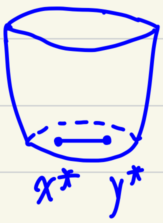
（1）假设f是一个凸函数，证明f的全局最小点的集合是一个凸集。



Denote S as the set of all global minima of f, we need to prove that S is convex.

$\forall x^*, y^* \in S$, since f is a convex function, $\forall \lambda \in (0,1)$

$$f(\lambda x^* + (1-\lambda) y^*) \le \lambda f(x^*) + (1-\lambda) f(y^*) \qquad (1)$$

Meanwhile, note that x^*, y^* are global minima, $\forall x, y \in \mathrm{dom}(f)$, $\exists M$

$$f(x^*) \le f(x),\ f(y^*) \le f(y),\quad f(x^*) = f(y^*) = M$$

Let $x = \lambda x^* + (1-\lambda) y^* \in \mathrm{dom}(f)$, $y = y^* \in \mathrm{dom}(f)$, then

$$f(\lambda x^* + (1-\lambda) y^*) \ge f(x^*)$$
$$= \lambda f(x^*) + (1-\lambda) f(y^*) \qquad (2)$$

From equation (1), (2) we know $f(\lambda x^* + (1-\lambda) y^*) = \lambda f(x^*) + (1-\lambda) f(y^*)$
$$= M$$

Meaning that $\lambda x^* + (1-\lambda) y^* \in S$, therefore S is convex. $\square$

（2）证明函数$f(x) = 8x_1 + 12x_2 + x_1^2 - 2x_2^2$只有一个稳定点，并且它既不是一个最大点也不是一个最小点，而是一个鞍点。

$$\frac{\partial f(x)}{\partial x_1} = 2x_1 + 8,\quad \frac{\partial f(x)}{\partial x_2} = -4x_2 + 12 \Rightarrow \nabla f(x) = (2x_1 + 8, -4x_2 + 12)$$

Let $\nabla f(x) = 0$, we have $\begin{cases} x_1 = -4 \\ x_2 = 3 \end{cases}$, so f has only 1 stationary point

Note that $\nabla^2 f(x) = H = \begin{pmatrix} 2 & 0 \\ 0 & -4 \end{pmatrix}$, Let $x = (1,1)$, then $x^T H x = -2 < 0$

Let $x = (1,0)$, then $x^T H x = 2 > 0$

Therefore $\nabla^2 f(x)$ is an indefinite matrix, so $(-4, 3)$ is a saddle point.

（3）考虑函数$f(x_1, x_2) = (x_1 + x_2^2)^2$,在点$x = (1,0)$的位置上考虑搜索方向$p = (-1,1)$，证明$p$为一个下降方向，并且找到这个方向上的精确线搜索步长。

$$\nabla f(x_1, x_2) = (2(x_1 + x_2^2), 4x_2(x_1 + x_2^2))$$

Note that $\nabla f(x_1, x_2)^T p \Big|_{(x_1, x_2) = (1,0)} = (2,0)^T(-1,1) = -2 < 0$, so $p = (-1,1)$ is a descent direction. To solve for step α_k, we aim to optimize:

$$\alpha^* = \arg\min_{\alpha} f(x + \alpha p)$$

Let $g(\alpha) \triangleq f(x + \alpha p) = f(1-\alpha, \alpha) = (1 - \alpha + \alpha^2)^2$, then

$$\frac{\partial g(\alpha)}{\partial \alpha} = 2(1 - \alpha + \alpha^2)(-1 + 2\alpha) = 0 \Rightarrow \alpha = \frac{1}{2}$$

Obviously $g(\alpha)$ is convex, so the step of exact linear search is $\alpha = \frac{1}{2}$

（4）已知序列$x_k = \frac{1}{k!}$，该序列是二次收敛，还是超线性收敛？

One can easily tell that $\lim_{k\to\infty} x_k = 0$. To study the convergence of $\{x_k\}$, we look into

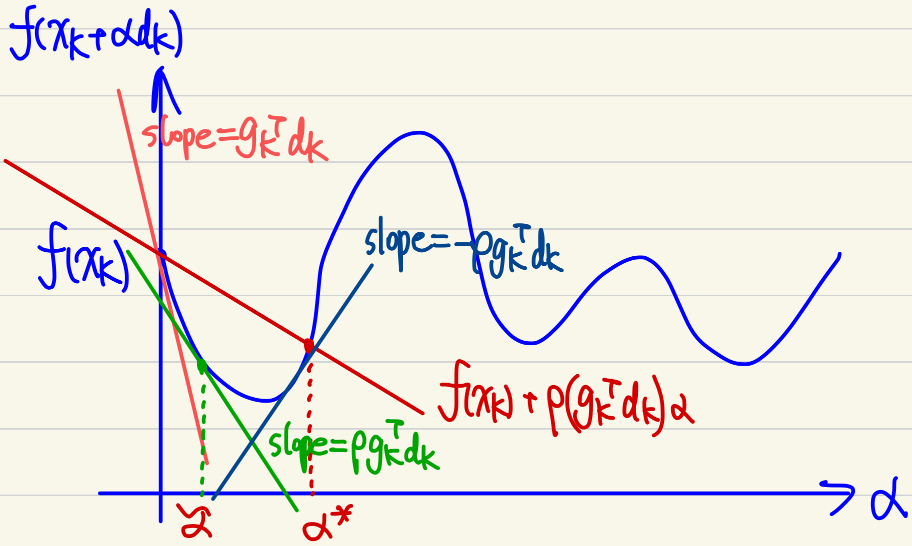
$$\lim_{k\to\infty}\frac{|x_{k+1}-0|}{|x_k-0|}=\lim_{k\to\infty}\frac{1}{k+1}=0$$

