Homework 7

陈远洋

Problem 1

Consider the problem

$$\max_{x_1,\cdots,x_n} x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$$
$$\text{s.t.} \sum_{i=1}^n x_i = 1,\ x_i \geq 0, i = 1,\cdots,n,$$

where a_i are given positive scalars. Find a global maximum and show that it is unique.

Solution

令 $\beta_i = \frac{a_i}{\sum_{i=1}^n a_i}$, $y_i = \frac{x_i}{\beta_i}$, $Q = x_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$。
则有 $\ln Q = \sum_{i=1}^n a_i \ln x_i = (\sum_{i=1}^n a_i)(\sum_{i=1}^n \beta_i \ln y_i + \sum_{i=1}^n \beta_i \ln \beta_i)$。所以要使得 Q 最大，只需使得 $\sum_{i=1}^n \beta_i \ln y_i$ 最大。而由 Jensen 不等式可知:

$$\begin{aligned}\sum_{i=1}^n \beta_i \ln y_i &\leq \ln(\sum_{i=1}^n \beta_i y_i)\\ &= \ln(\sum_{i=1}^n x_i)\\ &= 0\end{aligned}$$

当且仅当 $y_1 = \cdots = y_n = 1$, 即 $x_i = \frac{a_i}{\sum_{i=1}^n a_i}$ 时等号成立。此时 Q 最大可取得 $\prod_{i=1}^n (\frac{a_i}{\sum_{i=1}^n a_i})^{a_i}$。

其实由 Jensen 不等式已经可知上面取值的唯一性。但是下面我们对此严加证明。

首先，我们考虑 $g(x) = \ln x$ 的凹凸性：我们给出以下引理：

Theorem 1. $-g(x)$ 在 $\mathbb{R}^+$ 上是强凸的，即 $\forall x \neq y \in \mathbb{R}^+, \forall \lambda \in (0,1)$, $g(\lambda x + (1-\lambda)y) > \lambda g(x) + (1-\lambda)g(y)$。

证明. 固定 $y \in \mathbb{R}^+$, $\lambda \in (0,1)$。令 $f(x) = \lambda g(x) + (1-\lambda)g(y) - g(\lambda x + (1-\lambda)y)$。容易看出 $f(y) = 0$。对 $f(x)$ 求导，有 $f'(x) = \lambda g'(x) + \lambda g'(\lambda x + (1-\lambda)y) = \lambda(\frac{1}{x} - \frac{1}{\lambda x + (1-\lambda)y})$。可知

$\forall 0 < x < y,\ f'(x) > 0; \forall x > y,\ f'(x) < 0$。所以 f(x) 在 $(0, y)$ 严格单调递增，在 $(y, +\infty)$ 严格单调递减。由此可知 $f(x) < 0, \forall x \neq y \in \mathbb{R}^+$。 $\square$

进一步由于 $h(X) = \ln Q = \sum_{i=1}^n a_i \ln x_i$ 是若干个关于 x_i 的强凹函数凸组合，所以 $h(X)$ 也是强凹函数。所以如果 $\exists X_1 \neq X_2 \in \mathbb{R}^{\text{n}+}$, $s.t.\, h(X_1) = h(X_2)$ 并且二者都是最优解。那么 $\forall \lambda \in (0,1)$, $h(\lambda X_1 + (1-\lambda) X_2) > \lambda h(X_1) + (1-\lambda) h(X_2) = h(X_1)$。此时与最优解的条件矛盾，所以最优解一定是唯一的。

tips: 这题如果不用琴声不等式的话也可以用拉格朗日乘子法，对原函数取对数后加上 $\lambda \sum x_i$。再对 x_i 求导得到 $\frac{a_i}{x_i} - \lambda = 0$，而 $\sum x_i = 1$，所以 $x_i = \frac{a_i}{\sum_{i=1}^n a_i}$。

Problem 2

Show that if x^* is a local minimum of the twice continuously differentiable function $f : \mathbb{R}^n \mapsto \mathbb{R}$ over the convex set X, then

$$(x - x^*)^T \nabla^2 f(x^*)(x - x^*) \geq 0,$$

for all $x \in X$, such that $\nabla f(x^*)^T (x - x^*) = 0$.

