f'(a) < 0$, $f'(b) > 0$, $\epsilon > 0$.

步 1　$c = \frac{1}{2}(a+b)$, 转步2.

步 2　若$b - a \le \epsilon$, 则转步4; 否则，转步3.

步 3　计算$f'(c)$. 若$f'(c) = 0$, 则转步4; 若$f'(c) < 0$, 则令$a = c$, 转步1; 若$f'(c) > 0$, 则令$b = c$, 转步1.

步 4　令$x^* = c$. 停.

抛物线法——二次插值法

设在已知的三点$x_1 < x_0 < x_2$ 处对应的函数值$f_0 = f(x_0)$, $f_1 = f(x_1)$, $f_2 = f(x_2)$. 形成一个"高-低- 高" 的形状,即

$$f_1 > f_0,\ f_0 < f_2.$$

抛物线法——二次插值法

设在已知的三点$x_1 < x_0 < x_2$处对应的函数值$f_0 = f(x_0)$, $f_1 = f(x_1)$, $f_2 = f(x_2)$. 形成一个"高-低- 高"的形状,即

$$f_1 > f_0, \; f_0 < f_2.$$

考虑二次多项式

$$q(t) = at^2 + bt + c.$$

则q 在$t = -\frac{b}{2a}$处取得最小值.

抛物线法——二次插值法

设在已知的三点$x_1 < x_0 < x_2$ 处对应的函数值$f_0 = f(x_0)$, $f_1 = f(x_1)$, $f_2 = f(x_2)$. 形成一个"高-低- 高" 的形状,即

$$f_1 > f_0,\ f_0 < f_2.$$

考虑二次多项式

$$q(t) = at^2 + bt + c.$$

则q 在$t = -\frac{b}{2a}$ 处取得最小值.
现在要选择a, b 使得$q(t)$ 经过点$(x_i, f(x_i))$, $i = 0, 1, 2$.

将三点代入$q(t)$, 可求得

$$a = -\frac{(x_0 - x_2)f_1 + (x_2 - x_1)f_0 + (x_1 - x_0)f_2}{(x_1 - x_0)(x_0 - x_2)(x_2 - x_1)},$$

$$b = \frac{(x_0^2 - x_2^2)f_1 + (x_2^2 - x_1^2)f_0 + (x_1^2 - x_0^2)f_2}{(x_1 - x_0)(x_0 - x_2)(x_2 - x_1)}$$

于是，二次函数$q(t)$ 的极小点为

$$\bar{x} = -\frac{b}{2a} = \frac{1}{2}\frac{(x_0^2 - x_2^2)f_1 + (x_2^2 - x_1^2)f_0 + (x_1^2 - x_0^2)f_2}{(x_0 - x_2)f_1 + (x_2 - x_1)f_0 + (x_1 - x_0)f_2}. \quad (1)$$

于是，二次函数$q(t)$ 的极小点为

$$\bar{x} = -\frac{b}{2a} = \frac{1}{2}\frac{(x_0^2 - x_2^2)f_1 + (x_2^2 - x_1^2)f_0 + (x_1^2 - x_0^2)f_2}{(x_0 - x_2)f_1 + (x_2 - x_1)f_0 + (x_1 - x_0)f_2}. \quad (1)$$

当x_1, x_0, x_2 三点等距时，也就是

$$x_2 - x_0 = x_0 - x_1 = \Delta x, \Delta x > 0$$

时，则有

$$\bar{x} = x_0 + \frac{f_1 - f_2}{2(f_1 - 2f_0 + f_2)}\Delta x.$$

如图

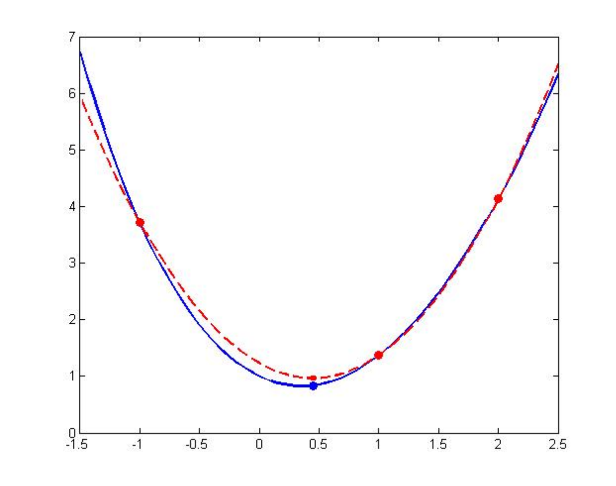


图: $f(x) = e^{-x} + x^2$ 及其二次插值

抛物线法——二次插值法

已知的三点$x_1 < x_0 < x_2$ 处对应的函数值$f_0 = f(x_0)$, $f_1 = f(x_1)$, $f_2 = f(x_2)$,

$$f_1 > f_0,\ f_0 < f_2,$$

控制误差$\epsilon_1 > 0$, $\epsilon_2 > 0$.

