数学分析：Weierstrass theorem

若 $f(x)$ 在闭区间 $[a,b]$ 上连续，则 $f(x)$ 在 $[a,b]$ 上必定能取得最大值和最小值.

$$f(x) = I_{[0,1]}(x) = \begin{cases} 0 & x\in[0,1] \\ +\infty & \text{otherwise} \end{cases}$$ indicator function

OR Note 3

Defn (实值函数) 映射 $f:\mathbb{R}^n \to \mathbb{R}$

(广义实值函数) 映射 $f:\mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R}\cup\{\pm\infty\}$

(适当函数 proper function) 给定广义实值函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$，非空集合 Ω.

① 若存在 $x\in\Omega$ 使得 $f(x)<+\infty$，且

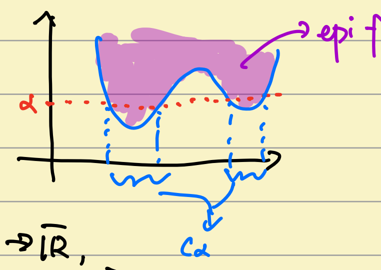
② 任意 $x\in\Omega$, $f(x) > -\infty$

则称 f 为适当函数.

Defn (下水平集 lower level set) 给定广义实值函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$,

$$C_\alpha := \{x \mid f(x) \le \alpha\}$$

称为 f 的下水平集.



Defn (上图 epigraph) 给定广义实值函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$,

$$\text{epi } f := \{(x,t)\in\mathbb{R}^{n+1} \mid f(x)\le t\}$$

称为 f 的上图.

Thm (下半连续函数与闭函数) 设 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$，以下命题等价.

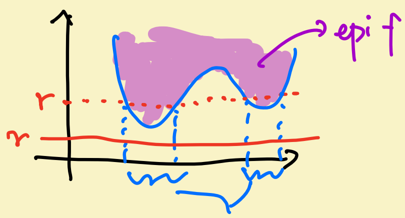
① $f(x)$ 任意 α-下水平集是闭集

② $f(x)$ 是下半连续的

③ $f(x)$ 是闭函数.

OR Note 3

拓展：定义域非紧集，且目标函数非连续



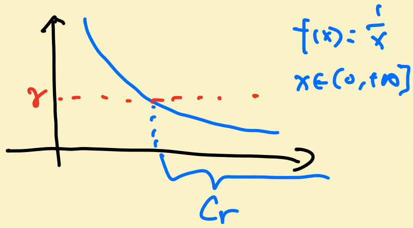
Thm (Weierstrass theorem) $f:\Omega \to (-\infty, +\infty]$ 适当且闭的函数，如果：

① $\mathrm{dom} f = \{x\in\Omega,\ f(x) < +\infty\}$ 是有界的。

② 存在一个 r 下水平集：$C_r = \{x\in\Omega,\ f(x) \le r\}$ 是非空且有界的

③ f 是强制的 (coersive)：$\forall \|x^k\| \to +\infty,\ x^k\in\Omega$，有 $\lim\limits_{k\to\infty} f(x^k) = +\infty$

以上三条中任意一个成立，则优化问题 $\min\limits_{x\in\Omega} f(x)$



的最小值点集 $\{x\in\Omega \mid f(x)\le f(y),\ \forall y\in\Omega\}$ 是非空且紧的。（有界闭集）

$\mathrm{epi} f = \{(x,t)\in\mathbb{R}^{n+1},\ f(x)\le t\}$

Proof: (2) ⇒ 结论。先证目标函数下确界存在。设下确界：$t := \inf f(x) = -\infty$

则存在点列 $\{x^k\}_{k=1}^{\infty} \subset C_r$，使得 $\lim\limits_{k\to+\infty} f(x^k) = t = -\infty$。因为 C_r 有界，

则 $\{x^k\}$ 存在聚点，记为 x^*。

因为 f 是下半连续的，所以 f 的上图为闭集。由于 $(x^*, t)\in \mathrm{epi} f \Rightarrow$

$f(x^*) \le t = -\infty$，与函数适当性矛盾。故 $t > -\infty$。

由于 t 是下确界，且 $f(x^*) \le t \Rightarrow f(x^*) = t$

根据前述定理，下水平集 C_r 是闭集，由 C_r 有界

⇒ 最小值点集为紧集

(1) ⇒ (2) ⇒ 结论。(1) ⇒ dom f 有界。由于 f 是适当函数，$\exists x_0\in\Omega$ 使得 $f(x_0) < +\infty$

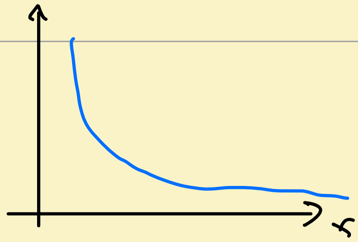
令 $r = f(x_0)$。由 dom f 有界，⇒ $C_r \subset \mathrm{dom} f$ 有界且非空 ⇒ (2) 成立 ⇒ 结论。

(3) ⇒ (2) ⇒ 结论。假设 C_r 无界，$\exists \{x^k\} \subset C_r$，s.t. $\lim\limits_{k\to\infty}\|x^k\| = +\infty$。根据 f 的强制性，有 $\lim\limits_{k\to\infty} f(x^k) = +\infty$。由于 C_r 是下水平集，$\forall x\in C_r$，$f(x)\le r$，与 $\lim\limits_{k\to\infty} f(x^k) = +\infty$ 矛盾。⇒ (2) 成立 ⇒ 结论。

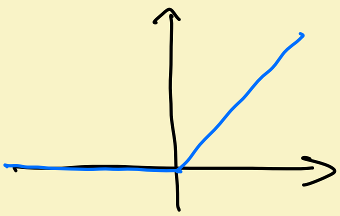
①. 三个条件都是保证 f(x) 最小值不能在无穷远处取到。

②. 可行域可以无界，只需要保证 (2)：即 f(x) 为闭适当函数，且下水平集非空且有界。

$f(x)=e^{-x}$



$f(x)=\max(x,0)$



E.g. $f(x)=x^2 \quad x\in\mathbb{R} \quad \Rightarrow$ 强制函数

$f(x)=e^{-x} \quad x\in\mathbb{R} \quad \Rightarrow$ 下水平集无界.

