基于Wolfe-Powell非精确一维搜索的拟牛顿算法

大数据学院 和泳毅 PB19010450

一维搜索过程是最优化方法的基本组成部分，精确一维搜索方法往往需要花费很大的工作量，特别是当迭代点远离问题的解时，精确地求解一个一维子问题通常不是十分有效的。实际上，很多最优化方法如牛顿法和拟牛顿法，其收敛速度并不依赖于精确一维搜索过程。因此，我们只需要保证目标函数$f(\mathbf{x})$在每一步都有一定的下降量，同时保证步长不太小，这样就可以大大节省工作量。

在牛顿法中，面临的主要困难一是牛顿法是局部收敛的，当初始点选择不当时，往往导致不收敛；二是当Hesse矩阵G_k不正定时，不能保证产生的方向是下降方向。但是牛顿法二阶收敛是非常好的性质，为此人们在克服上述缺点方面做了很多工作，提出了许多拟牛顿法。比如用对称秩二校正公式推出的**DFP**法与**BFGS**法。

本文将使用基于**Wolfe-Powell**非精确一维搜索的拟牛顿法来解决无约束优化问题。

一、问题描述

给定目标函数$f(\mathbf{x})$，求解以下无约束优化问题

$$\min_{\mathbf{x}} f(\mathbf{x}) \tag{1}$$

并给出以下三个目标函数供测试：

Rosenbrock函数：

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

Beale函数：

$$f(\mathbf{x}) = (1.5 - x_1 + x_1x_2)^2 + (2.25 - x_1 + x_1x_2^2)^2 + (2.625 - x_1 + x_1x_2^3)^2$$

McCormick函数：

$$f(\mathbf{x}) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$$

二、算法原理

（一）最速下降法

定义1　(1) 设$\bar{\mathbf{x}} \in S$，$\mathbf{d} \in \mathbb{R}^n$是非零向量。若存在$\delta > 0$，使得

$$\bar{\mathbf{x}} + \lambda \mathbf{d} \in S, \forall \lambda \in (0, \delta)$$

则称$\mathbf{d}$是S在$\bar{\mathbf{x}}$处的可行方向。

(2) 设$f(\mathbf{x})$是$\mathbb{R}^n$上的实函数，$\bar{\mathbf{x}} \in \mathbb{R}^n$，$\mathbf{d}$是非零向量。若存在$\delta > 0$，使得

$$f(\bar{\mathbf{x}} + \lambda \mathbf{d}) < f(\bar{\mathbf{x}}), \forall \lambda \in (0, \delta)$$

则称$\mathbf{d}$为函数$f(\mathbf{x})$在$\bar{\mathbf{x}}$处的下降方向。

注意到，满足$f(\bar{\mathbf{x}})^T \mathbf{d} < 0$的$\mathbf{d}$为$f(\mathbf{x})$在$\bar{\mathbf{x}}$处的下降方向。

考虑问题(1)，一个自然的想法是产生点列$\{\mathbf{x}^{(k)}\}$的函数列$\{f(\mathbf{x}^{(k)})\}$为严格单调下降数列，即对$\forall k$，$f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$。

对$f(\mathbf{x})$做Taylor展开

$$f(\mathbf{x}) = f(\mathbf{x}^{(k)}) + \bigtriangledown f(\mathbf{x}^{(k)})^T (\mathbf{x} - \mathbf{x}^{(k)}) + O(\|\mathbf{x} - \mathbf{x}^{(k)}\|^2).$$

取负梯度方向

$$\mathbf{d}^{(k)} = - \bigtriangledown f(\mathbf{x}^{(k)}),$$

则当α_k足够小时，总有

$$f(\mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}) < f(\mathbf{x}^{(k)}).$$

于是有最速下降法：

Gradient

输入：初始点$\mathbf{x}_0$，终止误差$\varepsilon > 0$。

(1) $k := 0$。

(2) 计算：$\mathbf{g}^{(k)} := \bigtriangledown f(\mathbf{x}^{(k)})$，如果$\|\mathbf{g}^{(k)}\| < \varepsilon$，停止迭代。

(3) 令搜索方向$\mathbf{d}^{(k)} = -\mathbf{g}^{(k)}$，并由一维搜索确定步长$\alpha_k$。

(4) 更新$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$，置$k := k+1$，返回步骤(2)。

输出：$\mathbf{x}^{(k)}$，$f(\mathbf{x}^{(k)})$。

定理1（最速下降法收敛定理）设$f \in C^1$，最速下降法中的一维搜索采用（非）精确一维搜索过程，则迭代点列$\{\mathbf{x}^{(k)}\}$的每一个聚点是驻点。

总体全局性并未能保证最速下降法是一个有效的方法，因为最速下降方向仅是算法的局部性质，对于许多问题，最速下降法并非“最速下降”，而是下降非常缓慢。

（二）牛顿法

牛顿法的基本思想是利用二次函数近似目标函数，把这个二次函数的极小点作为新的迭代点。

设$f(\mathbf{x})$二次连续可微，求解问题(1),在$\mathbf{x}^{(k)}$处做Taylor展开取前三项，得到

$$q^{(k)}(\mathbf{s}) = f(\mathbf{x}^{(k)}) + \mathbf{g}^{(k)^T}\mathbf{s} + \frac{1}{2}\mathbf{s}^T G_k \mathbf{s},$$

其中$\mathbf{s} = \mathbf{x} - \mathbf{x}^{(k)}$，$\mathbf{g}^{(k)} = \bigtriangledown f(\mathbf{x}^{(k)})$，$G_k = \bigtriangledown^2 f(\mathbf{x}^{(k)})$。设$G_k$正定，那么上式有唯一极小点，是$q^{(k)}(\mathbf{s})$的驻点，即使

$$\bigtriangledown q^{(k)}(\mathbf{s}) = \mathbf{g}^{(k)} + G_k\mathbf{s} = 0,$$

得到极小点

$$\mathbf{s}^{(k)} = -G_k^{-1}\mathbf{g}^{(k)}.$$

称搜索方向$\mathbf{s}^{(k)}$为牛顿方向，在实际计算中，$\mathbf{s}^{(k)}$的计算常常通过解方程

$$G_k\mathbf{s} = -\mathbf{g}^{(k)},$$

得到。以下给出经典牛顿法迭代格式：

Newton

输入：初始点$\mathbf{x}_0$，终止误差$\varepsilon > 0$。

(1) $k := 0$。

(2) 计算：$\mathbf{g}^{(k)} := \bigtriangledown f(\mathbf{x}^{(k)})$，如果$\|\mathbf{g}^{(k)}\| < \varepsilon$，停止迭代。

