第一章 介绍

优化的目标是为了"最大"和"最小"。比如对人类而言，最大的利益，最小的风险。对自然也是如此。很多自然系统能稳定存在和它的能量最小状态密切相关。从数学角度看，最大和最小并无区别，统一的，我们用

$$\min_{x \in \Omega} f(x)$$

的形式来表示。这里 f 被称为目标函数（objective function），x 被称为变量（variables），它一般是一个向量。Ω 表示 x 可以取值的范围，称为可行域（feasible region）。

这里 f 和 Ω 的不同取法会导致完全不同性质的模型（modeling）。比如，若 $\Omega \subseteq \mathbb{R}^n$ 是连续区域，f 在 Ω 上连续是一种典型的情况；而相对的，比如 Ω 是离散的点集，而 f 在这些离散的点上有定义（自然不可能连续）。甚至 f 未必是确定的映射，它可以包含一个在 Ω 中分布的随机变量。这三种典型模型，分别被称为连续优化问题、组合优化问题和随机优化问题。它们的数学基础和理论体系完全不同，因此分别有三门不同的专业课程来讲授。我们这门课，只限第一种模型：连续优化问题的算法。

为了能胜读本课程的学习，你必须确保你已经具备了如下基础知识：

- 数学分析（特别是微分学）
- 线性代数（特别是线性空间、特征分解）
- 计算机编程基础（任何一门语言）

如果你缺少以上特别是数学基础，我强烈建议你考虑一下调整你的课程选择。而本课程中的内容，将会在未来的绝大多数应用数学相关领域用到。

我们先进一步缩小讨论的范围，并建立符号体系。

形如：

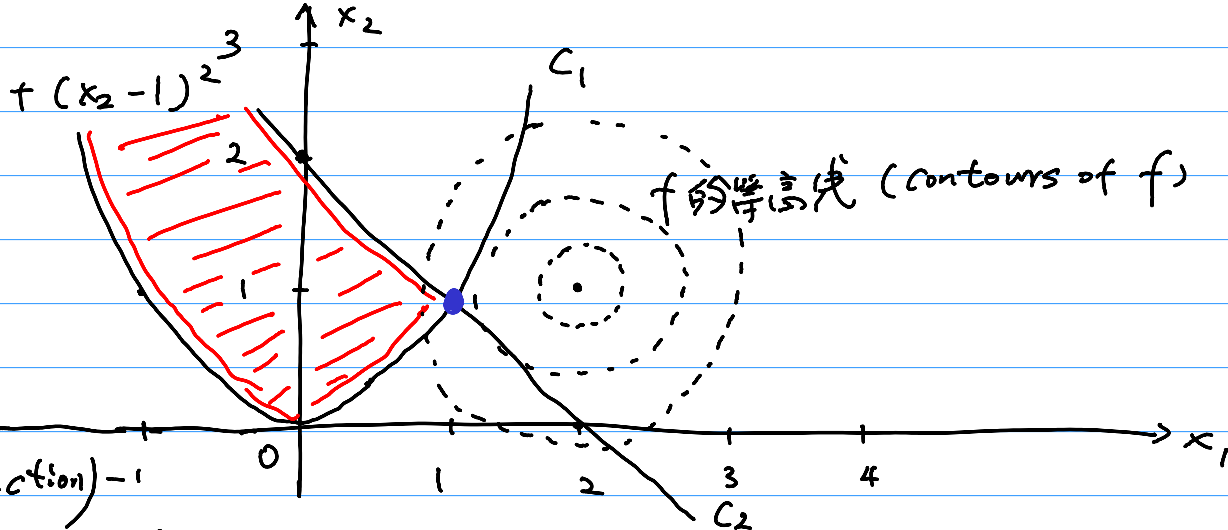
$$\begin{cases}\min\limits_{\vec{x}\in\mathbb{R}^n} f(\vec{x}) \\ \text{s.t.}\quad c_i(x)=0,\ i\in E \\ \qquad\ \ c_i(x)\geq 0,\ i\in I\end{cases}$$

Ω 的代数表示.

