混合变量优化的网格自适应直接搜索算法

林元莘

2023 年 12 月 31 日

摘要

这篇论文介绍了一种新的用于解决约束混合变量优化问题的无导数网格自适应直接搜索（MADS）算法类，其中变量可以是连续的或分类的。这种新的算法类被称为混合变量 MADS（MV-MADS），它既泛化了用于线性约束混合变量问题的混合变量模式搜索（MVPS）算法，又泛化了用于只包含连续变量的一般约束问题的 MADS 算法。收敛分析利用 Clarke 非光滑微积分，同时建立了一些合理的条件，以确保迭代的子序列收敛到在非光滑和混合变量意义下合适定义的稳定点。

关键词：混合变量优化;MADS 算法;Clarke 非光滑微积分

目录

1 简介 **1**

2 混合变量的 MADS 算法 **3**

2.1 混合变量的局部最优 . . . 3

2.2 搜索步骤 . . . 3

2.3 局部轮询步骤 . . . 4

2.4 扩展轮询步骤 . . . 6

2.5 网格参数的更新 . . . 6

2.6 算法伪代码 . . . 7

3 收敛性分析 **8**

3.1 前提条件 . . . 8

3.2 极限点的方向稳定性 . . . 9

3.3 关于离散变量局部极小点条件 . . . 10

3.4 稳定点条件 . . . 10

4 总结 **12**

1 简介

本篇论文《Mesh adaptive direct search algorithms for mixed variable optimization》（混合变量优化的网格自适应直接搜索算法）来自 Mark A. Abramson, Charles Audet, James W. Chrissis, Jennifer G. Walston 于 2009 年发布在《Optimization Letters》上的文章。

在我们所读的这篇论文中，我们将网格自适应直接搜索（MADS）算法推广到混合变量优化问题，并建立了一个统一的收敛理论。

混合变量优化问题的特点是连续变量和分类变量的组合，后者是离散变量，必须从有限预定义的列表或类别集中取值。分类变量可能是非数值的，比如颜色、形状或材料类型；混合变量优化问题会遇到以下困难：

- 混合变量优化问题会具有十分复杂的可行域和解空间，在算法设计上和计算复杂度上带来巨大挑战。
- 传统的优化算法，例如分支定界法，就不能直接运用在此类问题上。
- 在目标函数或者限制函数是类似黑箱的函数时，在一些情况下，重新制定变量的方法不可行。

为了使算法尽可能具有普遍性，我们允许分类变量值的更改可能带来问题变量维度的变化。我们用 n^c,n^d 来表示连续变量和离散变量的最大维度。我们记可行域为 Ω，目标函数为 $f:\Omega \to R\cup\{\infty\}$，x 为变量。我们令变量 $x=(x^c,x^d)$ 分为连续和类别的两部分，那么我们就有 $x^c\in\Omega^c\subseteq\mathbb{R}^{n^c}, x^d\in\Omega^d\subseteq\mathbb{Z}^{n^d}$。则优化问题为

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

$$\Omega=\bigcup_{x^d\in\Omega^d}(\Omega^c(x^d)\times\{x^d\})$$

在真实算法过程中，我们利用障碍法得到新的障碍目标函数 $f_\Omega=f+$

ψ_Ω，其中

$$\psi(x)=\begin{cases}0, & \text{if } x\in\Omega\\ \infty, & \text{if } x\notin\Omega\end{cases} \tag{1}$$

这种方法适用于一般的集合约束。

MADS 算法类是由 Audet 和 Dennis 引入并分析的，作为广义模式搜索（GPS）算法的扩展，用于解决非线性约束问题，但没有分类变量。模式搜索最早由 Torczon 作为一种用于非约束问题的无导数算法类引入和分析。Lewis 和 Torczon 将其扩展到有界和线性约束问题。在所有这些情况下，只要目标函数在包含所有迭代点的紧集上连续可微分，就可以证明迭代的子序列收敛到一阶稳定点。Audet 和 Dennis 使用 Clarke 非光滑微积分分析了广义模式搜索（GPS）算法类，用于具有更弱光滑性假设的目标函数的问题。