(3) 解线性方程组$G_k\mathbf{d} = -\mathbf{g}^{(k)}$，求出牛顿方向$\mathbf{d}^{(k)}$。

(4) 更新$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$，置$k := k+1$，返回步骤(2)。

输出：$\mathbf{x}^{(k)}$，$f(\mathbf{x}^{(k)})$。

定理2（牛顿法收敛定理）设$f \in C^2$，$\mathbf{x}^k$充分靠近$\mathbf{x}^*$，其中$\bigtriangledown f(\mathbf{x}^*) = 0$。如果$\bigtriangledown^2 f(\mathbf{x}^*)$正定，目标函数的Hesse矩阵$G(\mathbf{x})$满足**Lipschitz**条件，则对一切的$k$，牛顿迭代有定义，所得序列$\{\mathbf{x}^k\}$收敛到$\mathbf{x}^*$，且具有二阶收敛速率。

应该注意到，当初始点远离最优解时，G_k不一定正定。牛顿方向不一定是下降方向，其收敛性不能保证。这说明恒取步长为1是不合适的，应该在牛顿法中采用一维搜索来确定步长。但是应该强调，仅当步长$\{\alpha_k\}$收敛到1时，牛顿法才是二阶收敛的，于是产生了阻尼牛顿法。

（三）拟牛顿法

牛顿法的突出优点是局部收敛很快，但运用牛顿法要计算二阶导，而且目标函数的Hesse阵可能非正定，甚至奇异。为了克服这些缺点，人们提出了拟牛顿法。其基本思想是：用不含二阶导数的矩阵H_k近似牛顿法中的Hesse矩阵的逆$G(\mathbf{x}^{(k)})^{-1}$。

定 义 2 记 $\mathbf{s}^{(k)}=\mathbf{x}^{(k+1)}-\mathbf{x}^{(k)}$， $\mathbf{y}^{(k)}=\triangledown f(\mathbf{x}^{(k+1)})-\triangledown f(\mathbf{x}^{(k)})$ 则根据Taylor展开有

$$\triangledown^2 f(\mathbf{x}^{(k+1)})\mathbf{s}^{(k)}\approx\mathbf{y}^{(k)}\quad\text{or}\quad\triangledown^2 f(\mathbf{x}^{(k+1)})^{-1}\mathbf{y}^{(k)}\approx\mathbf{s}^{(k)}.$$

构造出Hesse矩阵的近似B_{k+1}，及其逆的近似H_{k+1}. 使其满足

$$B_{k+1}\mathbf{s}^{(k)}=\mathbf{y}^{(k)},\quad H_{k+1}\mathbf{y}^{(k)}=\mathbf{s}^{(k)}, \tag{2}$$

称为正割条件或拟牛顿条件。

定义3 (1)对称秩一校正(SR1)

$$H_{k+1}=H_k+a\mathbf{u}\mathbf{u}^T,(a\in\mathbb{R},\mathbf{u}\in\mathbb{R}^n). \tag{3}$$

(2)对称秩二校正(SR2)

$$H_{k+1}=H_k+a\mathbf{u}\mathbf{u}^T+b\mathbf{v}\mathbf{v}^T,(a,b\in\mathbb{R},\ \ \mathbf{u},\mathbf{v}\in\mathbb{R}^n). \tag{4}$$

由SR1可以推导出对称秩一校正迭代格式：

$$H_{k+1}=H_k+\frac{\left(\mathbf{s}^{(k)}-H_k\mathbf{y}^{(k)}\right)\left(\mathbf{s}^{(k)}-H_k\mathbf{y}^{(k)}\right)^T}{\left(\mathbf{s}^{(k)}-H_k\mathbf{y}^{(k)}\right)^T\mathbf{y}^{(k)}}. \tag{5}$$

该迭代格式具有二次终止性，对二次函数有遗传性。但不能保证H_{k+1}的正定性，仅当$\left(\mathbf{s}^{(k)}-H_k\mathbf{y}^{(k)}\right)^T\mathbf{y}^{(k)}>0$时，才能保证正定性。而这个条件往往很难保证。

引理(Sherman-Morrison定理) 设$A\in\mathbb{R}^{n\times n}$是非奇异矩阵，$\mathbf{u},\mathbf{v}\in\mathbb{R}^n$是任意向量。若$1+\mathbf{v}^TA^{-1}\mathbf{u}\neq 0$，则$A$的秩一校正$A+\mathbf{U}^T\mathbf{V}$非奇异，且其逆可以表示为

$$\left(A+\mathbf{u}\mathbf{v}^T\right)^{-1}=A^{-1}-\frac{A^{-1}\mathbf{u}\mathbf{v}^TA^{-1}}{1+\mathbf{v}^TA^{-1}\mathbf{u}}.$$

1. DFP法

由SR2以及拟牛顿条件(2)，有

$$H_{k+1}\mathbf{y}^{(k)} = H_k\mathbf{y}^{(k)} + \left(a\mathbf{u}^T\mathbf{y}^{(k)}\right)\mathbf{u} + \left(b\mathbf{v}^T\mathbf{y}^{(k)}\right)\mathbf{v} = \mathbf{s}^{(k)},$$

这里的$\mathbf{u}$和$\mathbf{v}$并不唯一确定，但有一种明确的选择是：

$$\begin{cases} \mathbf{u} = \mathbf{s}^{(k)}, & a\mathbf{u}^T\mathbf{y}^{(k)} = 1; \\ \mathbf{v} = H_k\mathbf{y}^{(k)}, & b\mathbf{v}^T\mathbf{y}^{(k)} = -1. \end{cases}$$

因此有

$$H_{k+1} = H_k + \frac{\mathbf{s}^{(k)}\mathbf{s}^{(k)^T}}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}} - \frac{H_k\mathbf{y}^{(k)}\mathbf{y}^{(k)^T}H_k}{\mathbf{y}^{(k)^T}H_k\mathbf{y}^{(k)}}. \tag{6}$$

