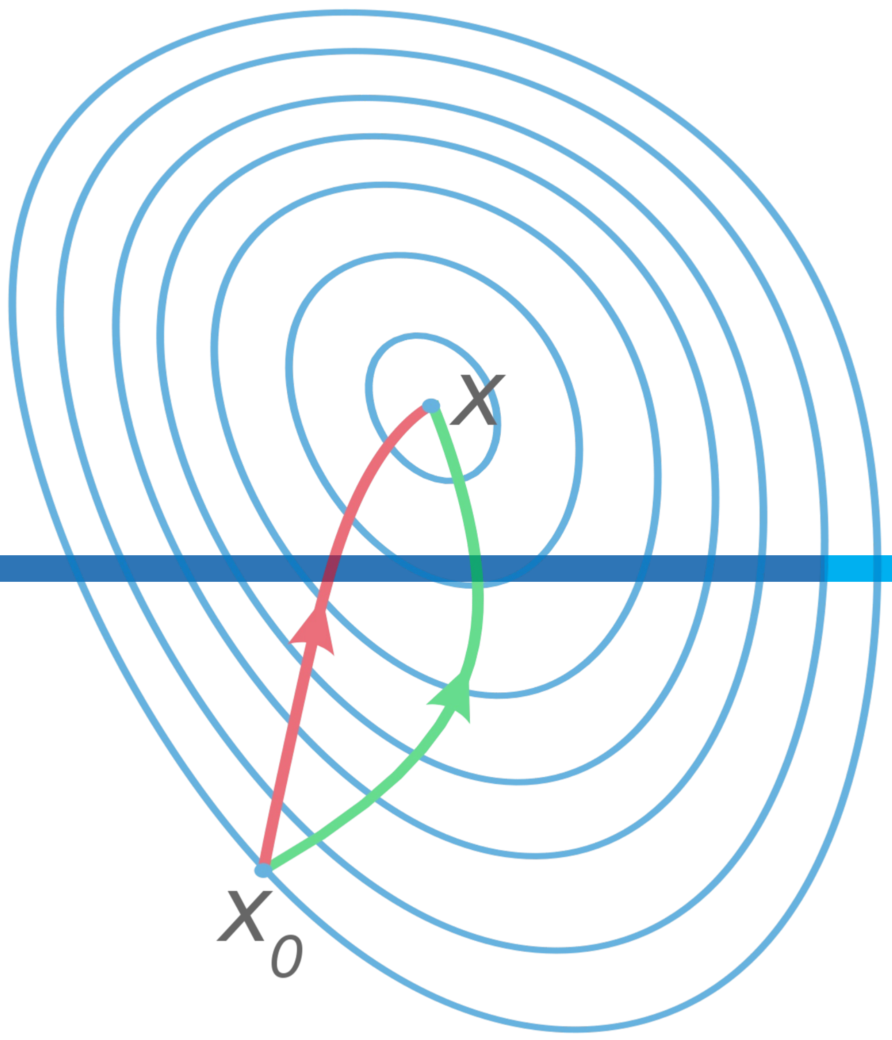
最优化方法

第四周



计算机学院
余皓然

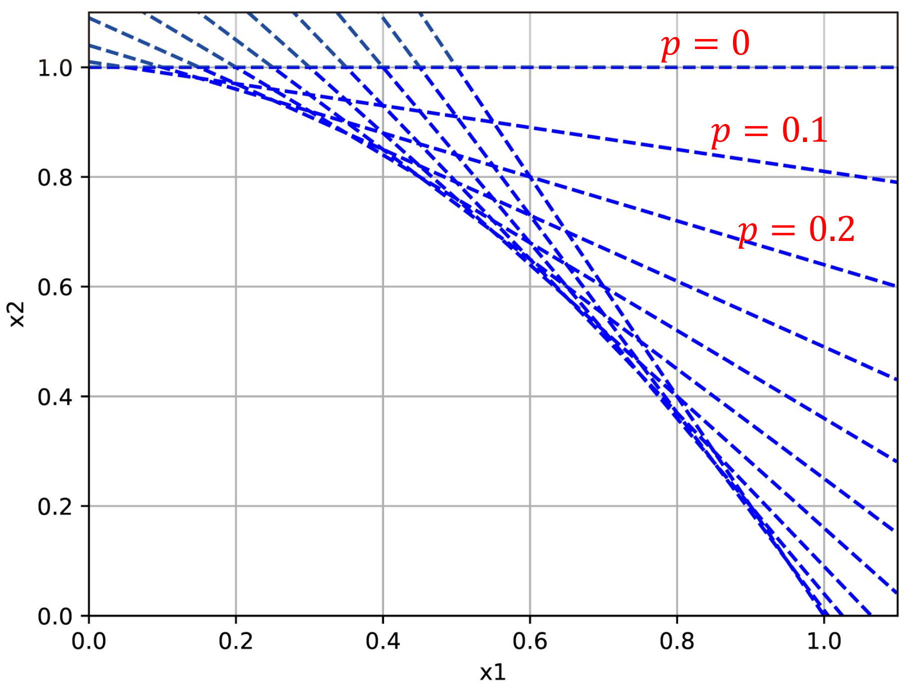
2024/3/18

一种简单的内点法（续）

Part1 如何从复杂度的角度衡量算法好坏？
Part2 针对一些有特定形式的最优化问题，如何找到最优解？
无约束凸优化问题、梯度下降法、牛顿法； **线性规划问题、单纯形法、内点法；** **有约束凸优化问题、KKT条件、内点法**
Part3 针对较复杂的最优化问题（如NP难问题），如何找到较好解？
Part4 针对较复杂的多目标最优化问题，如何找到较好解？

例子

$$\begin{aligned}
&\text{min} \quad -x_1 - x_2 \\
&\text{s.t.} \quad 2px_1 + x_2 \le p^2 + 1, \forall p \in [0,0.1,\dots,1], \\
&\text{var.} \quad x_1 \ge 0, x_2 \ge 0.
\end{aligned}$$



例子

内点法（牛顿法计算每个t下的解）

$$\min \quad t\boldsymbol{c}^T\boldsymbol{x} - \sum_{i=1}^{m} \log\left(-\sum_{j=1}^{n} A_{ij}x_j + b_i\right)$$
$$\text{var.} \quad \boldsymbol{x}.$$

$$x^{(k)} = x^{(k-1)} - \left(\nabla^2 f(x^{(k-1)})\right)^{-1} \nabla f(x^{(k-1)}),$$

设置：

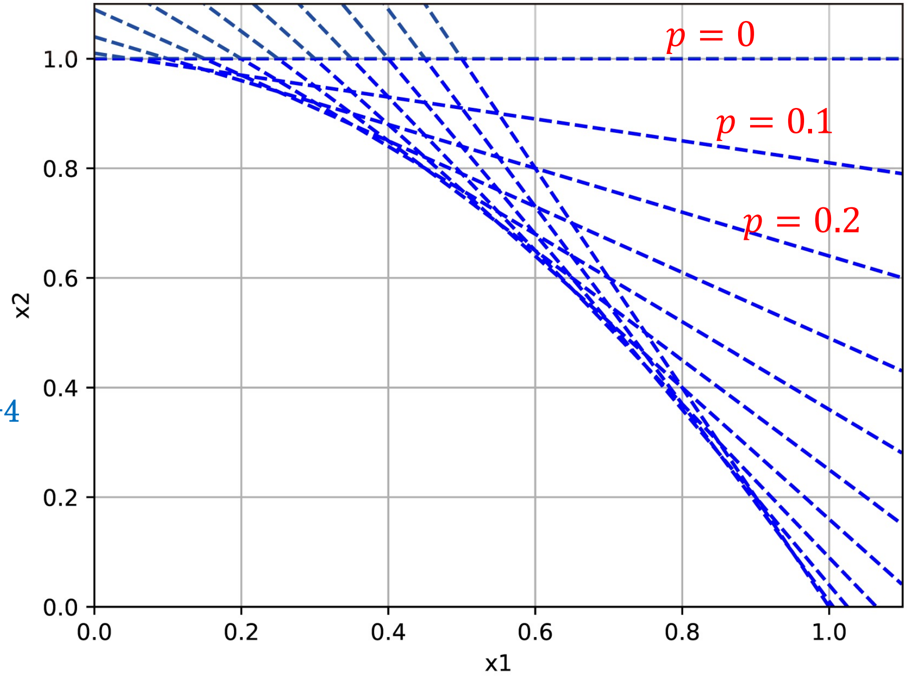
初始解$(10^{-4}, 10^{-4})$

初始t=1

参数μ=15　t更新规则改为 $t \leftarrow t + \mu$

外迭代终止条件　相邻t的两个解距离<10^{-4}

内迭代终止条件　相邻两个解距离<10^{-4}



例子

内点法（牛顿法计算每个t下的解）

$$\begin{aligned}&\min \quad t\boldsymbol{c}^T\boldsymbol{x}-\textstyle\sum_{i=1}^{m}\log\left(-\sum_{j=1}^{n}A_{ij}x_j+b_i\right)\\ &\text{var.} \quad \boldsymbol{x}.\end{aligned}$$

$$x^{(k)}=x^{(k-1)}-\left(\nabla^2 f(x^{(k-1)})\right)^{-1}\nabla f(x^{(k-1)}),$$

设置：

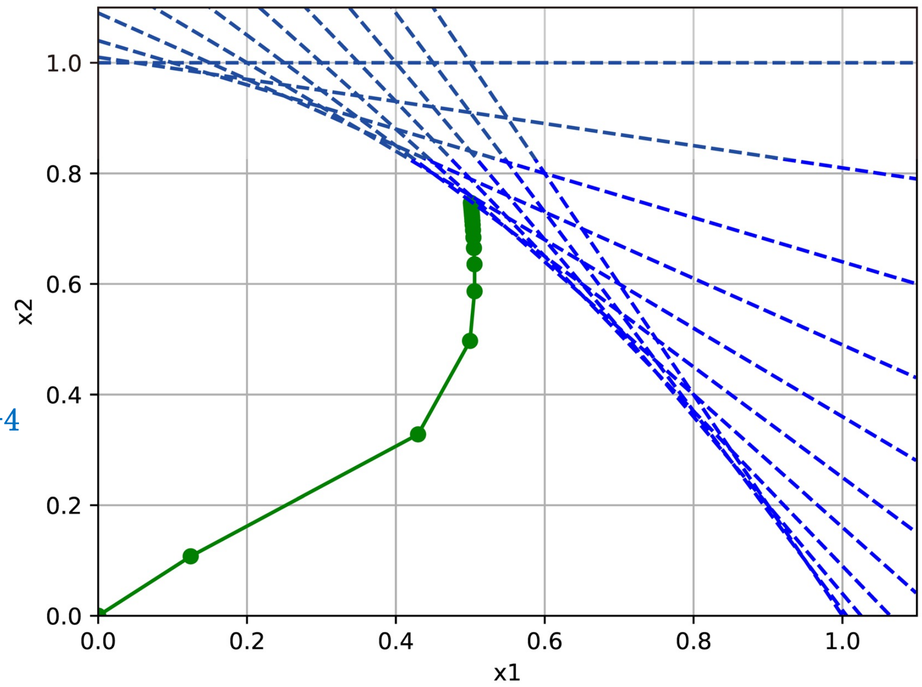
初始解$(10^{-4}, 10^{-4})$

初始t=1

参数μ=15　t更新规则改为$t \leftarrow t+\mu$

外迭代终止条件　相邻t的两个解距离$<10^{-4}$

内迭代终止条件　相邻两个解距离$<10^{-4}$



例子

内点法（梯度下降法计算每个t下的解）

$$\min \quad t\boldsymbol{c}^T\boldsymbol{x} - \sum_{i=1}^{m} \log\left(-\sum_{j=1}^{n} A_{ij}x_j + b_i\right)$$
$$\text{var.} \quad \boldsymbol{x}.$$

$$x^{(k)} = x^{(k-1)} - \alpha_k \nabla f(x^{(k-1)}),$$

设置：

初始解$(10^{-4}, 10^{-4})$

初始t=1

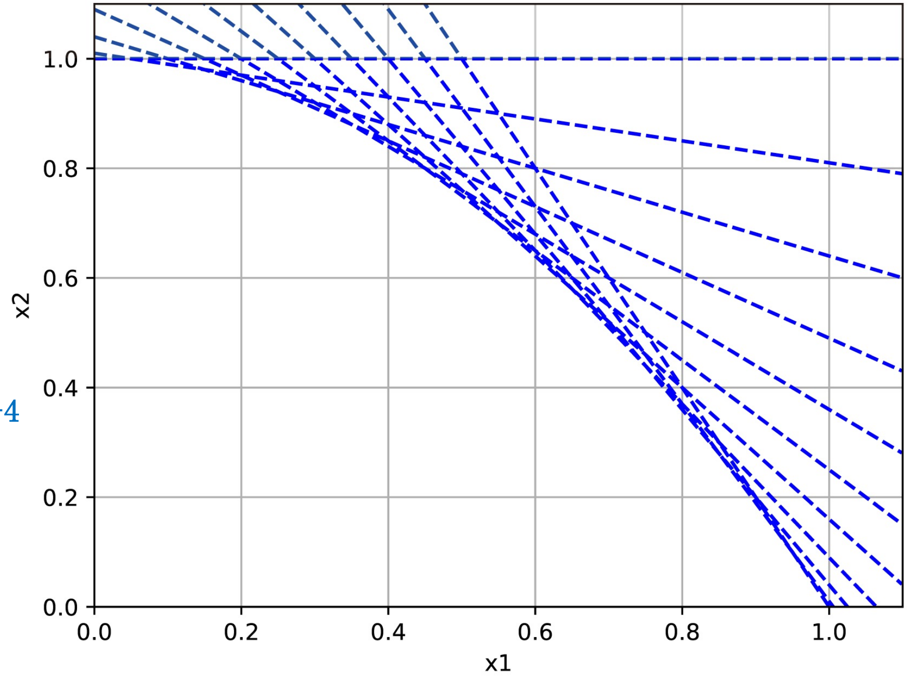
参数μ=15　t更新规则改为$t \leftarrow t + \mu$

外迭代终止条件　相邻t的两个解距离<10^{-4}

内迭代终止条件　相邻两个解距离<10^{-4}

或超5000次

参数$\alpha=10^{-5}$



例子

内点法（梯度下降法计算每个t下的解）

$$\min \quad t\boldsymbol{c}^T\boldsymbol{x} - \sum_{i=1}^{m}\log\left(-\sum_{j=1}^{n}A_{ij}x_j + b_i\right)$$
$$\text{var.} \quad \boldsymbol{x}.$$

