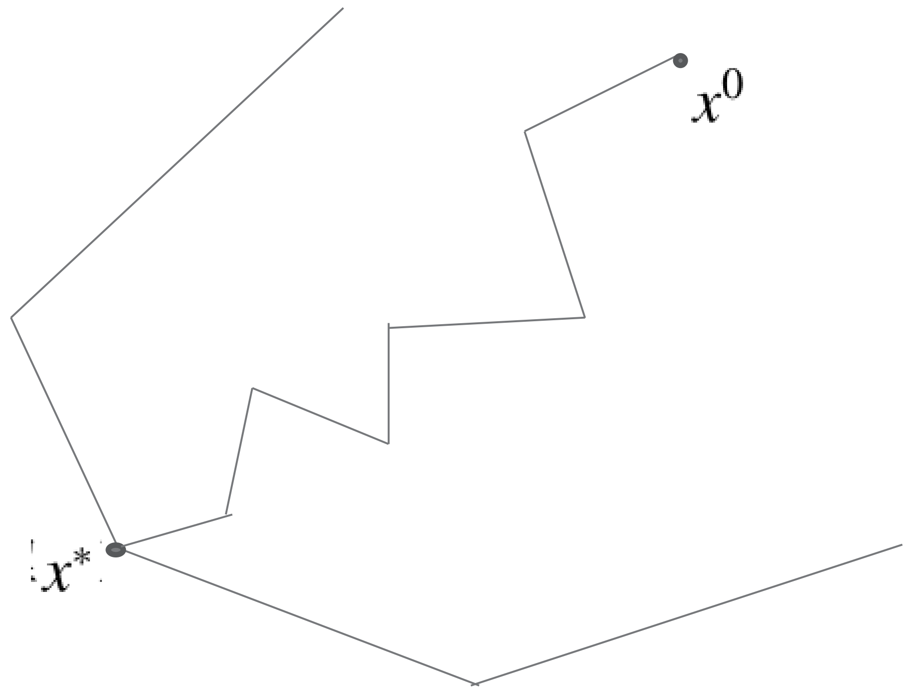
1.4 线搜索迭代算法概述 — 迭代算法

求解最优化问题(1.1.1)的算法多种多样，其中迭代算法是最主要的方法，其基本思想为：起始于某个点$x^0 \in \mathbb{R}^n$，按照某种规则依次产生一个迭代点列$\{x^k\}$，直到算法终止.

$$\min_{x \in \mathscr{F}} f(x)$$



最优解

1、可行性

2、最优性不等式

1.4 线搜索迭代算法概述 — 迭代算法

按照迭代点是否可行, 迭代算法可分为:

- **可行算法**: 若所有迭代点x^k $(k \in \{0, 1, 2, \ldots\})$均是问题(1.1.1)的可行点, 那么称该算法为可行算法;

- **不可行算法**: 若至少有一个迭代点x^k不是问题(1.1.1)的可行点, 那么称该算法为不可行算法.

按照目标函数值的下降特性, 常遇到以下两类算法:

- **单调下降算法**: 如果对任意的$k \in \{0, 1, 2, \ldots\}$恒有$f(x^{k+1}) \leq f(x^k)$, 那么称该算法为单调下降算法; 如果不等号是严格成立的, 那么称该算法为严格单调下降算法;

- **非单调下降算法**: 如果对于某个$k \in \{0, 1, 2, \ldots\}$有$f(x^{k+1}) > f(x^k)$, 但存在某个正整数$l_k$ 使得$f(x^{k+l_k}) < f(x^k)$, 那么称该算法为非单调下降算法.

1.4 线搜索迭代算法概述 — 线搜索迭代算法

在迭代算法中，线搜索迭代算法是最流行的迭代算法之一，其基本框架为

算法 1.4.1 (线搜索迭代算法的一般框架)

步1 选择一个起始点$x^0 \in \mathbb{R}^n$. 置$k := 0$.

步2 判断当前迭代点x^k是否满足终止条件.

步3 从当前点出发，选择沿什么方向(即迭代方向，不妨记为d^k)进行迭代，以及沿该方向走多远(即迭代步长，不妨记为λ_k)，确定下一迭代点$x^{k+1} := x^k + \lambda_k d^k$.

步4 置$k := k + 1$，转步2.