【步 1】 若$|x_1 - x_2| \le \epsilon_1$, 则转步10; 否则，转步2.

【步 2】 若$|(x_0 - x_2)f_1 + (x_2 - x_1)f_0 + (x_1 - x_0)f_2| \le \epsilon_2$, 则转步10; 否则，转步3.

【步 3】 按公式(1) 计算$\bar{x}$, 并计算$\bar{f} = f(\bar{x})$, 转步4.

【步 4】 若$f_0 < \bar{f}$, 则转步6; 若$f_0 = \bar{f}$, 转步7; 若$f_0 > \bar{f}$, 转步5.

抛物线法——二次插值法（续）

【步 5】 若$x_0 > \bar{x}$, 则令$x_2 = x_0$, $x_0 = \bar{x}$, $f_2 = f_0$, $f_0 = \bar{f}$, 转步1; 否则$x_0 < \bar{x}$, 则令$x_1 = x_0$, $x_0 = \bar{x}$, $f_1 = f_0$, $f_0 = \bar{f}$,转步1.

【步 6】 若$x_0 < \bar{x}$, 则令$x_2 = \bar{x}$, $f_2 = \bar{f}$, 转步1; 否则$x_0 > \bar{x}$, 则令$x_1 = \bar{x}$, $f_1 = \bar{f}$,转步1.

【步 7】 若$x_0 < \bar{x}$, 则$x_1 = x_0$, $x_2 = \bar{x}$, $x_0 = \frac{1}{2}(x_1 + x_2)$, $f_1 = f_0$, $f_2 = \bar{f}$, 计算$f_0 = f(x_0)$, 转步1; 若$x_0 = \bar{x}$, 则转步8; 若$x_0 > \bar{x}$, 则转步9.

抛物线法——二次插值法（续）

【步 8】 令$\hat{x} = \frac{1}{2}(x_1 + x_0)$, 计算$\hat{f} = f(\hat{x})$;

若$\hat{f} < f_0$, 则$x_2 = x_0$, $x_0 = \hat{x}$, $f_2 = f_0$, $f_0 = \hat{f}$, 转步1.

若$\hat{f} = f_0$, 则$x_1 = \hat{x}$, $x_2 = x_0$, $x_0 = \frac{1}{2}(x_1 + x_2)$, $f_1 = f(\hat{x})$, $f_2 = f_0$, 转步1.

若$\hat{f} > f_0$, 则$x_1 = \hat{x}$, $f_1 = \hat{f}$, 转步1.

【步 9】 令$x_1 = \bar{x}$, $x_2 = x_0$, $x_0 = \frac{1}{2}(x_1 + x_2)$, $f_1 = \bar{f}$, $f_2 = f_0$, 计算$f_0 = f(x_0)$, 转步1.

【步 10】 令$x^* = x_0$, $f^* = f_0$, 停.

进退法

如何确定包含极小点的一个区间

思想：从一点出发，按一定的步长，试图确定出函数值呈现“高-低-高”的三点，一个方向不成功，就退回来，再沿反方向寻找

进退法

进退法的步骤

步 1. 给定初始点$x_0 \in R$, 初始步长$h_0 > 0$, 令$h = h_0$, $x_1 = x_0$, 计算$f(x_1)$, 并令$k = 0$.

步 2. 令$x_4 = x_1 + h$, 计算$f(x_4)$, 令$k := k + 1$.

步 3. 若$f(x_4) < f(x_1)$, 则转步4, 否则, 转步5.

步 4. 令$x_2 = x_1$, $x_1 = x_4$, $f(x_2) = f(x_1)$, $f(x_1) = f(x_4)$, 令$h := 2h$, 转步2.

步 5. 若$k = 1$, 则转步6; 否则, 转步7.

步 6. 令$h = -h$, $x_2 = x_4$, $f(x_2) = f(x_4)$, 转步2.

步 7. 令$x_3 = x_2$, $x_2 = x_1$, $x_1 = x_4$, 停止计算.

区间$[x_1, x_3]$ 或$[x_3, x_1]$ 即为包含极小点的区间.