上式被称DFP(Davidon-Fletcher-Powell)校正公式。

定理3(DFP公式的正定性) 当且仅当$\mathbf{s}^{(k)^T}\mathbf{y}^{(k)} > 0$时，DFP校正公式(6)保持正定性。

证： 式(6)满足拟牛顿条件(2)。当且仅当对某个非奇异矩阵J_{k+1}，有

$$H_{k+1} = J_{k+1}J_{k+1}^T$$

时，H_{k+1}正定。

定理的必然性显然，事实上，若H_{k+1}正定，则存在非奇异矩阵J_{k+1}使得$J_{k+1}J_{k+1}^T\mathbf{y}^{(k)} = \mathbf{s}^{(k)}$。定义$\mathbf{w}^{(k)} = J_{k+1}^T\mathbf{y}^{(k)}$，则

$$\mathbf{s}^{(k)^T}\mathbf{y}^{(k)} = \mathbf{y}^{(k)^T}J_{k+1}J_{k+1}^T\mathbf{y}^{(k)} = \mathbf{w}^{(k)^T}\mathbf{w}^{(k)} > 0$$

必要性得证。下面证明充分性。设$H_k = L_kL_k^T$是正定矩阵H_k的Cholesky分解，式(6)可以写成

$$\begin{aligned} H_{k+1} =& H_k + \left(\frac{1}{\beta}\right)^{1/2}H_k\mathbf{y}^{(k)}\mathbf{s}^{(k)^T} - \frac{1}{\mathbf{y}^{(k)^T}H_k\mathbf{y}^{(k)}}H_k\mathbf{y}^{(k)}\mathbf{y}^{(k)^T}H_k \\ &+ \left(\frac{1}{\beta}\right)^{1/2}\mathbf{s}^{(k)}\mathbf{y}^{(k)^T}H_k - \frac{H_k\mathbf{y}^{(k)}\mathbf{y}^{(k)^T}H_k}{\mathbf{y}^{(k)^T}H_k\mathbf{y}^{(k)}} + \frac{1}{\mathbf{s}^{(k)^T}\mathbf{y}^{(k)}}\mathbf{s}^{(k)}\mathbf{s}^{(k)^T} \\ &- \left(\frac{1}{\beta}\right)^{1/2}\mathbf{s}^{(k)}\mathbf{y}^{(k)^T}H_k - \left(\frac{1}{\beta}\right)^{1/2}H_k\mathbf{y}^{(k)}\mathbf{s}^{(k)^T} + \frac{H_k\mathbf{y}^{(k)}\mathbf{y}^{(k)^T}H_k}{\mathbf{y}^{(k)^T}H_k\mathbf{y}^{(k)}} \\ =& J_{k+1}J_{k+1}^T, \end{aligned}$$

其中

$$
\begin{aligned}
J_{k+1} &= L_k + \frac{\left(\mathbf{s}^{(k)} - L_k \mathbf{w}^{(k)}\right) \mathbf{w}^{(k)T}}{\mathbf{w}^{(k)T} \mathbf{w}^{(k)}}, \\
w_k &= \left(\frac{\mathbf{s}^{(k)T} \mathbf{y}^{(k)}}{\mathbf{y}^{(k)T} H_k \mathbf{y}^{(k)}} \right)^{1/2} L_k^T \mathbf{y}^{(k)}, \\
\beta &= \left(\mathbf{s}^{(k)T} \mathbf{y}^{(k)} \right) \left(\mathbf{y}^{(k)T} H_k \mathbf{y}^{(k)} \right).
\end{aligned}
$$

注意到

$$
\begin{aligned}
& 1 + \frac{1}{\mathbf{w}^{(k)T} \mathbf{w}^{(k)}} \mathbf{w}^{(k)T} L_k^{-1} \cdot \left(\mathbf{s}^{(k)} - L_k \mathbf{w}^{(k)} \right) \\
=& 1 + \frac{1}{\mathbf{s}^{(k)T} \mathbf{y}^{(k)}} \left(\frac{\mathbf{s}^{(k)T} \mathbf{y}^{(k)}}{\mathbf{y}^{(k)T} H_k \mathbf{y}^{(k)}} \right)^{1/2} \mathbf{y}^{(k)T} L_k \cdot L_k^{-1} \cdot \left[\mathbf{s}^{(k)} - \left(\frac{\mathbf{s}^{(k)T} \mathbf{y}^{(k)}}{\mathbf{y}^{(k)T} H_k \mathbf{y}^{(k)}} \right)^{\frac{1}{2}} H_k \mathbf{y}^{(k)} \right] \\
=& \left(\frac{s_k^T \mathbf{y}^{(k)}}{\mathbf{y}^{(k)T} H_k \mathbf{y}^{(k)}} \right)^{1/2}
\end{aligned}
$$

以及$\mathbf{s}^{(k)T} \mathbf{y}^{(k)} > 0$，$H_k$正定，故上式大于零。从而再由引理可知秩一校正的矩阵J_{k+1}可逆，从而DFP公式产生的H_{k+1}正定。

以下给出**DFP**法迭代格式：

DFP

输入：初始点$\mathbf{x}_0$，终止误差$\varepsilon > 0$。

(1) 初始化$H_0 = I$，$k := 0$。

(2) 计算搜索方向：$\mathbf{d}^{(k)} := -H_k \bigtriangledown f(\mathbf{x}^{(k)})$，如果$\|\mathbf{g}^{(k)}\| < \varepsilon$，停止迭代。

(3) 一维搜索确定步长α_k，令$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$。

(4) 令$\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$，$\mathbf{y}^{(k)} = \bigtriangledown f(\mathbf{x}^{(k+1)}) - \bigtriangledown f(\mathbf{x}^{(k)})$，当$\mathbf{s}^{(k)T} \mathbf{y}^{(k)} > 0$作更新：
$H_{k+1} = H_k + \frac{\mathbf{s}^{(k)} \mathbf{s}^{(k)T}}{\mathbf{s}^{(k)T} \mathbf{y}^{(k)}} - \frac{H_k \mathbf{y}^{(k)} \mathbf{y}^{(k)T} H_k}{\mathbf{y}^{(k)T} H_k \mathbf{y}^{(k)}}$。置$k := k+1$，返回步骤(2)。

输出：$\mathbf{x}^{(k)}$，$f(\mathbf{x}^{(k)})$。

对于一般线性函数，校正保持正定性，因而$\mathbf{d}^{(k)}$总是下降方向，且具有超线性的收敛速度。但进一步研究发现，DFP方法具有数值不稳定性，以下给出的BFGS校正克服了这个缺点。

2. BFGS法

类似地，可以从拟牛顿条件(2)得到关于B_k的对称秩二校正公式

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

在对上式两次应用引理的求逆公式就得到了H_K的BFGS校正公式

$$
\begin{aligned}
H_{k+1}^{(BFGS)} = & H_k + \left(1 + \frac{\mathbf{y}^{(k)^T} H_k \mathbf{y}^{(k)}}{\mathbf{s}^{(k)^T} \mathbf{y}^{(k)}}\right) \frac{\mathbf{s}^{(k)} \mathbf{s}^{(k)^T}}{\mathbf{s}^{(k)^T} \mathbf{y}^{(k)}} \\
& - \frac{H_k \mathbf{y}^{(k)} \mathbf{s}^{(k)^T} + \mathbf{s}^{(k)} \mathbf{y}^{(k)^T} H_k}{\mathbf{s}^{(k)^T} \mathbf{y}^{(k)}}.
\end{aligned} \tag{7}
$$

