

全国大学生数学建模竞赛论文模板

摘要

摘要

对于问题一,

对于问题二,

对于问题三,

对于问题四,

最后,

关键字: 关键词 关键词 关键词 关键词 关键词

一、问题重述

1.1 问题背景

丝绸之路作为古代中西方文化交流的核心通道，玻璃是早期贸易往来的重要物证。早期西亚和埃及的玻璃多以珠形饰品传入我国，我国古代吸收其技术后，利用本土原料制作玻璃，虽外观与外来品相似，但因助熔剂差异（如铅矿石、草木灰等），化学成分截然不同，形成了铅钡玻璃（我国自创，以楚文化为代表）、高钾玻璃（流行于岭南及东南亚、印度等区域）等本土特色品种。古代玻璃因埋藏环境易风化，风化过程中元素交换导致成分比例改变，影响类别判断，而部分风化文物表面仍保留未风化区域，为成分研究提供了特殊样本，对于研究古代中国社会和玻璃工艺具有很高的价值。

1.2 问题要求

问题 1 分析玻璃文物的表面风化状态与其类型（高钾玻璃 / 铅钡玻璃）、纹饰、颜色之间的关联；结合玻璃类型，总结文物表面有无风化时化学成分含量的统计规律；并基于风化点的检测数据，预测其风化前的化学成分含量。

问题 2 依据附件数据，提炼高钾玻璃与铅钡玻璃的分类规律；针对这两类玻璃，分别选取合适的化学成分进行亚类划分，明确具体的划分方法及结果，并分析该分类结果的合理性与敏感性。

问题 3 对附件表单 3 中未知类别的玻璃文物，通过分析其化学成分鉴别其所属类型（高钾玻璃或铅钡玻璃），并对该分类结果的敏感性进行分析。

问题 4 针对高钾玻璃和铅钡玻璃这两类不同的文物样品，分别分析其内部化学成分之间的关联关系，并比较两类玻璃在化学成分关联关系上的差异性。

二、问题分析

2.1 问题一分析

对于问题一，

2.2 问题二分析

对于问题二，

2.3 问题三分析

对于问题三，

2.4 问题四分析

对于问题四，

三、 模型假设

为简化问题，本文做出以下假设：

- 假设 1
- 假设 2
- 假设 3

四、 符号说明

符号	说明	单位
m	质量	kg
V	体积	m^3

五、 问题一的模型的建立和求解

5.1 玻璃类型、颜色、纹饰与风化的关系

首先我们对表单 2 中各文物采样点的化学成分进行累加，其中样本编号为 15、17 的文物化学成分总和分别为 79.47%、71.89%，不满足题目对成分比例累加和介于 85% 105% 之间的要求，因此我们将其剔除。

为了分析表面风化与玻璃类型、纹饰、颜色之间的关系，我们分别统计（表面风化，玻璃类型）（表面风化，纹饰）（表面风化，颜色）这三个二元组的列联表数据，并进行了可视化。