作业

- 小结各种精确一维搜索算法;
- 证明: Fibonacci 数列$\{F_n\}$ 具有如下表达式:

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right], n = 0, 1, 2, \cdots$$

- 证明黄金分割法与Fibonacci 法都是线性收敛的, 收敛性为$\frac{\sqrt{5}-1}{2}$;

不精确搜索

实际应用中,

不精确搜索

实际应用中，

1. 真正的精确值很难达到，需要较大的计算量;

不精确搜索

实际应用中,

1. 真正的精确值很难达到,需要较大的计算量;
2. 应该将更多的计算资源配置到多维优化算法,而不是追求高精度的一维搜索.

Wolfe 准则

设$f(\boldsymbol{x})$ 可微，取$\mu \in (0, \frac{1}{2})$, $\sigma \in (\mu, 1)$, 选取$\alpha_k > 0$ 使得

- $f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \alpha_k d_k) \geq -\mu\alpha_k \nabla f_k^T d_k, \quad \text{(w.a)}$
- $\nabla f(\boldsymbol{x}_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f_k^T d_k \quad \text{(w.b)}$

成立.

不精确搜索

Wolfe 准则

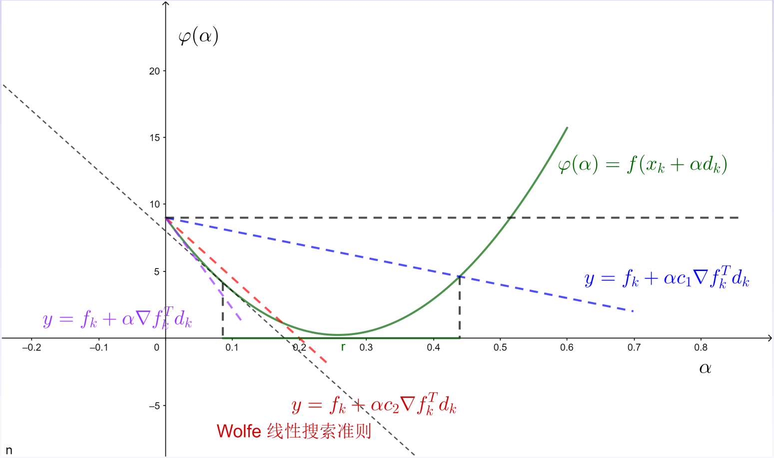
设$f(\boldsymbol{x})$ 可微，取$\mu \in (0, \frac{1}{2})$, $\sigma \in (\mu, 1)$, 选取$\alpha_k > 0$ 使得

- $f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \alpha_k d_k) \geq -\mu\alpha_k \nabla f_k^T d_k$, (w.a)
- $\nabla f(\boldsymbol{x}_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f_k^T d_k$ (w.b)

成立.

条件(w.a) 称为充分下降条件, 条件(w.b) 称为曲率条件.

Wolfe 准则



$\varphi(\alpha)$
20
15
10
5
−5
−0.2
−0.1
0
0.1
0.2
r
0.3
0.4
0.5
0.6
0.7
0.8
α
$\varphi(\alpha) = f(x_k + \alpha d_k)$
$y = f_k + \alpha c_1 \nabla f_k^T d_k$
$y = f_k + \alpha \nabla f_k^T d_k$
$y = f_k + \alpha c_2 \nabla f_k^T d_k$
Wolfe 线性搜索准则
n

不精确搜索

定理9 (α_k 的存在性)

设$f(\boldsymbol{x})$ 有下界且$\nabla f_k^T d_k < 0$. 令$\mu \in \left(0, \frac{1}{2}\right)$, $\sigma \in (\mu, 1)$, 则存在区间$[c_1, c_2]$, $0 < c_1 < c_2$, 使每个$\alpha \in [c_1, c_2]$ 均满足Wolfe准则的两个条件.

不精确搜索

算法描述

步 1 给定$\mu \in \left(0, \frac{1}{2}\right)$, $\sigma \in (\mu, 1)$, 令$a = 0$, $b = +\infty$, $\alpha = 1$, $j = 0$;

步 2 $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha d_k$, 计算$f_{k+1}$, ∇f_{k+1}.

若α 满足(w.a) 和(w.b), 则$\alpha_k = \alpha$, 停;

若α 不满足(w.a),(步长太大) , 转步3;

若α 不满足(w.b),(步长太小) , 转步4;

步 3 令$b = \alpha$, $\alpha = \frac{a+b}{2}$, 转步2.

步 4 令$a = \alpha$, $\alpha = \min\left\{2\alpha, \frac{a+b}{2}\right\}$, 转步2.

例题

例4.7

用不精确一维搜索求Rosenbrock 函数

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

在点$\boldsymbol{x}_k = (0,0)^T$ 沿方向$d_k = (1,0)^T$ 的近似步长α_k.