1.4 线搜索迭代算法概述 —线搜索迭代算法

在算法1.4.1中:

步1是关于算法的启动问题. 对于不可行算法, 任取一点即可; 而对于可行算法, 此步需要找问题的一个可行点, 这并不容易.

步2是关于算法的终止问题. 常用的终止规则有以下几类:

(a) $\|x^{k+1} - x^k\| < \epsilon$;

(b) $|f(x^{k+1}) - f(x^k)| < \epsilon$;

(c) $\dfrac{\|x^{k+1}-x^k\|}{\|x^k\|} < \epsilon$;

(d) $\dfrac{|f(x^{k+1})-f(x^k)|}{|f(x^k)|} < \epsilon$;

(e) $\|\nabla f(x^k)\| \le \epsilon$.

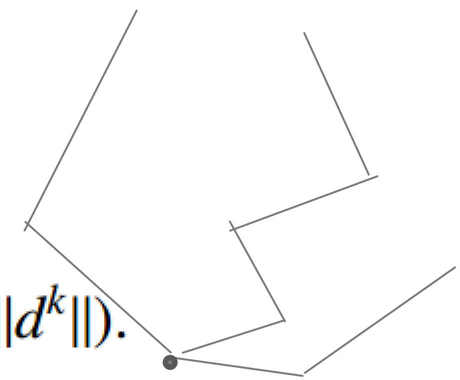
步3是线搜索迭代算法的核心, 将在以下分别介绍.

1.4 线搜索迭代算法概述 — 迭代方向

定义 1.4.1 在点x^k处, 对于向量$d^k \in \mathbb{R}^n \setminus \{0\}$, 若存在$\bar{\lambda} > 0$, 使得对任意的$\lambda \in (0, \bar{\lambda})$都有$f(x^k + \lambda d^k) < f(x^k)$, 则称$d^k$为函数$f$在$x^k$处的一个下降方向.

若函数f是可微的, 则下面的命题给出下降方向的一个判别条件.

命题 1.4.1 假设函数f一阶连续可微, 那么d^k为f在x^k处的下降方向当且仅当$\nabla f(x^k)^\top d^k < 0$.



证明 根据一阶Taylor展开式，有

$$f(x^k + \lambda d^k) = f(x^k) + \lambda \nabla f(x^k)^\top d^k + o(\lambda \|d^k\|).$$

结合定义1.4.1和上式, 命题的结论易证. □

如果$\nabla f(x^k) \neq 0$, 那么显然$-\nabla f(x^k)$是f在x^k处的一个下降方向.

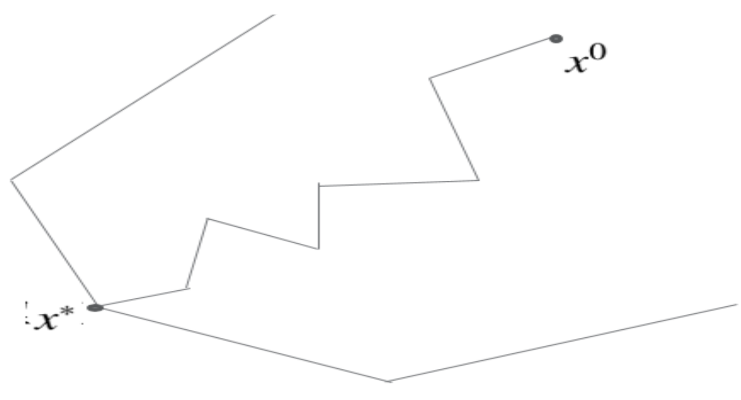
1.4 线搜索迭代算法概述 — 迭代方向

定义 1.4.2 给定非空集合$\mathscr{F} \subseteq \mathbb{R}^n$和点$x^k \in \mathscr{F}$. 对于向量$d^k \in \mathbb{R}^n \setminus \{0\}$, 若存在$\bar{\lambda} > 0$, 使得对任意的$\lambda \in (0, \bar{\lambda})$都有$x^k + \lambda d^k \in \mathscr{F}$, 则称$d^k$为点$x^k$处关于$\mathscr{F}$的一个可行方向.

对于$\mathscr{F}$的内点来说, 显然任意的非零向量都是其可行方向; 而对于$\mathscr{F}$的边界点来说, 指向$\mathscr{F}$内部的方向是其可行方向, 否则不是可行方向.

