优化模型与软件工具

主讲教师：董庆兴

华中师范大学 信息管理学院
qxdong@mail.ccnu.edu.cn

2017 年 11 月 21 日

大纲

最小元的对偶性质

- 我们可以用对偶广义不等式来刻画集合 $S \subseteq \mathbb{R}^m$（可能非凸）关于正常锥 K 导出的广义不等式的最小元和极小元
- $\mathbf{x}$ 是 S 上关于广义不等式 $\preceq_K$ 的最小元的充要条件是，对于所有的 $\lambda \succ_{K^*} 0$,$\mathbf{x}$ 是 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 的唯一最优解
- 几何上看，这意味着对于任意的 $\lambda \succ_{K^*} 0$ 超平面 $\{\mathbf{z}|\lambda^{\mathrm{T}}(\mathbf{z}-\mathbf{x})=0\}$ 是 $\mathbf{x}$ 处对 S 的一个严格支撑超平面（所谓严格就是与 S 只相交于 $\mathbf{x}$）
- 上述结论对 S 是不是凸集无要求

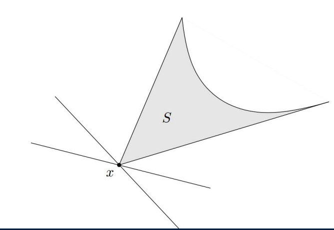
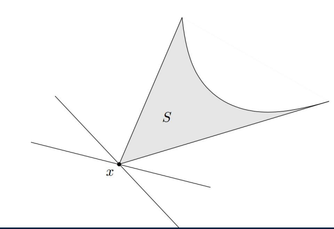


图: 最小元的对偶性质：点 $\mathbf{x}$ 是集合 S 中关于 $\mathbb{R}^2_+$ 的最小元 $\Longleftrightarrow$ 对于任意的 $\lambda \succ_{K^*} 0$ 超平面 $\{\mathbf{z}|\lambda^{\mathrm{T}}(\mathbf{z}-\mathbf{x})=0\}$ 在 $\mathbf{x}$ 处对 S 的一个严格支撑，即超平面规定的一个半空间包含了 S，且只在 $\mathbf{x}$ 处与 S 接触

最小元的对偶性质说明



- 设 $\mathbf{x}$ 是 S 的最小元，即对于任意 $\mathbf{z} \in S$ 有 $\mathbf{x} \preceq_K \mathbf{z}$，同时令 $\lambda \succ_{K^*} 0$，而 $\mathbf{z} \in S, \mathbf{z} \neq \mathbf{x}$。因为 x 是最小元，我们有 $\mathbf{z} - \mathbf{x} \succeq_K 0$

- 根据 $\lambda \succ_{K^*} 0, \mathbf{z} - \mathbf{x} \neq 0$，可以得到 $\lambda^{\mathrm{T}}(\mathbf{z} - \mathbf{x}) > 0$。因为 $\mathbf{z}$ 是 S 上任意一个不等于 $\mathbf{x}$ 的元素，所以 $\mathbf{x}$ 是在 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 的唯一解

- 反之，假设对于所有 $\lambda \succ_{K^*} 0$，$\mathbf{x}$ 是在 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 的唯一解，但 $\mathbf{x}$ 不是 S 的最小元，那么存在 $\mathbf{z} \in S$ 满足 $\mathbf{z} \not\succeq_K \mathbf{x}$。因为 $\mathbf{z} - \mathbf{x} \not\succeq_K 0$，存在 $\tilde{\lambda} \succ_{K^*} 0$ 且 $\tilde{\lambda}^{\mathrm{T}}(\mathbf{z} - \mathbf{x}) < 0$。这与 $\mathbf{x}$ 是 S 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 的唯一解矛盾

极小元的对偶性质

- 如果 $\lambda \succ_{K^*} 0$，$\mathbf{x}$ 在 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$，那么 $\mathbf{x}$ 是极小点
- 为说明这一点，假设 $\lambda \succ_{K^*} 0$ 并且 $\mathbf{x}$ 在 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$，但 $\mathbf{x}$ 不是极小元。也就是存在 $\mathbf{z} \in S$ 满足 $\mathbf{z} \neq x, \mathbf{z} \preceq_K \mathbf{x}$，那么有 $\lambda^{\mathrm{T}}(\mathbf{x} - \mathbf{z}) > 0$，这与我们的假设 $\mathbf{x}$ 在 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 矛盾