例题

例4.7

用不精确一维搜索求Rosenbrock 函数

$$f(\boldsymbol{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

在点$\boldsymbol{x}_k = (0,0)^T$ 沿方向$d_k = (1,0)^T$ 的近似步长α_k.

【解：】

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{pmatrix},$$

$$f_k = f(0,0) = 1,\ \nabla f_k = (-2,0)^T,\ \nabla f_k^T d_k = -2 < 0.$$

例题

给定$\mu = 0.1$, $\sigma = 0.5$. 令$a = 0$, $b = \infty$, $\alpha = 1$, $j = 0$.

例题

给定$\mu = 0.1$, $\sigma = 0.5$. 令$a = 0$, $b = \infty$, $\alpha = 1$, $j = 0$.
$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha d_k = (1,0)^T$, $f_{k+1} = f(1,0) = 100$. 因为

$$f_k - f_{k+1} = 1 - 100 = -99 < -\mu\alpha\nabla f_k^T d_k = 0.2,$$

所以式(w.a) 不成立.

例题

给定$\mu = 0.1$, $\sigma = 0.5$. 令$a = 0$, $b = \infty$, $\alpha = 1$, $j = 0$.
$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha d_k = (1, 0)^T$, $f_{k+1} = f(1, 0) = 100$. 因为

$$f_k - f_{k+1} = 1 - 100 = -99 < -\mu\alpha\nabla f_k^T d_k = 0.2,$$

所以式(w.a) 不成立.
则令$b = 1$, $\alpha = \frac{\alpha + a}{2} = 0.5$, 重新计算$\boldsymbol{x}_{k+1}$. 详细计算过程如下表.

例题

j	α	$\boldsymbol{x}_{k+1}$	f_{k+1}	(w.a)	(w.b)
0	1	$(1,0)^T$	100	不成立	
1	0.5	$(0.5,0)^T$	6.25	不成立	
2	0.25	$(0.25,0)^T$	0.953	不成立	
3	0.125	$(0.125,0)^T$	0.790	成立	成立

例题

j	α	$\boldsymbol{x}_{k+1}$	f_{k+1}	(w.a)	(w.b)
0	1	$(1,0)^T$	100	不成立	
1	0.5	$(0.5,0)^T$	6.25	不成立	
2	0.25	$(0.25,0)^T$	0.953	不成立	
3	0.125	$(0.125,0)^T$	0.790	成立	成立

从而得到满足Wolfe 条件的步长$\alpha_k = 0.125$, 则

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k = (0.125, 0).$$

Goldstein 准则

Goldstein 准则

设$f(\boldsymbol{x})$ 可微, $0 < \rho < \frac{1}{2}$, 选取$\alpha_k \in (0,1]$, 使得

$$f(\boldsymbol{x}_k + \alpha_k d_k) \le f(\boldsymbol{x}_k) + \rho\alpha_k \nabla f_k^T d_k, \tag{2}$$

$$f(\boldsymbol{x}_k + \alpha_k d_k) \ge f(\boldsymbol{x}_k) + (1-\rho)\alpha_k \nabla f_k^T d_k. \tag{3}$$

为简便起见，记$\varphi(\alpha) = f(\boldsymbol{x}_k + \alpha d_k)$.

Goldstein 线性搜索

步 1. 初始化. $\alpha_0 \in [0, \alpha_{\max}]$, $0 < \rho < \frac{1}{2}$, $t > 1$, $k = 0$, $a_0 = 0$, $b_0 = \alpha_{\max}$.

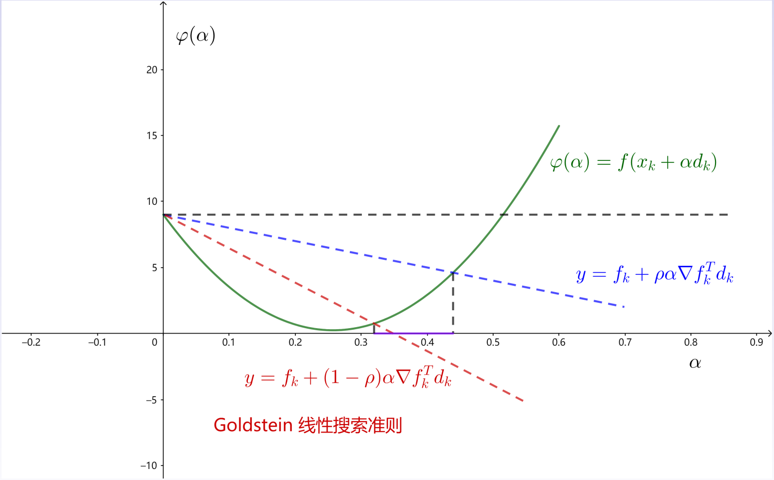
步 2. 若$\varphi(\alpha_k) \le \varphi(0) + \rho\alpha_k\varphi'(0)$, 转步3. 否则, 令$a_{k+1} = a_k$, $b_{k+1} = \alpha_k$. 转步4.