在 GPS 中，算法系统地评估位于当前迭代点周围的网格或格点上的试点。该网格由一组有限的正定向量定义，即可以将空间中的任何向量表示为该集合中向量的非负线性组合的方向。在处理有限数量的线性约束时，Lewis 和 Torczon 引入了一种选择这些方向的方案，以便它们生成相对于任何附近约束的切线锥。然而，对于非线性约束问题，生成切线锥需要使用无限多的正定方向，这违反了 GPS 理论。已经进行了两次尝试以克服这一限制，但都存在严重限制。例如 filter-GPS 算法，它在实践上表现良好但是存在反例使其收敛性无法保证。

MADS 算法类作为对 filter-GPS 的替代方法被引入，但具有更强的收敛性质。在 MADS 中，添加了一个新参数，使算法在选择正定方向时更加灵活，以至于变量空间在渐进意义上在一组方向中被探索。在合理的假设下，根据目标函数的光滑性假设，能够收敛到 Clarke 意义上的一阶和二阶稳定点。

2 混合变量的 MADS 算法

2.1 混合变量的局部最优

局部最优是根据局部邻域来定义的。连续变量的邻域是良定的，然而，由于分类变量没有很自然的度量，局部邻域需要使用者自己定义。我们可以通过一个函数 $N:\Omega\to 2^{\Omega}$ 来定义一个一般的局部邻域。按照惯例，我们假设所有的 $x\in\Omega$，使用者定义的 $N(x)$ 是个有限集，并且 $x\in N(x)$。例如，整数型变量常见的邻域定义为 $N(x)=\{y\in\Omega : y^c=x^c, ||y^d-x^d||_1\le 1\}$。但一般的类别变量可能就没有固定的度量，这种选择就会失效。

定义 2.1. *$x=(x^c,x^d)\in\Omega$ 是对于邻域 $N(x)\subset\Omega$, f 的局部最优点，如果存在 $\epsilon>0$ 使得对于所有的 v 在如下集合中，都有 $f(x)\le f(v)$.*

$$\Omega\bigcap\Big(\bigcup_{y\in N(x)}(B_\epsilon(y^c)\times\{y^d\})\Big)$$

MADS 算法的每一次迭代都包括一个可选的搜索步骤、一个局部轮询步骤、一个扩展轮询步骤。这些步骤都要在某个网格内去评估 f_Ω 的具体取值。每一次迭代的目的都是为了寻找一个更加的网格点，也就是说网格内有一点 y 满足 $f_\Omega(y)<f_\Omega(x_k)$，那么这个 y 就是我们要找到更优点。

2.2 搜索步骤

在确定搜索网格之前，我们要确定网格的搜索方向。对于所有离散变量可能取到的组合 $i=1,2,\ldots,i_{\max}$，我们定义一个搜索方向集合 $D^i=G^iZ^i$。其中 G^i 是一个非奇异生成矩阵，Z^i 是一个有限的整数向量集合。D 是一组正定方向，也就是其非负线性组合必然可以张成 $\mathbb{R}^{n^c}$。

每次迭代都要生成一个新的网格 M_k，它生成了一个关于 Ω^d 和 Ω^c 有限格点的直和。

$$M_k=\bigcup_{i=1}^{i_{max}}M_k^i\times\Omega^d$$

$$M_k^i = \bigcup_{x \in V_k} \{x^c + \Delta_k^m D^i z : z \in \mathbb{N}^{|D^i|}\}$$

其中网格大小参数 $\Delta_k^m > 0$ 以及 V_k 表示所有在第 k 步迭代已经被评估过的测试点。并且，邻域必须要能设计成在网格内。

这里搜索步骤可以使用任何策略来在这组网格点 M_k 上评估 f_Ω，因为这个步骤对收敛理论并没有贡献。但选择合适的搜索策略，例如使用替代模型，可以极大提高算法性能。

2.3 局部轮询步骤

在 MADS 轮询中，我们考虑第二网格参数————轮询大小参数 $\Delta_k^p > 0$，对于所有的 k，满足 $\Delta_k^m \le \Delta_k^p$，并且在任意迭代指标的子集 K 上 $\lim_{k\in K} \Delta_k^m = 0 \leftrightarrow \lim_{k\in K} \Delta_k^p = 0$。

