

第四章 求解非线性方程组的迭代法

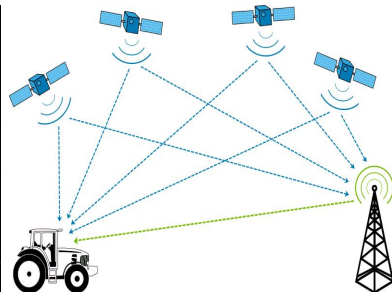
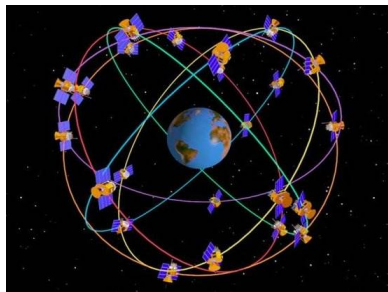
殷东生

yindongsheng@tsinghua.edu.cn

清华大学数学科学系

2023年秋季学期

GPS定位



每个卫星会有位置和距离数据 $\mathbf{s}_i = (x_i, y_i, z_i, \rho_i)^T$ ，定位的位置和距离为 $\mathbf{u} = (x, y, z, b)^T$ ，则可得非线性方程组：

$$\sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} + b = \rho_i, \quad i = 1, 2, 3, 4$$

无约束优化

设 $f: \mathbb{R}^n \rightarrow \mathbb{R}$, 求 f 在 \mathbb{R}^n 上的最小值 $f(\mathbf{x}^*)$, 亦即

$$f(\mathbf{x}^*) = \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

若目标函数 $f(\mathbf{x})$ 可微, 则 \mathbf{x}^* 是局部最优解的必要条件:

\mathbf{x}^* 是非线性方程组

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)^T = 0$$

的解。

进一步若 $f(\mathbf{x})$ 是凸函数, 则上述条件为充要条件。

梯度下降法 (gradient descent)

梯度下降法是一种常见的一阶优化方法，是求解无约束优化问题最简单、最经典的方法之一。

考虑 $\min_{\mathbf{x}} f(\mathbf{x})$ ，若能构造一个序列

$$\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots \text{ s.t. } f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)}), \quad k = 0, 1, 2, \dots$$

则不断迭代可收敛到局部极小点。由Taylor展开

$$f(\mathbf{x} + \mathbf{h}) \approx f(\mathbf{x}) + \mathbf{h}^T \nabla f(\mathbf{x}).$$

如要 $f(\mathbf{x} + \mathbf{h}) < f(\mathbf{x})$ ，则可选择向负梯度方向迭代，即

$$\mathbf{h} = -\gamma \nabla f(\mathbf{x})$$

此处 $\gamma > 0$ 是一个常数。于是有梯度下降法

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \gamma \nabla f(\mathbf{x}^{(k)}).$$

约束优化问题

给定向量值函数：

$$G(\mathbf{x}) \in \mathbb{R}^k, \quad H(\mathbf{x}) \in \mathbb{R}^l$$

求如下的带约束的优化问题：

$$f(\mathbf{x}^*) = \min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

$$\text{满足} \quad G(\mathbf{x}) \geq 0,$$

$$H(\mathbf{x}) = 0$$

若 f, G, H 可微，则存在Lagrange乘子 $\lambda^* \in \mathbb{R}^k$, $\mu^* \in \mathbb{R}^l$ 使得 $(\mathbf{x}^*, \lambda^*, \mu^*)$ 满足KKT条件：

$$\nabla f(\mathbf{x}) - \sum_{i=1}^k \lambda_i \nabla G_i(\mathbf{x}) + \sum_{i=1}^l \mu_i \nabla H_i(\mathbf{x}) = \mathbf{0}$$

$$\min(\lambda, G(\mathbf{x})) = \mathbf{0}$$

$$H(\mathbf{x}) = \mathbf{0}$$

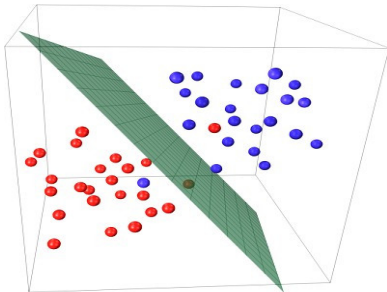
支持向量机(Support Vector Machine)

考虑训练集

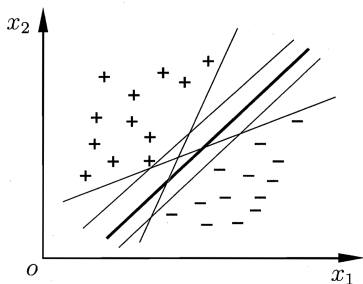
$$\{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), \dots, (\mathbf{x}_m, c_m)\}, \quad \mathbf{x}_i \in \mathbb{R}^n.$$

其中 $c_i \in \{-1, +1\}$ 是表明 \mathbf{x}_i 所属类的指标。

分类学习最基本的想法就是基于训练集 D 在样本空间中找到一个划分平面把不同类别的样本分开。



但是能将训练样本分开的划分超平面有很多个。



样本空间中，划分超平面可用如下的方程来描述

$$\mathbf{w}^T \mathbf{x} + b = 0.$$

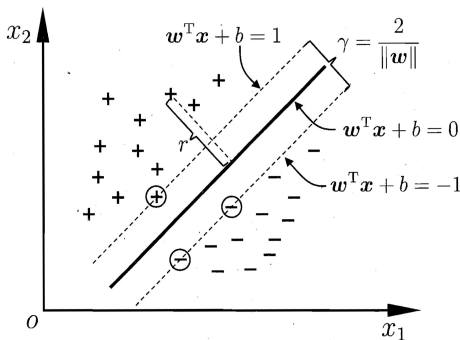
向量 \mathbf{w} 是划分超平面的法向量； b 是位移，决定了超平面和原点的距离。
样本空间任一点 \mathbf{x} 到划分超平面 (\mathbf{w}, b) 的距离

$$\gamma = \frac{|\mathbf{w}^T \mathbf{x} + b|}{\|\mathbf{w}\|}.$$

若超平面 (\mathbf{w}, b) 能将训练样本正确分类，即对 $(\mathbf{x}_i, c_i) \in D$ ，若 $c_i = +1$ ，则有 $\mathbf{w}^T \mathbf{x}_i + b > 0$ ；若 $c_i = -1$ ，则有 $\mathbf{w}^T \mathbf{x}_i + b < 0$ 。令

$$\begin{cases} \mathbf{w}^T \mathbf{x}_i + b \geq +1, & c_i = +1; \\ \mathbf{w}^T \mathbf{x}_i + b \leq -1, & c_i = -1 \end{cases}$$