$$x^{(k)} = x^{(k-1)} - \alpha_k \nabla f(x^{(k-1)}),$$

设置：

初始解$(10^{-4}, 10^{-4})$

初始t=1

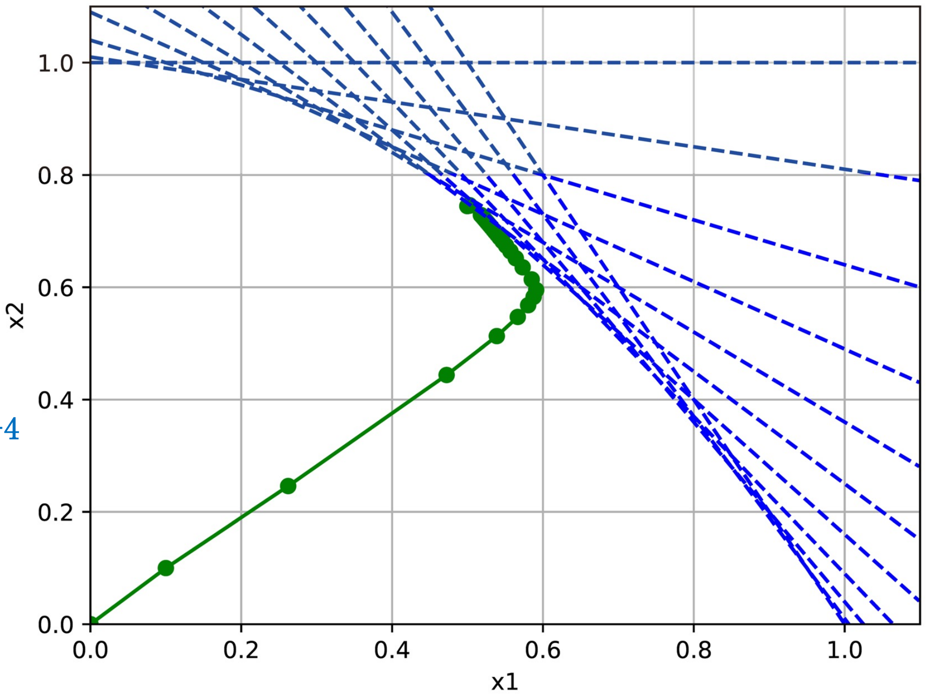
参数μ=15　t更新规则改为$t \leftarrow t + \mu$

外迭代终止条件　相邻t的两个解距离$<10^{-4}$

内迭代终止条件　相邻两个解距离$<10^{-4}$

或超5000次

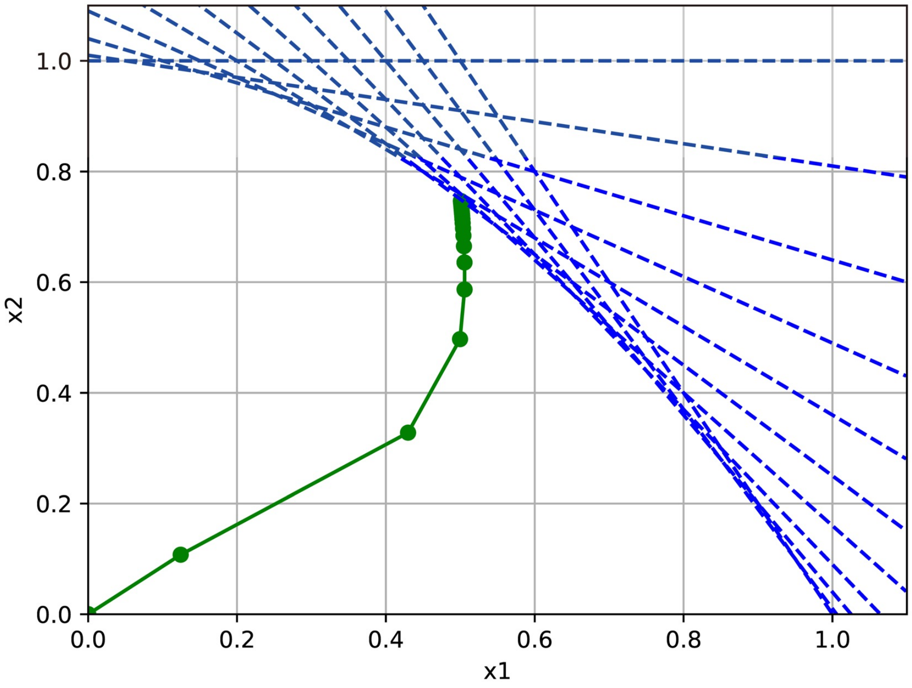
参数$\alpha=10^{-5}$



例子

内点法（牛顿法计算每个t下的解）

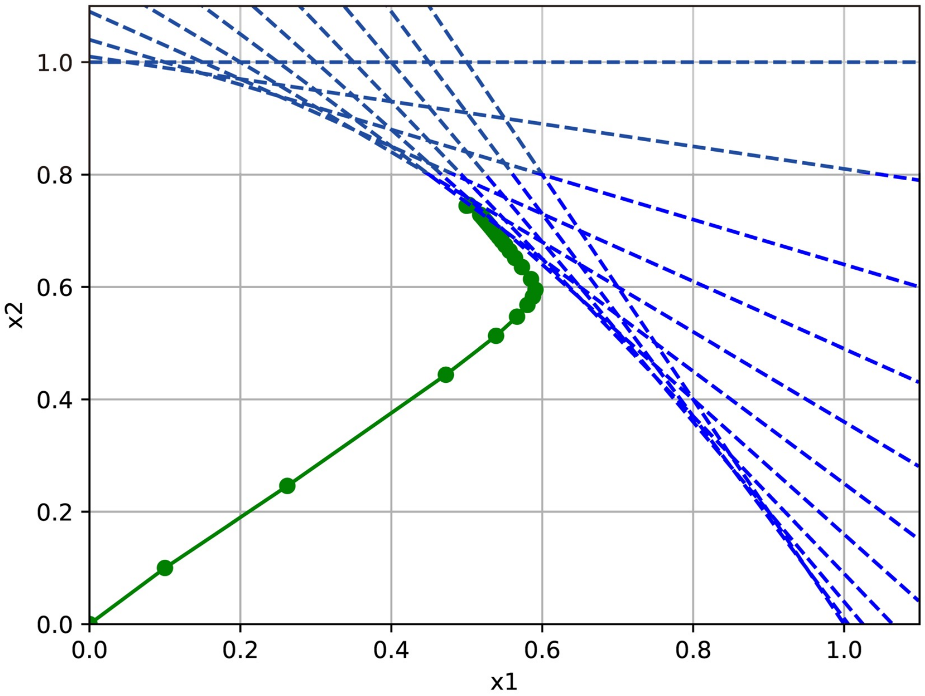
$$x^{(k)} = x^{(k-1)} - \left(\nabla^2 f(x^{(k-1)})\right)^{-1} \nabla f(x^{(k-1)}),$$



0.011s

内点法（梯度下降法计算每个t下的解）

$$x^{(k)} = x^{(k-1)} - \alpha_k \nabla f(x^{(k-1)}),$$



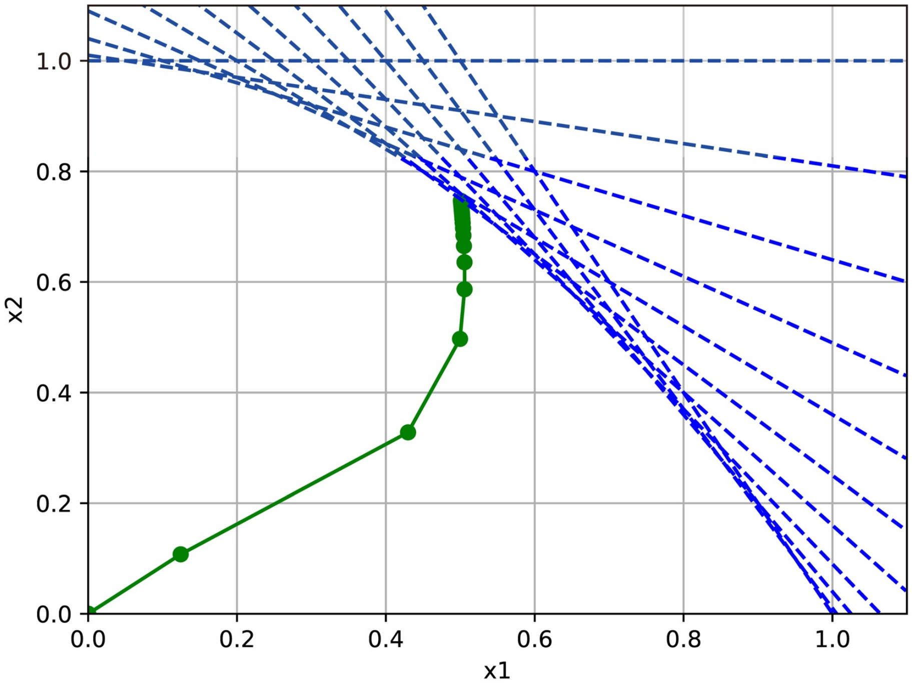
1.707s

同样t值下$\boldsymbol{x}^*(t)$应该相等，因为内迭代终止条件的影响，左右两图结果不同
若将内迭代终止条件改为“相邻两个解距离<10^{-5}”，左右图结果将更相似

例子

内点法（牛顿法计算每个t下的解）

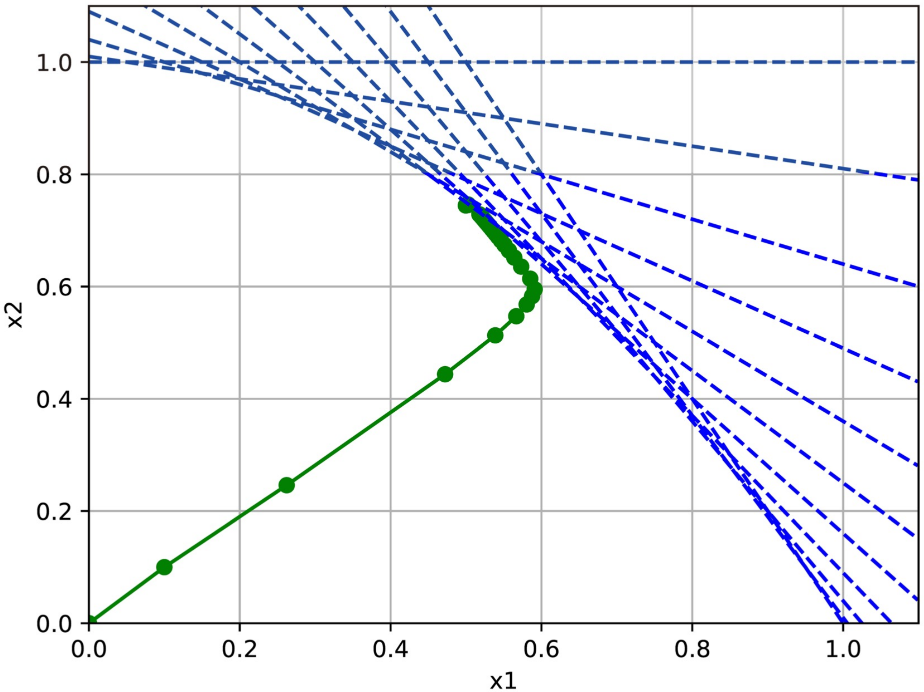
$$x^{(k)} = x^{(k-1)} - \left(\nabla^2 f(x^{(k-1)})\right)^{-1} \nabla f(x^{(k-1)}),$$



0.011s

内点法（梯度下降法计算每个t下的解）

$$x^{(k)} = x^{(k-1)} - \alpha_k \nabla f(x^{(k-1)}),$$



1.707s

线性规划

最优化问题的一般形式

$$
\begin{array}{ll}
\min & f(\boldsymbol{x}) \\
\text{s.t.} & g_i(\boldsymbol{x}) \leq 0, i = 1, \ldots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \ldots, p, \\
\text{var.} & \boldsymbol{x} \in \mathcal{D}.
\end{array}
$$

线性规划问题

$$
\begin{array}{ll}
\min & \boldsymbol{a}_0^T \boldsymbol{x} + b_0 \\
\text{s.t.} & \boldsymbol{a}_i^T \boldsymbol{x} + b_i \leq 0, i = 1, \ldots, m, \\
& \boldsymbol{d}_j^T \boldsymbol{x} + c_j = 0, j = 1, \ldots, p, \\
\text{var.} & \boldsymbol{x}.
\end{array}
$$

当目标和约束都是线性时

很多有约束的最优化问题不属于线性规划

有约束凸优化问题

Part1 如何从复杂度的角度衡量算法好坏？
Part2 针对一些有特定形式的最优化问题，如何找到最优解？
无约束凸优化问题、梯度下降法、牛顿法； **线性规划问题、单纯形法、内点法；** **有约束凸优化问题、KKT条件、内点法**
Part3 针对较复杂的最优化问题（如NP难问题），如何找到较好解？
Part4 针对较复杂的多目标最优化问题，如何找到较好解？

有约束凸优化

可行域是凸集

无约束凸优化问题
$\min \quad f(\boldsymbol{x})$
$\text{var.} \quad \boldsymbol{x} \in \mathcal{R}^n.$
其中，$f(\boldsymbol{x})$是凸函数

有约束凸优化问题
$\min \quad f(\boldsymbol{x})$
$\text{var.} \quad \boldsymbol{x} \in \mathcal{C}.$
其中，$f(\boldsymbol{x})$是凸函数，$\mathcal{C}$是凸集合

当$f(\boldsymbol{x})$一阶/二阶连续可微时，
可用梯度下降/牛顿法等求解

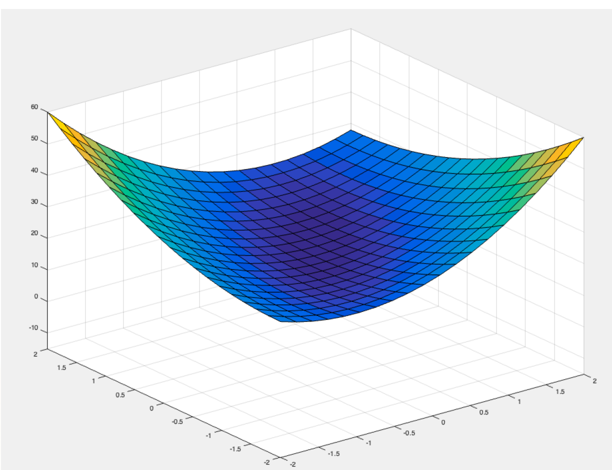
有约束凸优化

可行域是凸集

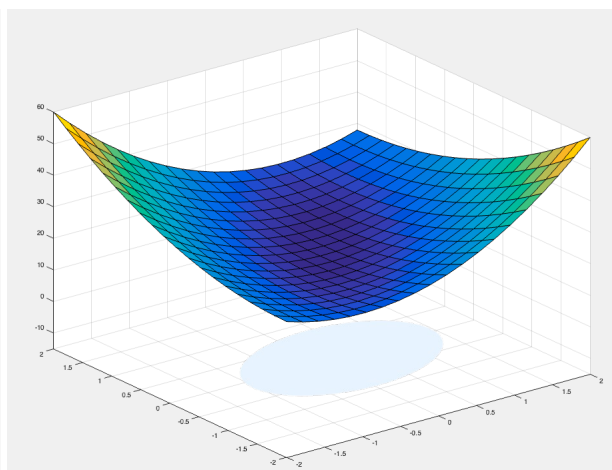
无约束凸优化问题
min $\quad f(\boldsymbol{x})$
var. $\quad \boldsymbol{x} \in \mathcal{R}^n$.
其中，$f(\boldsymbol{x})$是凸函数

有约束凸优化问题
min $\quad f(\boldsymbol{x})$
var. $\quad \boldsymbol{x} \in \mathcal{C}$.
其中，$f(\boldsymbol{x})$是凸函数，$\mathcal{C}$是凸集合