然后每一次迭代的轮询方向 $D_k(x) \subseteq D^{i_0}$ 是关于第 i_0 组离散变量的正定方向的子集

定义 2.2. *在每一步迭代时，MADS 框架如下定义：*

$$P_k(x) = \{(x^c + \Delta_k^m d, x^d) : d \in D_k(x)\} \subset M_k$$

其中 $D_k(x)$ 是一个正生成集。对于任意 $d \in D_k$ 满足：

- *$d \neq 0$ 可以写成 D 中方向的非负整数线性组合。*
- *轮询点的界会被轮询大小参数控制。即 $\Delta_k^m ||d|| \le \Delta_k^p \max\{||d'|| : d' \in D\}$。*
- *归一化集 $D_k(x)$ 的极限也是正生成集。*

局部轮询主要评估的便是 $P_k(x_k) \bigcup N(x_k)$ 上的点。如果局部轮询没有找到更小的目标函数值，那么由扩展轮询步骤来查找部分有希望的值的周边值。

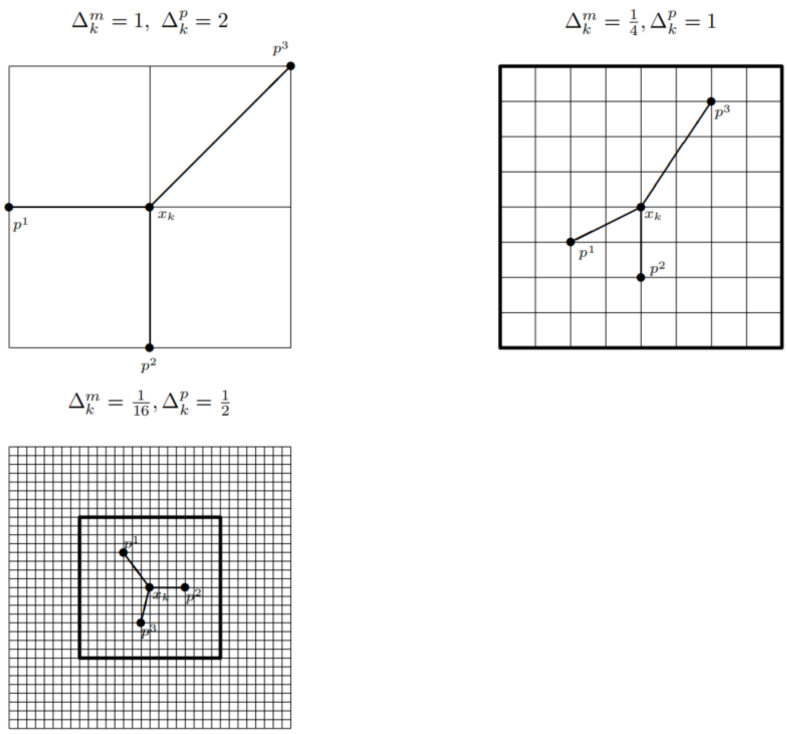

图 1: 当 MADS 框架 $P_k = \{x_k + \Delta_k^m d : d \in D_k\} = \{p^1, p^2, p^3\}$ 并且网格大小参数 Δ_k^m 和轮询大小参数 Δ_k^p 不同时的例子

2.4 扩展轮询步骤

在扩展轮询步骤中，我们放松了轮询条件，考虑一个容忍值 $\xi_k \geq \xi$，寻找那些邻域内满足 $f_\Omega(x_k) \leq f_\Omega(y) \leq f_\Omega(x_k) + \xi_k$ 的点 y_k，再对他们做轮询步骤。每次对 y_k 轮询都会产生一个新的迭代点，即形成序列 $\{y_k^j\}_{j=1}^{J_k}$，从 $y_k^0 = y_k$ 开始，扩展轮询的结束点 $z_k = y_k^{J_k}$ 出现时要满足存在 $d \in D_k(z_k)$ 使得 $f_\Omega(z_k^c + \Delta_k^m d, z_k^d) < f_\Omega(x_k)$ 或者对任意 $d \in D_k(z_k)$ 有 $d \in D_k(z_k)$ 使得 $f_\Omega(z_k^c + \Delta_k^m d, z_k^d) \geq f_\Omega(x_k)$。

扩展轮询点集表示如下：

$$X_k(\xi_k) = \bigcup_{y_k \in N_k^{\xi_k}} \bigcup_{j=1}^{J_k} P_k(y_k^j)$$

其中 $N_k^{\xi_k} = \{y \in N(x_k) : f_\Omega(x_k) \leq f_\Omega(y) \leq f_\Omega(x_k) + \xi_k\}$。在实践过程中，一般令 $\xi_k = \max\{\xi, 0.05|f(x_k)|\}$。太大的 ξ_k 会产生太多扩展轮询点，会造成太大的开销，但可能会找到更好的局部最优点。