表 1 表面风化与颜色的列联表

表面风化	颜色							
	浅绿	浅蓝	深绿	深蓝	紫	绿	蓝绿	黑
无风化	2	6	3	2	2	1	6	0
风化	1	12	4	0	2	0	9	2

表 2 表面风化与纹饰、玻璃类型的列联表

(a) 表面风化与纹饰的列联表

表面风化	纹饰		
	A	B	C
无风化	11	0	11
风化	11	6	17

(b) 表面风化与玻璃类型的列联表

表面风化	类型	
	铅钡	高钾
无风化	12	10
风化	28	6

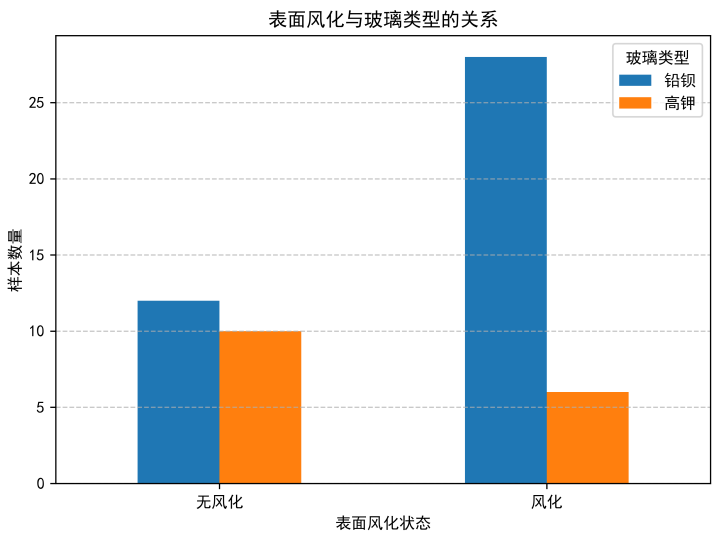


图 1 表面风化与玻璃类型

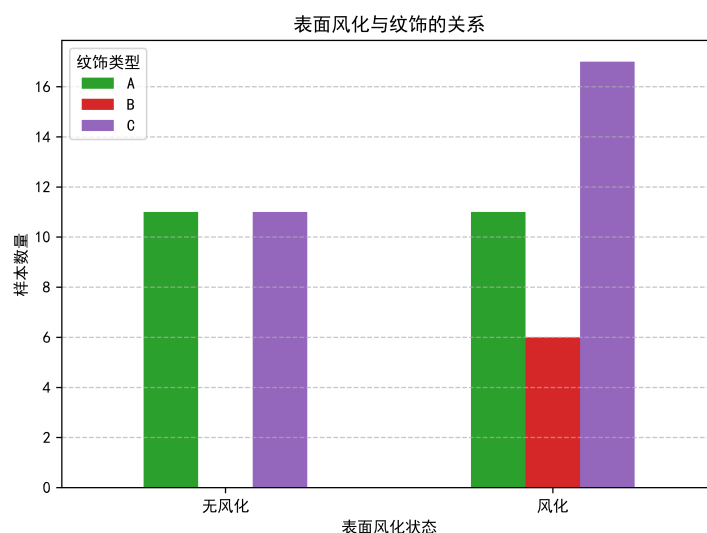


图 2 表面风化与纹饰

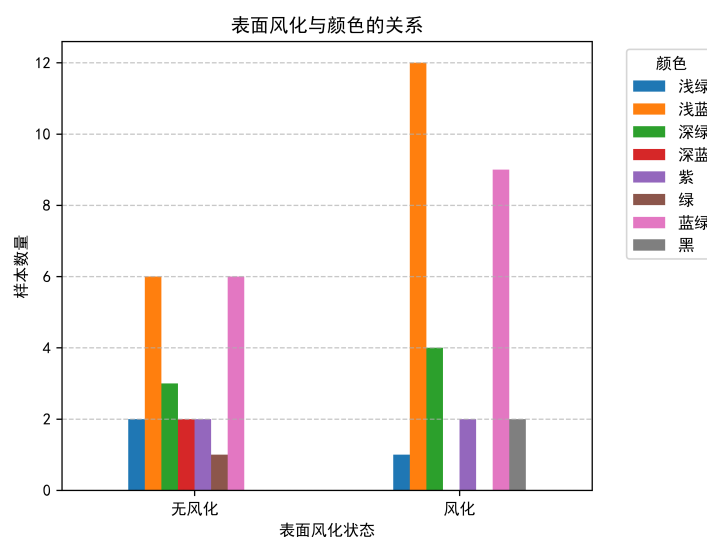


图 3 表面风化与颜色

为了量化表面风化与玻璃类型、纹饰、颜色之间的关系，我们引入了卡方检验。卡方检验用于检验两个分类变量是否独立，通过比较观测值与期望值的差异，用 χ^2 统计量判断关联是否显著，适用于计数数据。

$$\chi^2 = \sum \frac{(O - E)^2}{E} \quad (1)$$

其中： χ^2 ：卡方统计量； O ：实际观测频数； E ：理论期望频数； \sum ：对所有单元格求和。

分别带入（表面风化，玻璃类型）（表面风化，纹饰）（表面风化，颜色）的列联表数据可以求出 χ^2 值和 p 值，我们这里取 $p < 0.005$ 。从下面的表格中我们可以看出，是否风化与玻璃类型之间存在显著关系，而风化与纹饰、颜色之间则不存在显著关系。