以下给出**BFGS**法迭代格式：

BFGS

输入：初始点$\mathbf{x}_0$，终止误差$\varepsilon > 0$。

(1) 初始化$H_0 = I$，$k := 0$。

(2) 计算搜索方向：$\mathbf{d}^{(k)} := -H_k \bigtriangledown f(\mathbf{x}^{(k)})$，如果$\|\mathbf{g}^{(k)}\| < \varepsilon$，停止迭代。

(3) 一维搜索确定步长α_k，令$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{d}^{(k)}$。

(4) 令$\mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$，$\mathbf{y}^{(k)} = \bigtriangledown f(\mathbf{x}^{(k+1)}) - \bigtriangledown f(\mathbf{x}^{(k)})$，作更新：
$H_{k+1}^{(BFGS)} = H_k + \left(1 + \frac{\mathbf{y}^{(k)^T} H_k \mathbf{y}^{(k)}}{\mathbf{s}^{(k)^T} \mathbf{y}^{(k)}}\right) \frac{\mathbf{s}^{(k)} \mathbf{s}^{(k)^T}}{\mathbf{s}^{(k)^T} \mathbf{y}^{(k)}} - \frac{H_k \mathbf{y}^{(k)} \mathbf{s}^{(k)^T} + \mathbf{s}^{(k)} \mathbf{y}^{(k)^T} H_k}{\mathbf{s}^{(k)^T} \mathbf{y}^{(k)}}$，置$k := k+1$，返回步骤(2)。

输出：$\mathbf{x}^{(k)}$，$f(\mathbf{x}^{(k)})$。

（四）基于黄金分割准则的精确一维步长搜索

在迭代格式中，通过求解一维最优化问题

$$
\min_{\alpha \geq 0} \varphi(\alpha) = f(\mathbf{x}^{(k)} + \alpha \mathbf{d}^{(k)}), \tag{8}
$$

来确定步长。此时称为精确一维搜索。

黄金分割法是一种分割方法，基本思想是通过取试探点和进行函数值的比较，使包含极小点的搜索区间不断缩小，当区间缩短到一定程度时，区间上的各点均接近极小值，可以看作是极小值的近似。

详细迭代过程如下：

Gold_div_search

输入：迭代点$\mathbf{x}^{(k)}$，搜索区间$[a, b]$，精度要求$\varepsilon > 0$。

(1) 初始化试探点，$\alpha_1 = a + 0.382(b-a)$，$\alpha_2 = a + 0.618(b-a)$。计算$\varphi_1 = \varphi(\alpha_1)$，$\varphi_2 = \varphi(\alpha_2)$。

(2) 如果$b - a \leq \varepsilon$，算法结束。否则如果$\varphi_1 > \varphi_2$，进入步骤(3)；如果$\varphi_1 < \varphi_2$，进入步骤(4)；如果$\varphi_1 = \varphi_2$，进入步骤(5)。

(3) 令$a = \alpha_1$，$\alpha_1 = \alpha_2$，$\alpha_2 = a + 0.618(b-a)$，返回步骤(2)。

(4) 令$b = \alpha_2$，$\alpha_2 = \alpha_1$，$\alpha_1 = a + 0.382(b-a)$，返回步骤(2)。

(5) 令$a = \alpha_1$，$b = \alpha_2$，$\alpha_1 = a + 0.382(b-a)$，$\alpha_2 = a + 0.618(b-a)$，返回步骤(2)。

输出：α_1。

（五）基于Wolfe-Powell准则的非精确一维步长搜索

在实际计算中，往往不是求解一维最优化问题(8)，而是找出满足某些适当条件的粗略近似解作为步长，此时称为非精确一维搜索。

Wolfe-Powell准则给出条件

$$\varphi(\alpha) \le \varphi(0) + \rho\alpha\varphi'(0) \tag{9}$$

$$\varphi'(x) \ge \sigma\varphi'(0) \tag{10}$$

其中式(10)是精确线性搜索满足的正交条件${\mathbf{g}^{(k+1)}}^T\mathbf{d}^{(k)} = 0$的近似。以下给出迭代格式：

Wolfe_Powell

输入： 迭代点$\mathbf{x}^{(k)}$，搜索区间$[0,\bar{\alpha}]$，$\rho \in (0,\frac{1}{2})$，$\sigma \in (\rho, 1)$。

(1) 计算$\varphi_0 = \varphi(0)$，$\varphi_0' = \varphi'(0)$，令$a_1 = 0$，$a_2 = \bar{\alpha}$，$\varphi_1 = \varphi_0$，$\varphi_1' = \varphi_0'$。选取$\alpha \in (a_1, a_2)$。

(2) 计算$\varphi = \varphi(\alpha)$，若$\varphi(\alpha) \le \varphi(0) + \rho\alpha\varphi'(0)$，进入步骤(3)；否则，由$\varphi_1, \varphi_1', \varphi$构造多项式$P^{(1)}$并得到极小点$\hat{\alpha} = a_1 + \frac{1}{2}\frac{(a_1-\alpha)^2\varphi_1'}{(\varphi_1-\varphi)-(a_1-\alpha)\varphi_1'}$，令$a_2 = \alpha$，$\alpha = \hat{\alpha}$，重复步骤(2)。

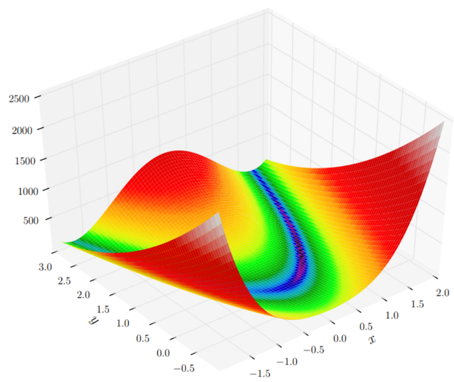
(3) 计算$\varphi' = \varphi'(\alpha)$，若$\varphi'(x) \ge \sigma\varphi'(0)$，则输出$\alpha_k = \alpha$，算法停止；否则，由$\varphi, \varphi', \varphi_1'$构造多项式$P^{(2)}$并得到极小点$\hat{\alpha} = \alpha - \frac{(a_1-\alpha)\varphi'}{\varphi_1'-\varphi'}$，令$a_1 = \alpha$，$\alpha = \hat{\alpha}$，$\varphi_1 = \varphi$，$\varphi_1' = \varphi'$，返回步骤(2)。