如下图所示，距离超平面最近的这几个训练样本点使得上式等号成立，它们被称为“**支持向量**”（support vector）。



两个不同类支持向量到超平面的距离之和为

$$\gamma = \frac{2}{\|\mathbf{w}\|}$$

它被称为“**间隔**” (margin)。

想找到具有“最大间隔”的划分超平面，亦即找到参数 \mathbf{w} 和 b 使得 γ 最大，即

$$\begin{aligned} \max_{\mathbf{w}, b} \quad & \frac{2}{\|\mathbf{w}\|} \\ \text{s.t.} \quad & c_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m \end{aligned}$$

显然，为了最大化间隔，仅需最大化 $\|\mathbf{w}\|^{-1}$ ，这等价于最小化 $\|\mathbf{w}\|^2$ 。则优化问题可改写为支持向量机 (SVM) 的基本型。

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & c_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m \end{aligned}$$

方程组求根

非线性方程和方程组

$$F(\mathbf{x}) = 0, \quad \mathbf{x} \in \mathbb{R}^n, \quad F(\mathbf{x}) \in \mathbb{R}^n.$$

当 $n = 1$ 时变为单变量的单个方程

$$f(x) = 0.$$

单个方程常见的有多项式形式的代数方程

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0,$$

代数方程只有 $n = 1, 2, 3, 4$ 时有解析的求根公式, $n \geq 5$ 时, 没有通用的求根公式。

除了代数方程外。另一种常见的方程是超越方程如

$$3x^2 - e^x = 0, \quad x = \frac{1}{2} + \sin x.$$

数学家简介 (Cardano)

卡尔达诺(1501-1576)意大利数学家、医学家、物理学家，是意大利文艺复兴时期百科全书式的学者。一生共写了各类文章、书籍200多种，现存材料就有约7000页。他的数学贡献体现在几本著作中：《论掷骰游戏》给出一些概率论的基本概念和定理。得到所谓“幂定理”等结果；他最重要的著作《大术》首次公布了三、四次代数方程的一般解法，确认了高于一次的代数方程多于一个根，已知方程的一个根可将方程降阶，指出方程的根与系数间的某些关系，利用反复实施代换的方法求得方程的近似解。在解方程中使用了虚数等。在其当中关于一般二次代数方程的求根公式今称“卡当公式”或“卡尔达诺公式”。他的另外两部著作《事物之精妙》与《世间万物》包含了大量力学、机械学、天文学、化学、生物学等自然科学与技术的知识，还有密码术、炼金术，以及占星术等内容。被誉为当时最好的百科全书，曾多次印刷，广泛流传于欧洲大陆。作为医生既精于诊断开方，也专于外科手术。他还在理论上第一个阐述了诊断斑疹伤寒病的方法。对生理学和心理学的问题提出自己的见解。



在科学和工程计算中经常遇到求解多变量非线性方程组的问题。非线性性给理论研究和数值算法的研究带来极大的困难。

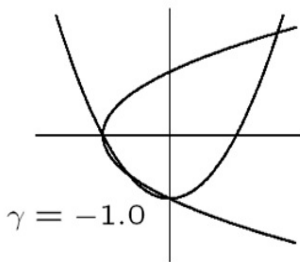
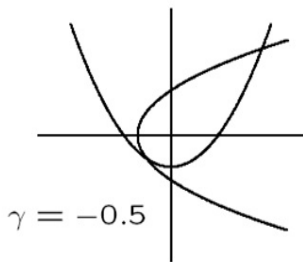
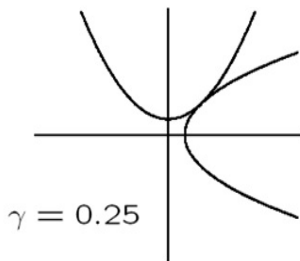
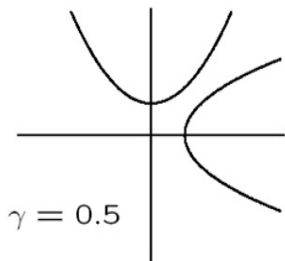
- 在理论上线性方程组 $A\mathbf{x} = \mathbf{b}$ 只要 A^{-1} 存在, 则解存在唯一, 而非线性方程组解的存在唯一性问题没有完全解决。

例

$$\begin{cases} f_1(x_1, x_2) = x_1^2 - x_2 + \gamma = 0, \\ f_2(x_1, x_2) = -x_1 + x_2^2 + \gamma = 0 \end{cases}$$

是 \mathbb{R}^2 上的两条抛物线, 它们的交点就是方程组的解, 对 $-1 \leq \gamma \leq 1$,

- ① $\gamma = 0.5$, 无解;
- ② $\gamma = \frac{1}{4}$, 一个解 $x_1 = x_2 = \frac{1}{2}$;
- ③ $\gamma = 0$, 两个解 $x_1 = x_2 = 0$; $x_1 = x_2 = 1$;
- ④ $\gamma = -1$, $x_1 = -1, x_2 = 0$; $x_1 = 0, x_2 = -1$; $x_1 = x_2 = \frac{1}{2}(1 \pm \sqrt{5})$ 。



还有无穷多解的方程组，如

例

$$\begin{cases} f_1(x_1, x_2) = \sin \frac{\pi}{2}x_1 - x_2 = 0 \\ f_2(x_1, x_2) = x_2 - 1/2 = 0 \end{cases}$$

所以对非线性方程组只能给出解在域 D 中存在唯一的充分条件，在此前提下讨论求解的方法。

注记

通常都用迭代法求解非线性方程，因为用直接消去法即使能消成一个变量的单个方程，还会出现大量增根，而且数值上也是不稳定的。只有小规模的特殊非线性方程才可能用直接法求解。

向量值函数是指映射 $F : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$

$$F(x) = \begin{pmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \dots \\ f_m(x_1, x_2, \dots, x_n) \end{pmatrix}$$

f_i 是一个多变量的标量函数。

定义 (向量函数的连续性)

设 $F : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{x}_0 \in D$ 。若 $\forall \epsilon > 0, \exists \delta > 0$, 使得

$$\forall \mathbf{x} \in B(\mathbf{x}_0, \delta) \subset D,$$

都有

$$\|F(\mathbf{x}) - F(\mathbf{x}_0)\| < \epsilon$$

则称函数 F 在 \mathbf{x}_0 点连续。设 $F : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ 在 D 中的每一点连续, 称 F 在 D 上连续。

定义 (一致连续性)

