无约束优化

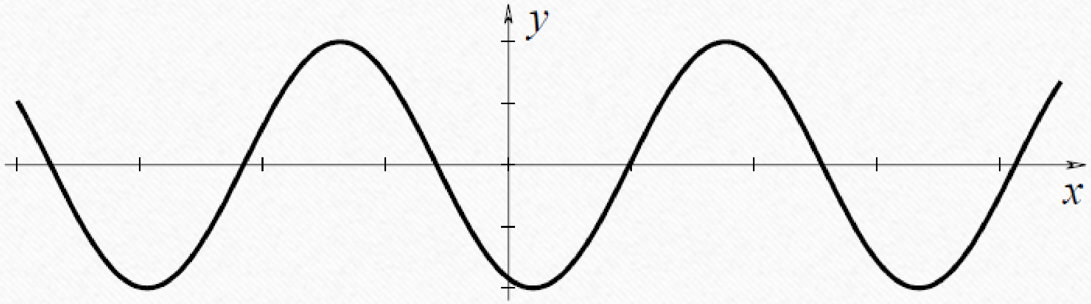
山东大学 辛士庆

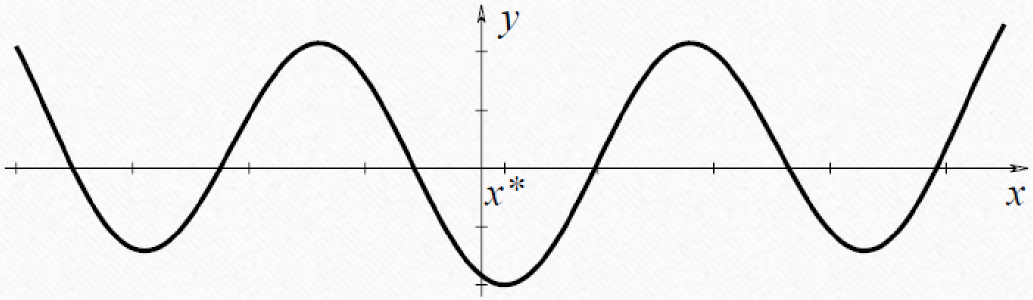
2018年8月26号 桂林

问题描述

$$\text{Find } \mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} f(\mathbf{x}), \quad \text{where } f : \mathbb{R}^n \mapsto \mathbb{R}.$$

极小值往往不唯一





局部最优解

$\mathbf{x}^*$ is a *local minimizer* for $f: \mathbb{R}^n \mapsto \mathbb{R}$ if

$$f(\mathbf{x}^*) \le f(\mathbf{x}) \quad \text{for } \|\mathbf{x}^* - \mathbf{x}\| \le \varepsilon \ (\varepsilon > 0).$$

必要条件

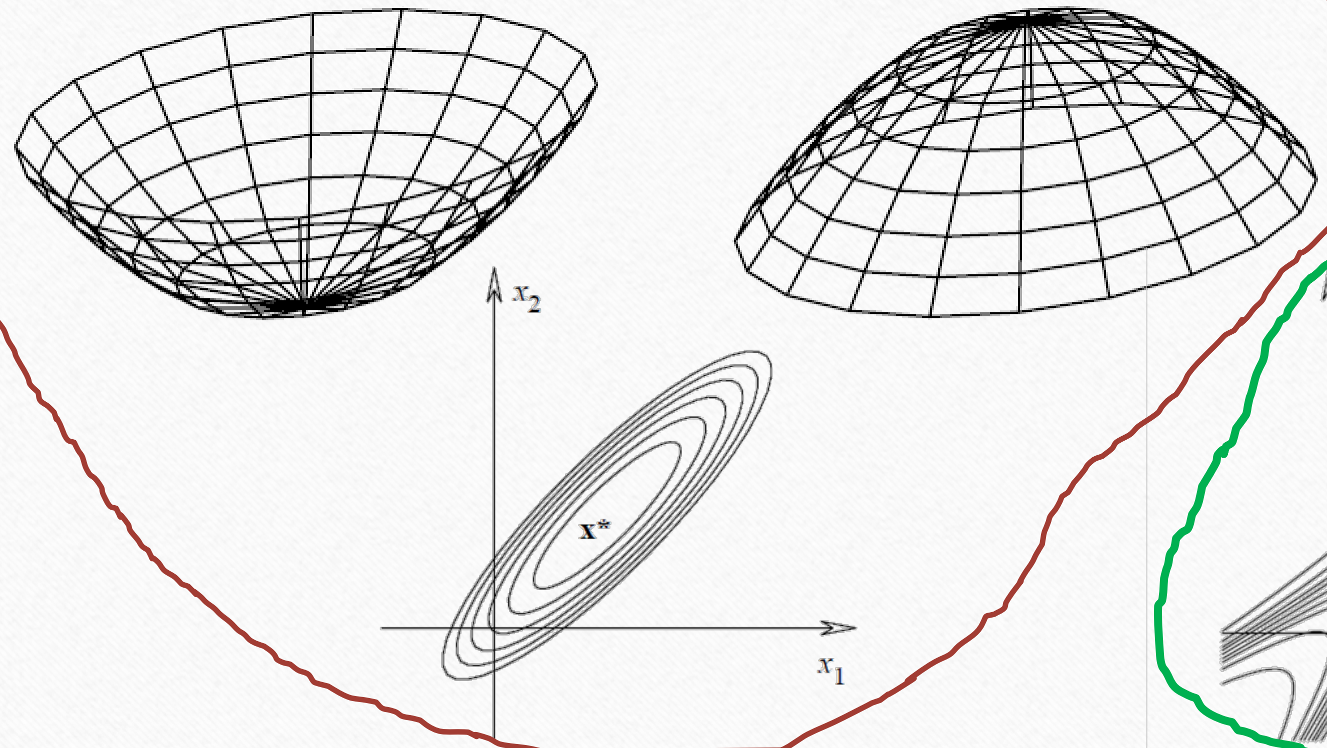
$$f(\mathbf{x}+\mathbf{h}) \;=\; f(\mathbf{x}) + \mathbf{h}^\top \mathbf{f}\,'(\mathbf{x}) + O(\|\mathbf{h}\|^2)\,,$$