输出： α。

三、测试函数说明

Rosenbrock函数：

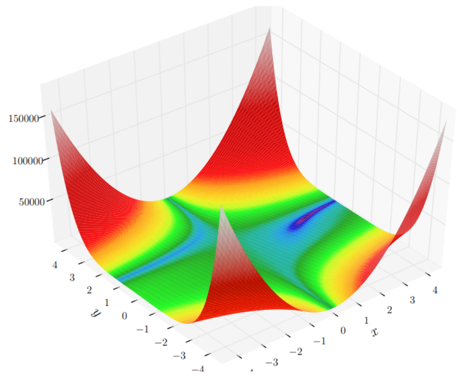
$$f(\mathbf{x}) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2$$



最优解为$f(1,1) = 0$。

Beale函数：

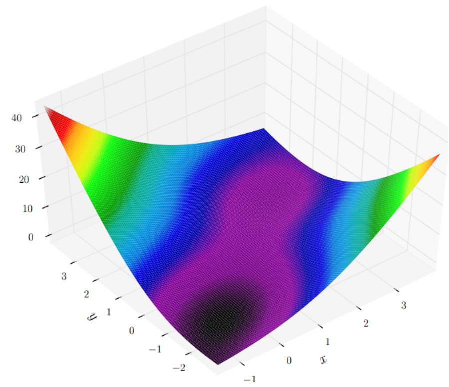
$$f(\mathbf{x}) = (1.5 - x_1 + x_1 x_2)^2 + (2.25 - x_1 + x_1 x_2^2)^2 + (2.625 - x_1 + x_1 x_2^3)^2$$



当搜索区间为$-4.5 \le x, y \le 4.5$时最优解为$f(3, 0.5) = 0$。

McCormick函数：

$$f(\mathbf{x}) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$$



当搜索区间为$-1.5 \le x \le 4$，$-3 \le y \le 4$时最优解为$f(-0.54719, -1.54719) = -1.9133$。

四、程序输入输出说明

1.

请输入需要测试的函数：
【1】*Rosenbrock*
【2】*Beale*
【3】*McCormick*

首先选择需要测试的函数，输入1\2\3以回车结束。

2.

请输入需要使用的方法：
【1】采用精确一维搜索的最速下降法
【2】经典牛顿法
【3】采用基于$Wolfe-Powell$非精确一维搜索的DFP法
【4】采用基于$Wolfe-Powell$非精确一维搜索的$BFGS$法

其次选择需要测试的方法，输入1\2\3\4以回车结束。

3.

请输入初始点：

最后输入初始点，两个坐标值以空格隔开，回车结束。

4.

迭代11次
最优解：$x1 = 2.9995974679750543, x2 = 0.49988731493905675$
最优值：$f = 2.9754691805977275e-08$