扩展轮询步骤在有限点集合里进行轮询，当出现存在一个方向使得存在更小的目标值或者任意方向任意点都不存在更小的目标函数值则停止。

以上三个步骤的任意一个找到了更好的网格点，那么迭代停止，形成新的迭代点 x_{k+1}。如果没有找到，则细化网格点。

2.5 网格参数的更新

我们设置一个确定的有理数 $\tau > 1$ 和两个整数 $w^- \leq -1, \quad w^+ \geq 0$，网格大小参数 Δ_k^m 如下更新：

$$\Delta_{k+1}^m = \tau^{w_k} \Delta_k^m$$

$$w_k \in \begin{cases} \{0, 1, \ldots, w^+\} & \text{若找到了更优的网格点} \\ \{w^-, w^- + 1, \ldots, -1\} & \text{其他情况} \end{cases}$$

即找到更优点时，我们更新网格中心点，并且放松网格大小，防止过多的搜索；如果没有找到新的更优点，我们保留网格中心点，通过缩小网格大小来尝试找到更优的点。

2.6 算法伪代码

Algorithm 1: 网络自适应方向搜索算法 (MADS)

Input: 目标函数 f，可行域 Ω，算法参数 ξ, Δ_0^m，起始点 x_0

Output: 最优点 x^*

```
设置参数 ξ_0 ≥ ξ，Δ_0^p ≥ Δ_0^m。初始迭代 k = 0;
while True do
    计算 M_k，在 M_k 上进行搜索步骤;
    if 找到更优点 then
        更新中心点和参数，进入下一步循环;
    else
        计算 P_k(x_k)，在 P_k(x_k) ⋃ N(x_k) 上进行局部轮询步骤;
        if 找到更优点 then
            更新中心点和参数，进入下一步循环;
        else
            计算 X(ξ_k)，在 X(ξ_k) 上进行扩展轮询步骤;
            if 找到更优点 then
                更新中心点和参数，进入下一步循环;
            else
                保留中心点，更新网格参数;
            end
        end
    end
    if 迭代次数达到上限或者中心点无法继续更新 then
        break
    end
    k ← k + 1;
end
return x_{k+1};
```

3 收敛性分析

3.1 前提条件

收敛性分析需要依靠一些标准的假设。

- 初始点 x_0 满足 $f_\Omega(x_0) < \infty$。
- 所有迭代点列 $\{x_k\}$ 都落在一个紧集内。
- 离散邻域的集合 $N(x_k)$ 都在网格 M_k 中。
- $\liminf_{k\to\infty} \Delta_k^p = \liminf_{k\to\infty} \Delta_k^m = 0$
- 存在一个最小框架中心的改良子序列 $\{x_k\}$,存在极限点 $\hat{x} = \lim_{k\in K} x_k$, $\hat{y} = \lim_{k\in K} y_k$。对于每次扩展轮询步骤从 y_k 开始生成的终止点 z_k 并且 $\lim_{k\in K} \Delta_k^p = 0$ 也有极限点 $\hat{z} = \lim_{k\in K} z_k$.
- $\hat{y} \in N(\hat{x})$

以下定义是 Clake 非光滑微积分的基础，也是下面收敛性证明的关键。

定义 3.1. 向量 $v \in R^{n^c}$ 是一个在点 x 上，对于 Ω 的连续变量的超切向量。如果其满足 $\exists\epsilon > 0$，使得对于所有 $y \in B_\epsilon(x^c)$，$w \in B_\epsilon(v)$ 和 $t \in (0,\epsilon)$，满足 $(y + tw, x^d) \in \Omega$。
Ω 在 x 点的所有超切向量形成的集合 $T_\Omega^H(x)$ 称为超切锥。

定义 3.2. 向量 $v \in R^{n^c}$ 是一个在点 x 上，对于 Ω 的连续变量的 *Clarke* 切向量。如果任意向量列 $\{y_k\}$ 收敛到 x^c，任意正实数列 $\{t_k\}$ 收敛到 *0*，存在收敛到 v 的向量列 $\{w_k\}$ 满足 $(y_k + t_k w_k, x^d) \in \Omega$。
所有 *Clarke* 切向量集合 $T_\Omega^{Cl}(x)$ 称为 *Clarke* 切锥。
记 x 点的切锥记作 $T_\Omega^{C_0}(x)$，如果 $T_\Omega^{Cl}(x) = T_\Omega^{C_0}(x)$，那么称 Ω 在 x 点正则。