Solution

采用反证法，假设这不是必要的。即有 x^* 是凸集 X 二阶连续可微函数 $f : \mathbb{R}^n \mapsto \mathbb{R}$ 的局部最小值，并且 $\exists x \in X, \nabla f(x^*)^T (x - x^*) = 0$，且有 $t = (x - x^*)^T \nabla^2 f(x^*)(x - x^*) < 0$。
令 $g(h) = f(x^* + h(x - x^*))$, $h \in [0, 1]$。由 X 是凸集可知 $\forall h \in [0,1]$, $x^* + h(x - x^*) \in X$。又因为 f 二阶连续可微，所以 g 也是二阶连续可微。将 $g(h)$ 对 h 在 0 点二阶泰勒展开，有

$$\begin{aligned}
g(h) &= g(0) + g'(0)h + g''(0)\frac{h^2}{2} + o(h^2) \\
&= f(x^*) + \nabla f(x^*)^T (x - x^*) h + (x - x^*)^T \nabla^2 f(x^*)(x - x^*)\frac{h^2}{2} + o(h^2) \\
&= f(x^*) + t\frac{h^2}{2} + o(h^2) \\
&= f(x^*) + \frac{h^2}{2}(t + o(1))
\end{aligned}$$

由 $o(1)$ 的性质可知 $\exists c > 0$, $s.t. \forall h \in [0, c]$ 对余项 $o(1)$ 有：$|o(1)| < \frac{|t|}{2}$。那么 $\forall h \in [0, c]$, $f(x + h(x - x^*)) = g(h) < f(x^*) + \frac{h^2}{2}(t + \frac{|t|}{2}) = f(x^*) + \frac{th^2}{4} < f(x^*)$。令 $D_1 = \{x_1 | x_1 = x^* + h(x - x^*),\ h \in [0, c]\}$. 所以对任意 x^* 的邻域 $D = B_\epsilon^o(x^*) \bigcap X$，$\exists x_2 \in D \bigcap D_1$，使得 $f(x_2) < f(x)$。这与 x^* 是局部最小点矛盾。

综上所述，如果 x^* 是凸集 X 二阶连续可微函数 $f:\mathbb{R}^n \mapsto \mathbb{R}$ 的局部最小值，并且有 $x \in X$, $s.t\, \nabla f(x^*)^T(x-x^*)=0$。那么必有 $(x-x^*)^T\nabla^2 f(x^*)(x-x^*) \geq 0$。

Problem 3

A farmer annually producing x_i units of a certain crop stores $(1-u_i)x_i$ units of his production, where $0 \leq u_i \leq 1$, and invests the remaining $u_i x_i$ units, thus increasing the next year's production to a level x_{i+1} given by

$$x_{i+1} = x_i + w u_i x_i, i = 0, 1, \cdots, N-1,$$

where w is a given positive scalar. The problem is to find the optimal investment sequence $u_0, \cdots, u_{N-1}$ that maximizes the total product stored over N years

$$x_N + \sum_{i=0}^{N-1}(1-u_i)x_i.$$

Show that one optimal sequence is given by:
(1) If $w > 1$, $u_0^* = \cdots = u_{N-1}^* = 1$.
(2) If $0 < w < \frac{1}{N}$, $u_0^* = \cdots = u_{N-1}^* = 0$.
(3) If $\frac{1}{N} \leq w \leq 1$,

$$u_0^* = \cdots = u_{N-\bar{i}-1}^* = 1,$$
$$u_{N-\bar{i}}^* = \cdots = u_{N-1}^* = 0,$$

where $\bar{i}$ is the integer such that $1/(\bar{i}+1) < w \leq 1/\bar{i}$.