结合定义1.4.1和定义1.4.2, 可知如果d^k为点x^k处的可行下降方向, 那么总存在一个λ_k使得

$$x^k + \lambda_k d^k \in \mathscr{F} \quad 且 \quad f(x^k + \lambda_k d^k) < f(x^k).$$



1.4 线搜索迭代算法概述 — 迭代步长

精确一维线搜索 假设在迭代点x^k处已得到函数f的下降方向d^k, 为了得到下一迭代点并且使得f的函数值有尽可能大的下降, 可以通过求解一维最优化问题

$$\lambda_* = \arg\min_{\lambda \geq 0} f(x^k + \lambda d^k) \tag{1.4.1}$$

来获得迭代步长$\lambda_k := \lambda_*$, 进而得到下一迭代点$x^{k+1} = x^k + \lambda_k d^k$. 通过求解问题(1.4.1) 获得迭代步长的方法也可以写为

$$f(x^k + \lambda_k d^k) = \min_{\lambda \geq 0} f(x^k + \lambda d^k). \tag{1.4.2}$$

称(1.4.2)式为精确一维线搜索公式.

实现精确一维线搜索的方法有很多, 包括Fibonacci法、黄金分割法、进退法、平分法、抛物线法等等. 黄金分割法又称为0.618法.

1.4 线搜索迭代算法概述 — 迭代步长

非精确一维线搜索 精确一维线搜索方法往往需要花费很大的工作量; 而精确一维线搜索的目的是沿迭代方向d^k寻找一个迭代步长使得f的函数值有尽可能大的下降. 因此, 在实际中, 为了降低计算费用, 只需要沿迭代方向d^k寻找一个迭代步长使得f的函数值有一个满意的下降量即可. 这样的线搜索被称为非精确一维线搜索或近似一维线搜索.

Goldstein型线搜索准则. Goldstein于1965年提出了以下的线搜索准则: 选取$\lambda_k > 0$使得

$$
\begin{aligned}
f(x^k + \lambda_k d^k) - f(x^k) &\le \sigma \lambda_k \nabla f(x^k)^\top d^k, \\
f(x^k + \lambda_k d^k) - f(x^k) &\ge (1 - \sigma) \lambda_k \nabla f(x^k)^\top d^k,
\end{aligned}
$$

其中$\sigma \in (0, 1/2)$.

1.4 线搜索迭代算法概述 — 迭代步长

Armijo型线搜索准则. Armijo于1966年提出了以下的线搜索准则: 选取$\lambda_k := \rho\gamma^{m_k}$使得$m_k$ 为满足下式的最小非负整数:

$$f(x^k + \rho\gamma^{m_k} d^k) - f(x^k) \le \sigma\rho\gamma^{m_k} \nabla f(x^k)^\top d^k,$$

其中$\rho > 0, \sigma, \gamma \in (0, 1)$.

Wolfe型线搜索准则. Wolfe于1969年提出了以下的线搜索准则: 选取$\lambda_k > 0$使得

$$f(x^k + \lambda_k d^k) - f(x^k) \le \sigma\lambda_k \nabla f(x^k)^\top d^k,$$

$$\nabla f(x^k + \lambda_k d^k)^\top d^k \ge \delta \nabla f(x^k)^\top d^k,$$

其中$\sigma \in (0, 1/2), \delta \in (\sigma, 1)$. Powell于1976年也提出这一线搜索准则, 因此, Wolfe型线搜索准则也称为Wolfe-Powell型线搜索准则.

1.4 线搜索迭代算法概述 — 迭代步长

非单调一维线搜索 求解一个最优化问题, 其最终的目的是使得目标函数在可行域上的函数值下降到不能再下降为止. 最优化问题多种多样, 单调下降算法并不是对求解所有的优化问题都有效. 在算法迭代过程中, 允许目标函数值有所回升但保持总的趋势是下降的, 很多时候会得到更好的计算效果, 之所谓"拳头缩回来再打出去更有力量", 这样的想法导致了非单调下降算法的出现. 在非单调下降算法中使用的线搜索被称为非单调一维线搜索. 众所周知: 对于所涉及函数高度非凸以及在最优解附近有"山谷"的优化问题, 非单调下降算法常常能提高找到全局最优解的可能性和改进算法的收敛速度. 非单调下降算法已经广为研究. 现如今已经提出了很多非单调一维线搜索.

下面介绍几个非单调一维线搜索准则.