设 $K \subset D$, 若 $\forall \epsilon > 0, \exists \delta > 0$, 使得

$$\forall \mathbf{x}, \mathbf{y} \in K, \|\mathbf{x} - \mathbf{y}\| < \delta$$

满足

$$\|F(\mathbf{x}) - F(\mathbf{y})\| < \epsilon$$

则称 F 在 K 上一致连续。

定义

设 $K \subset D$, 若 $\forall \mathbf{x}, \mathbf{y} \in K$, 满足

$$\|F(\mathbf{x}) - F(\mathbf{y})\| < L \|\mathbf{x} - \mathbf{y}\|^p$$

若 $0 < p < 1$, 称 F 在 K 上 Hölder 连续。 p 称为 指标。 若 $p = 1$, 称 F 在 K 上 Lipschitz 连续。

为讨论迭代法，有必要讨论一下 \mathbb{R}^n 中的收敛性(拓扑结构)。

定义

设 $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ 是赋范空间 $(\mathbb{R}^n, \|\cdot\|)$ 中的向量序列。若存在 $\mathbf{x} \in \mathbb{R}^n$ ，使得

$$\lim_{k \rightarrow \infty} \|\mathbf{x}^{(k)} - \mathbf{x}\| = 0,$$

则称 $\{\mathbf{x}^{(k)}\}$ 收敛于 \mathbf{x} ，记为 $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{x}$ 。

注记

上面定义的收敛性表面上看来依赖于 \mathbb{R}^n 中的范数，但由于有限维线性空间中的所有范数都是等价的，向量序列的收敛性不依赖于范数的选择。

定理

设 $\mathbf{x}^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^T$, $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$, 则

$$\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = \mathbf{x} \iff \lim_{k \rightarrow \infty} x_i^{(k)} = x_i, \quad \forall i = 1, 2, \dots, n.$$

亦即向量序列收敛等价于其分量序列收敛。

定义 (向量函数的可导性)

设 $F : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$, $\mathbf{x} \in D$ 。若存在 $A(\mathbf{x}) \in \mathbb{R}^{m \times n}$, 使得

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}, \mathbf{x} + \mathbf{h} \in D} \frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - A(\mathbf{x})\mathbf{h}\|}{\|\mathbf{h}\|} = 0.$$

则称 F 在 \mathbf{x} 点可导 (可微)。记作 $F'(\mathbf{x}) = A(\mathbf{x}) = \frac{\partial F}{\partial \mathbf{x}}(\mathbf{x})$ 。

若函数 F 在 $\mathbf{x} \in D$ 可微, 则 F 在 \mathbf{x} 点连续。

因为 $\forall \epsilon > 0$, 由导数存在知存在 $\delta > 0$, 使得对任意 \mathbf{h} , $\|\mathbf{h}\| < \delta$,

$$\frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - A(\mathbf{x})\mathbf{h}\|}{\|\mathbf{h}\|} < \epsilon$$

即

$$\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x})\| \leq \epsilon \|\mathbf{h}\| + \|A(\mathbf{x})\mathbf{h}\| \leq (\epsilon + \|A\|) \|\mathbf{h}\|.$$

若 F 在 $\mathbf{x} \in D$ 可微, 则 f_i 偏导数存在, 且其Jacobi矩阵

$$F'(\mathbf{x}) = A \stackrel{\text{def}}{=} \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

F 在 $\mathbf{x} \in D$ 可微, 则 f_i 在 $\mathbf{x} \in D$ 可微。

$$\frac{\max_i |f_i(\mathbf{x} + \mathbf{h}) - f_i(\mathbf{x}) - \nabla f_i \cdot \mathbf{h}|}{\|\mathbf{h}\|_\infty} \leq \frac{\|F(\mathbf{x} + \mathbf{h}) - F(\mathbf{x}) - A\mathbf{h}\|_\infty}{\|\mathbf{h}\|_\infty} \rightarrow 0$$

注记

各偏导数存在 \nRightarrow 可导!!!

例

$$\begin{cases} f(x_1, x_2) = \frac{x_1 x_2^2}{x_1^2 + x_2^4}, & (x_1, x_2) \neq \mathbf{0}, \\ f(0, 0) = 0, \end{cases}$$

解:

$$\frac{\partial f(\mathbf{0})}{\partial x_1} = \lim_{h_1 \rightarrow 0} \frac{f(h_1, 0) - f(\mathbf{0})}{h_1} = 0,$$

$$\frac{\partial f(\mathbf{0})}{\partial x_2} = \lim_{h_2 \rightarrow 0} \frac{f(0, h_2) - f(\mathbf{0})}{h_2} = 0,$$

取

$$\mathbf{h} = (h_1, h_2)$$

且 $h_2 = h_1 \neq 0$,

则

$$\begin{aligned}\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{h}) - f(\mathbf{0})}{\|\mathbf{h}\|_\infty} &= \lim_{h_1 \rightarrow 0} \frac{h_1^3}{h_1(h_1^2 + h_1^4)} \\ &= \lim_{h_1 \rightarrow 0} \frac{1}{1 + h_1^2} \neq 0.\end{aligned}$$

取不同的 \mathbf{h} ，极限不相等，则 $F(\mathbf{x})$ 在 $\mathbf{x} = \mathbf{0}$ 处的导数不存在，因为 $F(\mathbf{x})$ 在 \mathbf{x} 点可导是指定义对 $\forall \mathbf{h} \in \mathbb{R}^n$ ，当 $\mathbf{h} \rightarrow \mathbf{0}$ 时成立。

若取 $h_1 = h_2^2$ ，有

$$\begin{aligned}&\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{h}) - f(\mathbf{0})}{\|\mathbf{h}\|_\infty} \\ &= \lim_{h_2 \rightarrow 0} \frac{h_2^4}{\|h_2^2\|_\infty (h_2^4 + h_2^4)} = \infty.\end{aligned}$$

数学家简介 (Abel)

尼尔斯·亨利克·阿贝尔 (1802年8月5日—1829年4月6日)，挪威数学家，在很多数学领域做出了开创性的工作。他最著名的一个结果是首次完整给出了高于四次的一般代数方程没有一般形式的代数解的证明。这个问题是他那时最著名的未解决问题之一，悬疑达250多年。他也是椭圆函数领域的开拓者，阿贝尔函数的发现者。除了五次方程之外，他还研究了更广的一类代数方程，后人发现这是具有交换的伽罗瓦群的方程。为了纪念他，后人称交换群为阿贝尔群。阿贝尔还研究过无穷级数，得到了一些判别准则以及关于幂级数求和的定理。这些工作使他成为分析学严格化的推动者。埃尔米特曾说：阿贝尔留下的思想可供数学家们工作150年。尽管阿贝尔成就极高，却在生前没有得到认可，他的生活非常贫困，死时只有27岁。数学界的诺贝尔奖Abel奖。



注记

迭代法与直接法不同，它并不试图在有限计算步中得到问题的真解，而是从若干个初始向量出发，用一个适当的算法，得到一个近似解的序列 $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ 。

一般 $\mathbf{x}^{(k+1)}$ 的计算公式可写成如下形式

$$\mathbf{x}^{(k+1)} = F_k \left(\mathbf{x}^{(k)}, \mathbf{x}^{(k-1)}, \dots, \mathbf{x}^{(k-m)} \right), k = 0, 1, \dots$$