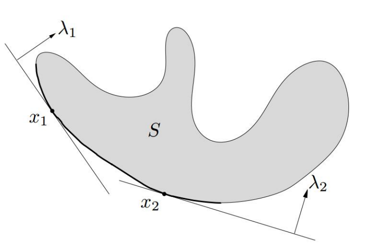


图: 集合 $S \subseteq \mathbb{R}^2$。其中关于 $\mathbb{R}^2_+$ 的极小点集合由其边界的（左下）的深色部分所示。S 上极小化 $\lambda_1^{\mathrm{T}}\mathbf{z}$ 的解为 $\mathbf{x}_1$，因为 $\lambda_1 \succ 0$，所以 $\mathbf{x}_1$ 是极小的。S 上极小化 $\lambda_2^{\mathrm{T}}\mathbf{z}$ 的解为 $\mathbf{x}_2$，因为 $\lambda_2 \succ 0$，所以它是另外一个极小点

极小元的对偶性质逆命题

- 极小元对偶性质的逆命题一般不成立：S 上的极小元 $\mathbf{x}$ 可以对于任何 λ 都不是 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 的解
- 当凸性成立的时候，该逆定理是成立的。假设 S 是凸集，那么对于任意极小元 $\mathbf{x}$，存在非零 $\lambda \succeq_{K^*} 0$ 使得 $\mathbf{x}$ 在 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$

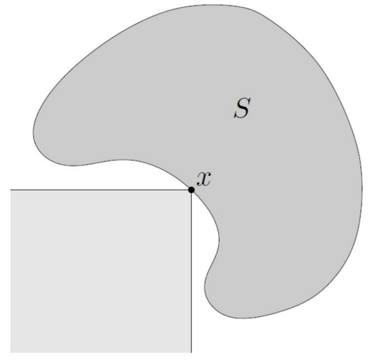
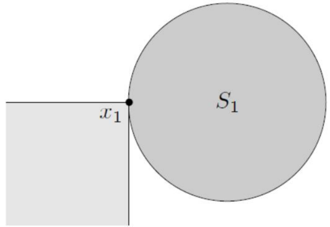


图: 点 $\mathbf{x}$ 是 $S \subseteq \mathbb{R}^2$ 关于 $\mathbb{R}^2_+$ 的极小元。但是不存在 λ 使得 $\mathbf{x}$ 在 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$

凸集极小元对偶性质逆命题 1

- 设 $\mathbf{x}$ 是 S 的极小元，也就是说 $[(\mathbf{x}-K)\setminus\{\mathbf{x}\}]\cap S=\emptyset$



图：点 $\mathbf{x}_1\in S_1$ 是极小的，但对于任意 $\lambda\succ 0$ ，它都没有在 S_1 上极小化 $\lambda^{\mathrm{T}}z$。但是对于 $\lambda=(1,0)$ 它确实在所有 $\mathbf{z}\in S_1$ 中极小化了 $\lambda^{\mathrm{T}}\mathbf{z}$

- 对凸集 $(\mathbf{x}-K)\setminus\{\mathbf{x}\}$ 和 S 应用超平面分离定理，我们可以得出：存在 $\lambda\neq 0$ 和 μ，使得对于所有 $\mathbf{y}\in K$ 有 $\lambda^{\mathrm{T}}(\mathbf{x}-\mathbf{y})\leq\mu$, 对于所有 $\mathbf{z}\in S$ 有 $\lambda^{\mathrm{T}}\mathbf{z}\geq\mu$
- 根据第一个不等式，可知 $\lambda\succeq_{K^*}0$ (广义不等式对偶性质得到)。由于 $\mathbf{x}\in S$ 和 $\mathbf{x}\in\mathbf{x}-K$，我们有 $\lambda^{\mathrm{T}}\mathbf{x}=\mu$, 所以第二个不等式表明 μ 是 S 上 $\lambda^{\mathrm{T}}\mathbf{z}$ 的最小值。因此 $\mathbf{x}$ 是 S 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 的一个解，这里 $\lambda\neq 0,\lambda\succeq_{K^*}0$
- 上述逆定理无法加强为 $\lambda\succ_{K^*}0$ 反例表明，凸集 S 上的极小元 $\mathbf{x}$ 可以对于任意 $\lambda\succ_{K^*}0$ 都不是极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 的解