表 3 卡方检验结果

关系	χ^2	df	p 值	是否显著
风化 × 颜色	7.0114	7	$p \approx 0.426$	否
风化 × 纹饰	4.9412	2	$p \approx 0.085$	否
风化 × 类型	5.0610	1	$p \approx 0.024$	是

5.2 玻璃是否风化化学成分含量的统计规律

以文物采样点为单位，以玻璃类型、是否风化为分组依据，将数据分为四个组别：

1. 无风化铅钡玻璃
2. 无风化高钾玻璃
3. 风化铅钡玻璃
4. 风化高钾玻璃

我们对预处理后的数据进行统计，计算出了每种组别的化学成分含量的均值、极差、方差、有效样本数。同时，为了更直观的看出化学成分的变化，我们将同一化学成分的风化前后含量做成了柱状图，以下进行部分展示。

表 4 铅钡无风化样本化学成分统计数据

化学成分	均值	极差	方差	有效样本数
二氧化硅 (SiO ₂)	53.4438	43.5700	212.7885	13
氧化钠 (Na ₂ O)	3.3433	2.0000	1.3008	3
氧化钾 (K ₂ O)	0.4356	1.4600	0.2193	9
氧化钙 (CaO)	1.3909	4.1100	2.2167	11
氧化镁 (MgO)	1.4812	4.9400	2.6958	8
氧化铝 (Al ₂ O ₃)	2.8915	3.5600	1.6320	13
氧化铁 (Fe ₂ O ₃)	2.1240	4.4200	2.8855	5
氧化铜 (CuO)	1.8400	8.3500	6.8727	11
氧化铅 (PbO)	23.5938	29.9200	82.7080	13
氧化钡 (BaO)	24.4662	23.2300	62.8734	13
五氧化二磷 (P ₂ O ₅)	1.0682	5.6500	2.7670	11
氧化锶 (SrO)	0.4825	0.6800	0.0664	8
氧化锡 (SnO ₂)	0.4000	0.0000	nan	1
二氧化硫 (SO ₂)	2.0500	3.2200	5.1842	2

表 5 铅钡风化样本化学成分统计数据

化学成分	均值	极差	方差	有效样本数
二氧化硅 (SiO ₂)	33.6147	64.3600	296.5795	36
氧化钠 (Na ₂ O)	3.1173	7.1200	5.4794	11
氧化钾 (K ₂ O)	0.3937	1.3000	0.1208	16
氧化钙 (CaO)	2.4835	6.0300	2.4015	34
氧化镁 (MgO)	1.0970	2.2600	0.2306	23
氧化铝 (Al ₂ O ₃)	3.8383	13.8900	11.6460	36
氧化铁 (Fe ₂ O ₃)	0.9529	2.5500	0.4407	21
氧化铜 (CuO)	2.1135	10.3800	6.3094	34
氧化铅 (PbO)	36.8719	57.9000	229.9385	36
氧化钡 (BaO)	34.5803	64.8600	297.0158	35
五氧化二磷 (P ₂ O ₅)	4.9863	14.0600	16.5052	30
氧化锶 (SrO)	0.4122	1.0000	0.0484	32
氧化锡 (SnO ₂)	0.7700	1.0800	0.5832	2
二氧化硫 (SO ₂)	7.1980	15.4800	57.9916	5

表 6 高钾无风化样本化学成分统计数据

化学成分	均值	极差	方差	有效样本数
二氧化硅 (SiO ₂)	67.9842	28.0400	76.6518	12
氧化钠 (Na ₂ O)	2.7800	1.2800	0.4144	3
氧化钾 (K ₂ O)	9.7233	9.8100	9.2269	12
氧化钙 (CaO)	6.0500	7.4800	6.6337	10
氧化镁 (MgO)	1.3033	1.4600	0.2784	9
氧化铝 (Al ₂ O ₃)	6.6200	8.1000	6.2076	12
氧化铁 (Fe ₂ O ₃)	2.3180	5.6200	2.4002	10
氧化铜 (CuO)	2.6755	4.6200	2.3750	11
氧化铅 (PbO)	0.7057	1.5100	0.3939	7
氧化钡 (BaO)	1.4360	2.8600	1.1488	5
五氧化二磷 (P ₂ O ₅)	1.5300	4.3400	2.0473	11
氧化锶 (SrO)	0.0833	0.0800	0.0010	6
氧化锡 (SnO ₂)	2.3600	0.0000	nan	1
二氧化硫 (SO ₂)	0.4067	0.1100	0.0032	3