这里 s.t. 是 subject to 的缩写，意味着 $\vec{x}$ 并非在 $\mathbb{R}^n$ 的全空间寻找，我们必须在一系列 $c_i(x)=0$ 和 $c_i(x)\geq 0$ 规定的闭区域中寻找我们的能使 $f(\vec{x})$ 取到最小值的 $\vec{x}$. 显然闭集对这种最小值的存在性是重要的。

例如：

$$\begin{cases}\min\ f(x_1,x_2)=(x_1-2)^2+(x_2-1)^2 \\ \text{s.t.}\quad c_1:\ x_1^2-x_2\leq 0 \\ \qquad\ \ c_2:\ x_1+x_2\leq 2\end{cases}$$



s.t.约束的闭集，称为可行域 (feasible region)，在此例中即加红色部分。我们需要在可行域中寻找使目标函数 (object function)

(我们以后都默认用 $\vec{x}^*$ 表示优化问题的解 (solution))

取极小值的点 $\vec{x}^*$. 此例中似乎看上去是蓝色的那点。在这里根据我们的约定，$E=\phi$, $I=\{1,2\}$.

基本上所有我们感兴趣的优化模型都可以化为上面这种形式，除了有时下标会再更复杂一些，比如两个变化的整数 $\vec{x}_{ij}$, $i,j\in\mathbb{Z}^+$. 或者我们求的不是极小 min 而是极大 max. 但后者很容易用一个替换 f 使其变为 min 问题.

运输问题 (transportation problem) 的例子：一个公司有两个工厂 F_1 和 F_2 和 12 个门市 $R_1, R_2, \cdots, R_{12}$。每个工厂 F_i 的每周出货量是 a_i，而每个门市 R_j 每周的需求是 b_j。而从工厂 F_i 到门市 R_j 的单位运费是 c_{ij}。请决定运输方案以满足所有要求并使总运输成本最低。令从 F_i 到 R_j 的每周运量是 x_{ij}，则

$$
\begin{cases}
\min \quad f(\{x_{ij}, i=1,2, j=1,2,\cdots,12\}) = \sum_{i=1,2,\, j=1,2,\cdots,12} c_{ij}\, x_{ij} & \text{(运输总成本)}. \\
\text{s.t.} \quad \sum_{j=1}^{12} x_{ij} \le a_i, \quad i=1,2 & \text{(每个工厂运到 12 个门市的总量)}
\end{cases}
$$

$$\sum_{i=1}^{2} x_{ij} \ge b_j, \quad j=1,2,\cdots,12 \quad \text{(每个门市的每周获得总量)}$$

$$x_{ij} \ge 0, \quad i=1,2, \quad j=1,2,\cdots,12 \quad \text{(运量不得为负数)}$$

考虑这里的约束是否合理？是否可能有其它可接受的设计？

上述模型的一个特点是目标函数和所有约束都是线性函数。这样的模型也被称为线性规划。它在实际生产中十分有用。(linear programming)

连续和离散 (continus and discrete) 优化 (optimization)

注意到上述问题中，我们的目标函数都是连续函数（更多情况下还会要求连续可微）。若 t 不在 $\mathbb{R}$ 中取值，而是在 $\mathbb{Z}$ 中取值，那么我们不再有连续性，而问题也相应的变为整数规划 (integer programming)。比如在运输问题中，我们运输的是某种固定单位的产品，比如冰箱，那么我们只能以整数为单位运输。有别于连续优化问题，这是一类离散优化问题。这里我请大家仔细想一下连续性的本质，造成这两

连续和离散的本质区别。

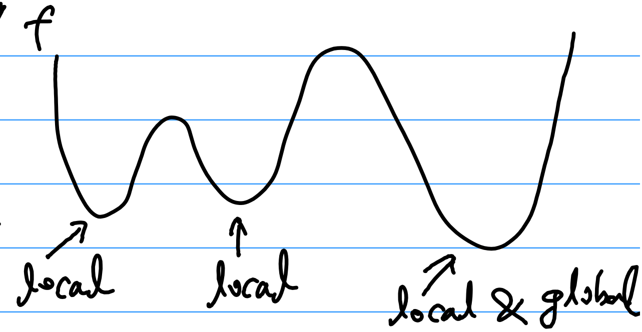
类问题在算法选择上有什么本质不同。这里我提两个例子供大家参考：①请寻找紫金港校区的最低点；②此时此刻，我们班里谁身上现金最多。

约束(constrained)和无约束(unconstrained)优化.

若 s.t. 后面的可行域是全体 $\mathbb{R}^n$，则此时我们可以无视 s.t.，这就是无约束优化问题，否则则是约束优化问题。当目标函数 f 连续可微时，无约束优化问题立刻变为求 $\nabla f=0$ 的非线性方程求根问题。而约束优化问题，我们一般要设法将其转为等价的无约束优化问题，例如 Lagrange 乘子法。我们前面已经看到了线性规化，若目标函数和约束中至少有一个非线性函数，则问题相应的被称为非线性规化(nonlinear programming)。这个是我们这门课所重点讨论的.