以下我们定义广义微积分，来研究不可导函数的微分性质。

定义 3.3. 局部 *Lipschitz* 函数 f 在点 $x=(x^c,x^d)\in\Omega$，关于方向 $v\in\mathbb{R}^{n^c}$ 的广义方向导数为：

$$f^\circ(x;v)=\limsup_{y\to x^c,(y,x^d)\in\Omega,t\downarrow 0,(y+tv,x^d)\in\Omega}\frac{f(y+tv,x^d)-f(y,x^d)}{t}$$

并且，如果满足 $T_\Omega^H(x)$ 非空、$v\in T_\Omega^{Cl}(x)$ 则有如下结论：

$$f^\circ(x;v)=\lim_{u\to v,\quad u\in T_\Omega^H(x)}f^\circ(x;u)$$

定义 3.4. 局部 *Lipschitz* 函数 f 在点 $x=(x^c,x^d)\in\Omega$ 的广义梯度集合为：

$$\partial f(x)=\{s\in\mathbb{R}^{n^c}:f^\circ(x;v)\geq v^Ts,\forall v\in\mathbb{R}^{n^c}\}$$

为了在混合变量的情况下阐明稳定点（即驻点）的概念，我们引入一些非光滑的术语来定义。

定义 3.5. 若 f 在 x^* 附近 *Lipschitz* 连续。如果对于任意方向 v 属于 *Clarke* 切锥或者切锥，满足 $f^\circ(x^*;v)\geq 0$。那么我们就称 x^* 为 *Clarke* 或者切稳定点。

若该点负梯度存在，并且属于 *Clarke* 切锥或者切锥的极线，就称为 *Clarke* 或切 *KKT* 稳定点

3.2 极限点的方向稳定性

定理 3.6. 令 $\hat{w}$ 表示迭代中的改良点子列或者扩展轮询步骤结束点子列的极限点。如果 f 在 $\hat{w}$ 上对于连续变量 *Lipschitz* 连续，则对于超切锥 $T_\Omega^H(\hat{w})$ 里的任意改良方向 v，满足 $f^\circ(\hat{w};v)\geq 0$

证明. 由离散变量的性质，可以记 $\hat{w}=\lim_{k\in K}w_k,w_k=(w_k^c,\hat{w}^d)$。可以取出 K 的子列 L 可以满足 $v=\lim_{k\in L}\frac{d_k}{||d_k||}$。并且 Δ_k^p 趋于 0 可以保证 $\Delta_k^m||d_k||$ 也是趋于 0 的。极限过程则可以进行拆解。

默认以下极限过程中涉及的点都落在 Ω 内。

$$\begin{aligned}f^\circ(\hat{w};v) &= \limsup_{y\to\hat{w}^c,t\downarrow 0,u\to v,u\in T_\Omega^H(\hat{w})} \frac{f(y+tu,\hat{w}^d)-f(y,\hat{w}^d)}{t}\\ &\geq \limsup_{k\in L} \frac{f(w_k^c+\Delta_k^m d_k,\hat{w}^d)-f(w_k^c,\hat{w}^d)}{\Delta_k^m\|d_k\|}\\ &\geq 0\end{aligned}$$

□

3.3 关于离散变量局部极小点条件

由上定义，$\{x_k\}$ 是框架最小中心点集合子序列，$\hat{x}$ 为其极限点。

定义 3.7. 如果 f 是对于 $\hat{x}$ 下半连续的，对于 $\hat{y}\in N(x_k)$ 上半连续。则有 $f(\hat{x})\leq f(\hat{y})$。

证明. 由 x_k 为框架中心点，那么必有

$$f(\hat{x})=\lim_{k\in K} f(x_k)\leq \lim_{k\in K} f(y_k)=f(\hat{y})$$

□

至于这是在离散变量下的极小点，因为并没有遍历连续变量空间的所有方向，而是只是控制了 $\hat{y}$ 在其邻域内，可以认为保持连续变量此处在邻域内遍历了离散变量，使其在离散变量空间会是局部极小值。

3.4 稳定点条件

定理 3.8. $\hat{x}$ 是所有框架的最小中心点 x_k 的极限点。假设 $T_\Omega^H(\hat{x})\neq\emptyset$ ，并且改良方向在其中渐近稠密。