凸集极小元对偶性质逆命题 2

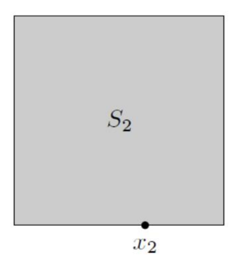


图: 点 $\mathbf{x}_2 \in S_2$ 不是极小的，但是对于 $\lambda = (1, 0)$ 它确实在所有 $\mathbf{z} \in S_2$ 中极小化了 $\lambda^{\mathrm{T}}\mathbf{z}$

- 同时，并不是对于任意 $\lambda \succeq_{K^*} 0$，在 $\mathbf{z} \in S$ 上极小化 $\lambda^{\mathrm{T}}\mathbf{z}$ 的解都一定是极小的

Pareto 最优制造前沿

- 考虑安排一个产品的生产，需要 n 种资源，有多种制造方式。用资源向量$\mathbf{x} \in \mathbb{R}^n$ 表示各种制造方法，其中 $x_i \geq 0$ 表示制造产品时消耗资源 i 的数量并且越少越好，生产集合$P \subseteq \mathbb{R}^n$ 定义为所有资源向量 $\mathbf{x}$ 的集合,P 上的极小元（在分量不等式的意义下）对应的制造方法称为Pareto 最优（有效）,P 的极小元构成的集合叫做有效制造前沿
- 资源向量 $\mathbf{x}$ 比与资源向量 $\mathbf{y}$更好意味着 $\forall i, x_i \leq y_i$ 并且存在某些 $i, x_i < y_i$，即 $\mathbf{x} \preceq \mathbf{y}, \mathbf{x} \neq \mathbf{y}$
- 我们可以通过在 P 上对任意满足 $\lambda \succ 0$ 的 λ 极小化 $\lambda^{\mathrm{T}}\mathbf{x}$ 来得到 Pareto 最优制造方法。这里 λ 有一个简单解释：λ_i 是资源 i 的价格

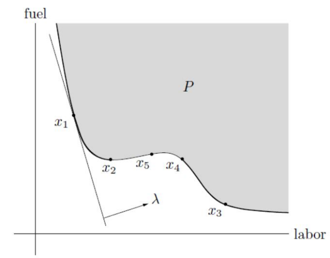


图: 制造集合 P 如阴影所示，表示制造产品所需要的劳动力和燃料。两端深色曲线表示了有效制造前沿。点 $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ 是有效的。点 $\mathbf{x}_4, \mathbf{x}_5$ 不是（$\mathbf{x}_2$ 对应了更少燃料更少人力的方法）。点 $\mathbf{x}_1$ 是对应于（正的）价格向量 λ 的最小成本制造方法。点 $\mathbf{x}_2$ 是有效的，但是对于任意价格向量 $\lambda \succeq 0$ 都无法通过极小化总成本 $\lambda^{\mathrm{T}}\mathbf{x}$ 找到 $\mathbf{x}_2$

极小值定义

局部极小值

假设 $f:\mathbb{R}^n\to\mathbb{R}$，对于 $\mathbf{x}^*\in\mathbf{dom}\ f$，如果存在某个 $\epsilon>0$。使得所有与 $\mathbf{x}^*$ 距离小于 ϵ（即 $\|\mathbf{x}-\mathbf{x}^*\|\le\epsilon$）的 $\mathbf{x}\in\mathbf{dom}\ f$, 均满足不等式 $f(\mathbf{x})\ge f(\mathbf{x}^*)$，则称 $\mathbf{x}^*$ 为局部极小点，$f(\mathbf{x}^*)$ 为局部极小值

全局极小值

假设 $f:\mathbb{R}^n\to\mathbb{R}$，对于 $\mathbf{x}^*\in\mathbf{dom}\ f$，如果对于所有的的 $\mathbf{x}\neq\mathbf{x}^*\in\mathbf{dom}\ f$, 均满足不等式 $f(\mathbf{x})\ge f(\mathbf{x}^*)$，则称 $\mathbf{x}^*$ 为全局极小点，$f(\mathbf{x}^*)$ 为全局极小值