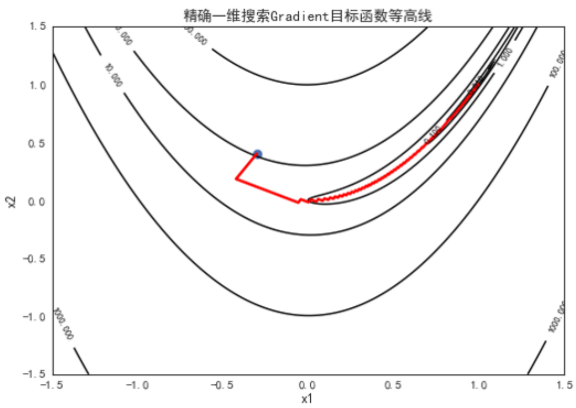
程序输出迭代次数，最优解，最优值。

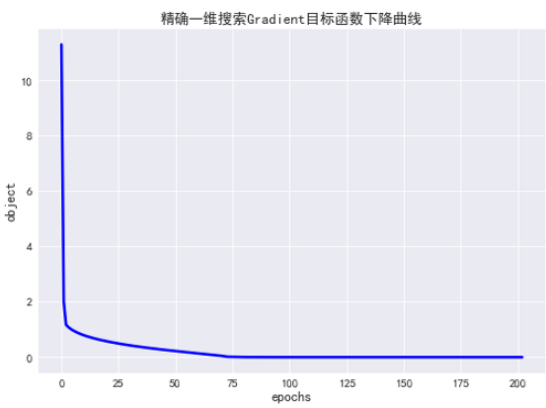
五、程序测试结果

1. Rosenbrock函数

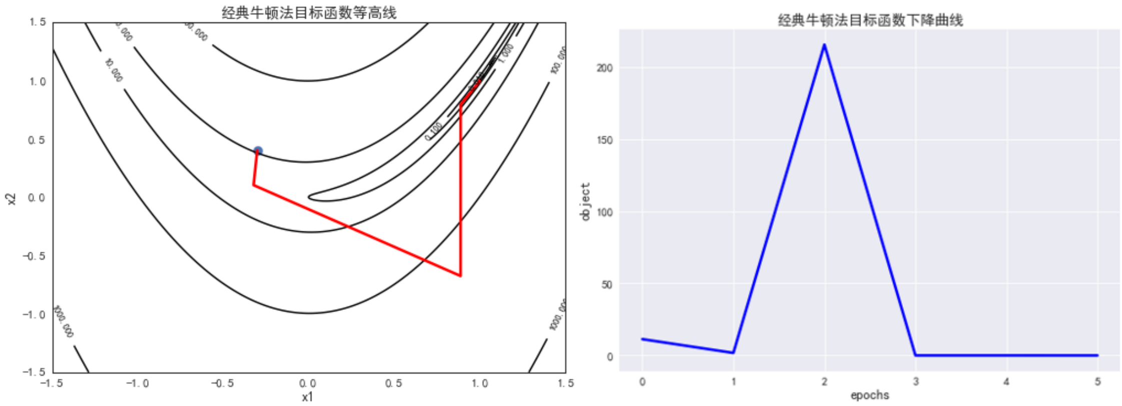
考察在同一初始点$(-0.3, 0.4)$下不同算法的效果

采用精确一维搜索的最速下降法：

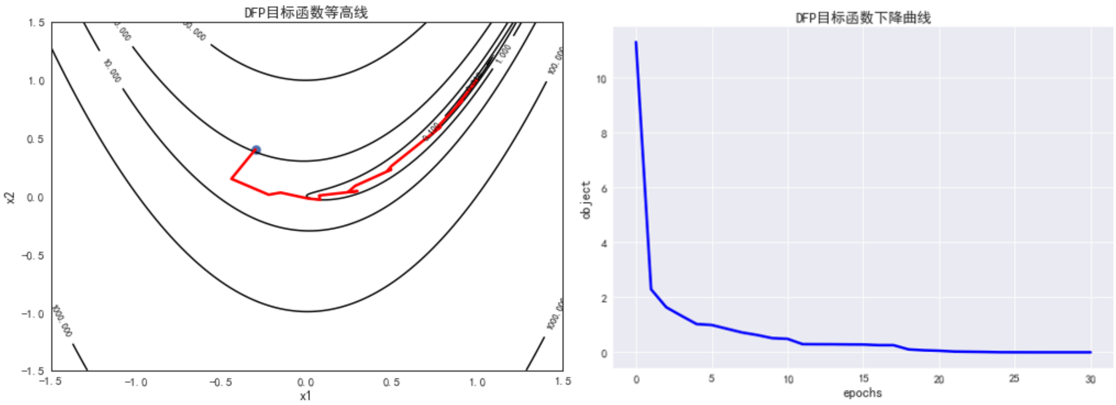




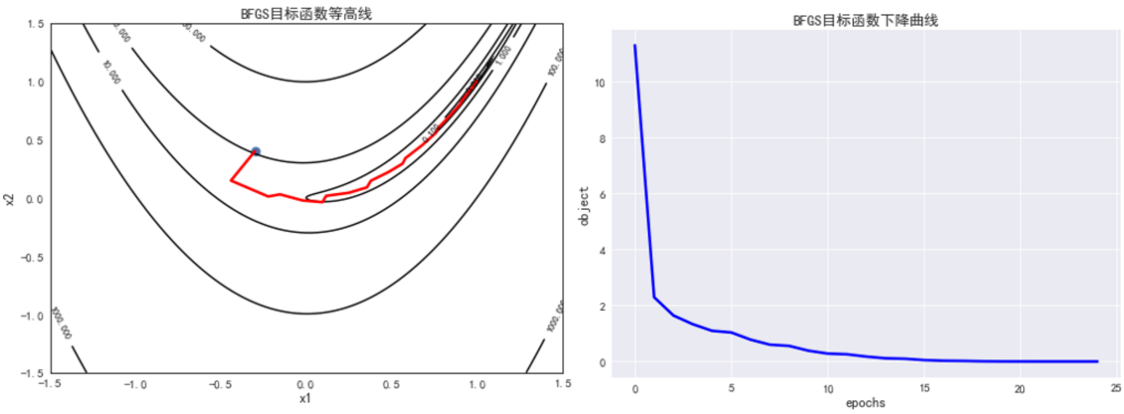
采用经典牛顿法：



采用基于**Wolfe-Powell**非精确一维搜索的**DFP**法：



采用基于**Wolfe-Powell**非精确一维搜索的**BFGS**法：



算法	最优解	最优值	次数
最速下降法	(0.9999999368454756, 0.9999999368454756)	4.754338744371732e-15	203
经典牛顿法	(0.9999999999999997, 0.9999999999953213)	2.1884018353647033e-21	6
DFP	(0.9999999999999961, 0.9999999999999775)	2.1818475153974118e-26	31
BFGS	(1.0000000003482363, 1.0000000007088101)	1.364898433414235e-19	25

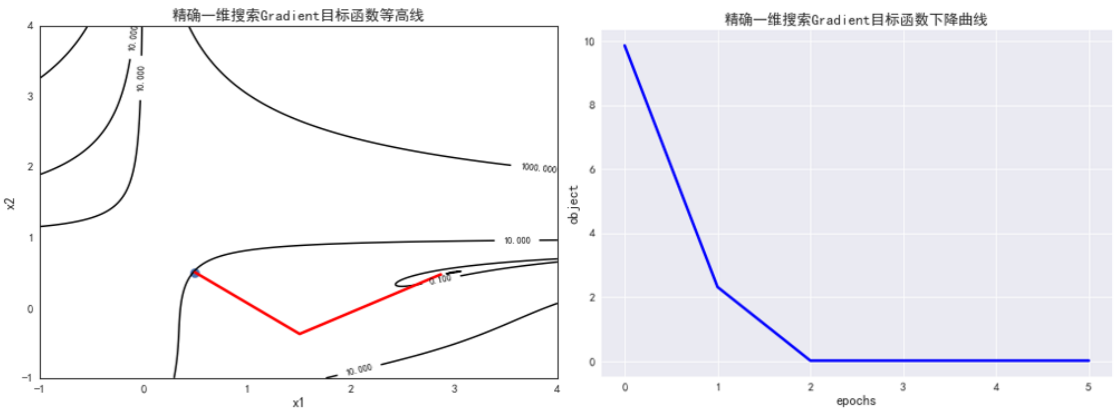
考察基于**Wolfe-Powell**非精确一维搜索的**BFGS**法在不同初始点下的效果：

初始点	最优解	最优值	次数
$(0.5, 0.5)$	(0. 9999999999645257, 0. 9999999999304009)	1. 4405338339947363e-21	15
$(-0.5, -0.5)$	(1. 0000000001169451, 1. 0000000002594174)	7. 88396281863918e-20	25
$(-2, -5)$	(1. 0000000023260547, 1. 0000000048258566)	8. 429340684293242e-18	38
$(10, -5)$	(1. 0000000000009632, 1. 0000000000026739)	5. 678880697649583e-23	65
$(10, -50)$	(0. 9999999997694025, 0. 9999999995469737)	5. 984815360609098e-20	66
$(100, -50)$	(1. 0000000007796892, 1. 0000000013279087)	5. 965738947584368e-18	288
$(100, -500)$	(1. 0000000000001712, 1. 000000000000408)	4. 583745808153962e-25	287