- 若 $m \geq 1$ ，则称为**多步迭代法**(multistep iterative method)。
- 若 $m = 0$ ，即 $\mathbf{x}^{(k+1)}$ 的计算只与 $\mathbf{x}^{(k)}$ 相关

$$\mathbf{x}^{(k+1)} = F_k(\mathbf{x}^{(k)}), k = 0, 1, \dots$$

则称其为**单步法**(one step method)。

- 如果 F_k 是线性映射，即

$$\mathbf{x}^{(k+1)} = B_k \mathbf{x}^{(k)} + \mathbf{f}_k, \quad k = 0, 1, \dots$$

其中

$$B_k \in \mathbb{R}^{n \times n}, \quad \mathbf{f}_k \in \mathbb{R}^n,$$

则称该方法为线性单步迭代法， B_k 称为迭代矩阵(iterative matrix)。

- 进一步，若 B_k 和 \mathbf{f}_k 与 k 无关，即

$$\mathbf{x}^{(k+1)} = B \mathbf{x}^{(k)} + \mathbf{f}, \quad k = 0, 1, \dots$$

则称该方法为线性定常单步迭代法。

定义 (收敛阶)

迭代法称为是 p 阶收敛到 \mathbf{x}^* 的, 若算法产生的序列 $\{\mathbf{x}^{(k)}\}$ 满足

$$\exists C > 0 : \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^p} \leq C, \forall k \geq k_0.$$

注记

- 若 $p = 1$, 算法收敛, 则必有 $C < 1$, 此时 C 称为**收敛因子**。
- 线性方程组的迭代法的收敛性不依赖于初值的选取, 具有全局收敛性。
- 而非线性方程的迭代法的收敛性依赖于初值的选取, 一般只有局部收敛性(即要求初值 $\mathbf{x}^{(0)}$ 在 \mathbf{x}^* 的一个小邻域内)。相应的对区间内任意初值收敛的算法称为全局收敛到 \mathbf{x}^* 。

① $p = 1$ 线性收敛。若

$$\lim_{k \rightarrow \infty} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} = 0$$

称超线性收敛。

② $p = 2$ 平方收敛；

③ 若 $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ 超线性收敛于 \mathbf{x}^* ，则

$$\lim_{k \rightarrow \infty} \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} = 1.$$

因为

$$\left| \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} - 1 \right| \leq \frac{\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\|}{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|} \rightarrow 0.$$

不动点与压缩映射

定义 (不动点)

设 $G : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, 若 $\mathbf{x}^* \in D$ 满足 $\mathbf{x}^* = G(\mathbf{x}^*)$, 则称 \mathbf{x}^* 为映射 G 的不动点。

$$\begin{array}{ccc} F(\mathbf{x}) = 0 & \iff & \mathbf{x} = G(\mathbf{x}) \\ \text{求根} & & \text{求不动点} \end{array}$$

定义 (压缩映射)

假定 $G : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, 若 $\alpha \in (0, 1)$, 有

$$\|G(\mathbf{x}) - G(\mathbf{y})\| \leq \alpha \|\mathbf{x} - \mathbf{y}\|, \forall \mathbf{x}, \mathbf{y} \in D_0 \subset D$$

则称 G 在 D_0 上为压缩映射。 α 为 压缩常数。

压缩映射对范数的依赖性

映射压缩性依赖于范数的选择。 $G(\mathbf{x}) = A\mathbf{x}$

$$A = \begin{pmatrix} 0.1 & 0.9 \\ 0 & 0.1 \end{pmatrix}$$

由于 $\|A\|_2 = \sqrt{\rho(A^T A)} \leq 0.83$, 于是

$$\|G(\mathbf{x}) - G(\mathbf{y})\|_2 \leq \|A\|_2 \|\mathbf{x} - \mathbf{y}\|_2 \leq 0.83 \|\mathbf{x} - \mathbf{y}\|_2$$

是压缩的。但

$$\|G([1, 1]) - G([0, 0])\|_\infty = \|[1, 0.1]\|_\infty = 1 = \|[1, 1] - [0, 0]\|_\infty$$

因而在无穷范数下不压缩。

定理 (压缩映射原理)

设 $G : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ 在闭集 $D_0 \subset D$ 上是压缩映射, 且 G 把 D_0 映入自身, 即 $GD_0 \subset D_0$, 则 G 在 D_0 上有唯一的不动点。

设有如下不动点问题

$$\mathbf{x} = G(\mathbf{x})$$

构造

$$\begin{cases} \mathbf{x}^{(0)} \in \mathbb{R}^n, \\ \mathbf{x}^{(k+1)} = G(\mathbf{x}^{(k)}). \end{cases}$$

其为单步迭代法。

问题: $\lim_{k \rightarrow \infty} \mathbf{x}^{(k)} = ?$

定义 (局部收敛)

设 $\mathbf{x}^* = G(\mathbf{x}^*)$ 。若 \mathbf{x}^* 的邻域 $S \subset D$, 使得

$$\forall \mathbf{x}^{(0)} \in S, \mathbf{x}^{(k)} \in S, \mathbf{x}^{(k)} \rightarrow \mathbf{x}^*$$

称 \mathbf{x}^* 是 S 的吸引点, S 为 \mathbf{x}^* 的吸引域。

若 \mathbf{x}^* 有吸引域, 则称其为局部收敛的。

定理

设 $G: D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, $\mathbf{x}^* = G(\mathbf{x}^*)$ 。若有开球 $S = S(\mathbf{x}^*, \delta)$ 和常数 $\alpha \in (0, 1)$, 使得

$$\|G(\mathbf{x}) - G(\mathbf{x}^*)\| = \|G(\mathbf{x}) - \mathbf{x}^*\| \leq \alpha \|\mathbf{x} - \mathbf{x}^*\|, \quad \forall \mathbf{x} \in S,$$

则 $\forall \mathbf{x}^{(0)} \in S$, \mathbf{x}^* 是迭代序列 $\mathbf{x}^{(k)}$ 的吸引点(收敛极限)。

注记

前面的压缩映射原理是在整个区域 $D_0 \subset D$ 上满足压缩条件, 它的收敛性是大范围收敛, 且定理本身包含了不动点 \mathbf{x}^* 的存在唯一性。而上面的定理是假定 $G(\mathbf{x})$ 的不动点 \mathbf{x}^* 已知的前提下, 要求在 \mathbf{x}^* 的邻域 $S(\mathbf{x}^*, \delta)$ 上满足压缩条件, 因而它只是一个局部收敛性。

在实际应用的时候用 $G(x)$ 在 x^* 点的导数 $G'(x^*)$ 比压缩条件更方便。

定理

设 $G : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$ 在 $x^* = G(x^*)$ 处可导, 若 $\rho(G'(x^*)) = \sigma < 1$, 则存在开球 $S = S(x^*, \delta) \subset D$ 使得 $\forall x^{(0)} \in S$, 由不动点迭代产生的迭代序列收敛到 x^* 。