1.4 线搜索迭代算法概述 — 迭代步长

基于前若干步迭代点函数值极大的非单调线搜索准则. Grippo, Lampariello和Lucidi (见文献[14]) 最早提出了一个非单调一维线搜索准则: 寻找步长$\lambda_k = \rho\gamma^{h_k}$使得$h_k$是满足下式的最小非负整数:

$$f(x^k + \rho\gamma^{h_k} d^k) \le \max_{0\le j\le m_k} f(x^{k-j}) + \sigma\rho\gamma^{h_k} \nabla f(x^k)^\top d^k, \qquad (1.4.3)$$

其中, $m_0 = 0$且对于$k \ge 1$有: m_k是一个正整数且满足$0 \le m_k \le \min\{m_{k-1}, M\}$, ρ是一个给定的正实数, σ和γ是满足$\sigma, \gamma \in (0,1)$的两个常数. 当$f(x^k) = \max_{0\le j\le m_k} f(x^{k-j})$时, (1.4.3)式归结为单调的Armijo型线搜索公式.

1.4 线搜索迭代算法概述 — 迭代步长

基于之前所有步迭代点函数值凸组合的非单调线搜索准则. 2004年, Zhang和Hager (见文献[24]) 提出了如下的非单调线搜索准则: 寻找步长$\lambda_k > 0$使得以下两式成立:

$$f(x^k + \lambda_k d^k) \leq C_k + \delta\lambda_k \nabla f(x^k)^\top d^k, \tag{1.4.4}$$

$$\nabla f(x^k + \lambda_k d^k)^\top d^k \geq \sigma \nabla f(x^k)^\top d^k, \tag{1.4.5}$$

其中, δ和σ是满足$0 < \delta < \sigma < 1$的两个常数且C_k选取方式如下:

$$Q_{k+1} = \eta_k Q_k + 1 \quad \text{且} \quad C_{k+1} = (\eta_k Q_k C_k + f(x^{k+1}))/Q_{k+1},$$

其中$C_0 = f(x^0)$, $Q_0 = 1$, $\eta_k \in [\eta_{\min}, \eta_{\max}]$且$0 \leq \eta_{\min} \leq \eta_{\max} \leq 1$.

容易看到: C_{k+1}是C_k和$f(x^{k+1})$的凸组合. 由于$C_0 = f(x^0)$, 所以C_k是函数值$f(x^0), f(x^1), \ldots, f(x^k)$的凸组合. 因此, 在Zhang-Hager的方法中, C_k含有第k步迭代之前所有迭代点对应函数值的信息. η_k的选取目的是控制“非单调性”的度. 如果对所有的k有$\eta_k = 0$, 那么Zhang-Hager的非单调线搜索可归结为单调的Wolfe线搜索.

1.4 线搜索迭代算法概述 — 迭代步长

基于前若干步迭代点函数值凸组合的非单调线搜索准则. 2010年, Hu, Huang和Lu (见文献[15]) 提出了如下的非单调线搜索准则: 寻找步长$\lambda_k > 0$使得(1.4.4)式和(1.4.5)式成立, 其中δ, σ是满足$0 < \delta < \sigma < 1$的常数且C_k按如下方式选取:

$$Q_k = 1 + \eta_k \sum_{i=1}^{m_k-1} \eta_{k-i} \quad 且 \quad C_k = \frac{\eta_k \sum_{i=1}^{m_k-1} \eta_{k-i} f(x^{k-i}) + f(x^k)}{Q_k},$$

其中, $\eta_k \in [\eta_{\min}, \eta_{\max}]$且满足

$$0 \leq \eta_{\min} \leq \eta_{\max} \leq 1 \quad 和 \quad f(x^k) - C_k \leq \frac{\delta l_k |\nabla f(x^k)^\top d^k|}{2},$$

这里, 如果∇f是Lipschitz连续的且具有Lipschitz常数$\mathcal{L}$, 那么可以置$l_k := \frac{(1-\sigma)|\nabla f(x_k)^\top d_k|}{\mathcal{L}\|d_k\|^2}$; 否则置$l_k := 0$.