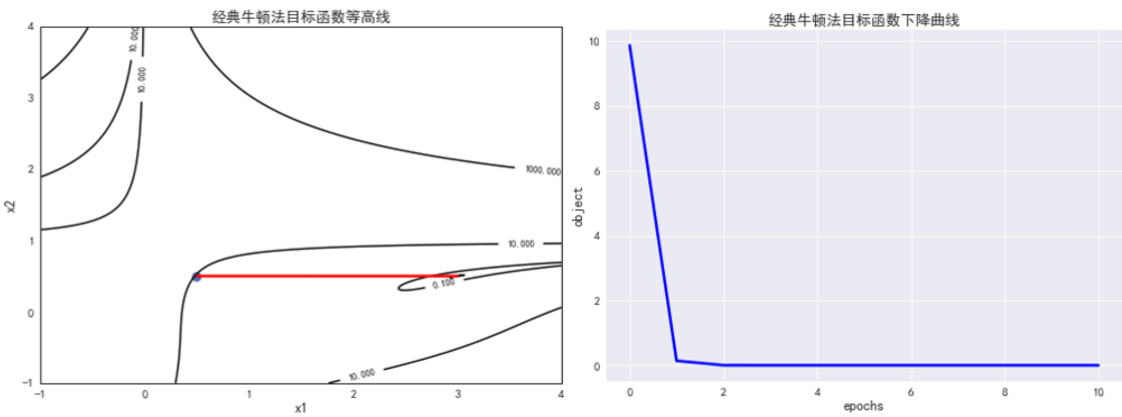
2. Beale函数

考察在同一初始点$(0.5, 0.5)$下不同算法的效果

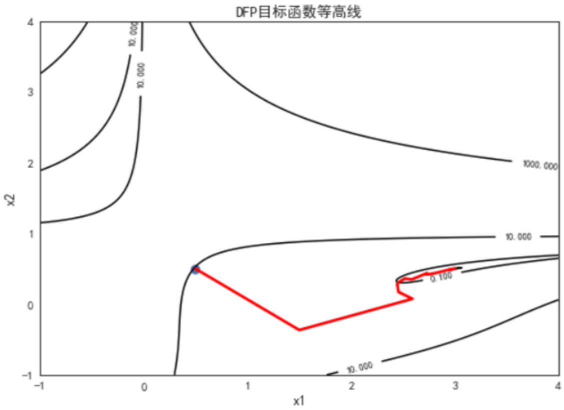
采用精确一维搜索的最速下降法：

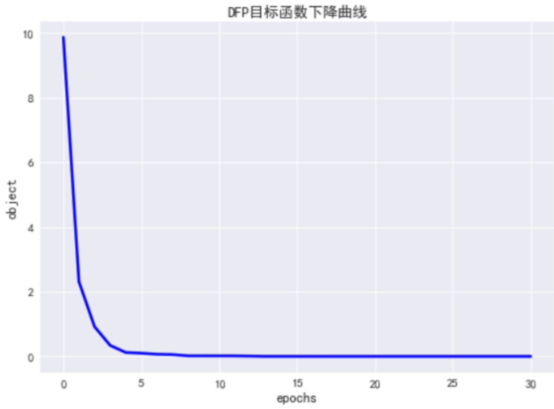


采用经典牛顿法：

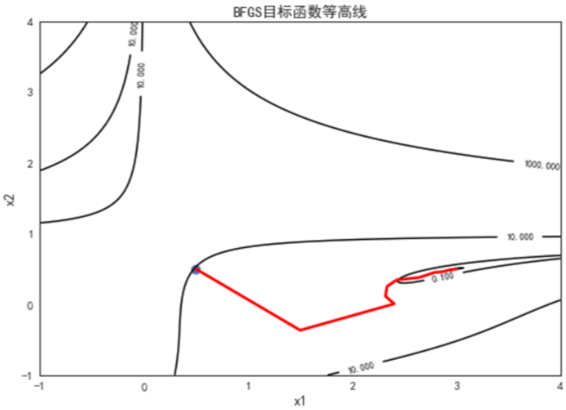


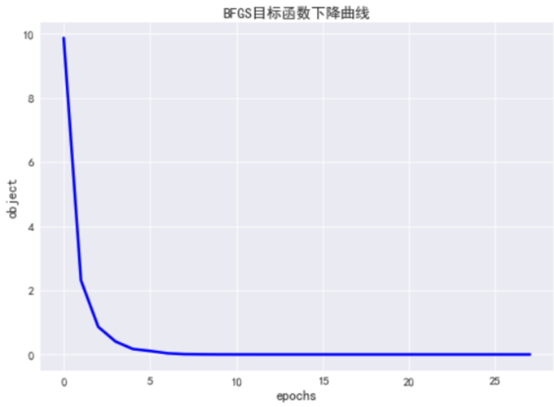
采用基于**Wolfe-Powell**非精确一维搜索的**DFP**法：





采用基于**Wolfe-Powell**非精确一维搜索的**BFGS**法：





算法	最优解	最优值	次数
最速下降法	(2.8755968600817567, 0.46705811823227883)	0.002857305563428932	6
经典牛顿法	(2.9995974679750543, 0.49988731493905675)	2.9754691805977275e-08	11
DFP	(3.000000006963388, 0.5000000049195799)	2.428723111796065e-16	31
BFGS	(3.000000043525999, 0.5000000149328366)	6.984822518127638e-16	28

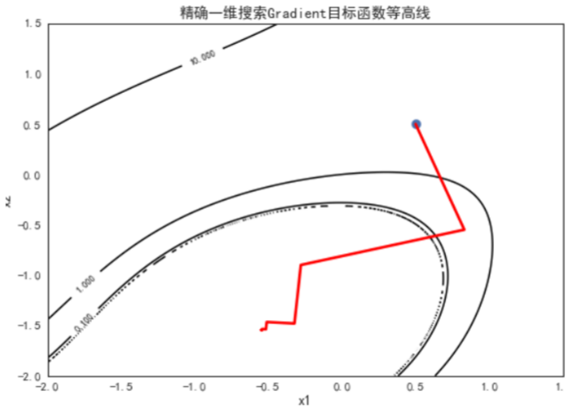
考察基于**Wolfe-Powell**非精确一维搜索的**BFGS**法在不同初始点下的效果：

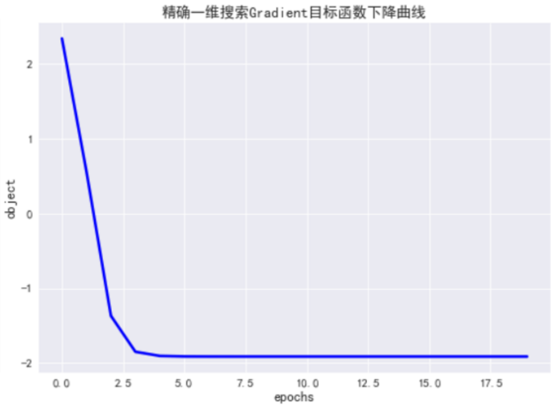
初始点	最优解	最优值	次数
$(0.5, 0.5)$	(3.0000000673826563, 0.5000000185873378)	8.081377965410317e-16	19
$(-0.5, -0.5)$	(2.99999997187749, 0.4999999934602392)	1.3092642229480963e-16	29
$(2, 0)$	(3.0000000686409236, 0.5000000185604835)	8.08817943744345e-16	21
$(0, 0)$	(3.00000058152033, 0.5000000189682727)	1.0183999029410446e-15	20
$(2, 2)$	无法收敛	\	\
$(-2, 2)$	解非最优	\	\
$(-2, -2)$	解非最优	\	\