当$f(\boldsymbol{x})$一阶/二阶连续可微时，
可用梯度下降/牛顿法等求解



对$\boldsymbol{x}$取值范围没有限制



限制$\boldsymbol{x}$从蓝色区域取值

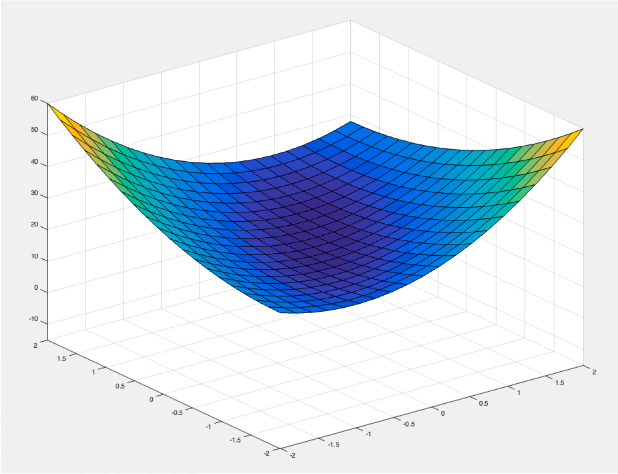
有约束凸优化

可行域是凸集

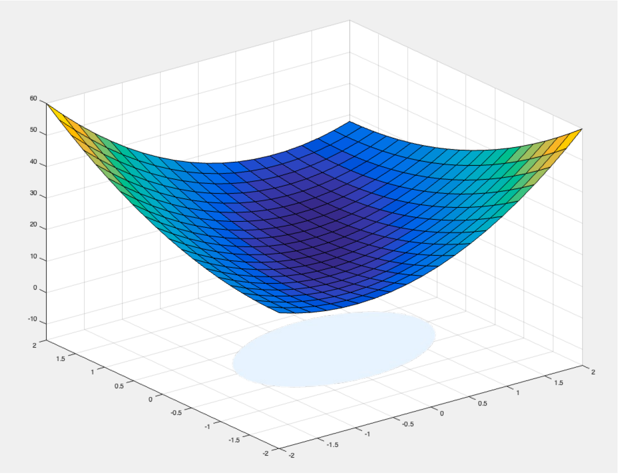
无约束凸优化问题
min　$f(\boldsymbol{x})$
var.　$\boldsymbol{x} \in \mathcal{R}^n$.
其中，$f(\boldsymbol{x})$是凸函数

有约束凸优化问题
min　$f(\boldsymbol{x})$
var.　$\boldsymbol{x} \in \mathcal{C}$.
其中，$f(\boldsymbol{x})$是凸函数，$\mathcal{C}$是凸集合

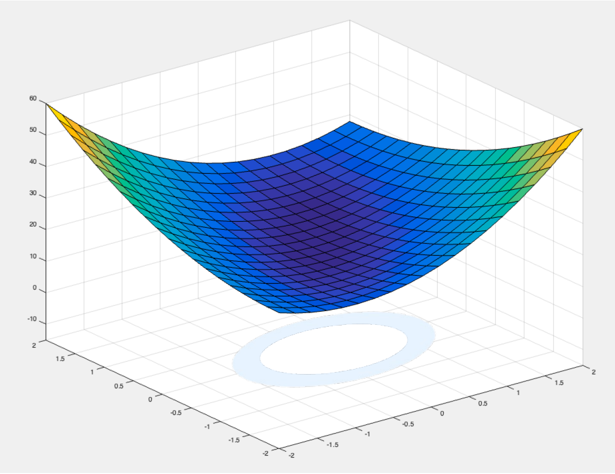
当$f(\boldsymbol{x})$一阶/二阶连续可微时，
可用梯度下降/牛顿法等求解



对$\boldsymbol{x}$取值范围没有限制



限制$\boldsymbol{x}$从蓝色区域取值



非凸集合时

有约束凸优化

一般可整理为

有约束凸优化问题

$$\min \quad f(\boldsymbol{x})$$
$$\text{var.} \quad \boldsymbol{x} \in \mathcal{C}.$$

其中，$f(\boldsymbol{x})$是凸函数，$\mathcal{C}$是凸集合

有约束凸优化问题的标准形式

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \leq 0, i = 1, \dots, m, \\ & h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\ \text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n. \end{aligned}$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是凸函数，$h_j(\boldsymbol{x})$是仿射函数

仿射函数：可写成$\boldsymbol{a}_j^T \boldsymbol{x} + b_j$的形式

有约束凸优化

一般可整理为

有约束凸优化问题

$$\min \quad f(\boldsymbol{x})$$
$$\text{var.} \quad \boldsymbol{x} \in \mathcal{C}.$$

其中，$f(\boldsymbol{x})$是凸函数，$\mathcal{C}$是凸集合

有约束凸优化问题的标准形式

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \leq 0, i = 1, \dots, m, \\ & h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\ \text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n. \end{aligned}$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是凸函数， $h_j(\boldsymbol{x})$是仿射函数

仿射函数：可写成$\boldsymbol{a}_j^T\boldsymbol{x} + b_j$的形式

容易证明 满足$g_i(\boldsymbol{x}) \leq 0$及$h_j(\boldsymbol{x}) = 0$的集合是凸集（构造$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C}$，再根据凸集定义证明）

注意$h_j(\boldsymbol{x})$需要是仿射函数，若仅是凸函数，约束条件不一定构成凸集

（考虑$h_j(\boldsymbol{x}) = 0 \Leftrightarrow h_j(\boldsymbol{x}) \leq 0, -h_j(\boldsymbol{x}) \leq 0$）

有约束凸优化

有约束凸优化问题的标准形式

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\ & h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\ \text{var.} & \boldsymbol{x} \in \mathcal{R}^n. \end{array}$$

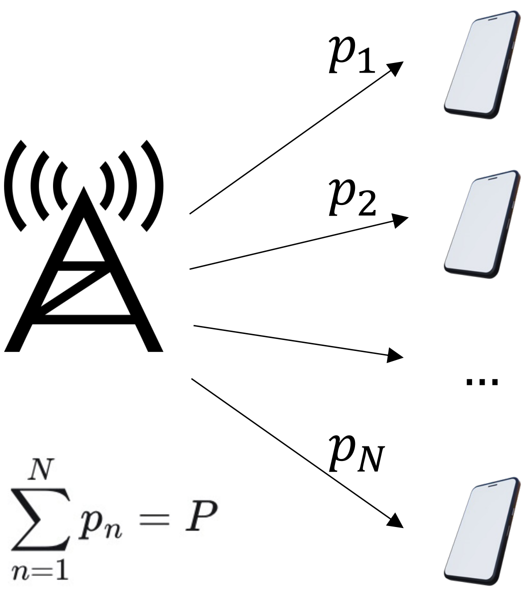
其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是凸函数，$h_j(\boldsymbol{x})$是仿射函数

包含各种特例：无约束凸优化问题、线性规划问题等

对于这些特例，已有多项式时间复杂度的求解算法

发射功率分配问题

移动通信基站如何分配用于传输数据的**功率**，从而**最大化总传输速率**

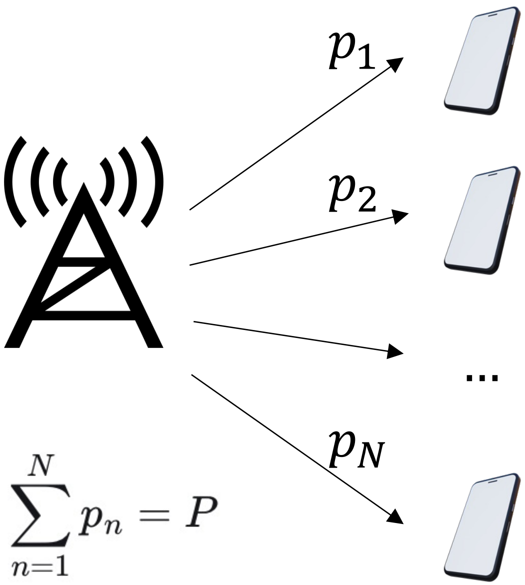


$$\sum_{n=1}^{N} p_n = P$$

$$p_n \geq 0$$

发射功率分配问题

移动通信基站如何分配用于传输数据的**功率**，从而**最大化总传输速率**



$$\sum_{n=1}^{N} p_n = P$$

$$p_n \geq 0$$

g_1

g_2

...

g_N

信道增益
（受传输距离、障碍物等影响）

$$\log_2\left(1+\frac{g_1 p_1}{N_0}\right)$$

$$\log_2\left(1+\frac{g_2 p_2}{N_0}\right)$$

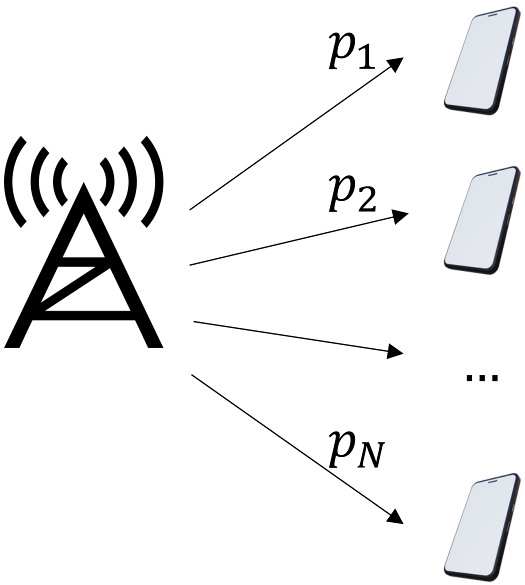
...

$$\log_2\left(1+\frac{g_N p_N}{N_0}\right)$$

传输速率
（N_0是噪声）

发射功率分配问题

移动通信基站如何分配用于传输数据的功率，从而最大化总传输速率



$$
\begin{aligned}
\max_{p_n}. \quad & \sum_{n=1}^{N} \log_2\left(1+\frac{g_n p_n}{N_0}\right) \\
\text{s.t.} \quad & \sum_{n=1}^{N} p_n = P \\
& p_n \geq 0
\end{aligned}
$$

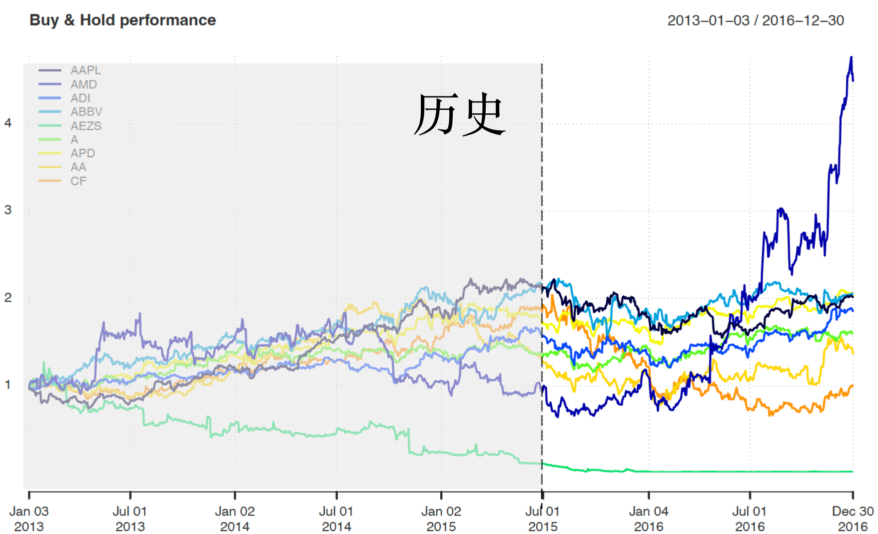
是有约束的凸优化问题

（改写成最小化问题、分析海森矩阵）

* 在实际通信领域，考虑时延约束下的动态功率分配、多基站通信等

投资组合优化问题

如何对不同资产进行**投资**，从而**最大化收益的期望、最小化风险（方差）**



投资策略记为$\boldsymbol{w} = (w_1, w_2, \ldots, w_N)^T$

将资产投资收益向量建模成随机向量，向量的期望和协方差矩阵记为$\boldsymbol{\mu}$和$\boldsymbol{\Sigma}$（根据历史得到）

内容取自Daniel P. Palomar <ELEC5470/IEDA6100A - Convex Optimization>

投资组合优化问题

如何对不同资产进行**投资**，从而**最大化收益的期望**、**最小化风险（方差）**

The idea of **Markowitz's mean-variance portfolio (MVP)** (Markowitz 1952) is to find a trade-off between the expected return $\mathbf{w}^T\boldsymbol{\mu}$ and the risk of the portfolio measured by the variance $\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w}$:

$$\begin{array}{ll} \underset{\mathbf{w}}{\text{maximize}} & \mathbf{w}^T\boldsymbol{\mu} - \lambda\mathbf{w}^T\boldsymbol{\Sigma}\mathbf{w} \\ \text{subject to} & \mathbf{1}^T\mathbf{w} = 1 \quad \boxed{\mathbf{w} \geq 0.} \end{array}$$

where $\mathbf{w}^T\mathbf{1} = 1$ is the capital budget constraint and λ is a parameter that controls how risk-averse the investor is.

是有约束的凸优化问题

（改写成最小化问题、随机向量的协方差矩阵一定是半正定）

$$\boldsymbol{\Sigma} = \mathrm{var}(\boldsymbol{r}) = \mathbb{E}\left[\big(\boldsymbol{r} - \mathbb{E}(\boldsymbol{r})\big)\big(\boldsymbol{r} - \mathbb{E}(\boldsymbol{r})\big)^T\right] \succcurlyeq \mathbf{0}$$

内容取自Daniel P. Palomar <ELEC5470/IEDA6100A - Convex Optimization>

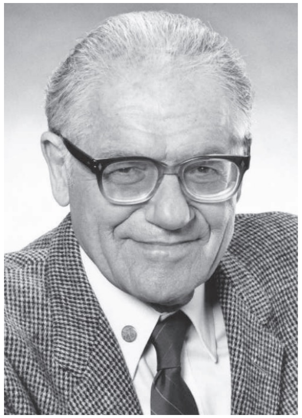
KKT条件

Part1 如何从复杂度的角度衡量算法好坏？
Part2 针对一些有特定形式的最优化问题，如何找到最优解？
无约束凸优化问题、梯度下降法、牛顿法； 线性规划问题、单纯形法、内点法； 有约束凸优化问题、**KKT条件**、内点法
Part3 针对较复杂的最优化问题（如NP难问题），如何找到较好解？
Part4 针对较复杂的多目标最优化问题，如何找到较好解？

KKT条件

（一定条件下）凸优化问题的全局最优解的充要条件：KKT条件

KKT（Karush-Kuhn-Tucker）条件



William Karush , 1987

Harold Kuhn and Albert Tucker, 1980

The KKT conditions were originally named after Harold W. Kuhn and Albert W. Tucker, who first published the conditions in 1951.[2] Later scholars discovered that the necessary conditions for this problem had been stated by William Karush in his master's thesis in 1939.[3][4]

KKT条件

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是凸函数，$h_j(\boldsymbol{x})$是仿射函数

定理

对前述凸优化问题，若$f(\boldsymbol{x})$和$g_i(\boldsymbol{x})$都是连续可微函数且**Slater条件成立**，那么$\boldsymbol{x}^*$是全局最优解的**充要条件**是$\boldsymbol{x}^*$满足：

$$
\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},
$$

$$
g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,
$$

$$
\lambda_i^* \ge 0, i = 1, \dots, m,
$$

$$
\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.
$$

其中，λ_i^*, ν_j^*是对偶问题的最优解。

参见Stephen Boyd, Lieven Vandenberghe <Convex Optimization> 5.5.3 KKT optimality conditions

KKT条件

$$
\begin{aligned}
\text{min} \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是凸函数，$h_j(\boldsymbol{x})$是仿射函数

定理

对前述凸优化问题，若$f(\boldsymbol{x})$和$g_i(\boldsymbol{x})$都是连续可微函数且Slater条件成立，那么$\boldsymbol{x}^*$是全局最优解的充要条件是$\boldsymbol{x}^*$满足：

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} v_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0}, \qquad \text{稳定性条件}$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p, \qquad \text{原问题可行条件}$$

$$\lambda_i^* \ge 0, i = 1, \dots, m, \qquad \text{对偶问题可行条件}$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m. \qquad \text{互补松弛条件}$$