Solution

记 $P_n = \prod_{i=0}^{n}(1+wu_i) > 0$, $P_{-1} = 1$。由题意总的收益函数 $Q_N(\vec{u}) = x_0((\sum_{i=0}^{N-1}(1-u_i)P_{i-1}) + P_{N-1})$ 要使得 $Q_N(\vec{u})$ 最大，即要使得 $Q_N^*(\vec{u}) = P_{N-1} + \sum_{i=0}^{N-1}(1-u_i)P_{i-1}$ 最大。

现在来看 $Q_N^*(\vec{u})$ 关于 u_{N-1} 的项：$(1-u_{N-1})P_{N-2}+(1+wu_{N-1})P_{N-2} = (2+(w-1)u_{N-1})P_{N-2}$。由于其它项与 u_{N-1} 无关，所以 $Q_N^*(\vec{u})$ 要取得最大值，必须要这两项取得最大值。而这两项是关于 u_{N-1} 的线性函数可以知道当系数大于 0 时，u_{N-1} 越大越好（这里最多取 1）；反之，u_{N-1} 越小越好（这里最小取 0）。由于系数与前面的 u_i 无关，所以无论前面哪个 u_i 取什么值，u_{N-1} 按照这样的取法都可以使收益函数取到最大值。即有 u_{N-1} 的最优反应: $\begin{cases} u_{N-1}^* = 0, & w \leq 1 \\ u_{N-1}^* = 1, & w > 1 \end{cases}$

像这样从后往前分析，下面我们来证明两个命题。

Lemma 1. *当对 $u_{N-1}, \cdots, u_{N-k-1}$ 的最优反应均为 0 时，u_{N-k}^* 的最优反应为 0 当且仅当 $w \leq \frac{1}{k}$。*

证明. u_{N-k} 代表着第 $N-k$ 年的投资比例，所以其仅与后面 $k+1$ 年的收益有关。而给定后面 $k-1$ 个最优反应均为 0 时。此时总的收益函数关于 u_{N-k} 的项变为

$$\begin{aligned} f(u_{N-k}) &= P_{N-1} + \sum_{i=N-k}^{N-1}(1-u_i)P_{i-1} \\ &= P_{N-k} + (k-1)P_{N-k} + (1-u_{N-k})P_{N-K-1} \\ &= (k(1+wu_{N-k}) + (1-u_{N-k}))P_{N-k-1} \\ &= (k+1+(kw-1)u_{N-k})P_{N-k-1} \end{aligned}$$

跟前面关于 u_{N-1} 的分析一样，当系数大于 0 时，u_{N-k} 越大越好；反之，u_{N-k} 越小越好。并且系数与前面的 $u_0,\cdots,u_{N-k-1}$ 无关，所以无论前面哪个 u_i 取什么值，u_{N-k} 按照这样的取法都可以使收益函数取到最大值。所以有 u_{N-k} 的最优反应：$\begin{cases} u^*_{N-k}=0, & kw-1\le 0 \Leftrightarrow w\le \frac{1}{k} \\ u^*_{N-k}=1, & kw-1>0 \Leftrightarrow w>\frac{1}{k} \end{cases}$ □

Lemma 2. *如果在从后往前分析时 u_{N-k+1} 的最优反应分析出来为 1，那么 u_{N-k} 的最优反应为 1。*

证明. 由上此时必有 $w>\frac{1}{k}$。此时关于 u_{N-k} 的项 $f(u_{N-k})$ 大于 $u_{N-k},\cdots,u_{N-1}$ 全部取 0 时后面项的和。即 $f(u_{N-k})$ 关于 P_{N-k-1} 前面的系数 h 大于 $k+1$。那么此时对于总的收益函数中关于 u_{N-k-1} 的项有：

$$\begin{aligned} f(u_{N-k-1}) &= h(1+wu_{N-k-1})P_{N-k-2} + (1-u_{N-k-1})P_{N-k-2} \\ &= (h+1+(hw-1)u_{N-k-1})P_{N-k-2} \end{aligned}$$

此时关于 u_{N-k-1} 的系数 $hw-1>(k+1)w-1>(k+1)\frac{1}{k}-1>0$，所以 u_{N-k-1} 的最优反应为 1。 □

综上所述我们对 (1),(2),(3) 的最优反应有了如下的结论：

1. 当 $w>1$ 时，此时首先对 u_{N-1} 的最优反应为 1，而由 Lemma2 可推知前面所有 u_i 的最优反应均为 1，所以 $u^*_0=\cdots=u^*_{N-1}=1$。
2. 当 $0<w<\frac{1}{N}$ 时，此时首先对 u_{N-1} 的最优反应为 0，而由 Lemma1 可推知前面所有 u_i 的最优反应均为 0，所以 $u^*_0=\cdots=u^*_{N-1}=0$。
3. 当 $\frac{1}{N}\le w\le 1$ 时，此时首先对 u_{N-1} 的最优反应为 0，而由 Lemma1 可推知直到 $u_{N-\bar{i}}$ 的最优反应为 0，到 $u_{N-\bar{i}-1}$ 时的最优反应变为 1，再由 Lemma2 可推知前面所有 u_i 的最优反应均为 1，所以 $u^*_0=\cdots=u^*_{N-\bar{i}-1}=1$，而 $u^*_{N-\bar{i}}=\cdots=u^*_{N-1}=0$。