- 如果 f 在 $\hat{x}$ 附近 *Lipschitz* 连续，那么 $\hat{x}$ 是一个 *Clarke* 稳定点。
- 如果 f 在 $\hat{x}$ 处严格可导，那么 $\hat{x}$ 是一个 *clarke KKT* 稳定点。

若 Ω 在 $\hat{x}$ 上正则，

- 如果 f 在 $\hat{x}$ 附近 *Lipschitz* 连续，那么 $\hat{x}$ 是一个切稳定点。
- 如果 f 在 $\hat{x}$ 处严格可导，那么 $\hat{x}$ 是一个切 *KKT* 稳定点。

证明. 如果超切锥不空，Rockafellar 已经证明了 $T_{\Omega}^{Cl}(\hat{x}) = cl(T_{\Omega}^{H}(\hat{x}))$。那么 f 在 $\hat{X}$ 的改良方向集合 S 会是这两个切锥的稠密子集，那么任意 $v \in T_{\Omega}^{Cl}(\hat{x})$ 都可以被一列改良方向逼近。由定义3.5和定理3.6易证得。

当 f 严格可导时，存在 $\nabla f(\hat{x})$，对于任意 $v \in T_{\Omega}^{Cl}(\hat{x})$，有 $-(\nabla f(\hat{x}))^T v = -f^{\circ}(\hat{x}; v) \leq 0$。由定义3.5即可证得。

当 Ω 在 $\hat{x}$ 处正则时，$T_{\Omega}^{Cl}(\hat{x}) = T_{\Omega}^{Co}(\hat{x})$。如上操作同理可证。 □

下面考虑扩展轮询步骤的终止点 z_k。其中子列极限为 $\hat{z}$。

定理 3.9. 假设 $T_{\Omega}^{H}(\hat{z}) \neq \emptyset$ ，并且改良方向在其中渐近稠密。

- 如果 f 在 $\hat{z}$ 附近 *Lipschitz* 连续，那么 $\hat{z}$ 对于连续变量是一个 *Clarke* 稳定点。
- 如果 f 在 $\hat{z}$ 处严格可导，那么 $\hat{z}$ 对于连续变量是一个 *clarke KKT* 稳定点。

若 Ω 在 $\hat{z}$ 上正则，

- 如果 f 在 $\hat{z}$ 附近 *Lipschitz* 连续，那么 $\hat{z}$ 对于连续变量是一个切稳定点。
- 如果 f 在 $\hat{z}$ 处严格可导，那么 $\hat{z}$ 对于连续变量是一个切 *KKT* 稳定点。

证明如上定理3.8。由于 z_k 为扩展轮询的结束点满足存在 $d \in D_k(z_k)$ 使得 $f_{\Omega}(z_k^c + \Delta_k^m d, z_k^d) < f_{\Omega}(x_k)$ 或者对任意 $d \in D_k(z_k)$ 有 $d \in D_k(z_k)$ 使得 $f_{\Omega}(z_k^c + \Delta_k^m d, z_k^d) \geq f_{\Omega}(x_k)$，和 x_k 有相似的极限动作。以上证明过程就可以同样施加给 z_k。

4 总结

- 本文主要基于混合变量优化问题，提出了 MADS 网格自适应方向搜索算法。
- 解释了 MADS 算法组成的三个步骤，并且给出算法流程。
- 基于 Clarke 非光滑微积分的理论，证明算法具有收敛性。

本篇论文填补了一类 MADS 算法的收敛性理论。并且较先前算法扩展了方法，可以处理混合变量问题，并且给出收敛性证明。但网格搜索计算开销巨大，需要较大存储空间和运算量。

希望将来对 MADS 这一普适性算法继续改进，并且可以将其扩展到随机噪声或者多目标问题上。

本文的最后，要感谢本次研讨课范金燕老师，一学期教给我们的优化知识将受用终身。感谢助教的辛勤付出，整理论文和安排课堂，让我们期末大作业顺利进行。研读完这篇论文感触良多，一方面拓宽了我对优化的认知，一方面也教给我各种混合变量的优化方法，将来在处理实际问题中得以运用。阅读英文论文也增长了我的见识，让我学会了如何查阅专业名词和相关文献，增长了我的学术素养，让我在以后的学习中受益匪浅。