必要条件 $\nabla f(\mathbf{x})=\mathbf{0}$，此点被称为平稳点或者驻点，极小则需 $\mathbf{H}(\mathbf{x})$ 半正定。不充分：$f(x)=x^3, f'(x)=0, f''(x)=0$ 但不是极值点

充分条件 $\nabla f(\mathbf{x})=\mathbf{0}$ 且 $\mathbf{H}(\mathbf{X})$ 正定, 则为严格局部极小点。不必要：$f(x)=x^4, \bar{x}=0$ 是极值点，但是 $f''(x)=0$

凸函数的极值

凸函数的局部极值即全局极值

假设 $f:\mathbb{R}^n\to\mathbb{R}$ 为凸函数，对于 $\mathbf{x}^*\in \textbf{dom}\ f$，如果 $\mathbf{x}^*$ 为 f 的局部极小点，则它就是 f 的全局极小点

证明：设 $\mathbf{x}^*$ 为 f 的局部极小点，则对于充分小的邻域 $N_\delta(\mathbf{x}^*)$ 中的一切 $\mathbf{x}$，均有 $f(\mathbf{x})\geq f(\mathbf{x}^*)$.$\forall\ \mathbf{y}\in \textbf{dom}\ f$，对于充分小的 $\lambda\in(0,1)$，有

$$[(1-\lambda)\mathbf{x}^*+\lambda\mathbf{y}]\in N_\delta(\mathbf{x}^*)$$

从而有

$$f[(1-\lambda)\mathbf{x}^*+\lambda\mathbf{y}]\geq f(\mathbf{x}^*)$$

由函数凸性可得

$$(1-\lambda)f(\mathbf{x}^*)+\lambda f(\mathbf{y})\geq f[(1-\lambda)\mathbf{x}^*+\lambda\mathbf{y}]\geq f(\mathbf{x}^*)$$

从而可得 $f(\mathbf{y})\geq f(\mathbf{x})$ ■

凸函数的极值判定

凸函数全局极值判定条件

假设 $f:\mathbb{R}^n\rightarrow\mathbb{R}$ 为凸函数，对于 $\mathbf{x}^*\in\mathbf{dom}\ f$，使得对于任意 $\mathbf{x}\in\mathbf{dom}\ f$, 有

$$\nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x}-\mathbf{x}^*)\geq 0$$

则 $\mathbf{x}^*$ 为 f 的全局极小点

证明: 由凸函数的一阶判定条件可知

$$f(\mathbf{x})\geq f(\mathbf{x}^*)+\nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x}-\mathbf{x}^*)$$

因此如果有 $\nabla f(\mathbf{x}^*)^{\mathrm{T}}(\mathbf{x}-\mathbf{x}^*)\geq 0$, 则有

$$f(\mathbf{x})\geq f(\mathbf{x}^*),\forall\ \mathbf{x}\in\mathbf{dom}\ f$$

■

当 $\mathbf{x}^*\in\mathbf{dom}\ f$ 是内点时，意味着 $\nabla f(\mathbf{x}^*)^{\mathrm{T}}=0$

利用极值条件求解极值点

例题

利用极值条件求解：$\min f(\mathbf{x}) = \frac{1}{3}x_1^3 + \frac{1}{3}x_2^3 - x_1^2 - x_2$

解：直接计算，

$$\nabla f(\mathbf{x}) = \begin{pmatrix} x_1^2 - 2x_1 \\ x_2^2 - 1 \end{pmatrix} \quad \nabla^2 f(\mathbf{x}) = \begin{pmatrix} 2x_1 - 2 & 0 \\ 0 & 2x_2 \end{pmatrix}$$