例

求解方程组

$$\begin{cases} x_1^2 - 10x_1 + x_2^2 + 8 = 0, \\ x_1x_2^2 + x_1 - 10x_2 + 8 = 0. \end{cases}$$

数学家简介 (Galois)

埃瓦里斯特·伽罗瓦，1811年10月25日生，法国数学家。现代数学中的分支学科群论的创立者。用群论彻底解决了根式求解代数方程的问题，而且由此发展了一整套关于群和域的理论，人们称之为伽罗瓦群和伽罗瓦理论。在世时在数学上研究成果的重要意义没被人们所认识，曾呈送科学院3篇学术论文，均被退回或遗失。后转向政治，支持共和党，曾两次被捕。21岁时死于一次决斗。他系统化地阐释了为何五次以上之方程式没有公式解，而四次以下有公式解。他漂亮地证明高斯的论断：若用尺规作图能作出正 p 边形， p 为质数的充要条件为 $p = 2^{2^k} + 1$ 。（所以正十七边形可做图）。他解决了古代三大作图问题中的两个：“不能任意三等分角”，“倍立方不可能”。



Newton迭代

求解 $F(\mathbf{x}) = 0$.

因为

$$F(\mathbf{x}) \approx F(\mathbf{x}^{(k)}) + F'(\mathbf{x}^{(k)})(\mathbf{x} - \mathbf{x}^{(k)}),$$

得不动点迭代格式

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - F'(\mathbf{x}^{(k)})^{-1}F(\mathbf{x}^{(k)})$$

称为Newton迭代法。

Newton迭代法

具体的计算步骤

- ① 给定初值 $\mathbf{x}^{(0)}$, 计算 $F'(\mathbf{x}^{(0)})$;
- ② 对 $k = 1, 2, \dots$, 求解线性方程组

$$F'(\mathbf{x}^{(k)})\Delta\mathbf{x}^{(k)} = -F(\mathbf{x}^{(k)}), \quad \text{牛顿方程},$$

得到 $\Delta\mathbf{x}^{(k)}$;

- ③ 更新

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta\mathbf{x}^{(k)},$$

计算 $F'(\mathbf{x}^{(k+1)})$, 返回第2步。

用Newton 迭代法求解非线性方程组

$$F(\mathbf{x}) = \begin{bmatrix} x_1^2 - 10x_1 + x_2^2 + 8 \\ x_1x_2^2 + x_1 - 10x_2 + 8 \end{bmatrix} = \mathbf{0}.$$

则其Jacobi矩阵为

$$F'(\mathbf{x}) = \begin{bmatrix} 2x_1 - 10 & 2x_2 \\ x_2^2 + 1 & 2x_1x_2 - 10 \end{bmatrix}.$$

选取初值 $\mathbf{x}^{(0)} = (0, 0)^T$, 求解方程组

$$F'(\mathbf{x}^{(0)})\Delta\mathbf{x}^{(0)} = -F(\mathbf{x}^{(0)}) \iff \begin{bmatrix} -10 & 0 \\ 1 & -10 \end{bmatrix} \Delta\mathbf{x}^{(0)} = \begin{bmatrix} -8 \\ -8 \end{bmatrix}$$

得 $\Delta\mathbf{x}^{(0)} = (0.8, 0.88)^T \Rightarrow \mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \Delta\mathbf{x}^{(0)} = (0.8, 0.88)^T$ 。继续计算有 $\mathbf{x}^{(2)} = (0.991787, 0.991812)^T$, $\mathbf{x}^{(3)} = (0.999975, 0.999969)^T$, $\mathbf{x}^{(4)} = (1.000000, 1.000000)^T$ 。

定理 (收敛性定理)

设 $F : D \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, $F(\mathbf{x}^*) = 0$, $F(\mathbf{x})$ 在 \mathbf{x}^* 的开邻域 $S_0 \subset D$ 上可导, 且 $F'(\mathbf{x})$ 连续, $F'(\mathbf{x}^*)$ 非奇异, 则存在闭球 $S = S(\mathbf{x}^*, \delta) \subset S_0$, 使映射

$$G(\mathbf{x}) = \mathbf{x} - F'(\mathbf{x})^{-1}F(\mathbf{x}),$$

对所有 $\mathbf{x} \in S$ 有意义, 且牛顿法生成的序列超线性收敛于 \mathbf{x}^* 。若

$$\|F'(\mathbf{x}) - F'(\mathbf{x}^*)\| \leq L\|\mathbf{x} - \mathbf{x}^*\|, \forall \mathbf{x} \in S.$$

则迭代序列 $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ 至少平方收敛。

牛顿法的优缺点

- 收敛快;
- 局部收敛性, 计算量大, 每步要计算 $F(\mathbf{x})$ 及 $F'(\mathbf{x})$ 的值, 并解线性方程组。

实际计算的时候, 为了得到 $\mathbf{x}^{(k+1)}$, 首先必须估计 $F(\mathbf{x}^{(k)})$ 以决定算法是否终止。如果要继续迭代, 则要计算Jacobi矩阵 $F'(\mathbf{x}^{(k)})$, 接着对 $F'(\mathbf{x})$ 做分解, 计算

$$F'(\mathbf{x}^{(k)})\mathbf{s} = -F(\mathbf{x}^{(k)}) \Rightarrow \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{s}.$$

注记

在这些步骤中, 计算 $F'(\mathbf{x}^{(k)})$ 和对它作分解的计算量最大。如果 $F'(\mathbf{x}^{(k)})$ 是稠密阵, 那么作分解的计算量为 $O(n^3)$ 。而用数值微商近似 $F'(\mathbf{x}^{(k)})$ 的计算量为计算 $F(\mathbf{x}^{(k)})$ 的计算量的 n 倍; 则当 n 很大时计算量很大。如果 $F'(\mathbf{x}^{(k)})$ 是稀疏的, 则计算量可以大大减少。

Newton迭代算法

Newton迭代法

- 1 $r_0 = \|F(\mathbf{x})\|,$
- 2 do while $\|F(\mathbf{x})\| > \tau_r r_0 + \tau_a;$
 - (a) 计算 $F'(\mathbf{x})$;
 - (b) 分解 $F'(\mathbf{x}) = LU$;
 - (c) 求解 $LU\Delta\mathbf{x} = -F(\mathbf{x})$;
 - (d) $\mathbf{x} = \mathbf{x} + \Delta\mathbf{x}$;
 - (e) 估计分析 $F(\mathbf{x})$ 。