$f(x)=I_{[0,1]} \quad x\in\mathbb{R} \quad \Rightarrow \mathrm{dom} f=[0,1]$ 有界.

唯一性：

① 若最优点唯一，所有算法最终收敛到 x^*，此时，可以比较不同算法到 x^* 的收敛速度.

② 若最优点不唯一，但最优值唯一，此时，可以比较不同算法达到最优值的收敛速度

最优性理论

Defn（下降方向）对于可微函数 f，如果存在向量 d 满足

$$\nabla f(x)^T d < 0$$

那么称 d 为 f 在点 x 处的一个下降方向.

如果 f 在 x 处存在一个下降方向 d，那么对于任意 $T>0$，存在 $t\in(0,T]$，使得

$$f(x+td) < f(x)$$

线搜索方法

$$\text{Iterate:} \quad x^{k+1}=x^k+\alpha_k d^k$$

d^k：搜索方向，α_k：步长（学习率）

精确搜索步长： $\alpha_k=\underset{\alpha>0}{\arg\min}\ \phi(\alpha)=f(x^k+\alpha d^k)$

① 计算量大，实际中很少用

② 实际中用非精确搜索步长，仅需求 $\phi(\alpha)$ 满足某些不等式性质

梯度下降法

① 无约束优化问题
② 只需知道一阶导.
③ 每步迭代成本很低.

考虑 $\phi(\alpha) = f(x^k + \alpha d^k)$

$$= f(x^k) + \nabla f(x^k)^T (x^k + \alpha d^k - x^k) + O(\alpha^2 \|d^k\|^2)$$

$$= f(x^k) + \alpha \nabla f(x^k)^T d^k + O(\alpha^2 \|d^k\|^2)$$

也称为最速下降法

当 α 足够小时, $d^k = -\nabla f(x^k)$ 会使得函数值下降速度最快

iff $\|\nabla f(x^k)\| = \|d^k\|$

$$\phi(\alpha) - f(x^k) \approx \alpha \nabla f(x^k)^T d^k \le \alpha \|\nabla f(x^k)\| \|d^k\|$$

Thm: 考虑二次, 正定(对称)问题 $f(x) = \frac{1}{2} x^T A x - b^T x$

① 精确搜索步长: $\alpha_k = \|\nabla f(x^k)\|_2^2 / \nabla f(x^k)^T A \nabla f(x^k)$

② Q-线性收敛: $\|x^{k+1} - x^*\|_A \le \left(\frac{\lambda_1 - \lambda_n}{\lambda_1 + \lambda_n}\right) \|x^k - x^*\|_A$

其中: λ_1, λ_n 对应 min, max eigenvalue of A. $\|x\|_A = \sqrt{x^T A x}$

Proof: 梯度: $\nabla f(x^k) = Ax^k - b$

迭代: $x^{k+1} = x^k - \alpha_k \nabla f(x^k)$

精确搜索步长: $\alpha^* = \arg\min_{\alpha > 0} \phi(\alpha) := f(x^k + \alpha d^k)$

$$\phi(\alpha) := f(x^k + \alpha d^k) = f(x^k - \alpha \nabla f(x^k))$$

$$= \frac{1}{2}(x^k - \alpha \nabla f(x^k))^T A (x^k - \alpha \nabla f(x^k)) - b^T (x^k - \alpha \nabla f(x^k))$$

$$= \frac{1}{2}(x^{kT} A x^k - 2\alpha x^{kT} A \nabla f(x^k) + \alpha^2 \nabla f(x^k)^T A \nabla f(x^k)) - b^T x^k + \alpha b^T \nabla f(x^k)$$

对 $\phi(\alpha)$ 求导

$A = A^T$

$$\phi'(\alpha) = -x^{kT} A \nabla f(x^k) + \alpha \nabla f(x^k)^T A \nabla f(x^k) + b^T \nabla f(x^k) = 0$$

$$\Rightarrow \alpha \nabla f(x^k)^T A \nabla f(x^k) = x^{kT} A \nabla f(x^k) - b^T \nabla f(x^k)$$

$$= (Ax^k - b)^T \nabla f(x^k)$$

$$= \nabla f(x^k)^T \nabla f(x^k) = \|\nabla f(x^k)\|_2^2$$

$$\Rightarrow \alpha = \frac{\|\nabla f(x^k)\|_2^2}{\nabla f(x^k)^T A \nabla f(x^k)}$$

② Q-线性收敛：$\|x^{k+1}-x^*\|_A \le \sqrt{\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)}\,\|x^k-x^*\|_A$

$$\Rightarrow \frac{\|x^{k+1}-x^*\|_A}{\|x^k-x^*\|_A} \le \sqrt{\left(\frac{\lambda_1-\lambda_n}{\lambda_1+\lambda_n}\right)^2}$$

$\lambda_1 \ll \lambda_n$ ① 比例越大 → 收敛速度越慢

② 比例越小 → 收敛速度越快

与 Hessian 矩阵的条件数有关

Defn (Condition number) 对可逆矩阵 A，其条件数为

$$\kappa(A) = \|A\|\cdot\|A\|^{-1}$$

→ 任意范数