由一阶必要条件 $\nabla f(\mathbf{x}) = 0$ 可得驻点 $\mathbf{x}^{(1)} = (0,1)^{\mathrm{T}}, \mathbf{x}^{(2)} = (0,-1)^{\mathrm{T}}, \mathbf{x}^{(3)} = (2,1)^{\mathrm{T}}$, $\mathbf{x}^{(4)} = (2,-1)^{\mathrm{T}}$, 对应的 Hessian 矩阵为

$$\nabla^2 f(\mathbf{x}^{(1)}) = \begin{pmatrix} -2 & 0 \\ 0 & 2 \end{pmatrix}, \nabla^2 f(\mathbf{x}^{(2)}) = \begin{pmatrix} -2 & 0 \\ 0 & -2 \end{pmatrix},$$

$$\nabla^2 f(\mathbf{x}^{(3)}) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \nabla^2 f(\mathbf{x}^{(4)}) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$$

$\nabla^2 f(\mathbf{x}^{(3)})$ 正定，因此 $\mathbf{x}^{(3)}$ 是一个严格局部最优解，其余点的 Hesse 矩阵都不是半正定的

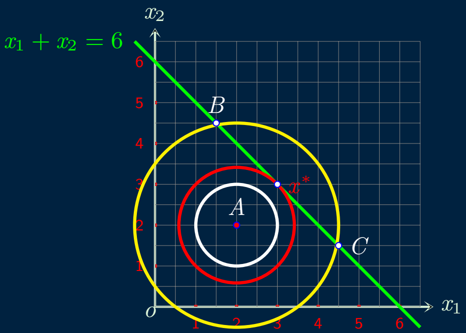
微积分方法的局限性

- 实际问题中，函数可能是不连续或者不可微的
- 需要解复杂的方程组，而方程组到目前仍没有有效的算法
- 实际的问题可能含有不等式约束，微积分的方法不易处理

图解法

例题

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) = (x_1-2)^2 + (x_2-2)^2 \\ \text{subject to} \quad & f_1(\mathbf{x}) = x_1 + x_2 - 6 = 0 \end{aligned}$$



下降方向

下降方向

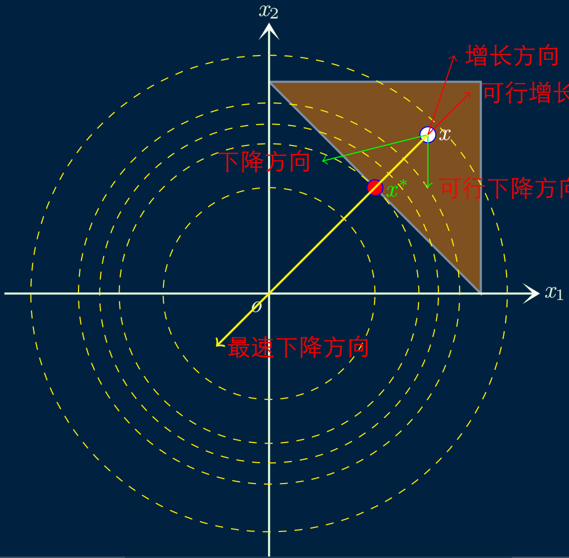
假设 $f:\mathbb{R}^n \to \mathbb{R}$，对于 $\mathbf{x} \in \mathbf{dom}\ f$，使得对于任意 $\bar{\alpha} > 0, \mathbf{d} \in \mathbb{R}^n$, 有 $f(\mathbf{x}+\alpha\mathbf{d}) < f(\mathbf{x}), \alpha \in (0,\bar{\alpha})$ 则 $\mathbf{d}$ 为 f 的一个下降方向

由泰勒展开可知 $f(\mathbf{x}+\alpha\mathbf{d}) = f(\mathbf{x}) + \alpha\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{d} + o(\alpha)$，因此满足 $\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{d} < 0$ 的 $\mathbf{d}$ 为 f 的一个下降方向

可行方向

假设 $f:\mathbb{R}^n \to \mathbb{R}$，对于 $\mathbf{x} \in \mathbf{dom}\ f$，若存在 $\alpha > 0, \mathbf{d} \in \mathbb{R}^n$, 使得 $f(\mathbf{x}+\alpha\mathbf{d}) \in \mathbf{dom}\ f$ 则 $\mathbf{d}$ 为 f 的一个可行方向