步 3. 若$\varphi(\alpha_k) \ge \varphi(0) + (1-\rho)\alpha_k\varphi'(0)$, 停止, 输出$\alpha_k$. 否则, 令$a_{k+1} = \alpha_k$, $b_{k+1} = b_k$.

- 若$b_{k+1} < \alpha_{\max}$, 转步4.
- 否则, 令$\alpha_{k+1} = t\alpha_k$, $k = k+1$, 转步2.

步 4. 令$\alpha_{k+1} = \frac{a_{k+1}+b_{k+1}}{2}$, $k = k+1$, 转步2.

Goldstein 准则



Armijo 准则

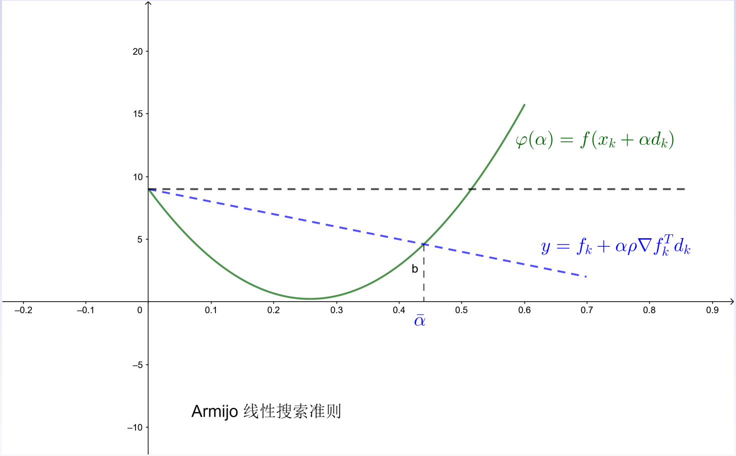
设d_k 满足$\nabla f_k^T d_k < 0$.

给定$\beta \in (0,1)$, $\rho \in \left(0, \frac{1}{2}\right)$. 设$m_k$ 是使得下述不等式

$$f(\boldsymbol{x}_k + \beta^m d_k) \leq f_k + \rho \beta^m \nabla f_k^T d_k, \qquad (Ar)$$

成立的最小非负整数，则令$\alpha_k = \beta^{m_k}$.

Armijo 准则



$\varphi(\alpha) = f(x_k + \alpha d_k)$
$y = f_k + \alpha\rho\nabla f_k^T d_k$
b
$\bar{\alpha}$
20
15
10
5
0
−5
−10
−0.2
−0.1
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Armijo 线性搜索准则

Armijo 准则

后退法

给出$\rho \in \left(0, \frac{1}{2}\right)$, $0 < \beta < 1$.

步 1 取$\alpha = 1$.

步 2 检验

$$f(\boldsymbol{x}_k + \alpha d_k) \le f(\boldsymbol{x}_k) + \rho\alpha \nabla f_k^T d_k$$

是否满足.

步 3 如果上式不满足, 取$\alpha = \beta\alpha$, 转步2. 否则, 取$\alpha_k = \alpha$, $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k d_k$.

假设

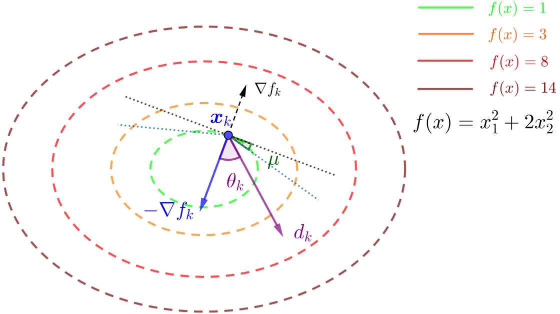
首先假设

$$\cos\theta_k = \cos < -g_k, d_k > = \frac{-\nabla f_k^T d_k}{\|\nabla f_k\| \|d_k\|},$$

$$\theta_k \le \frac{\pi}{2} - \mu,\ \ \mu > 0,\ \ \theta_k \in \left[0, \frac{\pi}{2}\right)$$

即: θ_k 不会无限地逼近直角.