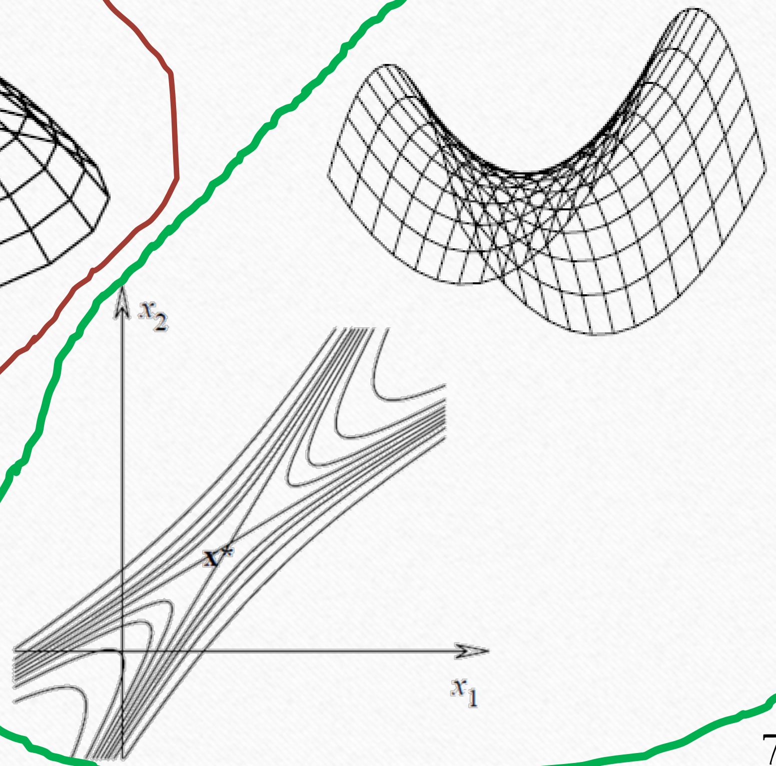
$$\mathbf{f}\,'(\mathbf{x}) \;\equiv\; \begin{bmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \dfrac{\partial f}{\partial x_n}(\mathbf{x}) \end{bmatrix} = 0$$

充分条件

$$\mathbf{f}''(\mathbf{x}) \equiv \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})\right] \qquad \text{正定}$$

举例说明





收敛阶

- 定义：$\mathbf{e}_k \equiv \mathbf{x}_k - \mathbf{x}^*$,
- 一般要求：$\|\mathbf{e}_{k+1}\| < \|\mathbf{e}_k\|$ for $k > K$.
- 线性收敛：$\|\mathbf{e}_{k+1}\| \le c_1 \|\mathbf{e}_k\|$ with $0 < c_1 < 1$ and $\mathbf{x}_k$ close to $\mathbf{x}^*$
- 二阶收敛：$\|\mathbf{e}_{k+1}\| \le c_2 \|\mathbf{e}_k\|^2$ with $c_2 > 0$ and $\mathbf{x}_k$ close to $\mathbf{x}^*$.
- 超线性收敛：$\dfrac{\|\mathbf{e}_{k+1}\|}{\|\mathbf{e}_k\|} \to 0$ for $k \to \infty$.