在此方法中, C_k是目前迭代点的函数值和之前若干个迭代点对应函数值的凸组合. 这一方法综合使用了Grippo-Lampariello-Lucidi的非单调线搜索准则和Zhang-Hager的非单调线搜索准则的思想. 一个重要特例是: 如果$\sum_{i=1}^{m_k-1} \eta_{k-i} f(x^{k-i}) \geq \sum_{i=1}^{m_k-1} \eta_{k-i} f(x^k)$, 那么选择$\eta_k \in (0, 1]$; 否则选择$\eta_k = 0$, 于是在每步迭代中, 或者使用单调线搜索(即: $\eta_k = 0$时); 或者使用非单调线搜索(即: $\eta_k \neq 0$时). 因此, 它可以被看成一个单调线搜索和非单调线搜索的混合线搜索方法.

1.4 线搜索迭代算法概述 — 算法收敛性

一个算法设计的好坏, 从理论上有一些判别标准. 算法的适定性(well-defined)和算法的收敛性(convergence)是一个算法的基本要求, 收敛率(rate of convergence)是衡量一个算法是否快速有效的重要方面.

适定性 设计一个优化算法, 首先要保证每一步都是适定的, 使算法能执行下去. 例如: 如果算法中有线搜索, 那么要保证线搜索有限终止; 如果算法中需要求解线性方程组, 那么要保证线性方程组可解等等. 如果一个优化算法的每一步都是适定的, 那么称这个算法是适定的.

1.4 线搜索迭代算法概述 — 算法收敛性

全局收敛性 假设一个优化算法是适定的并且产生迭代序列为$\{x^k\}$. 如果迭代序列$\{x^k\}$能在有限步内得到优化问题的最优解, 那么称该算法是有限终止的. 如果一个优化算法是有限终止于优化问题的最优解或所产生迭代序列$\{x^k\}$的每个聚点都是优化问题的最优解, 那么称该算法是收敛的. 显然收敛的算法才有意义. 一些算法是否收敛, 与初始迭代点的选取有关. 如果一个优化算法只有当初始迭代点充分靠近最优解时才收敛, 那么称该算法是局部收敛性的. 如果一个优化算法不要求初始迭代点充分靠近最优解(初始迭代点在一定程度上可以任意选取), 那么称这样的收敛算法是全局收敛性的. 由于一般情况下最优解是事先未知的, 因此具有全局收敛性的算法更具有实际的应用价值.

1.4 线搜索迭代算法概述 — 算法收敛性

收敛率 作为一个好算法, 还必须要求算法以较快的速度收敛到问题的一个最优解. 以下假设算法产生一个无穷迭代序列$\{x^k\}$并且x^*是$\{x^k\}$的一个聚点. 不妨假设$\{x^k\}$收敛到x^*.

- 全局Q-线性收敛性. 如果存在$\beta \in (0,1)$ 使得

  $$\|x^{k+1} - x^*\| \le \beta \|x^k - x^*\|, \quad \forall k \in \{0, 1, 2, \ldots\}$$

  那么称该算法是全局Q-线性收敛性的.

- 局部Q-收敛率. 如果存在$\beta > 0$和常数$\xi \ge 0$使得

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

  那么称该算法是局部Q-β阶收敛性的. 特别地,

  (a) 若$\beta = 1, \xi \in (0,1)$, 则称该算法是局部Q-线性收敛性的;

  (b) 若$\beta \in (1,2), \xi > 0$或$\beta = 1, \xi = 0$, 则称该算法是局部Q-超线性收敛性的;

  (c) 若$\beta = 2, \xi > 0$, 则称该算法是局部Q-二阶收敛性的.

1.4 线搜索迭代算法概述 — 小结与作业

小结：

本节介绍了线搜索迭代算法的一般框架和算法收敛性的概念。前者重点掌握迭代方向的概念以及精确与非精确一维线搜索的准则；后者重点掌握适定性、全局收敛性及收敛率的概念。

作业：

1.17 设方向d是函数f在点x处的一个下降方向. 记

$$M = I - \frac{dd^\top}{d^\top \nabla f(x)} - \frac{\nabla f(x) \nabla f(x)^\top}{\nabla f(x)^\top \nabla f(x)}.$$

试证: $d_M := -M\nabla f(x)$也是函数f在点x处的一个下降方向.

1.21 证明: 如果序列$\{x^k\} \subseteq \mathbb{R}^n$是$Q$-超线性收敛到$x^*$的, 那么

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

但反之不一定成立.

【提示: 反之部分可考虑数列$\{a^k\} \subseteq \mathbb{R}$, 其中, 对任意的$k \in \{1, 2, \cdots\}$有$a^{2k-1} = \frac{1}{k!}$和$a^{2k} = 2a^{2k-1}$.】