如下图所示:


$f(x)=1$
$f(x)=3$
$f(x)=8$
$f(x)=14$
$f(x)=x_1^2+2x_2^2$
∇f_k
$\boldsymbol{x}_k$
μ
θ_k
$-\nabla f_k$
d_k

引理1

设$f(x)$ 连续可微, 其梯度满足*Lipschitz* 连续条件

$$\|\nabla f(\boldsymbol{y}) - \nabla f(\boldsymbol{x})\| \leq M\|\boldsymbol{y} - \boldsymbol{x}\|. \tag{4}$$

如果$f(\boldsymbol{x}_k + \alpha d_k)$, $(\alpha > 0)$ 下有界. 对于满足*Wolfe* 准则

$$f(\boldsymbol{x}_k + \alpha_k d_k) \leq f(\boldsymbol{x}_k) + \rho\alpha_k \nabla f_k^T d_k, \tag{5}$$

$$\nabla f(\boldsymbol{x}_k + \alpha_k d_k)^T d_k \geq \sigma \nabla f_k^T d_k,\ 0 < \rho < \sigma < 1, \tag{6}$$

的任何$\alpha_k > 0$, 均有

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_k + \alpha_k d_k) \geq \frac{\rho(1-\sigma)}{M}\|\nabla f_k\|^2 \cos^2\theta_k. \tag{7}$$

证明

证明：

$$(\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k))^T d_k \leq \|\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k)\|\|d_k\|.$$

证明

证明：

$$(\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k))^T d_k \le \|\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k)\|\|d_k\|.$$

由Lipschitz 条件

$$\|\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k)\| \le M\|x_k+\alpha_k d_k-x_k\| = \alpha_k M\|d_k\|.$$

证明

证明：

$$(\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k))^T d_k \leq \|\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k)\|\|d_k\|.$$

由Lipschitz 条件

$$\|\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k)\| \leq M\|x_k+\alpha_k d_k-x_k\| = \alpha_k M\|d_k\|.$$

所以

$$(\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k))^T d_k \leq \alpha_k M\|d_k\|^2$$

证明

证明：

$$(\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k))^T d_k \leq \|\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k)\|\|d_k\|.$$

由Lipschitz 条件

$$\|\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k)\| \leq M\|x_k+\alpha_k d_k-x_k\| = \alpha_k M\|d_k\|.$$

所以

$$(\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k))^T d_k \leq \alpha_k M\|d_k\|^2$$

又由(6) 可得

$$\begin{aligned}&(\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f(\boldsymbol{x}_k))^T d_k\\ \geq&\sigma\nabla f(\boldsymbol{x}_k)^T d_k-\nabla f(\boldsymbol{x}_k)^T d_k\\ =&(\sigma-1)\nabla f(\boldsymbol{x}_k)^T d_k.\end{aligned}$$

综上可知, $\alpha_k M\|d_k\|^2 \geq (\sigma - 1)\nabla f(x_k)^T d_k$, 即

$$\begin{aligned}\alpha_k M\|d_k\|^2 &\geq (\sigma-1)\nabla f(\boldsymbol{x}_k)^T d_k \\ &=(\sigma-1)\|\nabla f_k\|\|d_k\|\cos\theta_k.\end{aligned}$$

综上可知，$\alpha_k M\|d_k\|^2 \geq (\sigma-1)\nabla f(x_k)^T d_k$，即

$$\begin{aligned}\alpha_k M\|d_k\|^2 &\geq (\sigma-1)\nabla f(\boldsymbol{x}_k)^T d_k \\ &=(\sigma-1)\|\nabla f_k\|\|d_k\|\cos\theta_k.\end{aligned}$$

$$\alpha_k\|d_k\| \geq \frac{1-\sigma}{M}\|\nabla f_k\|\cos\theta_k.$$

综上可知，$\alpha_k M\|d_k\|^2 \geq (\sigma-1)\nabla f(x_k)^T d_k$，即

$$\begin{aligned}\alpha_k M\|d_k\|^2 &\geq (\sigma-1)\nabla f(\boldsymbol{x}_k)^T d_k \\ &=(\sigma-1)\|\nabla f_k\|\|d_k\|\cos\theta_k.\end{aligned}$$

$$\alpha_k\|d_k\| \geq \frac{1-\sigma}{M}\|\nabla f_k\|\cos\theta_k.$$

由(5) 知，

$$f(\boldsymbol{x}_k)-f(\boldsymbol{x}_k+\alpha_k d_k) \geq -\rho\alpha_k\nabla f_k^T d_k$$
$$=\rho\alpha_k\|\nabla f_k\|\|d_k\|\cos\theta_k.$$