函数值下降的方向

下降方向： If $\mathbf{f}'(\mathbf{x}) \neq \mathbf{0}$ and $\mathbf{B}$ is a symmetric, positive definite matrix, then

$$\mathbf{h}_1 = -\mathbf{B}\mathbf{f}'(\mathbf{x}) \quad \text{and} \quad \mathbf{h}_2 = -\mathbf{B}^{-1}\mathbf{f}'(\mathbf{x})$$

are descent directions.

梯度下降法

梯度下降法的一般版本

◇ Step 0. Select a very small $\epsilon > 0$ for being used in the stopping criterion. Start at an arbitrary initial point x^0 and set $k = 0$.

◇ Step 1. Optimality check. If

$$\|\nabla f(x^k)\| \leq \epsilon$$

stop and $x^* = x^k$; otherwise go to Step 2.

梯度下降法的一般版本

- Step 2. Updating procedure.

$$x^{k+1} = x^k - \alpha_k g_k$$

where the n-dimensional column vector $g_k = \nabla f(x^k)^T$ and α_k is a nonnegative scalar minimizing $f(x^k - \alpha g_k)$. Set $k = k + 1$. Go back to Step 1.

共轭梯度法的核心思想

$$x^{k+1} = x^k + \alpha^k d^k,$$

$$d^k = -g^k + \sum_{j=0}^{k-1} \frac{g^{k\prime} Q d^j}{d^{j\prime} Q d^j} d^j.$$

共轭梯度法的一般版本

$r_0 := b - Qx_0$ (r_i为第i次迭代的误差)

$d_0 := r_0$ (d_i是我们要求的共轭向量)

$k := 0$ (k表示第几次迭代)

repeat

$\alpha_k := \frac{r_k^T r_k}{d_k^T Q d_k}$ (该项为学习率，是求出来的，对应的是之前说的a_i)

$x_{k+1} := x_k + \alpha_k d_k$

$r_{k+1} := r_k - \alpha_k Q d_k$

共轭梯度法的一般版本

如果r_{k+1}足够小，则提前退出循环 （也就是认为已经找到最优解了）

$\beta_k := \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$

$d_{k+1} := r_{k+1} + \beta_k d_k$ （Gram-Schmidt过程求d_k）

k:=k+1

end repeat

The result is x_{k+1}

牛顿法及其变种

牛顿法的依据

$$f(\mathbf{x}+\mathbf{h}) \simeq q(\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^{\top}\mathbf{f}'(\mathbf{x}) + \tfrac{1}{2}\mathbf{h}^{\top}\mathbf{f}''(\mathbf{x})\mathbf{h}\,.$$

$$\mathbf{f}'(\mathbf{x}) + \mathbf{f}''(\mathbf{x})\mathbf{h} = \mathbf{0}\,.$$

牛顿法迭代过程

$$
\begin{array}{ll}
\textbf{begin} & \\
\quad x := x_0; & \{\text{Initialisation}\} \\
\quad \textbf{repeat} & \\
\qquad \text{Solve } f''(x)h_n = -f'(x) & \{\text{find step}\} \\
\qquad x := x + h_n & \{\ldots \text{ and next iterate}\} \\
\quad \textbf{until} \text{ stopping criteria satisfied} & \\
\textbf{end} &
\end{array}
$$

BFGS

1. 给定初值 $\mathbf{x}_0$ 和精度阀值 ϵ, 并令 $B_0 = I$, $k := 0$.

2. 确定搜索方向 $\mathbf{d}_k = -B_k^{-1} \cdot \mathbf{g}_k$.

3. 利用 (1.13) 得到步长 λ_k, 令 $\mathbf{s}_k = \lambda_k \mathbf{d}_k$, $\mathbf{x}_{k+1} := \mathbf{x}_k + \mathbf{s}_k$.

4. 若 $\|\mathbf{g}_{k+1}\| < \epsilon$, 则算法结束.

5. 计算 $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$.

6. 计算 $B_{k+1} = B_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{B_k \mathbf{s}_k \mathbf{s}_k^T B_k}{\mathbf{s}_k^T B_k \mathbf{s}_k}$. 对Hessian矩阵的近似

7. 令 $k := k + 1$, 转至步 2.

DFP算法

$$D_{k+1} = D_k + \Delta D_k$$ 对Hessian逆矩阵的近似

$$\Delta D_k = \frac{s_k s_k^T}{s_k^T y_k} - \frac{D_k y_k y_k^T D_k}{y_k^T D_k y_k}$$

L-BFGS

$$D_{k+1} = V_k^T D_k V_k + \rho_k s_k s_k^T$$

$$\rho_k = \frac{1}{y_k^T s_k}, V_k = I - \rho_k y_k s_k^T$$

最小化Dirichlet能量

$$E[u] = \frac{1}{2}\int_{\Omega} \|\nabla u(x)\|^2 \, dV,$$

$$\Delta u(x) = 0$$

最小化Dirichlet能量

$$\frac{1}{2}\sum_{i,j} a_{ij}(f_j - f_i)^2$$

$$Lf = 0$$