其中，λ_i^*, v_j^*是对偶问题的最优解。

类似于《高等数学》中拉格朗日乘子法中的乘子

KKT条件

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是凸函数，$h_j(\boldsymbol{x})$是仿射函数

定理

对前述凸优化问题，若$f(\boldsymbol{x})$和$g_i(\boldsymbol{x})$都是连续可微函数且**Slater条件成立**，那么$\boldsymbol{x}^*$是全局最优解的**充要条件**是$\boldsymbol{x}^*$满足：

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} v_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0},$$

对比无约束下的一阶条件
额外考虑约束条件的影响

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

考虑$\min f(\boldsymbol{x}) + \lambda_i g_i(\boldsymbol{x})$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

考虑$g_i(\boldsymbol{x}^*) < 0$和$g_i(\boldsymbol{x}^*) = 0$两种情况

其中，λ_i^*, v_j^*是对偶问题的最优解。

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\ & h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\ \text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n. \end{aligned}$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是凸函数， $h_j(\boldsymbol{x})$是仿射函数

KKT条件

定理

对前述凸优化问题，若$f(\boldsymbol{x})$和$g_i(\boldsymbol{x})$都是连续可微函数且**Slater条件成立**，那么$\boldsymbol{x}^*$是全局最优解的**充要条件**是$\boldsymbol{x}^*$满足：

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

考虑$g_i(\boldsymbol{x}^*) < 0$和$g_i(\boldsymbol{x}^*) = 0$两种情况

其中，λ_i^*, ν_j^*是对偶问题的最优解。

$\min x^2$
$\text{s.t. } x - 3 \le 0$

$\min x^2$
$\text{s.t. } x + 3 \le 0$

Slater条件

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是凸函数，$h_j(\boldsymbol{x})$是仿射函数

Slater条件成立：存在$\boldsymbol{x}$使$g_i(\boldsymbol{x}) < 0, i = 1, \dots, m$且$h_j(\boldsymbol{x}) = 0, j = 1, \dots, p.$

（即存在$\boldsymbol{x}$处于可行域的相对内部）

以下凸优化问题不满足Slater条件，最优解不满足KKT条件：

$$
\begin{aligned}
\min \quad & x \\
\text{s.t.} \quad & x^2 \le 0, \\
\text{var.} \quad & x \in \mathcal{R}.
\end{aligned}
$$

易见$x^* = 0$，而KKT条件（稳定性条件）要求$1 + \lambda^* \cdot 2 \cdot 0 = 0$

例子

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} v_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \ldots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \ldots, p,$$

$$\lambda_i^* \ge 0, i = 1, \ldots, m,$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \ldots, m.$$

$$\begin{aligned} \min \quad & x_2 - x_1 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 4 \le 0, \\ & -x_1 - \sqrt{3}x_2 \le 0, \\ \text{var.} \quad & \boldsymbol{x} \in \mathcal{R}. \end{aligned}$$

（1）判断是凸优化问题且Slater条件成立

（2）列出KKT条件

（3）分析$\lambda_i^* > 0$或$\lambda_i^* = 0$的可能性

a. 若$\lambda_1^* = \lambda_2^* = 0$，稳定性条件不满足
b. 若$\lambda_1^* = 0, \lambda_2^* \ne 0$，稳定性条件不满足
c. 若$\lambda_1^* \ne 0, \lambda_2^* = 0$，推出$x_1^* = \sqrt{2}, x_2^* = -\sqrt{2}$，原问题可行条件不满足
d. 若$\lambda_1^* \ne 0, \lambda_2^* \ne 0$，推出$x_1^* = \sqrt{3}, x_2^* = -1, \lambda_1^* = \frac{\sqrt{3}+1}{8}, \lambda_2^* = \frac{\sqrt{3}-1}{4}$，满足KKT条件

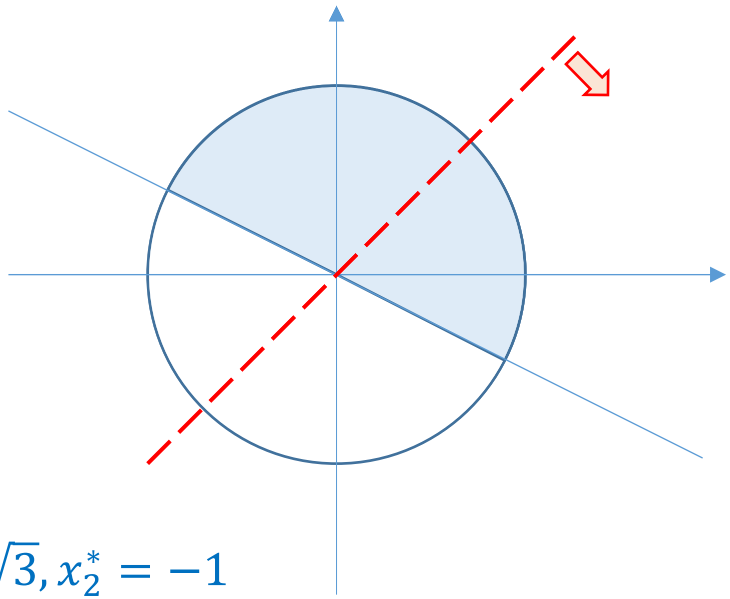
例子

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0},$$
$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$
$$\lambda_i^* \ge 0, i = 1, \dots, m,$$
$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

$$\begin{array}{ll} \min & x_2 - x_1 \\ \text{s.t.} & x_1^2 + x_2^2 - 4 \le 0, \\ & -x_1 - \sqrt{3}x_2 \le 0, \\ \text{var.} & \boldsymbol{x} \in \mathcal{R}. \end{array}$$

稳定性条件$\begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda_1^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} + \lambda_2^* \begin{bmatrix} -1 \\ -\sqrt{3} \end{bmatrix} = \mathbf{0}$ 即$\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \lambda_1^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} + \lambda_2^* \begin{bmatrix} -1 \\ -\sqrt{3} \end{bmatrix}$ $\quad \lambda_1^* \ge 0, \lambda_2^* \ge 0$

可视化：



$x_1^* = \sqrt{3}, x_2^* = -1$

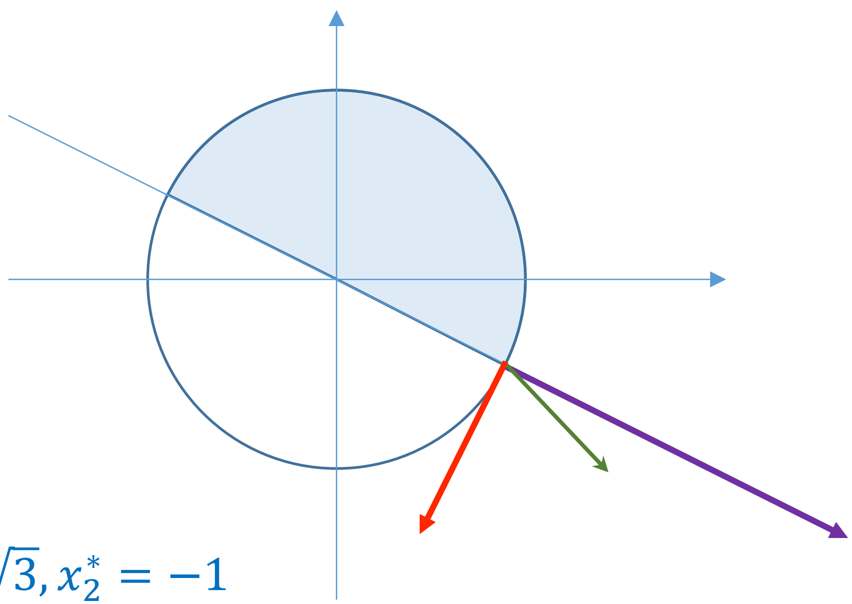
例子

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},$$
$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$
$$\lambda_i^* \ge 0, i = 1, \dots, m,$$
$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

$$\begin{aligned} \min \quad & x_2 - x_1 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 4 \le 0, \\ & -x_1 - \sqrt{3}x_2 \le 0, \\ \text{var.} \quad & \boldsymbol{x} \in \mathcal{R}. \end{aligned}$$

稳定性条件$\begin{bmatrix}-1\\1\end{bmatrix} + \lambda_1^*\begin{bmatrix}2x_1^*\\2x_2^*\end{bmatrix} + \lambda_2^*\begin{bmatrix}-1\\-\sqrt{3}\end{bmatrix} = \boldsymbol{0}$ 即$\begin{bmatrix}1\\-1\end{bmatrix} = \lambda_1^*\begin{bmatrix}2x_1^*\\2x_2^*\end{bmatrix} + \lambda_2^*\begin{bmatrix}-1\\-\sqrt{3}\end{bmatrix}$ $\quad \lambda_1^* \ge 0, \lambda_2^* \ge 0$

可视化：



$x_1^* = \sqrt{3}, x_2^* = -1$

该点同时激活了两个不等式约束，有$\lambda_1^* \ge 0, \lambda_2^* \ge 0$，即绿色向量的方向需要在红色和紫色之间

绿色向量：目标函数$f(\boldsymbol{x})$下降方向
紫色向量：$g_1(\boldsymbol{x})$上升方向
红色向量：$g_2(\boldsymbol{x})$上升方向

例子

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

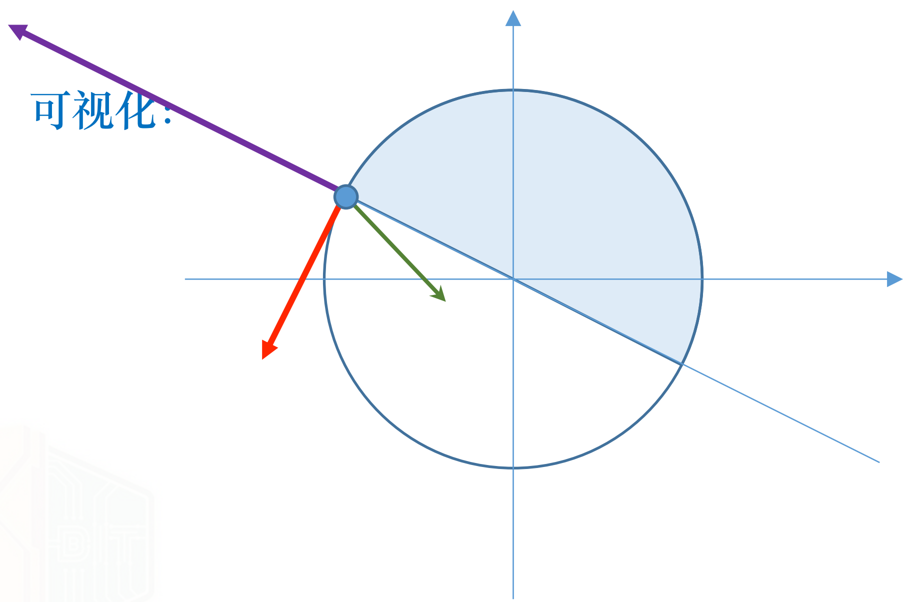
$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

$$\begin{aligned} \min \quad & x_2 - x_1 \\ \text{s.t.} \quad & x_1^2 + x_2^2 - 4 \le 0, \\ & -x_1 - \sqrt{3} x_2 \le 0, \\ \text{var.} \quad & \boldsymbol{x} \in \mathcal{R}. \end{aligned}$$

稳定性条件$\begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda_1^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} + \lambda_2^* \begin{bmatrix} -1 \\ -\sqrt{3} \end{bmatrix} = \mathbf{0}$ 即$\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \lambda_1^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} + \lambda_2^* \begin{bmatrix} -1 \\ -\sqrt{3} \end{bmatrix}$ $\lambda_1^* \ge 0, \lambda_2^* \ge 0$

可视化：



考虑蓝色点，绿色向量的方向无法在红色和紫色之间

故该点不是最优解

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

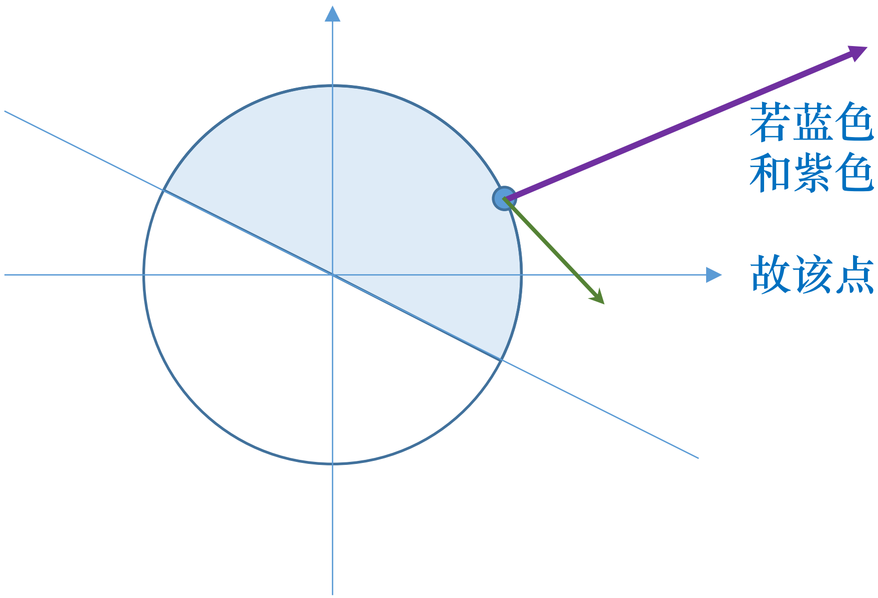
$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

例子

$$\begin{array}{ll} \min & x_2 - x_1 \\ \text{s.t.} & x_1^2 + x_2^2 - 4 \le 0, \\ & -x_1 - \sqrt{3}x_2 \le 0, \\ \text{var.} & \boldsymbol{x} \in \mathcal{R}. \end{array}$$

稳定性条件$\begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda_1^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} + \lambda_2^* \begin{bmatrix} -1 \\ -\sqrt{3} \end{bmatrix} = \boldsymbol{0}$ 即$\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \lambda_1^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} + \lambda_2^* \begin{bmatrix} -1 \\ -\sqrt{3} \end{bmatrix}$ $\quad \lambda_1^* \ge 0, \lambda_2^* \ge 0$

可视化：



若蓝色点是最优解：$\lambda_2^* = 0$，绿色向量和紫色向量的方向无法相同

故该点不是最优解

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

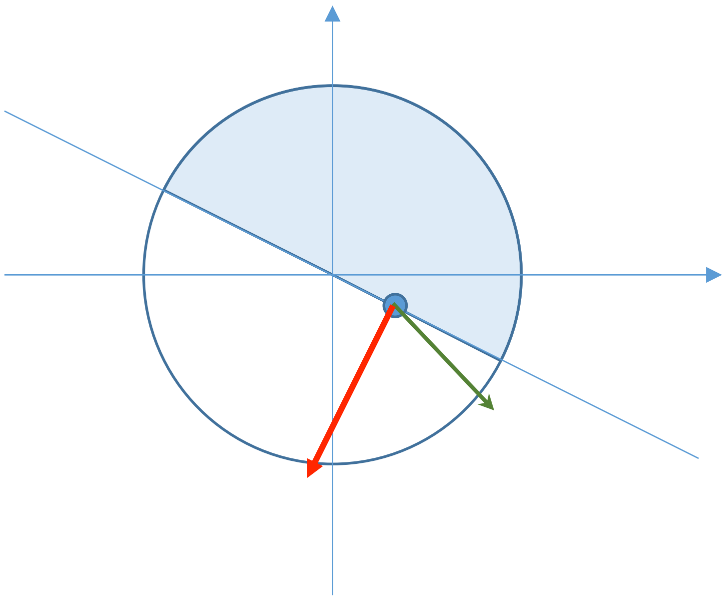
$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