表 7 高钾风化样本化学成分统计数据

化学成分	均值	极差	方差	有效样本数
二氧化硅 (SiO ₂)	93.9633	4.4200	3.0054	6
氧化钾 (K ₂ O)	0.7040	0.7500	0.0879	5
氧化钙 (CaO)	0.8700	1.4500	0.2379	6
氧化镁 (MgO)	0.5900	0.1000	0.0050	2
氧化铝 (Al ₂ O ₃)	1.9300	2.6900	0.9302	6
氧化铁 (Fe ₂ O ₃)	0.2650	0.1800	0.0048	6
氧化铜 (CuO)	1.5617	2.6900	0.8739	6
五氧化二磷 (P ₂ O ₅)	0.3360	0.4600	0.0316	5

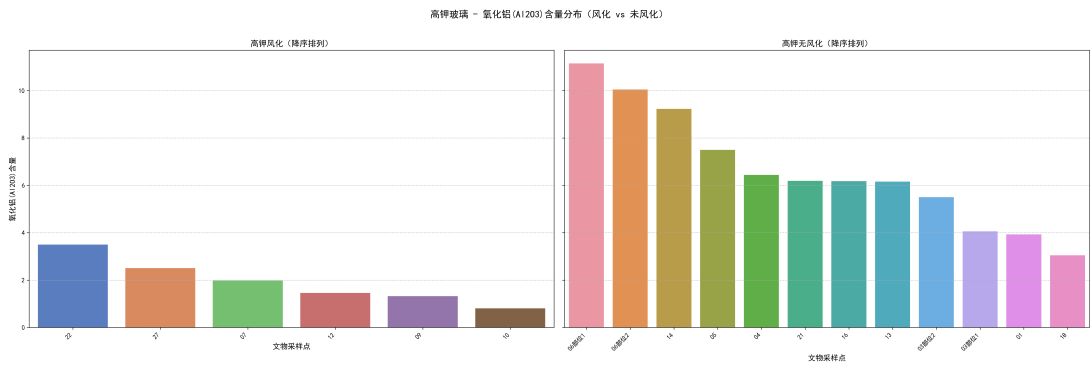


图 4 高钾玻璃 - 氧化铝 (Al₂O₃) 风化前后含量分布

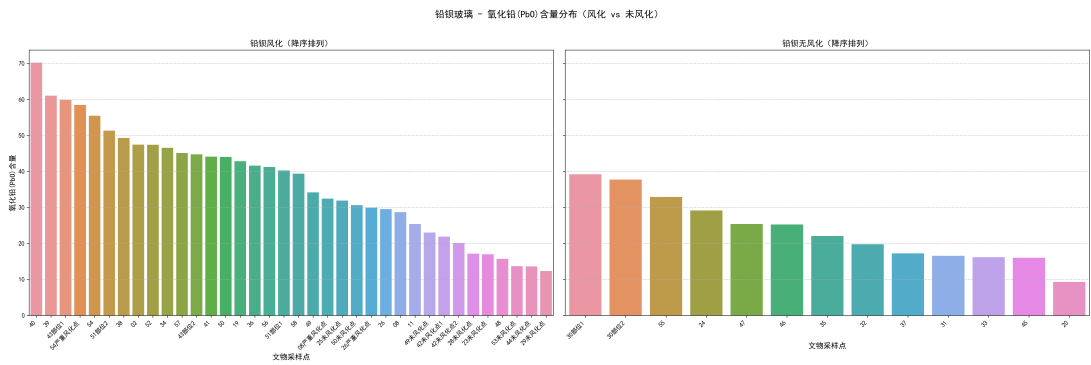


图 5 铅钡玻璃 - 氧化铅 (PbO) 风化前后含量分布

通过这些图表，可以直观地观察到：

- 铅钡玻璃在风化后，SiO₂ 含量略微增加，而 SrO、Fe₂O₃、CuO、Na₂O、MgO、CaO、K₂O、Al₂O₃、PbO、BaO、P₂O₅ 和 SO₂ 含量均有不同程度的下降；
- 高钾玻璃在风化后，SiO₂ 含量显著降低，而 SrO、Fe₂O₃、CuO、Na₂O、MgO、CaO、K₂O、Al₂O₃、PbO、BaO、P₂O₅ 和 SO₂ 含量明显增加，SnO₂ 含量基本不变；