3. McCormick函数

考察在同一初始点(0.5, 0.5)下不同算法的效果

采用精确一维搜索的最速下降法：

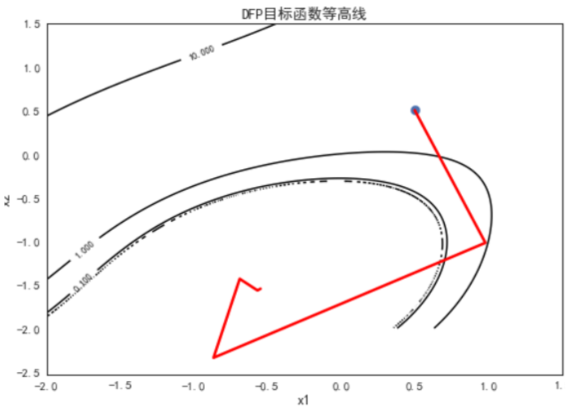


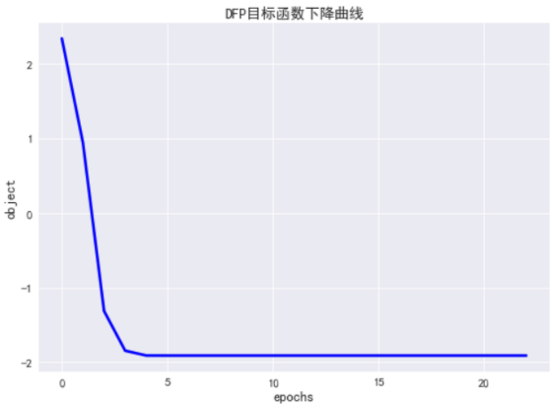


采用经典牛顿法：

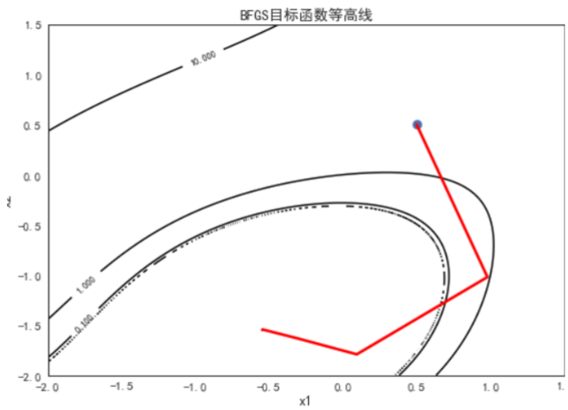
Hesse矩阵奇异，无法求解

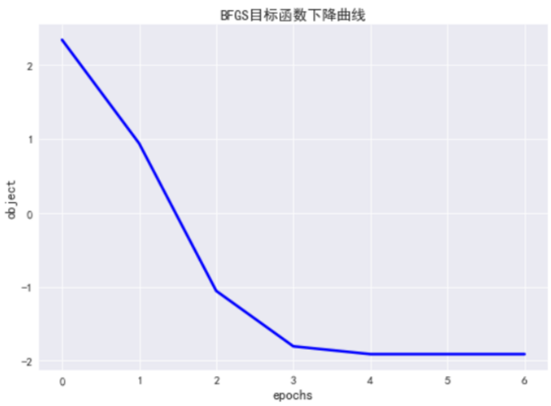
采用基于**Wolfe-Powell**非精确一维搜索的**DFP**法：





采用基于**Wolfe-Powell**非精确一维搜索的**BFGS**法：





算法	最优解	最优值	次数
最速下降法	(-0.547197539167360, -1.547197532106908)	-1.9132229549810358	20
经典牛顿法	无法求解	\	\
DFP	(-0.547197565531484, -1.547197564933061)	-1.9132229549810362	23
BFGS	(-0.547197551201126, -1.547197551033523)	-1.9132229549810362	7

考察基于**Wolfe-Powell**非精确一维搜索的**BFGS**法在不同初始点下的效果：

初始点	最优解	最优值	次数
$(0.5, 0.5)$	(-0.547197551201126, -1.547197551033523)	-1.9132229549810362	7
$(-0.5, -0.5)$	(-0.547197527042834, -1.547197531988847)	-1.9132229549810353	5
$(-2, 0)$	(-0.547197550723614, -1.547197551524310)	-1.9132229549810367	4
$(0, 0)$	(-0.547197544313902, -1.547197577894429)	-1.9132229549810353	25
$(2, 2)$	解非最优	\	\
$(-2, 2)$	(-0.547197551227754, -1.547197551268341)	-1.9132229549810362	5
$(-4, 4)$	(-0.547197551164487, -1.547197551138525)	-1.9132229549810367	5

六、分析总结

在目标函数足够“好”的情况下，牛顿法以及拟牛顿法的收敛速度都要比最速下降法快很多。并且牛顿法也比拟牛顿法收敛得更快，但可能存在一定的震荡现象。而实际情况中，目标函数往往不具有很好的性质，这时候牛顿法可能不再适用，需要使用拟牛顿法来完成求解。

同时，不同的初始点也可能产生截然不同的结果，甚至无法收敛。这与目标函数的性质有关系。例如，如果目标函数存在局部梯度很小或很大的情况，将对该区域的初始点造成很大的影响，所以选择一个恰当的初始点也是一个需要认真对待的工作。所以建议求解问题之前，可以观察目标函数的图像等方式来推测其性质，从而选择合适的算法，并在可能的最优解附近选取合适的初始点开始迭代。

最后，经过三个测试函数发现，**BFGS**法往往具有比最速下降法更快的收敛速度，比经典牛顿法跟广泛的适用对象，比**DFP**法更稳定的数值，是一个很好的迭代算法，如果配合合适的初始点，往往产生很好的效果。

七、参考文献

[1] 吴祈宗, 侯福均. 运筹学与最优化方法(第2版)[M]. 机械工业出版社, 2013:105-108, 112-117, 122-125.

[2] 袁亚湘, 孙文瑜. 最优化理论与方法[M]. 科学出版社, 1991:96-100, 108-110, 121-124, 224-233.