Newton迭代的Matlab代码

```
function [x, nit] = newtonsys(F, J, x0, toll, nmax, p)
[n,m]=size(F); nit=0; Fxn=zeros(n,1); x=x0; err=toll+1;
for i=1:n, for j=1:n, Jxn(i,j)=eval(J((i-1)*n+j,:)); end; end
[L,U,P]=lu(Jxn); step=0;
while err > toll
    if step == p
        step = 0;
        for i=1:n;
            Fxn(i)=eval(F(i,:));
            for j=1:n; Jxn(i,j)=eval(J((i-1)*n+j,:)); end
        end
        [L,U,P]=lu(Jxn);
    else
        for i=1:n, Fxn(i)=eval(F(i,:)); end
    end
    nit=nit+1; step=step+1; Fxn=-P*Fxn; y=forward_col(L,Fxn);
    deltax=backward_col(U,y); x = x + deltax; err=norm(deltax);
    if nit > nmax
        disp(' Fails to converge within maximum number of iterations ');
        break
    end
end
end
```

Newton法的变形

原始的Newton法由于每一步都需要计算 $F'(\mathbf{x})\mathbf{s} = -F(\mathbf{x})$ ，计算量很大，为了减少计算量，必须对Newton法进行修改。最简单的方法就是用 $F'(\mathbf{x}^{(0)})$ 代替所有的 $F'(\mathbf{x}^{(k)})$ ，这样就可以得到简化牛顿法(chord 方法):

- ① $r_0 = \|F(\mathbf{x})\|;$
- ② 计算 $F'(\mathbf{x})$;
- ③ 分解 $F'(\mathbf{x}) = LU$;
- ④ do while $\|F(\mathbf{x})\| > \tau_r r_0 + \tau_a$;
 - (c) 求解 $LU\Delta\mathbf{x} = -F(\mathbf{x})$;
 - (d) $\mathbf{x} = \mathbf{x} + \Delta\mathbf{x}$;
 - (e) 估计分析 $F(\mathbf{x})$ 。

虽然chord方法计算量小但是只有线性收敛，为了结合chord法计算量少和Newton法 收敛快的优点可以设计修正牛顿法(Shamanskii方法):

$$\begin{cases} \mathbf{x}^{(k,i+1)} = \mathbf{x}^{(k,i)} - F'(\mathbf{x}^{(k)})^{-1}F(\mathbf{x}^{(k,i)}), i = 0, 1, \dots, m-1, \\ \mathbf{x}^{(k)} = \mathbf{x}^{(k,0)}, \mathbf{x}^{(k+1)} = \mathbf{x}^{(k,m)}, k = 0, 1, \dots \end{cases}$$

注记

当 $m = 1$ 时就是Newton法， $m = \infty$ 时为chord方法。

定理 (Shamanskii方法的收敛性)

设 $F(\mathbf{x}) = 0$ 满足有唯一解 \mathbf{x}^* ， $F'(\mathbf{x})$ 在区域 D 上Lipschitz连续且 $F'(\mathbf{x}^*)$ 非奇异，令 $m \geq 1$ 则存在 $K_S > 0, \delta > 0$ 使得如果 $\mathbf{x}^{(0)} \in S = S(\mathbf{x}^*, \delta)$ ，则Shamanskii方法 超线性收敛到 \mathbf{x}^* ，收敛阶为 $m+1$ 且

$$\|\mathbf{e}_{k+1}\| \leq K_S \|\mathbf{e}_k\|^{m+1}.$$

阻尼Newton法

在使用Newton法求解的时候要求 $F'(\mathbf{x}^{(k)})$ 非奇异，所以当 $F'(\mathbf{x}^{(k)})$ 奇异或严重病态时迭代法无法进行下去。为了克服这一困难，可在原来的Newton法中引入参数 λ_k 将迭代序列改为

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \left[F'(\mathbf{x}^{(k)}) + \lambda_k I \right]^{-1} F(\mathbf{x}^{(k)}), \quad k = 0, 1, \dots$$

λ_k 称为阻尼参数，上面的方法称为阻尼Newton法。适当选取 λ_k 可以使 $F'(\mathbf{x}^{(k)}) + \lambda_k I$ 非奇异，非病态，从而使迭代序列 $\{\mathbf{x}^{(k)}\}_0^\infty$ 有意义。此时如果 $\mathbf{x}^{(k)} \rightarrow \mathbf{x}^*$ ，只有线性收敛速度。

注记

迭代法的计算量和迭代次数以及每次迭代的计算量有关，所以收敛快的方法不一定是最好的算法。

离散牛顿法

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - J(\mathbf{x}^{(k)}, \mathbf{h}^{(k)})^{-1} F(\mathbf{x}^{(k)}), \quad k = 0, 1, \dots$$

其中

$$J(\mathbf{x}^{(k)}, \mathbf{h}^{(k)}) = \left(\frac{1}{h_1^{(k)}} [F(\mathbf{x}^{(k)} + h_1^{(k)} \mathbf{e}_1)], \dots, \right. \\ \left. \frac{1}{h_1^{(n)}} [F(\mathbf{x}^{(k)} + h_1^{(n)} \mathbf{e}_n)] \right)$$

若取 $\mathbf{h}^{(k)} = F(\mathbf{x}^{(k)})$ ，则其为Newton-Steffensen方法。它的收敛速度和Newton法一样是2阶的。

但是在数值微分的时候，条件数是和步长 \mathbf{h} 成反比的，所以 $\mathbf{h}^{(k)}$ 不能太小，否则误差会很大。

拟 Newton 法

牛顿法在理论上和实践中均取得很好的效果。然而对于大规模问题，函数的雅克比矩阵计算代价特别大或者难以得到，即便得到雅克比矩阵我们还需 要求解一个大规模线性方程组。那么能否使用雅克比矩阵或其逆矩阵的近似来进行牛顿迭代呢？

拟牛顿法便是这样的算法，它能够在每一步以较小的计算代价生成近似矩阵，并且使用近似矩阵代替雅克比矩阵而产生的迭代序列仍具有超线性收敛的性质。

拟牛顿方法不计算雅克比矩阵 $\nabla F(\mathbf{x}^{(k)})$ ，而是构造其近似矩阵 B_k 或其逆的近似矩阵 H_k 。我们希望 B_k 或 H_k 仍然保留原矩阵的部分性质。那么拟牛顿矩阵应该满足一些什么性质？如何构造它们呢？

割线方程

设 $F(\mathbf{x})$ 在 $\mathbf{x}^{(k+1)}$ 处的近似为

$$F(\mathbf{x}) = F(\mathbf{x}^{(k+1)}) + \nabla F(\mathbf{x}^{(k+1)})(\mathbf{x} - \mathbf{x}^{(k+1)}) + \mathcal{O}(\|\mathbf{x} - \mathbf{x}^{(k+1)}\|^2).$$

令 $\mathbf{x} = \mathbf{x}^{(k)}$, $\mathbf{s}^k = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ 以及 $\mathbf{y}^k = F(\mathbf{x}^{(k+1)}) - F(\mathbf{x}^{(k)})$, 可得

$$\nabla F(\mathbf{x}^{(k+1)})\mathbf{s}^k + \mathcal{O}(\|\mathbf{s}^k\|^2) = \mathbf{y}^k.$$