So $\{x_k\}$ converges to 0 superlinearly

$$\lim_{k\to\infty}\frac{|x_{k+1}-0|}{|x_k-0|^2}=\lim_{k\to\infty}\frac{(k!)^2}{(k+1)!}=\lim_{k\to\infty}\frac{k!}{k+1}=\infty$$

So $\{x_k\}$ does not converge quadratically

（5）证明：假设f连续可微，$f(x_k+\alpha d_k)$在$\alpha > 0$时有下界，且$g_k^T d_k < 0$，则必存在α_k使得$x_k+\alpha_k d_k$满足 Wolfe 准则和强 Wolfe 准则。



By Taylor expansion, $f(x_k+\alpha d_k)=f(x_k)+\alpha\nabla f(x_k)^T d_k+o(\alpha\|d_k\|)$

at $\alpha=0$, the tangent slope of $f(x_k+\alpha d_k)$ is

$$\lim_{\alpha\to 0}\frac{f(x_k+\alpha d_k)-f(x_k)}{\alpha}=\nabla f(x_k)^T d_k$$

Since $f\in C^1$, $\forall\varepsilon>0$, $\exists\delta>0$, s.t.

$$\left|\frac{f(x_k+\alpha' d_k)-f(x_k)}{\alpha'}-\nabla f(x_k)^T d_k\right|\le\varepsilon, \quad \forall\alpha'\in O(0,\delta)$$

i.e., $f(x_k+\alpha' d_k)<\alpha'\varepsilon+\alpha'\nabla f(x_k)^T d_k+f(x_k)$, omitting the $\alpha'\varepsilon$ term, for $0<\rho<1$ we have

$$f(x_k+\alpha' d_k)<f(x_k)+\alpha'\nabla f(x_k)^T d_k<f(x_k)+\rho(\nabla f(x_k)^T d_k)\alpha'$$

Note that when $\alpha'\to\infty$, $f(x_k)+\rho\nabla f(x_k)^T d_k\,\alpha'\to-\infty$ since $\nabla f(x_k)^T d_k<0$

while $f(x_k+\alpha d_k)$ has a lower bound when $\alpha>0$, i.e. $\exists M\in\mathbb{R}$ s.t.

$$f(x_k+\alpha d_k)\ge M, \quad \forall\alpha\in(0,+\infty)$$

therefore, $\exists$ sufficient large $\alpha''\in(\alpha',\infty)$, s.t.

$$f(x_k+\alpha'' d_k) > f(x_k) + \rho \nabla f(x_k)^T d_k \alpha''$$

By the Intermediate value theroem, $\exists\, \alpha^* \in (\alpha', \alpha'')$, s.t.

$$f(x_k+\alpha^* d_k) = f(x_k) + \rho \nabla f(x_k)^T d_k \alpha^*$$

Denote $R(\alpha) = f(x_k+\alpha d_k) - [f(x_k) + \rho \nabla f(x_k)^T d_k \alpha]$, note that $R(0)=0$, $R(\alpha^*)=0$, by Rolle's theorem, $\exists\, \tilde{\alpha} \in (0, \alpha^*)$ s.t.

$$R'(\tilde{\alpha}^*) = \nabla f(x_k+\tilde{\alpha} d_k)^T d_k - \rho \nabla f(x_k)^T d_k = 0$$

$$\Rightarrow \quad \nabla f(x_k+\tilde{\alpha} d_k)^T d_k = \rho \nabla f(x_k)^T d_k$$

Therefore, for $0<\rho<\sigma<1$, given that $\nabla f(x_k)^T d_k < 0$

$$\begin{aligned} |\nabla f(x_k+\tilde{\alpha} d_k)^T d_k| &= \rho |\nabla f(x_k)^T d_k| \\ &< \sigma |\nabla f(x_k)^T d_k| \\ &= -\sigma \nabla f(x_k) d_k \end{aligned}$$

Meaning that $\exists\, \alpha_k \in (0, \alpha^*)$ s.t. $x_k + \alpha_k d_k$ satisfies the strong Wolfe condition (Wolfe condition included) $\square$