例子

$$\begin{array}{ll} \min & x_2 - x_1 \\ \text{s.t.} & x_1^2 + x_2^2 - 4 \le 0, \\ & -x_1 - \sqrt{3}x_2 \le 0, \\ \text{var.} & \boldsymbol{x} \in \mathcal{R}. \end{array}$$

稳定性条件 $\begin{bmatrix} -1 \\ 1 \end{bmatrix} + \lambda_1^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} + \lambda_2^* \begin{bmatrix} -1 \\ -\sqrt{3} \end{bmatrix} = \mathbf{0}$ 即 $\begin{bmatrix} 1 \\ -1 \end{bmatrix} = \lambda_1^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} + \lambda_2^* \begin{bmatrix} -1 \\ -\sqrt{3} \end{bmatrix}$ $\qquad \lambda_1^* \ge 0, \lambda_2^* \ge 0$

可视化：



若蓝色点是最优解：$\lambda_1^* = 0$，绿色向量和红色向量的方向无法相同

故该点不是最优解

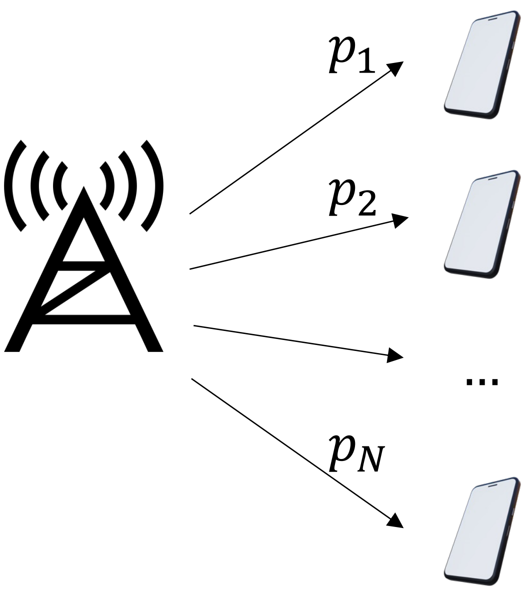
发射功率分配问题

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$



$$\begin{aligned} \max_{p_n}. \quad & \sum_{n=1}^{N} \log_2\left(1 + \frac{g_n p_n}{N_0}\right) \\ \text{s.t.} \quad & \sum_{n=1}^{N} p_n = P \\ & p_n \ge 0 \end{aligned}$$

发射功率分配问题

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \mathbf{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

KKT条件:

$$-\frac{\frac{g_n}{N_0}}{1 + \frac{g_n p_n^*}{N_0}} - \lambda_n^* + \nu^* = 0, n = 1, \dots, N,$$

$$p_n^* \ge 0, n = 1, \dots, N, \sum_{n=1}^{N} p_n^* = P,$$

$$\lambda_n^* \ge 0, n = 1, \dots, N,$$

$$-\lambda_n^* p_n^* = 0, n = 1, \dots, N.$$

发射功率分配问题

KKT条件：

$$-\frac{\frac{g_n}{N_0}}{1+\frac{g_n p_n^*}{N_0}} - \lambda_n^* + \nu^* = 0, n = 1, \dots, N,$$

$$p_n^* \geq 0, n = 1, \dots, N, \sum_{n=1}^{N} p_n^* = P,$$

$$\lambda_n^* \geq 0, n = 1, \dots, N,$$

$$-\lambda_n^* p_n^* = 0, n = 1, \dots, N.$$

把λ_n^*消去

$$p_n^* \geq 0, n = 1, \dots, N, \sum_{n=1}^{N} p_n^* = P,$$

$$\nu^* - \frac{\frac{g_n}{N_0}}{1+\frac{g_n p_n^*}{N_0}} \geq 0, n = 1, \dots, N,$$

$$-\left(\nu^* - \frac{\frac{g_n}{N_0}}{1+\frac{g_n p_n^*}{N_0}}\right) p_n^* = 0, n = 1, \dots, N.$$

发射功率分配问题

$$p_n^* \geq 0, n = 1, \dots, N, \sum_{n=1}^{N} p_n^* = P,$$

$$\nu^* - \frac{\frac{g_n}{N_0}}{1 + \frac{g_n p_n^*}{N_0}} \geq 0, n = 1, \dots, N,$$

$$-\left(\nu^* - \frac{\frac{g_n}{N_0}}{1 + \frac{g_n p_n^*}{N_0}}\right) p_n^* = 0, n = 1, \dots, N.$$

讨论ν^*与$\frac{g_n}{N_0}$的关系

若$\nu^* < \frac{g_n}{N_0}$： p_n^*需要大于0，即$\nu^* = \frac{\frac{g_n}{N_0}}{1+\frac{g_n p_n^*}{N_0}}$

若$\nu^* \geq \frac{g_n}{N_0}$： p_n^*需要等于0

总结： $$p_n^* = \begin{cases} \frac{1}{\nu^*} - \frac{N_0}{g_n}, & 若\nu^* < \frac{g_n}{N_0}, \\ 0, & 若\nu^* \geq \frac{g_n}{N_0}. \end{cases}$$

发射功率分配问题

$$p_n^* \geq 0, n = 1, \dots, N, \sum_{n=1}^{N} p_n^* = P,$$

$$\nu^* - \frac{\frac{g_n}{N_0}}{1 + \frac{g_n p_n^*}{N_0}} \geq 0, n = 1, \dots, N,$$

$$-\left(\nu^* - \frac{\frac{g_n}{N_0}}{1 + \frac{g_n p_n^*}{N_0}}\right) p_n^* = 0, n = 1, \dots, N.$$

$$p_n^* = \begin{cases} \frac{1}{\nu^*} - \frac{N_0}{g_n}, & \text{若}\nu^* < \frac{g_n}{N_0}, \\ 0, & \text{若}\nu^* \geq \frac{g_n}{N_0}. \end{cases}$$

等价于：$p_n^* = \left(\frac{1}{\nu^*} - \frac{N_0}{g_n}\right)^+$

$$\sum_{n=1}^{N} \left(\frac{1}{\nu^*} - \frac{N_0}{g_n}\right)^+ = P$$

发射功率分配问题

$$\sum_{n=1}^{N}\left(\frac{1}{\nu^*}-\frac{N_0}{g_n}\right)^+ = P$$

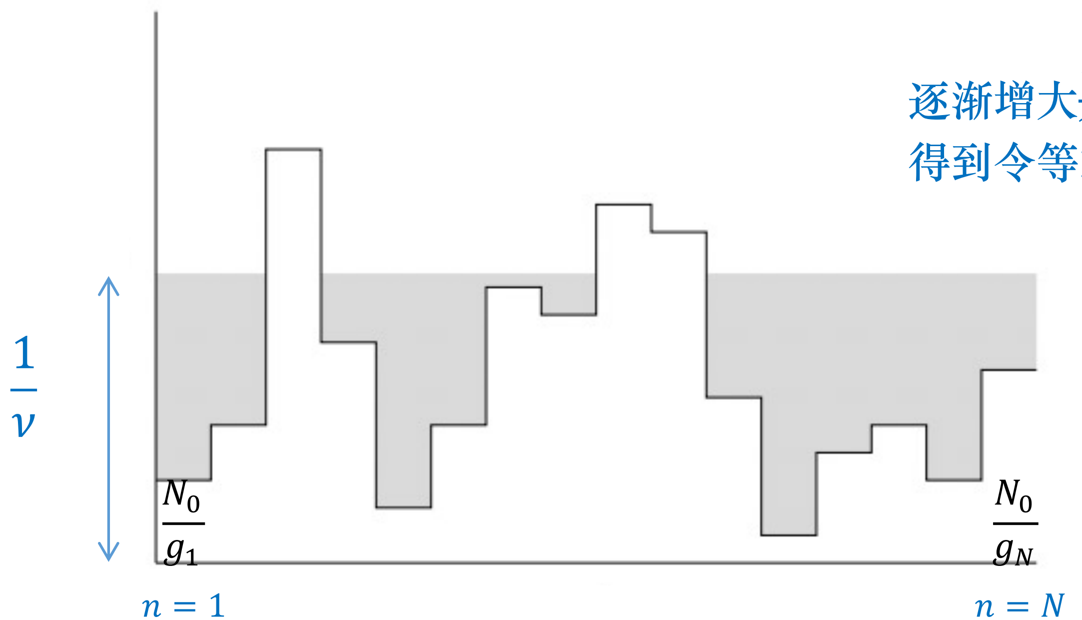
左侧随ν^*单调减，且$\nu^* = \frac{\max_n\{g_n\}}{N_0}$时$\sum_{n=1}^{N}\left(\frac{1}{\nu^*}-\frac{N_0}{g_n}\right)^+ = 0$

发射功率分配问题

$$\sum_{n=1}^{N}\left(\frac{1}{\nu^*}-\frac{N_0}{g_n}\right)^+ = P$$

左侧随ν^*单调减，且$\nu^* = \frac{\max_n\{g_n\}}{N_0}$时$\sum_{n=1}^{N}\left(\frac{1}{\nu^*}-\frac{N_0}{g_n}\right)^+ = 0$

注水算法



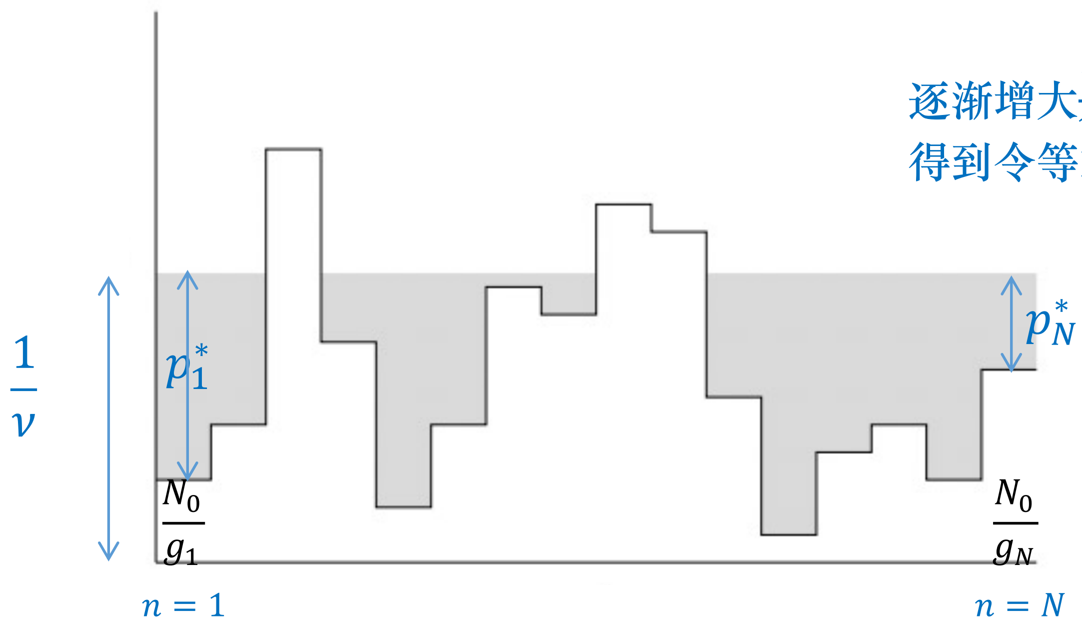
逐渐增大$\frac{1}{\nu}$，直到“水”面积等于P
得到令等式成立的ν^*

发射功率分配问题

$$\sum_{n=1}^{N}\left(\frac{1}{\nu^*}-\frac{N_0}{g_n}\right)^+ = P$$

左侧随ν^*单调减，且$\nu^* = \frac{\max_n\{g_n\}}{N_0}$时$\sum_{n=1}^{N}\left(\frac{1}{\nu^*}-\frac{N_0}{g_n}\right)^+ = 0$

注水算法

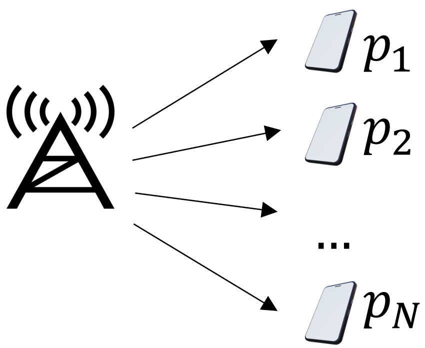


逐渐增大$\frac{1}{\nu}$，直到“水”面积等于P
得到令等式成立的ν^*

$$p_n^* = \left(\frac{1}{\nu^*}-\frac{N_0}{g_n}\right)^+$$

每条水柱高度即最优功率

发射功率分配问题



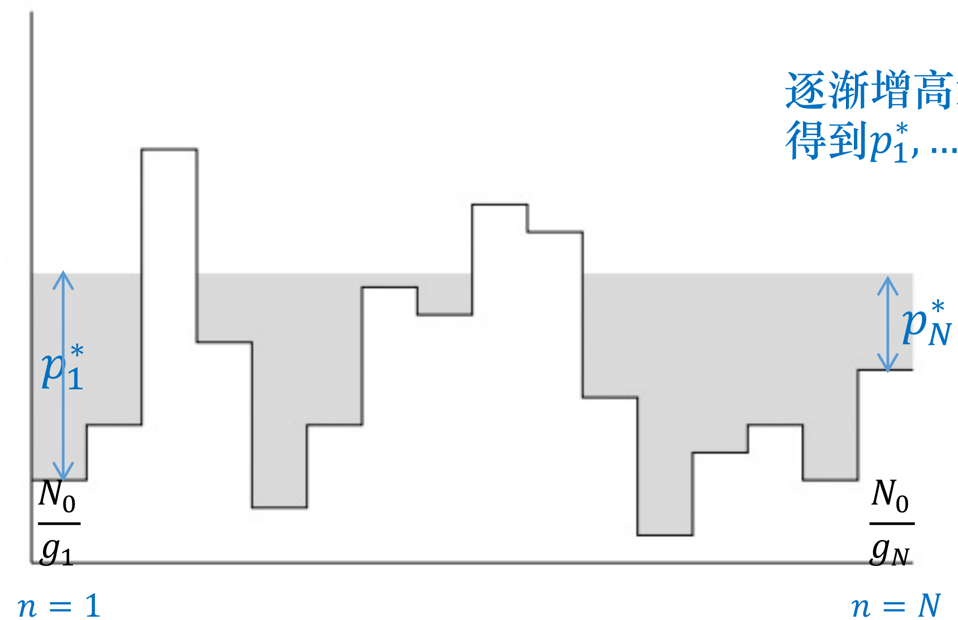
$$\max_{p_n} \quad \sum_{n=1}^{N} \log_2\left(1 + \frac{g_n p_n}{N_0}\right)$$

$$\text{s.t.} \quad \sum_{n=1}^{N} p_n = P$$

$$p_n \geq 0$$

注水算法

逐渐增高水面，直到面积等于P
得到$p_1^*, \ldots, p_N^*$等



KKT条件

定理

对前述凸优化问题，若$f(\boldsymbol{x})$和$g_i(\boldsymbol{x})$都是连续可微函数且Slater条件成立，那么$\boldsymbol{x}^*$是全局最优解的充要条件是$\boldsymbol{x}^*$满足：

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

其中，λ_i^*, ν_j^*是对偶问题的最优解。

对无约束凸优化问题，充要条件可简化为$\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$

对非凸优化问题，（在一定条件下）KKT条件是全局最优解的必要条件

求解有约束凸优化问题的内点法

Part1 如何从复杂度的角度衡量算法好坏？

Part2 针对一些有特定形式的最优化问题，如何找到最优解？

无约束凸优化问题、梯度下降法、牛顿法；
线性规划问题、单纯形法、内点法；
有约束凸优化问题、KKT条件、内点法

Part3 针对较复杂的最优化问题（如NP难问题），如何找到较好解？

Part4 针对较复杂的多目标最优化问题，如何找到较好解？

障碍函数

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数，$h_j(\boldsymbol{x})$是仿射函数，Slater条件成立

KKT条件：

$$\nabla f(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(\boldsymbol{x}^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(\boldsymbol{x}^*) = \boldsymbol{0},$$

$$g_i(\boldsymbol{x}^*) \le 0, i = 1, \dots, m, \qquad h_j(\boldsymbol{x}^*) = 0, j = 1, \dots, p,$$

$$\lambda_i^* \ge 0, i = 1, \dots, m,$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0, i = 1, \dots, m.$$

互补松弛条件难以处理，需分情况讨论（回顾发射功率分配问题）

障碍函数

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数，$h_j(\boldsymbol{x})$是仿射函数，Slater条件成立

⬇

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) + \sum_{i=1}^{m} {\color{red}I_t}\big(g_i(\boldsymbol{x})\big) \\
\text{s.t.} \quad & h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数，$h_j(\boldsymbol{x})$是仿射函数，Slater条件成立

$$
I_t(u) = -\frac{1}{t}\log(-u)
$$

$t > 0$控制障碍函数的形状，当$t \to \infty$时，新问题的最优解趋近于原问题最优解

障碍函数

$$
\begin{aligned}
\min \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数， $h_j(\boldsymbol{x})$是仿射函数，Slater条件成立

⇩

$$
\begin{aligned}
\min \quad & tf(\boldsymbol{x}) - \textstyle\sum_{i=1}^{m} \log\big(-g_i(\boldsymbol{x})\big) \\
\text{s.t.} \quad & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \\
\text{var.} \quad & \boldsymbol{x} \in \mathcal{R}^n.
\end{aligned}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数

该问题是不是凸优化问题?