- 风化过程对不同类型玻璃的化学成分影响存在明显差异。

5.3 风化前的化学成分含量的预测

5.3.1 数据分析

通过阅读数据，我们发现，除了编号为 49、50 的样本具有成对的 (风化前, 风化后) 的化学成分数据，其他的样本都是基于风化前后两群体的横截面数据，可以抽象为 (风化前, NULL) 或 (NULL, 风化后) 的形式。同时，考虑到需要预测的是化学成分，是总和为 1 的成分数据，各成分之间彼此依赖，不适宜使用一般的回归模型。

因此，我们基于成分数据分析 CoDA 模型，参考文献 [?]，在模型中添加先验正则项，最终建立风化前化学含量预测模型。

5.3.2 模型记号与数据映射

- 模型记号
 - \mathbf{a}_i (原始成分百分比)
 - \mathbf{p}_i (闭合后，和为 1 的比例)
 - \mathbf{z}_i (clr 后的向量)
 - $\bar{\mathbf{z}}_w, \bar{\mathbf{z}}_u$
 - δ_0 (按玻璃类型用文献导向值初始化)
 - \mathbf{w} (按成分设置权重)
 - λ (正则强度)
 - n (样本数)
- 令成分列为：

$$\mathcal{P} = \{\text{SiO}_2, \text{Na}_2\text{O}, \text{K}_2\text{O}, \dots, \text{SO}_2, \text{unknown}\}$$

共 $D = 15$ 个成分

- 将每行按百分比除以 100 进行闭合，得到组成比例矩阵 $\mathbf{p} \in \mathbb{R}^{n \times D}$ 。
 - 样本 i 的成分向量为 \mathbf{a}_i ，闭合后

$$\mathbf{p}_i = \mathcal{C}(\mathbf{a}_i) = \frac{\mathbf{a}_i}{\sum_{j=1}^D a_{ij}}, \quad \sum_{j=1}^D p_{ij} = 1.$$

- 为避免 $\log(0)$ ，在闭合前/或闭合后对零值做微小替换（伪计数） ε 。
- CLR 变换（把单纯形映到实向量空间）：对每行 i

$$\mathbf{z}_i = \text{clr}(\mathbf{p}_i) = \left(\ln \frac{p_{i1}}{g_i}, \dots, \ln \frac{p_{iD}}{g_i} \right), \quad g_i = \left(\prod_{j=1}^D p_{ij} \right)^{1/D},$$

5.3.3 风化模型

我们在 CLR 空间做建模，主要假设（能使问题可解且可解释）：

线性位移假设 对同类玻璃，风化后与风化前在 CLR 空间上满足近似的平移关系（均值差）：

$$\mathbf{z}_w \approx \mathbf{z}_{\text{pre}} + \boldsymbol{\delta} + \boldsymbol{\varepsilon},$$

其中 $\boldsymbol{\delta}$ 是同类玻璃的平均风化位移向量（在 CLR 空间）， $\boldsymbol{\varepsilon}$ 是噪声。

这样就可以得到基于群体差的回推策略而不需要成对样本。

- 设 $\bar{\mathbf{z}}_w$ 、 $\bar{\mathbf{z}}_u$ 分别为风化与未风化样本在 CLR 空间的均值，观测到的平均位移为：

$$\hat{\boldsymbol{\delta}}_{\text{obs}} = \bar{\mathbf{z}}_w - \bar{\mathbf{z}}_u.$$

- 对任一风化样点 i ，其风化前的 CLR 估计为：

$$\hat{\mathbf{z}}_{\text{pre},i}^{(\text{pure})} = \mathbf{z}_{w,i} - \hat{\boldsymbol{\delta}}_{\text{obs}}.$$