局部(local)和全局(global)优化.

对于连续问题，一个邻域内的最优值被称为问题的局部最优解。而我们往往更关心的是全局最优解。当存在多个局部最优解时，确定全局最优解是困难的，特别是当问题可行域复杂或目标函数变化剧烈时。尽管复杂的问题有其实际背景，但数学家们往往喜欢先捡软柿子捏，选出可行区域是凸集，且目标函数也是凸函数的问题。在这样的假设下，我们往往有一些很好的结论(尽管未必实用)，比如有且仅有一个局部最优解，因此它必然也是全局最优。



随机(stochastic)和确定(deterministic)优化.

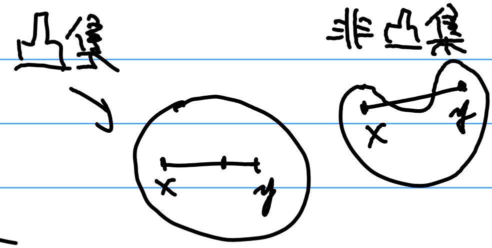
随机优化问题指目标函数中含有随机项或随机变量。这类问题往往有其实际背景，特别是金融领域。(一个有趣的结果是金融海啸)。尽管这个问题很热，我们在这个学期不会去研究。但我们不排除会用一些随机和统计的算法.

优化算法 (algorithms)

优化算法一般是迭代算法. 也即我们需要从一个初值 x_0 开始, 构造一个序列 x_k, 使 $x_k \to x^*$. 和大多数算法一样, 我们在构造算法的时候要考虑:

- 鲁棒性(或健壮性, Robustness), 问题对参数和初值要有一定的容忍性, 不能太敏感.
- 效率 (Efficiency), 用尽量少的内存, 用尽量少的机器时间.
- 准确性 (Accuracy), 数值解要有一定的精度.

上述三个要素必须结合在一起考虑, 单独追求其中一项是没有意义的.



凸性 (Convexity)

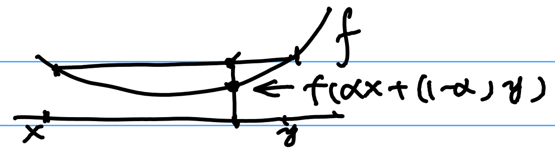
凸集: 称 $S \in R^n$ 是凸集, 若S中任意两点的连线也属于S. 用公式表达, 则是

$$\forall x, y \in S, \text{有} \quad \alpha x + (1-\alpha) y \in S, \quad \forall \alpha \in [0,1].$$

凸函数: 称定义在凸集上的函数f是凸函数, 若定义域中任意两点, 它们像之间的连线都在f的图像之上.
用公式表达, 则是

$$\forall x, y \in S, \text{有} \quad f(\alpha x + (1-\alpha) y) \le \alpha f(x) + (1-\alpha) f(y). \quad \forall \alpha \in [0,1].$$

称f是凹的 (concave) 若 $-f$ 是凸的.



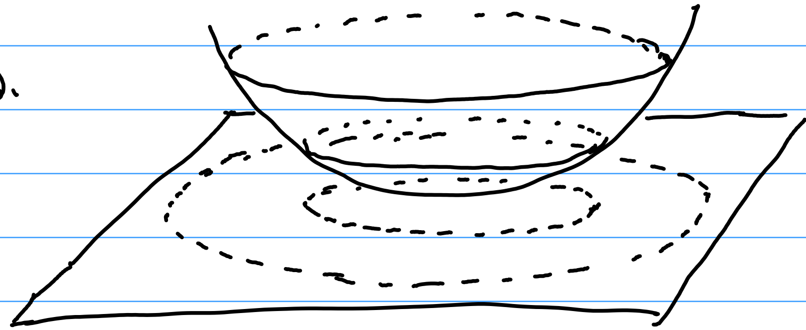
于是我们发现凸函数基本上就是一只碗的形状, 所以它的无约束优化具有全局最优解就很明显了. 而有一类特殊的优化问题称为凸优化 (convex programming), 若:

i) 目标函数是凸函数;

ii) 等式(equality)约束函数 $c_i(\cdot)$, $i \in E$, 都是线性的;

iii) 不等式 (inequality) 约束函数 $c_i(\cdot)$, $i\in\mathcal{I}$, 都是凹的.

(ii)、iii) 说的其实是约束区域可以构成一个封闭的凸区域.)



二维凸函数示意.