综上可知, $\alpha_k M\|d_k\|^2 \geq (\sigma-1)\nabla f(x_k)^T d_k$, 即

$$\begin{aligned}\alpha_k M\|d_k\|^2 \geq & (\sigma-1)\nabla f(\boldsymbol{x}_k)^T d_k \\ = & (\sigma-1)\|\nabla f_k\|\|d_k\|\cos\theta_k.\end{aligned}$$

$$\alpha_k\|d_k\| \geq \frac{1-\sigma}{M}\|\nabla f_k\|\cos\theta_k.$$

由(5) 知,

$$\begin{aligned}f(\boldsymbol{x}_k)-f(\boldsymbol{x}_k+\alpha_k d_k) \geq & -\rho\alpha_k\nabla f_k^T d_k \\ = & \rho\alpha_k\|\nabla f_k\|\|d_k\|\cos\theta_k.\end{aligned}$$

由此得到,

$$f(\boldsymbol{x}_k)-f(\boldsymbol{x}_k+\alpha_k d_k) \geq \rho\frac{1-\sigma}{M}\|\nabla f_k\|^2\cos^2\theta_k.$$

Wolfe 准则的收敛定理

定理10

设$f(x)$ 连续可微和下有界, $\nabla f(x)$ 在水平集$\Omega=\{x \mid f(x)\leq f(x_0)\}$ 上一致连续. 那么采用不精确搜索 *Wolfe* 准则(5) 和(6) 得到的点列满足

$$\lim_{k\to+\infty}\|\nabla f_k\|\cos\theta_k=0. \tag{8}$$

如果夹角条件满足

$$\begin{aligned}\cos\theta_k &= \cos<-\nabla f_k, d_k> \\ &\geq \cos\left(\frac{\pi}{2}-\mu\right)=\sin\mu>0.\end{aligned}$$

则

$$\lim_{k\to+\infty}\|\nabla f_k\|=0. \tag{9}$$

证明： 注意, $\nabla f_k^T(\alpha_k d_k) < 0$, 所以$\{\boldsymbol{x}_k\}$ 在水平集Ω 中.

证明： 注意, $\nabla f_k^T(\alpha_k d_k) < 0$, 所以$\{\boldsymbol{x}_k\}$ 在水平集Ω 中. 假设(8) 不成立, 则存在$\epsilon > 0$ 和子序列$\{\boldsymbol{x}_k\}_{k\in K}$, 使得

$$\|\nabla f_k\| \cos\theta_k \geq \epsilon, \ \forall k \in K.$$

证明： 注意，$\nabla f_k^T(\alpha_k d_k) < 0$，所以$\{\boldsymbol{x}_k\}$ 在水平集Ω 中. 假设(8) 不成立，则存在$\epsilon > 0$ 和子序列$\{\boldsymbol{x}_k\}_{k\in K}$，使得

$$\|\nabla f_k\| \cos\theta_k \geq \epsilon, \ \forall k \in K.$$

即

$$\|\nabla f_k\| \frac{-\nabla f_k^T(\alpha_k d_k)}{\|\nabla f_k\| \|\alpha_k d_k\|} = \frac{-\nabla f_k^T(\alpha_k d_k)}{\|\alpha_k d_k\|} \geq \epsilon, \forall k \in K.$$

于是由(5) 可得

$$
\begin{aligned}
f_k - f(\boldsymbol{x}_k + \alpha_k d_k) &\geq -\rho\alpha_k \nabla f_k^T d_k \\
&= \rho\|\alpha_k d_k\| \left(-\frac{\nabla f_k^T(\alpha_k d_k)}{\|\alpha_k d_k\|}\right) \\
&\geq \rho\epsilon\|\alpha_k d_k\|.
\end{aligned}
$$

于是由(5) 可得

$$
\begin{aligned}
f_k - f(\boldsymbol{x}_k + \alpha_k d_k) &\geq -\rho\alpha_k \nabla f_k^T d_k \\
&= \rho\|\alpha_k d_k\| \left(-\frac{\nabla f_k^T(\alpha_k d_k)}{\|\alpha_k d_k\|}\right) \\
&\geq \rho\epsilon\|\alpha_k d_k\|.
\end{aligned}
$$

由于$f(\boldsymbol{x}_k)$ 单调下降并且有下界, 故收敛.

于是由(5) 可得

$$
\begin{aligned}
f_k - f(\boldsymbol{x}_k + \alpha_k d_k) &\geq -\rho\alpha_k \nabla f_k^T d_k \\
&= \rho\|\alpha_k d_k\| \left(-\frac{\nabla f_k^T(\alpha_k d_k)}{\|\alpha_k d_k\|}\right) \\
&\geq \rho\epsilon\|\alpha_k d_k\|.
\end{aligned}
$$

由于$f(\boldsymbol{x}_k)$ 单调下降并且有下界, 故收敛.
所以由(5) 可知$\alpha_k d_k \to 0$.