tips: 其实这题仔细分析后（类似于 Lemma2 的分析）可知最优的反应一定是前面每个 u_i 取 1，而后面所有 u_i 取 0。那么假设在 k 截断此时收益函数变为：$x_0\sum_{i=k}^{N}\prod_{j=0}^{k-1}(1+wu_j)=x_0(N-k+1)(1+w)^k$，对 $k=s,s+1$ 的收益函数做商即可推出最优的截断点为 $N-\bar{i}$。

Problem 4

Consider the problem

$$\min_x \frac{f(x)}{g(x)}$$
$$\text{s.t.} x \in X,$$

where $f:\mathbb{R}^n \mapsto \mathbb{R}$ and $g:\mathbb{R}^n \mapsto \mathbb{R}$ are given functions and X is a given subset such that $g(x)>0$ for all $x\in X$. For $\lambda\in\mathbb{R}$, define

$$Q(\lambda)=\inf_{x\in X}\{f(x)-\lambda g(x)\},$$

and suppose that a scalar λ^* and a vector $x^*\in X$ satisfy $Q(\lambda^*)=0$ and

$$x^*\in \operatorname*{argmin}_{x\in X}\{f(x)-\lambda g(x)\}.$$

Show that x^* is an optimal solution of the original problem. Use this observation to suggest a solution method that does not require dealing with fractions of functions.

Solution

若存在这样的 λ^*，则有 $Q(\lambda^*)=\inf_{x\in X}\{f(x)-\lambda^* g(x)\}=0$，也即 $f(x)\geq\lambda^* g(x),\forall x\in X$。而 $g(x)>0\Rightarrow\frac{f(x)}{g(x)}\geq\lambda^*,\forall x\in X$。而 $x^*\in X, s.t. f(x)-\lambda^* g(x)$ 能取到最小值 0，即 $\frac{f(x^*)}{g(x^*)}=\lambda^*$。所以 $\forall x\in X, \frac{f(x)}{g(x)}\geq\frac{f(x^*)}{g(x^*)}\Rightarrow x^*$ 是最优解。

现在考虑如何将其抽离为算法。首先，引进 λ 作为新变量。对于每一个 $\lambda=\frac{f(x_1)}{g(x_1)}$，我们可以求出 $f(x)-\lambda g(x)$ 在 X 上的最优点，设之为 x_i 如果 $f(x_i)-\lambda g(x_i)=0$，则由上面的分析说明 x_i 已经是最优点。否则一定有 $f(x_i)-\lambda g(x_i)<0$, 因为 $f(x_1)-\lambda g(x_1)=0$，而最优点小于等于之。这个时候我们可以用 $\lambda'=\frac{f(x_i)}{g(x_i)}<\lambda$ 代替之作为新的迭代变量。重复这个过程，直到收敛。写成算法：

关于这个算法的收敛性，有一种说法是如果 $f(x),g(x)$ 连续且 X 是紧集，或者 $f(x),g(x)$ 凸，且 $g(x)>0$ 则该算法收敛到全局最优解。但是我还没验证过。

Algorithm 1 Fractional programming(Dinkelbach)

Input: $x_0 \in X, \epsilon > 0$;
Initialize $\lambda_0 = \frac{f(x_0)}{g(x_0)}$, $Q_0 = -\infty$, $i = 0$;
while $Q_i < -\epsilon$ **do**
    Solve $x_{i+1} = \text{argmin}_{x \in X}\{f(x) - \lambda_i g(x)\}$;
    Compute $Q_{i+1} = f(x_{i+1}) - \lambda_i g(x_{i+1})$;
    Update $\lambda_{i+1} = \frac{f(x_{i+1})}{g(x_{i+1})}$, $i = i + 1$;
end while
return x_i, λ_i;