障碍函数

分析新目标函数是否是凸函数：$tf(\boldsymbol{x})-\sum_{i=1}^{m}\log\left(-g_i(\boldsymbol{x})\right)$

分析$-\log\left(-g_i(\boldsymbol{x})\right)$是否是凸函数：

梯度：$-\frac{\nabla g_i(\boldsymbol{x})}{g_i(\boldsymbol{x})}$ 海森矩阵：$\boxed{-\frac{1}{g_i(\boldsymbol{x})}}\nabla^2 g_i(\boldsymbol{x})+\frac{1}{\left(g_i(\boldsymbol{x})\right)^2}\nabla g_i(\boldsymbol{x})\left(\nabla g_i(\boldsymbol{x})\right)^T$

在满足$g_i(\boldsymbol{x})\leq 0$的区域，海森矩阵为半正定矩阵

对$n\times n$实对称矩阵$\mathbf{A}$，若对任意向量$\boldsymbol{u}\in\mathbb{R}^n$有$\boldsymbol{u}^T\mathbf{A}\boldsymbol{u}\geq 0$，$\mathbf{A}$为半正定矩阵。

半正定矩阵之和依然是半正定矩阵。

障碍函数

$$
\begin{array}{ll}
\min & f(\boldsymbol{x}) \\
\text{s.t.} & g_i(\boldsymbol{x}) \le 0, i = 1, \dots, m, \\
& h_j(\boldsymbol{x}) = 0, j = 1, \dots, p, \\
\text{var.} & \boldsymbol{x} \in \mathcal{R}^n.
\end{array}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数，$h_j(\boldsymbol{x})$是仿射函数，Slater条件成立

$$
\begin{array}{ll}
\min & tf(\boldsymbol{x}) - \sum_{i=1}^{m} \log\big(-g_i(\boldsymbol{x})\big) \\
\text{s.t.} & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \\
\text{var.} & \boldsymbol{x} \in \mathcal{R}^n.
\end{array}
$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数

该问题是凸优化问题

如何求解?

障碍函数

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & g_i(\boldsymbol{x}) \leq 0, i = 1, \ldots, m, \\ & h_j(\boldsymbol{x}) = 0, j = 1, \ldots, p, \\ \text{var.} & \boldsymbol{x} \in \mathcal{R}^n. \end{array}$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数， $h_j(\boldsymbol{x})$是仿射函数，Slater条件成立

$$\begin{array}{ll} \min & tf(\boldsymbol{x}) - \sum_{i=1}^m \log(-g_i(\boldsymbol{x})) \\ \text{s.t.} & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \\ \text{var.} & \boldsymbol{x} \in \mathcal{R}^n. \end{array}$$

其中，$f(\boldsymbol{x}), g_i(\boldsymbol{x})$是连续可微凸函数

新问题的KKT条件

$$t\nabla f(\boldsymbol{x}^*) - \sum_{i=1}^m \frac{\nabla g_i(\boldsymbol{x}^*)}{g_i(\boldsymbol{x}^*)} + \sum_{j=1}^p v_j^* \boldsymbol{a}_j^T = \mathbf{0},$$

$$g_i(\boldsymbol{x}^*) \leq 0, i = 1, \ldots, m, \mathbf{A}\boldsymbol{x}^* = \boldsymbol{b}.$$

$\boldsymbol{a}_j$是$\mathbf{A}$的第j行

可以通过消元将原问题化为无约束问题再用牛顿法等，或直接对KKT条件的方程组进行求解

内点法

内点法

0 确定一个初始可行解$\tilde{\boldsymbol{x}}$，初始参数值$t > 0$

1 由$\tilde{\boldsymbol{x}}$出发，在当前t取值下，计算$\boldsymbol{x}^*(t)$

$$\begin{array}{ll} \min & tf(\boldsymbol{x}) - \sum_{i=1}^{m} \log\big(-g_i(\boldsymbol{x})\big) \\ \text{s.t.} & \mathbf{A}\boldsymbol{x} = \boldsymbol{b}, \\ \text{var.} & \boldsymbol{x} \in \mathcal{R}^n. \end{array}$$

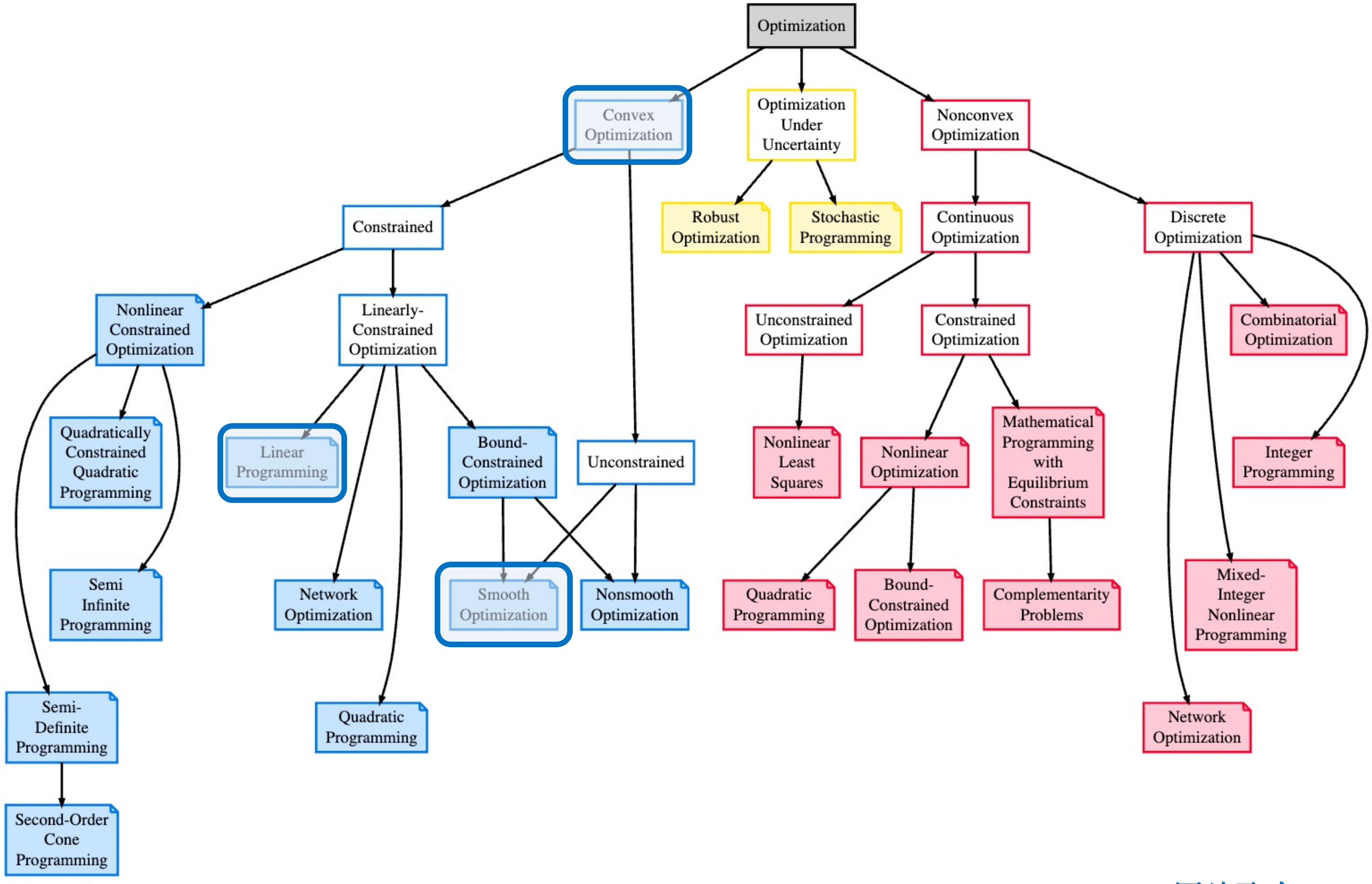
2 将$\boldsymbol{x}^*(t)$赋值给$\tilde{\boldsymbol{x}}$

3 如果循环结束条件（如t大于一定阈值）满足，则结束；否则，增加t取值（如令$t \leftarrow \mu t$），返回1

最优化问题的类别

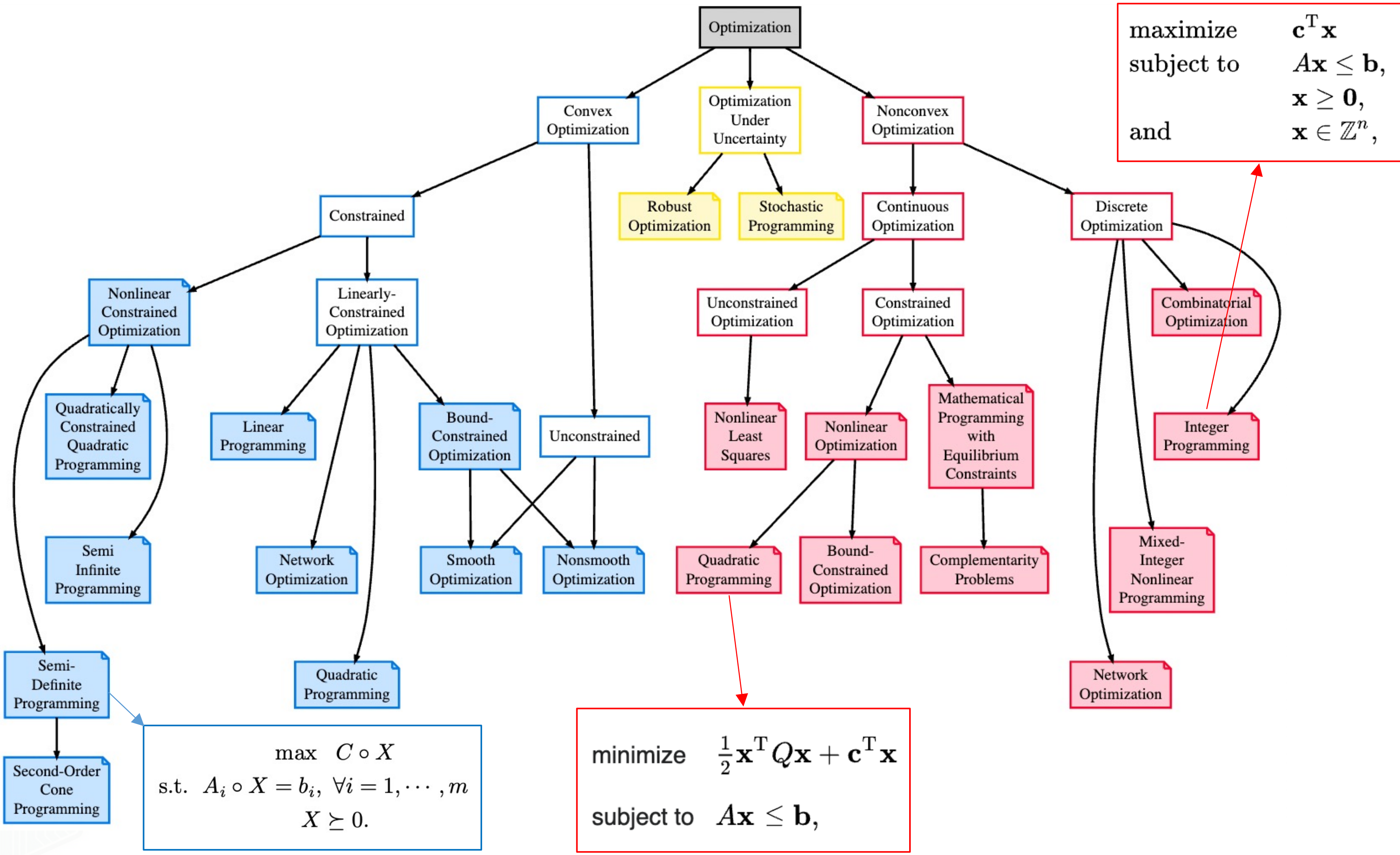
Part1 如何从复杂度的角度衡量算法好坏？
算法复杂度、P问题与NP问题与NP难问题
Part2 针对一些有特定形式的最优化问题，如何找到最优解？
无约束凸优化问题、梯度下降法、牛顿法； 线性规划问题、单纯形法、内点法； 有约束凸优化问题、KKT条件、内点法
Part3 针对较复杂的最优化问题（如NP难问题），如何找到较好解？
局部搜索、模拟退火、遗传算法、差分进化算法、蚁群算法、粒子群优化算法、人工蜂群算法
Part4 针对较复杂的多目标最优化问题，如何找到较好解？
多目标优化问题、NSGA-II算法

最优化问题的类别



图片取自neos-guide.org

最优化问题的类别



Optimization
Convex Optimization
Optimization Under Uncertainty
Nonconvex Optimization
Constrained
Robust Optimization
Stochastic Programming
Continuous Optimization
Discrete Optimization
Nonlinear Constrained Optimization
Linearly-Constrained Optimization
Unconstrained Optimization
Constrained Optimization
Combinatorial Optimization
Quadratically Constrained Quadratic Programming
Linear Programming
Bound-Constrained Optimization
Unconstrained
Nonlinear Least Squares
Nonlinear Optimization
Mathematical Programming with Equilibrium Constraints
Integer Programming
Semi Infinite Programming
Network Optimization
Smooth Optimization
Nonsmooth Optimization
Quadratic Programming
Bound-Constrained Optimization
Complementarity Problems
Mixed-Integer Nonlinear Programming
Semi-Definite Programming
Quadratic Programming
Network Optimization
Second-Order Cone Programming
maximize $\mathbf{c}^{\mathrm{T}}\mathbf{x}$
subject to $A\mathbf{x} \le \mathbf{b}$,
$\mathbf{x} \ge \mathbf{0}$,
and $\mathbf{x} \in \mathbb{Z}^n$,
$\max \quad C \circ X$
s.t. $A_i \circ X = b_i, \ \forall i = 1, \cdots, m$
$X \succeq 0.$
minimize $\frac{1}{2}\mathbf{x}^{\mathrm{T}} Q \mathbf{x} + \mathbf{c}^{\mathrm{T}}\mathbf{x}$
subject to $A\mathbf{x} \le \mathbf{b}$,