下降方向



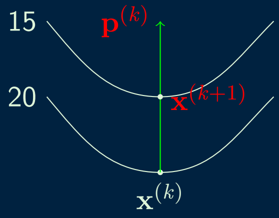
梯度方向

- 由微积分的基本知识可知，$\nabla f(\mathbf{x})$ 的方向是 $f(\mathbf{x})$ 的等值面（等值线）在点 $\mathbf{x}$ 处的法线方向
- 例：$f(x,y,z)=a_1x+a_2y+a_3z$ 的梯度为 $\nabla f(x,y,z)=(a_1,a_2,a_3)^{\mathrm{T}}$，恰好是 w 的一个等值面 $c=a_1x+a_2y+a_3z$ 的法线
- 梯度方向是函数具有最大变化率的方向（负梯度方向也叫最速下降方向）

数值最优化方法的基本思路

- 从某个初始点 $\mathbf{x}^{(0)}$ 出发，按某种算法找出点列 $\{\mathbf{x}^{(k)}\}$，对于最小值问题来讲满足 $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)}), \forall k = 0, 1, \cdots$
- 如果 $\{\mathbf{x}^{(k)}\}$ 是有限的，则 $\{\mathbf{x}^{(k)}\}$ 最后一点为最优解
- 如果 $\{\mathbf{x}^{(k)}\}$ 是无限的且收敛于 $\mathbf{x}^*$，即 $\lim\limits_{k\to\infty} \|\mathbf{x}^{(k)} - \mathbf{x}^*\| = 0$，则以此极限点为最优解（近似最优解）

下降算法



- 假定已经迭代到 $\mathbf{x}^{(k)}$，如果此时没有下降方向（沿任何方向移动都无法使目标函数值减小），则 $\mathbf{x}^{(k)}$ 是一个局部极小点，迭代停止
- 如果从 $\mathbf{x}^{(k)}$ 出发至少有一个方向是下降方向 $\mathbf{p}^{(k)}$，则沿该方向迈进适当一步，得到下一个迭代点 $\mathbf{x}^{(k+1)}$ 并使得 $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$
- 相当于在射线 $\mathbf{x} = \mathbf{x}^{(k)} + \lambda\mathbf{p}^{(k)}$ 上选定新点 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k\mathbf{p}^{(k)}$，其中 λ_k 叫做步长因子,$\mathbf{p}^{(k)}$ 为搜索方向

下降迭代算法步骤

1. 给出初始点 $\mathbf{x}^{(0)}$， $k \leftarrow 0$
2. 判断 $\mathbf{x}^{(k)}$ 是否为极小点或者近似极小点，是则停止，否则转第 3 步
3. 按照某种规则确定搜索方向 $\mathbf{p}^{(k)}$
4. 按照某种规则确定搜索步长 λ_k, 得到 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \lambda_k \mathbf{p}^{(k)}$， 使得 $f(\mathbf{x}^{(k)} + \lambda_k \mathbf{p}^{(k)}) < f(\mathbf{x}^{(k)})$， 转第 2 步

由上述步骤可知，确定搜索方向和搜索步长是非常关键的步骤，遵循不同的规则，就可以得到不同的算法

线性搜索

- 确定步长的一种符合直觉的做法是求使得目标函数值下降最多的 λ_k，也就是 $\lambda_k = \arg\min f(\mathbf{x}^{(k)} + \lambda \mathbf{p}^{(k)})$
- 由于这一过程是求解以 λ 为变量的一元函数 $f(\mathbf{x}^{(k)} + \lambda \mathbf{p}^{(k)})$ 的极小点 λ_k, 因此本方法被称作线性搜索或者一维搜索，这样确定的步长即为最优步长
- 线性搜索性质：在搜索方向上所得最优点处的梯度与搜索方向正交，即 $\nabla f(\mathbf{x}^{(k+1)})^{\mathrm{T}} \mathbf{p}^{(k)} = 0$

证明: 构造函数 $\phi(\lambda) = f(\mathbf{x}^{(k)} + \lambda \mathbf{p}^{(k)})$ ，从而由最优解的一阶条件可得 $\phi'(\lambda) = \nabla f(\mathbf{x}^{(k)} + \lambda \mathbf{p}^{(k)})^{\mathrm{T}} \mathbf{p}^{(k)} = \nabla f(\mathbf{x}^{(k+1)})^{\mathrm{T}} \mathbf{p}^{(k)} = 0$ ■