忽略高阶项 $\|\mathbf{s}^k\|^2$, 希望近似矩阵 B_{k+1} 满足方程

$$\mathbf{y}^k = B_{k+1}\mathbf{s}^k,$$

或者其逆的近似矩阵 H_{k+1} 满足方程

$$\mathbf{s}^k = H_{k+1}\mathbf{y}^k.$$

上述两个方程称为割线方程。

曲率条件

在求解优化问题时, 近似矩阵 B_k 的正定性是个很关键的因素, 在割线方程的两边同时左乘 $(s^k)^T$ 可得

$$(s^k)^T B_{k+1} s^k = (s^k)^T y^k,$$

因此条件

$$(s^k)^T y^k > 0,$$

为 B_{k+1} 正定的一个必要条件。很多时候额外要求这个条件在迭代过程中始终满足, 这个条件也称为曲率条件。

在通常情况下, 近似矩阵 B_{k+1} 或 H_{k+1} 是由上一步迭代加上一个修正得到的, 并且要求满足割线方程。

算法 (拟牛顿法算法框架)

- 1 给定 $\mathbf{x}^{(0)} \in \mathbb{R}^n$, 初始矩阵 $B_0 \in \mathbb{R}^{n \times n}$ (或 H_0), 令 $k = 0$.
- 2 **while** 未达到停机准则 **do**
- 3 计算方向 $\mathbf{d}^{(k)} = -B_k^{-1}F(\mathbf{x}^{(k)})$ 或 $\mathbf{d}^{(k)} = -H_k F(\mathbf{x}^{(k)})$.
- 4 通过线搜索找到合适的步长 $\alpha_k > 0$, 令 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$.
- 5 更新近似矩阵 B_{k+1} 或其逆矩阵的近似 H_{k+1} .
- 6 $k \leftarrow k + 1$.
- 7 **end while**

在实际应用中基于 H_k 的拟牛顿法更加实用, 这是因为根据 H_k 计算下降方向 $\mathbf{d}^{(k)}$ 不需要求解线性方程组, 而求解线性方程组在大规模问题上是非常耗时的. 但基于 B_k 的拟牛顿法有比较好的理论性质, 产生的迭代序列比较稳定. 如果有办法快速求解线性方程组, 我们也可采用基于 B_k 的拟牛顿法.

秩一更新 (SR1) 公式是结构最简单的拟牛顿矩阵更新公式。设 B_k 是第 k 步的近似矩阵, 希望通过对 B_k 进行秩一修正得到 B_{k+1} 使其满足割线方程。设

$$B_{k+1} = B_k + \mathbf{u}_k \mathbf{v}_k^T, \quad \mathbf{u}_k, \mathbf{v}_k \text{ 未知.}$$

其满足割线方程

$$B_{k+1} \mathbf{s}^k = \mathbf{y}^k \implies (B_k + \mathbf{u}_k \mathbf{v}_k^T) \mathbf{s}^k = \mathbf{y}^k,$$

亦即

$$\mathbf{u}_k \mathbf{v}_k^T \mathbf{s}^k = \mathbf{y}^k - B_k \mathbf{s}^k$$

因为 $\mathbf{v}_k^T \mathbf{s}^k \in \mathbb{R}$, 因此 \mathbf{u}_k 与 $\mathbf{y}^k - B_k \mathbf{s}^k$ 方向相同, 可得

$$\mathbf{u}_k = \mathbf{y}^k - B_k \mathbf{s}^k / \mathbf{v}_k^T \mathbf{s}^k,$$

所以

$$B_{k+1} = B_k + (\mathbf{y}^k - B_k \mathbf{s}^k) \frac{\mathbf{v}_k^T}{\mathbf{v}_k^T \mathbf{s}^k}$$

Broyden方法

若取 $\mathbf{v}_k = \mathbf{s}^k$, 得Broyden秩1方法

$$\begin{cases} \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - B_k^{-1} F(\mathbf{x}^{(k)}), & k = 0, 1, \dots \\ B_{k+1} = B_k + \frac{(\mathbf{y}^k - B_k \mathbf{s}^k)(\mathbf{s}^k)^T}{(\mathbf{s}^k)^T \mathbf{s}^k}. \end{cases}$$

若取 $\mathbf{v}_k = F(\mathbf{x}^{(k+1)})$, 则

$$\begin{cases} \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - B_k^{-1} F(\mathbf{x}^{(k)}), & k = 0, 1, \dots \\ B_{k+1} = B_k + \frac{(\mathbf{y}^k - B_k \mathbf{s}^k)(\mathbf{y}^k - B_k \mathbf{s}^k)^T}{(\mathbf{y}^k - B_k \mathbf{s}^k)^T \mathbf{s}^k}. \end{cases}$$

称为Broyden对称秩1方法。秩一方法结构简单, 但是不能保证矩阵在迭代过程中保持正定。

定理 (Sherman-Morrison 公式)

设 $A \in \mathbb{R}^{n \times n}$, 且 A 非奇异, $u, v \in \mathbb{R}^n$, 当且仅当 $1 + v^T A^{-1} u \neq 0$ 时, $A + uv^T$ 可逆, 且有

$$(A + uv^T)^{-1} = \left(I - \frac{A^{-1}uv^T}{1 + v^T A^{-1}u} \right) A^{-1}$$

逆Broyden方法

可以避免求逆得

- 逆Broyden秩1方法

$$\begin{cases} \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - H_k F(\mathbf{x}^{(k)}), & k = 0, 1, \dots \\ H_{k+1} = H_k + \frac{(\mathbf{s}^k - H_k \mathbf{y}^k)(\mathbf{s}^k)^T H_k}{(\mathbf{s}^k)^T H_k \mathbf{y}^k}. \end{cases}$$

- 逆Broyden对称秩1方法

$$\begin{cases} \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - H_k F(\mathbf{x}^{(k)}), & k = 0, 1, \dots \\ H_{k+1} = H_k + \frac{(\mathbf{s}^k - H_k \mathbf{y}^k)(\mathbf{s}^k - H_k \mathbf{y}^k)^T}{(\mathbf{s}^k - H_k \mathbf{y}^k)^T \mathbf{y}^k}. \end{cases}$$

C. T. Kelley, Iterative Methods for Linear and Nonlinear Equations, Society for Industrial and Applied Mathematics(SIAM), Philadelphia 1995.