凸优化

Convex Optimization

Stephen Boyd
Department of Electrical Engineering
Stanford University

Lieven Vandenberghe
Electrical Engineering Department
University of California, Los Angeles

CAMBRIDGE UNIVERSITY PRESS

介绍凸优化理论的经典书籍

完整电子版 及 Stanford课程幻灯（免费）

web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

web.stanford.edu/~boyd/cvxbook/bv_cvxslides.pdf

[BOOK] **Convex optimization**

S Boyd, SP Boyd, L Vandenberghe - 2004 - books.google.com

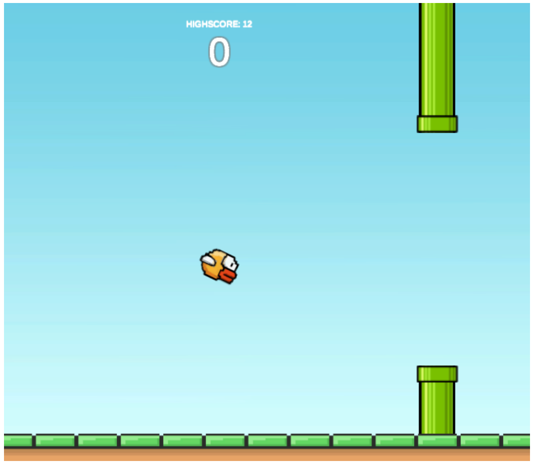
… general **convex optimization** … **convex** analysis, or the mathematics of **convex optimization**; several existing texts cover these topics well. Nor is the book a survey of algorithms for **convex** …

自然进化策略

Part1 如何从复杂度的角度衡量算法好坏？
Part2 针对一些有特定形式的最优化问题，如何找到最优解？
Part3 针对较复杂的最优化问题（如NP难问题），如何找到较好解？
Part4 针对较复杂的多目标最优化问题，如何找到较好解？

梯度信息

有时难以获得梯度信息或梯度不存在



https://flappy-bird.co/flappy-bird-2d-game

梯度信息

建模成一个关于控制策略的优化问题

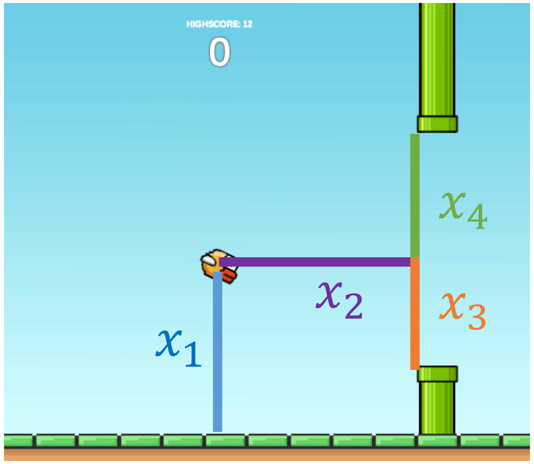
（1）建模控制策略

一种选择是定义变量$\boldsymbol{w} = (w_1, w_2, w_3, w_4)$

每0.5秒根据当前$\boldsymbol{x}$的值做决策：

当$\boldsymbol{w}^T\boldsymbol{x} \geq 0$时，选择“跳跃”；

当$\boldsymbol{w}^T\boldsymbol{x} < 0$时，不“跳跃”



（2）建模优化问题

$\max f(\boldsymbol{w})$ 飞行距离（即存活时长）

$\text{var } \boldsymbol{w} \in \mathcal{R}^4$

梯度信息

建模成一个关于控制策略的优化问题

（1）建模控制策略

一种选择是定义变量$\boldsymbol{w} = (w_1, w_2, w_3, w_4)$

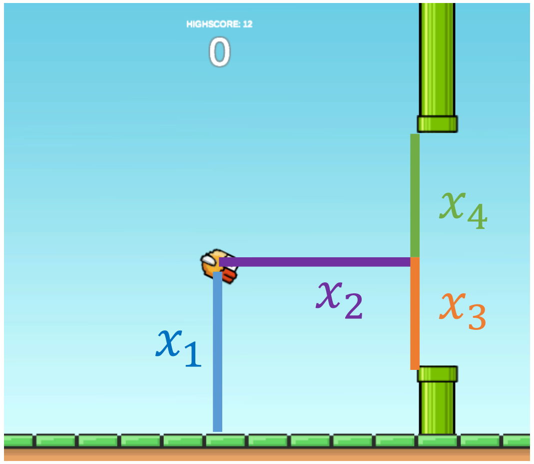
每0.5秒根据当前$\boldsymbol{x}$的值做决策：

当$\boldsymbol{w}^T\boldsymbol{x} \geq 0$时，选择“跳跃”；

当$\boldsymbol{w}^T\boldsymbol{x} < 0$时，不“跳跃”

（2）建模优化问题

$\max f(\boldsymbol{w})$ 飞行距离（即存活时长）
$\text{var } \boldsymbol{w} \in \mathcal{R}^4$

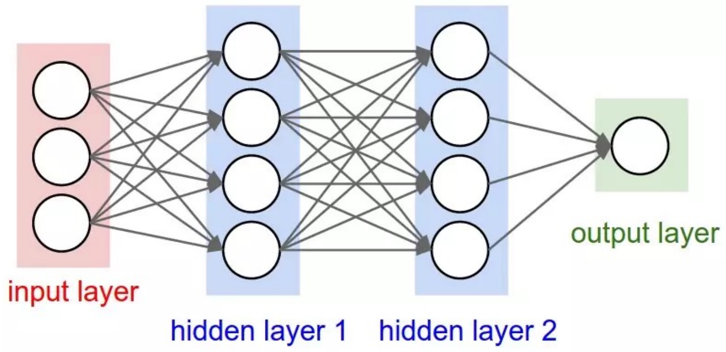


* 设定$\boldsymbol{w}$后，可通过游戏得$f(\boldsymbol{w})$，但无法得到$\nabla f(\boldsymbol{w})$

* 如果改进控制策略模型，可以得到目标函数对变量$\boldsymbol{w}$的梯度。参见强化学习中的“策略梯度”

梯度信息

有时难以获得梯度信息或梯度不存在



神经网络及训练过程的优化:

（1）网络权重：有梯度信息；

（2）超参数（学习率、丢弃率、网络层数、神经元个数）：无梯度信息

自然进化策略

求解最优化问题的主要思路:

（1）有梯度信息 → 利用梯度信息（前4周内容）

（2）没有梯度信息 → 估计函数下降方向，代替梯度信息（例：自然进化策略）

（3）没有梯度信息 → 不借助梯度信息（后4周内容）

自然进化策略

考虑如下最优化问题（假设无约束、变量取连续值）

$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

若可以求得$\nabla f(\boldsymbol{x})$，可利用梯度下降即$\boldsymbol{x}^{t+1} = \boldsymbol{x}^t - \alpha^t \nabla f(\boldsymbol{x}^t)$搜索最优解

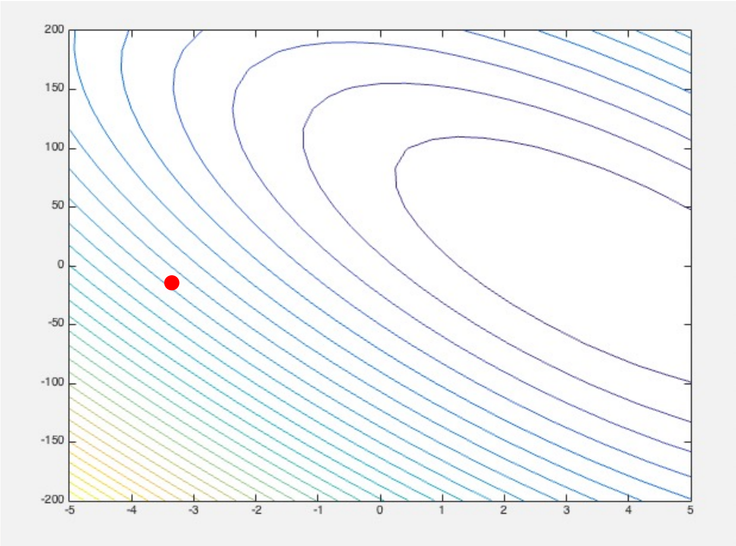
若无法求得$\nabla f(\boldsymbol{x})$，可以估计函数下降方向，代替梯度信息

自然进化策略

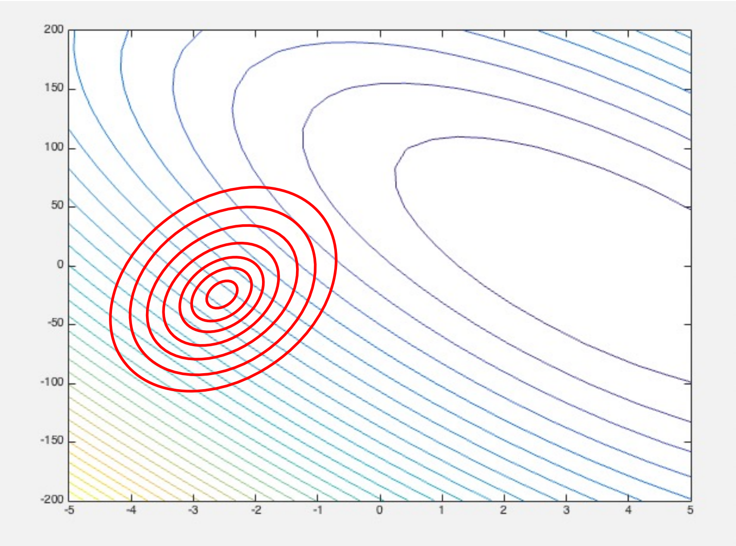
$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

把寻找$\boldsymbol{x}$的过程改为寻找关于$\boldsymbol{x}$分布的过程：

设$\boldsymbol{x}$服从概率密度函数$p_{\boldsymbol{\theta}}(\boldsymbol{x})$刻画的分布，$\boldsymbol{\theta}$是概率密度函数的参数



寻找$\boldsymbol{x}$



寻找$\boldsymbol{x}$的分布
（若$p_{\boldsymbol{\theta}}(\boldsymbol{x})$是高斯分布）

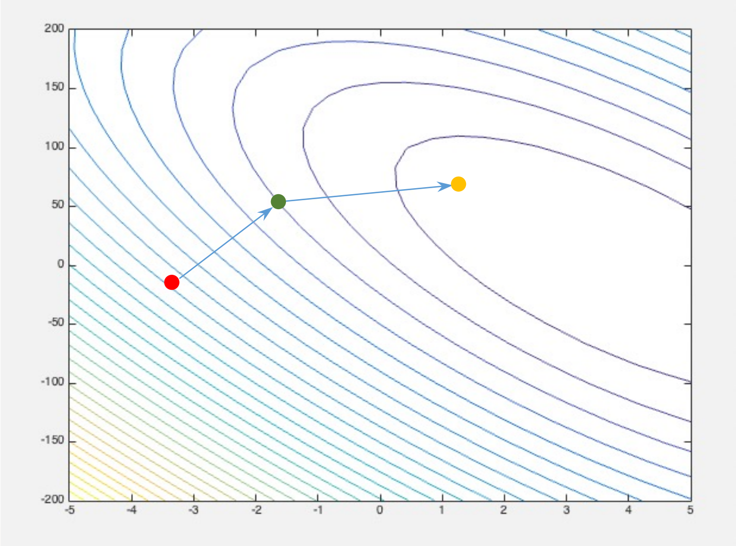
* 为展示方便，假设$f(\boldsymbol{x})$是凸函数，其实可以是任意函数

自然进化策略

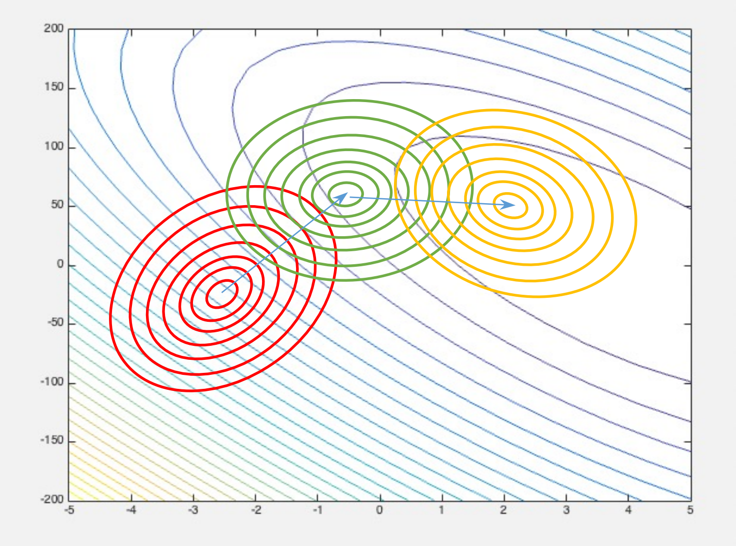
$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

把寻找$\boldsymbol{x}$的过程改为寻找关于$\boldsymbol{x}$分布的过程：

设$\boldsymbol{x}$服从概率密度函数$p_{\boldsymbol{\theta}}(\boldsymbol{x})$刻画的分布，$\boldsymbol{\theta}$是概率密度函数的参数



寻找$\boldsymbol{x}$



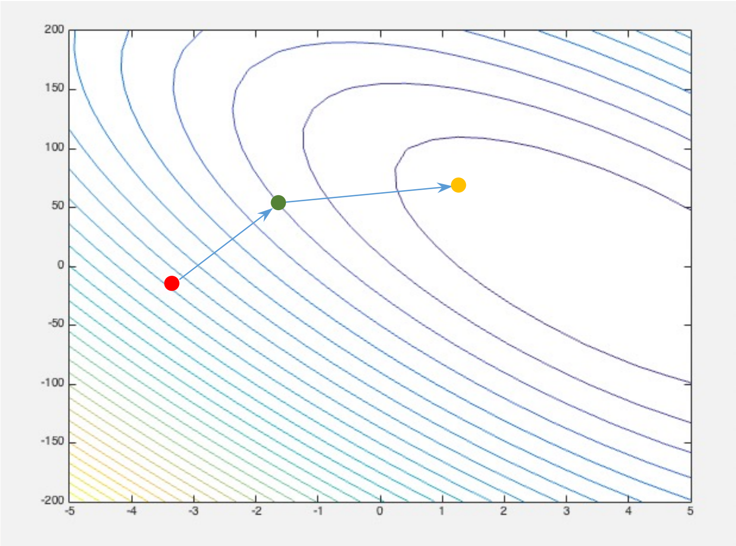
寻找$\boldsymbol{x}$的分布
（若$p_{\boldsymbol{\theta}}(\boldsymbol{x})$是高斯分布）