又由(6) 有

$$(\nabla f(\boldsymbol{x}_k + \alpha_k d_k) - \nabla f_k)^T d_k \geq -(1-\sigma)\nabla f_k^T d_k.$$

又由(6) 有

$$(\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f_k)^T d_k \geq -(1-\sigma)\nabla f_k^T d_k.$$

从而

$$\begin{aligned}
&\|\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f_k\|\|d_k\| \\
&\quad \geq (\nabla f(\boldsymbol{x}_k+\alpha_k d_k)-\nabla f_k)^T d_k \\
&\quad \geq -(1-\sigma)\nabla f_k^T d_k.
\end{aligned}$$

又由(6) 有

$$(\nabla f(\boldsymbol{x}_k + \alpha_k d_k) - \nabla f_k)^T d_k \geq -(1-\sigma)\nabla f_k^T d_k.$$

从而

$$\begin{aligned}
&\|\nabla f(\boldsymbol{x}_k + \alpha_k d_k) - \nabla f_k\| \|d_k\| \\
&\quad \geq (\nabla f(\boldsymbol{x}_k + \alpha_k d_k) - \nabla f_k)^T d_k \\
&\quad \geq -(1-\sigma)\nabla f_k^T d_k.
\end{aligned}$$

故

$$\|\nabla f(\boldsymbol{x}_k + \alpha_k d_k) - \nabla f_k\| \|\alpha_k d_k\| \geq -(1-\sigma)\nabla f_k^T (\alpha_k d_k).$$

因此

$$\epsilon \leq -\frac{\nabla f_k^T(\alpha_k d_k)}{\|\alpha_k d_k\|} \leq \frac{1}{1-\sigma}\|\nabla f(\boldsymbol{x}_k + \alpha_k d_k) - \nabla f_k\|.$$

由于$\alpha_k d_k \to 0, k \in K$, 而且$\nabla f(x)$ 在水平集上一致连续, 所以上式右端趋向于0, 从而产生矛盾. 这样, (8) 得证. 根据夹角条件, 很容易推导出(9).

一点说明

注意到上述的非精确搜索中均含有充分下降性条件，而收敛性证明也要求θ_k不能任意逼近直角，这在理论上是必要的，在实践中也是有意义的.

例4.8 (下降但不收敛到极小点的例子)

求解

$$\min_{x \in R} \quad f(x) = x^2.$$

显然该问题有唯一解$x^* = 0$. 设某算法产生迭代点列

$$\left\{ x_k: \ x_k = \left(1 + \frac{1}{k}\right)^{-k}, k = 1, 2, \cdots \right\}.$$

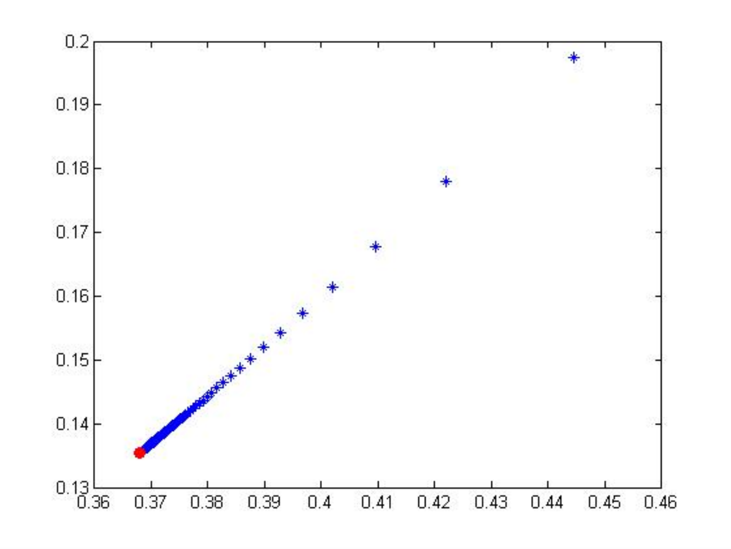
容易验证

$$f(x_k) > f(x_{k+1}),$$

但是

$$\lim_{k \to +\infty} f(x_k) = \frac{1}{e^2} \neq f(x^*).$$

如下图所示


0.2
0.19
0.18
0.17
0.16
0.15
0.14
0.13
0.36
0.37
0.38
0.39
0.4
0.41
0.42
0.43
0.44
0.45
0.46