BFGS公式

为克服秩一方法的缺陷, 可以考虑 B_k 的秩二更新, 类似设

$$B_{k+1} = B_k + a u u^T + b v v^T, \quad u, v \in \mathbb{R}^n, a, b \in \mathbb{R} \text{ 待定.}$$

根据割线方程有

$$B_{k+1} s^k = (B_k + a u u^T + b v v^T) s^k = y^k,$$

整理可得

$$(a \cdot u^T s^k) u + (b \cdot v^T s^k) v = y^k - B_k s^k$$

通过选取适当的 u 和 v 使得上式成立即可。一种最直接的取法就是让上面等式左右两边的两项分别对应相等, 即

$$\begin{aligned} u &= y^k, & a \cdot u^T s^k &= 1, \\ v &= B_k s^k, & b \cdot v^T s^k &= -1. \end{aligned}$$

由此得到更新方式

$$B_{k+1} = B_k + \frac{\mathbf{y}^k(\mathbf{y}^k)^T}{(\mathbf{s}^k)^T \mathbf{y}^k} - \frac{B_k \mathbf{s}^k (B_k \mathbf{s}^k)^T}{(\mathbf{s}^k)^T B_k \mathbf{s}^k}.$$

这就是基于 B_k 的 BFGS 公式, 是由 Broyden, Fletcher, Goldfarb, Shanno 四人名字的首字母组成。

根据 Sherman-Morrison 公式并假设 $H_k = B_k^{-1}$, 可得基于 H_k 的 BFGS 公式:

$$H_{k+1} = (I - \rho_k \mathbf{s}^k (\mathbf{y}^k)^T)^T H_k (I - \rho_k \mathbf{s}^k (\mathbf{y}^k)^T) + \rho_k \mathbf{s}^k (\mathbf{s}^k)^T,$$

其中 $\rho_k = 1/(\mathbf{s}^k)^T \mathbf{y}^k$ 。

若要 BFGS 公式更新产生的矩阵 H_{k+1} 正定, 一个充分条件是:

$$(\mathbf{s}^k)^T \mathbf{y}^k > 0, \quad \text{且 } H_k \text{ 正定}.$$

BFGS 公式是目前最有效的拟牛顿更新格式之一, 它有比较好的理论性质, 实现起来也并不复杂。

Broyden, Fletcher, Goldfarb, Shanno



什么是混沌？

- 差之毫厘，失之千里
- 解对初值的敏感性
 - 经典理论：微小的误差不会影响事物前进的方向。
 - 混沌：初始条件的小扰动，会对解产生致命的影响。
- 蝴蝶效应(Butterfly Effect)
 - 南美洲的蝴蝶煽动翅膀，会在中国南海诱发一场风暴吗？
 - 洛伦兹(Edward N. Lorenz)的气象小玩具。
- 不规则之中仍存在秩序
 - 天气看似无法准确预测，但又呈现年复一年相当程度的规律性。
 - 可以对短期的行为作出有效的预测。

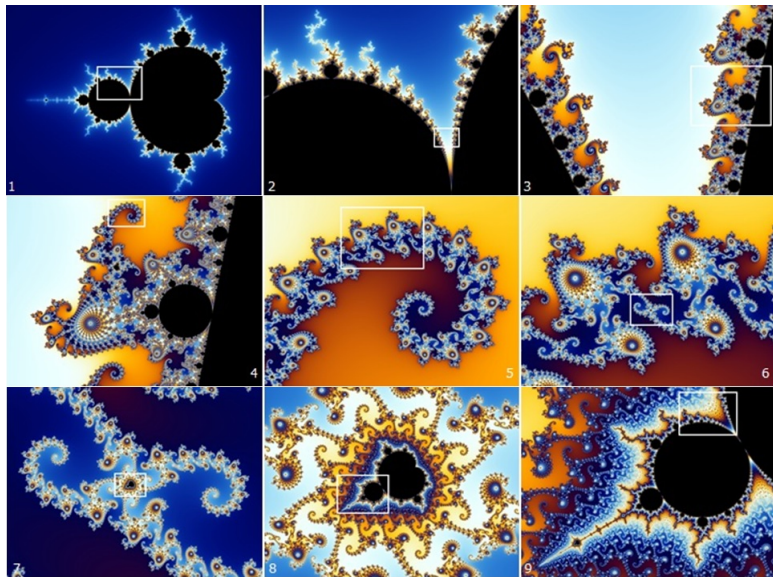
分形：混沌世界的秩序

分形结构：由不断的图形迭代生成

利用简单的规则让系统复杂；从复杂不可解的系统中找到简单美妙的秩序。



Mandelbrot 集合: $z_{n+1} = z_n^2 + c, c \in \mathbb{C}$



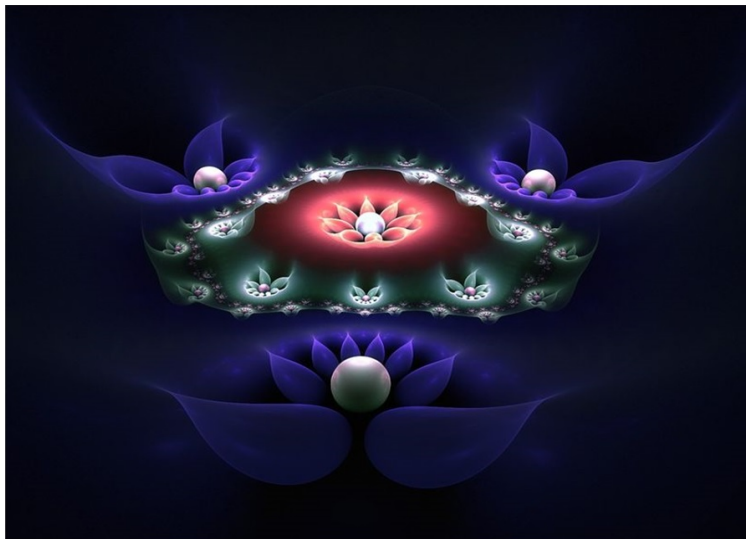
分形图



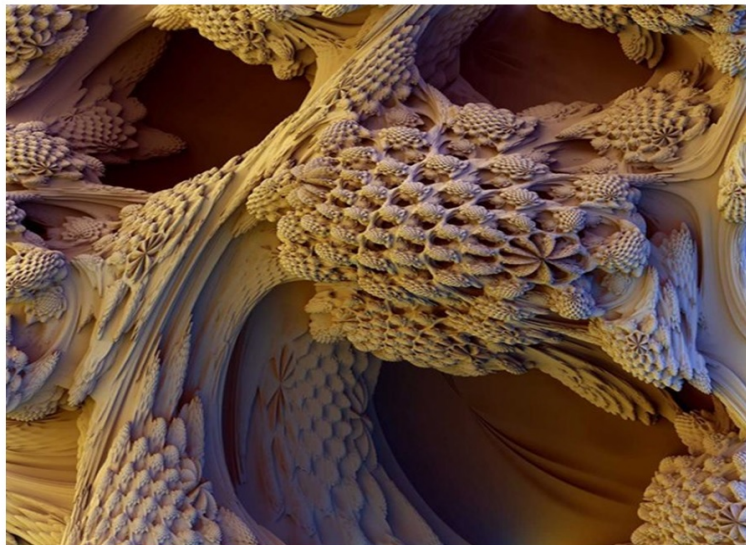
分形图



分形图



分形图



花椰菜