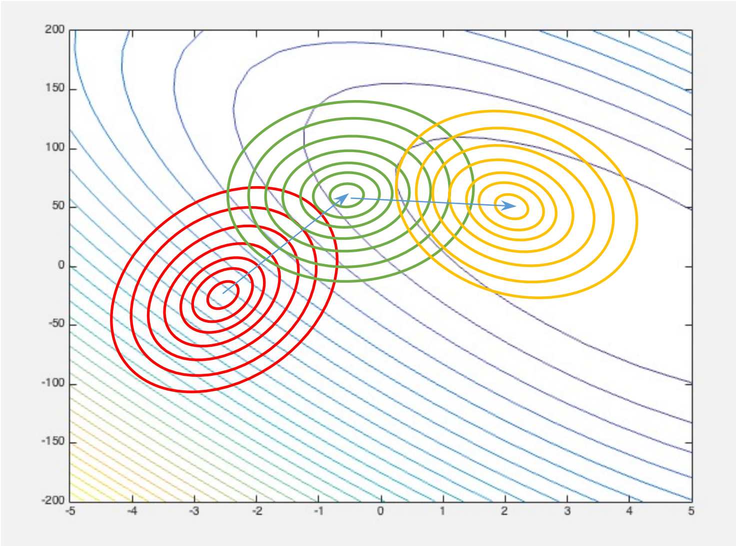
自然进化策略

$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

把寻找$\boldsymbol{x}$的过程改为寻找关于$\boldsymbol{x}$分布的过程：

设$\boldsymbol{x}$服从概率密度函数$p_{\boldsymbol{\theta}}(\boldsymbol{x})$刻画的分布，$\boldsymbol{\theta}$是概率密度函数的参数





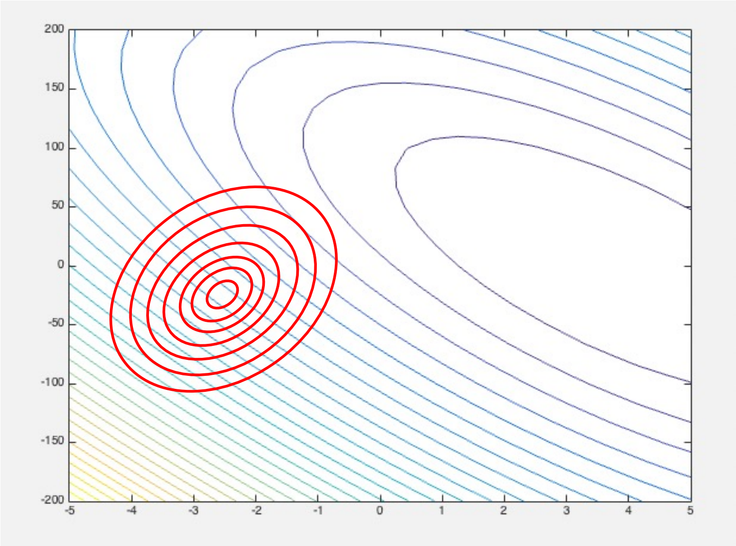
现在要优化的不是$\boldsymbol{x}$，而是$p_{\boldsymbol{\theta}}(\boldsymbol{x})$的参数$\boldsymbol{\theta}$。目标函数对$\boldsymbol{\theta}$的梯度是可能存在的

自然进化策略

$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

设$\boldsymbol{x} \sim p_{\boldsymbol{\theta}}$，改写为关于分布参数$\boldsymbol{\theta}$的优化问题

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \triangleq ?$$



自然进化策略

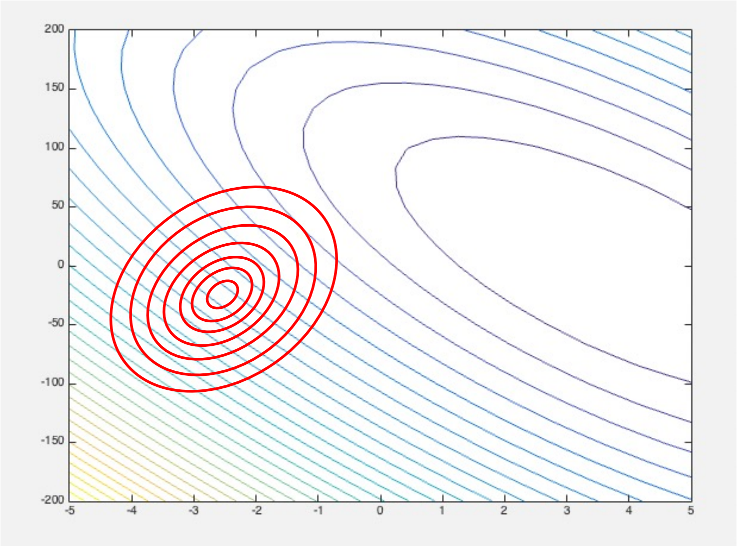
$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

设$\boldsymbol{x}\sim p_{\boldsymbol{\theta}}$，改写为关于分布参数$\boldsymbol{\theta}$的优化问题

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{x}\sim p_{\boldsymbol{\theta}}}[f(\boldsymbol{x})] = \int p_{\boldsymbol{\theta}}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$$

* 若 $p_{\boldsymbol{\theta}}(\boldsymbol{x})$ 取Delta分布，完全等价于原问题

为了利用$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \alpha^t \nabla J(\boldsymbol{\theta}^t)$优化$\boldsymbol{\theta}$，如何求$J(\boldsymbol{\theta})$关于$\boldsymbol{\theta}$的梯度?



自然进化策略

$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

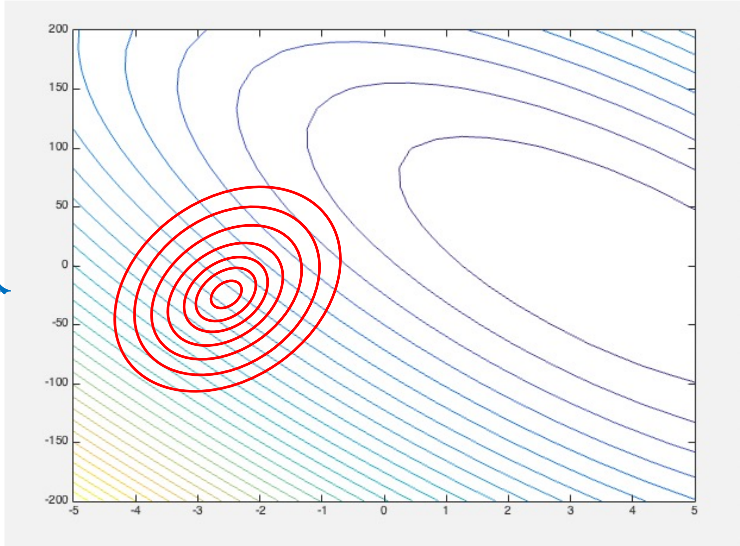
设$\boldsymbol{x} \sim p_{\boldsymbol{\theta}}$，改写为关于分布参数$\boldsymbol{\theta}$的优化问题

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{x} \sim p_{\boldsymbol{\theta}}}[f(\boldsymbol{x})] = \int p_{\boldsymbol{\theta}}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$$

为了利用$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \alpha^t \nabla J(\boldsymbol{\theta}^t)$优化$\boldsymbol{\theta}$，如何求$J(\boldsymbol{\theta})$关于$\boldsymbol{\theta}$的梯度?

$$\nabla J(\boldsymbol{\theta}) = \int \nabla p_{\boldsymbol{\theta}}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$$

因为通常缺少$f(\boldsymbol{x})$的表达式，难以直接计算积分



自然进化策略

$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

设$\boldsymbol{x} \sim p_{\boldsymbol{\theta}}$，改写为关于分布参数$\boldsymbol{\theta}$的优化问题

$$\min_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \triangleq \mathbb{E}_{\boldsymbol{x} \sim p_{\boldsymbol{\theta}}}[f(\boldsymbol{x})] = \int p_{\boldsymbol{\theta}}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$$

为了利用$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \alpha^t \nabla J(\boldsymbol{\theta}^t)$优化$\boldsymbol{\theta}$，如何求$J(\boldsymbol{\theta})$关于$\boldsymbol{\theta}$的梯度?

$$\nabla J(\boldsymbol{\theta}) = \int \nabla p_{\boldsymbol{\theta}}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x} = \int \nabla p_{\boldsymbol{\theta}}(\boldsymbol{x}) \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x})}{p_{\boldsymbol{\theta}}(\boldsymbol{x})} f(\boldsymbol{x}) d\boldsymbol{x} = \int p_{\boldsymbol{\theta}}(\boldsymbol{x}) \nabla \log p_{\boldsymbol{\theta}}(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$$

$$= \mathbb{E}_{\boldsymbol{x} \sim p_{\boldsymbol{\theta}}}[\nabla \log p_{\boldsymbol{\theta}}(\boldsymbol{x}) f(\boldsymbol{x})] \approx \frac{1}{M} \sum_m \nabla \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_m) f(\boldsymbol{x}_m),$$ 其中$\boldsymbol{x}_m$根据$p_{\boldsymbol{\theta}}(\boldsymbol{x})$采样

核心思想：构造成期望的形式，再通过蒙特卡洛方法近似

自然进化策略

$\min f(\boldsymbol{x})$
$\text{var } \boldsymbol{x} \in \mathcal{R}^n$

自然进化策略

0 初始化当前解$\boldsymbol{\theta}^0$及$t = 0$

1 确定步长$\alpha^t > 0$

2 根据当前分布$p_{\boldsymbol{\theta}^t}(\boldsymbol{x})$采样$M$个$\boldsymbol{x}_m$

3 更新：$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \alpha^t \frac{1}{M} \sum_m \nabla \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_m) f(\boldsymbol{x}_m)$

4 若满足收敛条件则结束；否则，$t \leftarrow t + 1$且返回1

自然进化策略

$\min f(\boldsymbol{x})$
var $\boldsymbol{x} \in \mathcal{R}^n$

自然进化策略

0 初始化当前解$\boldsymbol{\theta}^0$及$t = 0$

1 确定步长$\alpha^t > 0$

2 根据当前分布$p_{\boldsymbol{\theta}^t}(\boldsymbol{x})$采样$M$个$\boldsymbol{x}_m$

3 更新：$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \alpha^t \frac{1}{M} \sum_m \nabla \log p_{\boldsymbol{\theta}}(\boldsymbol{x}_m) f(\boldsymbol{x}_m)$

4 若满足收敛条件则结束；否则，$t \leftarrow t + 1$且返回1

例子，考虑$x \in \mathcal{R}$及x服从以θ为均值、σ为标准差（假设取值固定）的一维高斯分布

$$p_{\theta^t}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2} \frac{(x - \theta^t)^2}{\sigma^2}\right)$$

$$\frac{d \log p_{\theta^t}(x)}{d\theta^t} = \frac{x - \theta^t}{\sigma^2}$$

自然进化策略

$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

自然进化策略

0 初始化当前解θ^0及$t=0$，选择σ

1 确定步长$\alpha^t > 0$

2 根据当前分布$\mathcal{N}(\theta^t, \sigma^2)$采样$M$个$x_m$

3 更新：$\theta^{t+1} = \theta^t - \alpha^t \frac{1}{M} \sum_m \frac{x_m - \theta^t}{\sigma^2} f(x_m)$

4 若满足收敛条件则结束；否则，$t \leftarrow t+1$且返回1

参考 Ashwin Rao <CME241 - Evolutionary Strategies> 幻灯

自然进化策略

$$\min f(\boldsymbol{x})$$
$$\text{var } \boldsymbol{x} \in \mathcal{R}^n$$

自然进化策略

0 初始化当前解θ^0及$t=0$，选择σ

1 确定步长$\alpha^t > 0$

2 根据当前分布$\mathcal{N}(\theta^t, \sigma^2)$采样$M$个$x_m$

3 更新：$\theta^{t+1} = \theta^t - \alpha^t \frac{1}{M} \sum_m \frac{x_m - \theta^t}{\sigma^2} f(x_m)$

4 若满足收敛条件则结束；否则，$t \leftarrow t+1$且返回1

等价写法

0 初始化当前解θ^0及$t=0$，选择σ

1 确定步长$\alpha^t > 0$

2 根据分布$\mathcal{N}(0,1)$采样M个ϵ_m

3 更新：$\theta^{t+1} = \theta^t - \alpha^t \frac{1}{M\sigma} \sum_m \epsilon_m f(\theta^t + \epsilon_m \sigma)$

4 若满足收敛条件则结束；否则，$t \leftarrow t+1$且返回1

不需要$f(\boldsymbol{x})$的梯度信息

自然进化策略

Evolution strategies as a scalable alternative to reinforcement learning

T Salimans, J Ho, X Chen, S Sidor… - arXiv preprint arXiv …, 2017 - arxiv.org

We explore the use of Evolution Strategies (ES), a class of black box optimization algorithms, as an alternative to popular MDP-based RL techniques such as Q-learning and …

☆ Save 99 Cite Cited by 1622 Related articles »

Evolution Strategies as a Scalable Alternative to Reinforcement Learning

Tim Salimans **Jonathan Ho** **Xi Chen** **Szymon Sidor** **Ilya Sutskever**

OpenAI

Abstract

We explore the use of Evolution Strategies (ES), a class of black box optimization algorithms, as an alternative to popular MDP-based RL techniques such as Q-learning and Policy Gradients. Experiments on MuJoCo and Atari show that ES is a viable solution strategy that scales extremely well with the number of CPUs available: By using a novel communication strategy based on common random numbers, our ES implementation only needs to communicate scalars, making it possible to scale to over a thousand parallel workers. This allows us to solve 3D humanoid walking in 10 minutes and obtain competitive results on most Atari games after one hour of training. In addition, we highlight several advantages of ES as a black box optimization technique: it is invariant to action frequency and delayed rewards, tolerant of extremely long horizons, and does not need temporal discounting or value function approximation.

自然进化策略

Algorithm 1 Evolution Strategies

Input: Learning rate α, noise standard deviation σ, initial policy parameters θ_0
for $t = 0, 1, 2, \ldots$ **do**
  Sample $\epsilon_1, \ldots \epsilon_n \sim \mathcal{N}(0, I)$
  Compute returns $F_i = F(\theta_t + \sigma\epsilon_i)$ for $i = 1, \ldots, n$
  Set $\theta_{t+1} \leftarrow \theta_t + \alpha\frac{1}{n\sigma}\sum_{i=1}^{n} F_i\epsilon_i$
end for

Algorithm 2 Parallelized Evolution Strategies

Input: Learning rate α, noise standard deviation σ, initial policy parameters θ_0
Initialize: n workers with known random seeds, and initial parameters θ_0
for $t = 0, 1, 2, \ldots$ **do**
  for each worker $i = 1, \ldots, n$ **do**
    Sample $\epsilon_i \sim \mathcal{N}(0, I)$
    Compute returns $F_i = F(\theta_t + \sigma\epsilon_i)$
  end for
  Send all scalar returns F_i from each worker to every other worker
  for each worker $i = 1, \ldots, n$ **do**
    Reconstruct all perturbations ϵ_j for $j = 1, \ldots, n$ using known random seeds
    Set $\theta_{t+1} \leftarrow \theta_t + \alpha\frac{1}{n\sigma}\sum_{j=1}^{n} F_j\epsilon_j$
  end for
end for

本讲小结

- 有约束凸优化问题、KKT条件、内点法
- 自然进化策略

主要参考资料

Daniel P. Palomar <ELEC5470/IEDA6100A - Convex Optimization> Slides

Stephen Boyd, Lieven Vandenberghe <Convex Optimization> Book

neos-guide.org <Guide to Optimization>

Ashwin Rao <CME241 - Evolutionary Strategies> Slides

Salimans et al., <Evolution Strategies as a Scalable Alternative to Reinforcement Learning> Paper

谢谢！