- 然后逆 CLR 得到比例并乘以 100 得到百分比：

$$\hat{\mathbf{p}}_{\text{pre},i} = \text{clr}^{-1}(\hat{\mathbf{z}}_{\text{pre},i}) = \frac{\exp(\hat{\mathbf{z}}_{\text{pre},i})}{\sum_{k=1}^D \exp(\hat{\mathbf{z}}_{\text{pre},i,k})}.$$

5.3.4 将文献先验融合进 CoDA 模型

为了把“机理知识”引入（例如高钾玻璃倾向于 K_2O 严重流失、 SiO_2 相对富集；铅钡体系可能出现 Pb/Ba 比例变化与硫酸盐富集等），我们在 CLR 空间对 $\boldsymbol{\delta}$ 加带方向性的岭惩罚：

- 设观测到的 $\mathbf{d} = \hat{\boldsymbol{\delta}}_{\text{obs}}$ 。我们引入先验中心向量 $\boldsymbol{\delta}_0$ （来源于机理/文献方向性），以及对每个成分的权重向量 $\mathbf{w} = (w_1, \dots, w_D)$ （表示你对该分量先验的信心和强度），并设正则强度为 $\lambda \geq 0$ 。最小化目标：

$$\min_{\boldsymbol{\delta}} \|\boldsymbol{\delta} - \mathbf{d}\|_2^2 + \lambda \|\mathbf{W}(\boldsymbol{\delta} - \boldsymbol{\delta}_0)\|_2^2,$$

其中 $\mathbf{W} = \text{diag}(\mathbf{w})$ 。这是二次型问题，有闭式解：

$$\hat{\boldsymbol{\delta}}_{\text{reg}} = (\mathbf{I} + \lambda \mathbf{W}^2)^{-1}(\mathbf{d} + \lambda \mathbf{W}^2 \boldsymbol{\delta}_0).$$

当 $\lambda = 0$ 或者 $\mathbf{w} = 0$ 时，退化为纯数据驱动（ $\hat{\boldsymbol{\delta}} = \mathbf{d}$ ）。当 λ 较大且某些 w_j 很大时， $\hat{\boldsymbol{\delta}}$ 会更靠近 $\boldsymbol{\delta}_0$ （先验主导）。

5.3.5 先验权重 λ 和噪声 ϵ 的调优

基于已知文物编号 49、50 风化前后的化学成分变化，我们将其作为测试集，通过网格搜索来寻找先验权重 λ 和噪声 ϵ 的最优值。我们使用以下公式检验模型性能：

$$\text{MAE}_{\text{prior_percent}} = \frac{\text{MAE}_{\text{model}}}{\text{MAE}_{\text{prior}}} \times 100\%$$

$$\text{其中 } \text{MAE}_{\text{model}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad \text{MAE}_{\text{prior}} = \frac{1}{n} \sum_{i=1}^n |y_i - p_i|$$

经过网格搜索，最终得到化学成分预测模型如下：

$$\hat{\mathbf{a}}_{i,\text{pre}}^{\text{prior}}(\%) = 100 \cdot \frac{\exp(\mathbf{z}_i - \hat{\boldsymbol{\delta}}_g^{\text{prior}})}{\mathbf{1}^\top \exp(\mathbf{z}_i - \hat{\boldsymbol{\delta}}_g^{\text{prior}})}, \quad (2)$$

$$\text{其中 } \hat{\boldsymbol{\delta}}_g^{\text{prior}} = (\mathbf{I} + \lambda \mathbf{W}_g^2)^{-1} [\mathbf{d}_g + \lambda \mathbf{W}_g^2 \boldsymbol{\delta}_{0,g}], \quad (3)$$

$$\text{s.t.} \begin{cases} \lambda \geq 0, \quad \mathbf{w}_g \in \mathbb{R}_{\geq 0}^D, \\ \boldsymbol{\delta}_{0,g} \in \mathcal{H}, \quad \mathbf{d}_g \in \mathcal{H}, \quad \mathbf{z}_i \in \mathcal{H}, \\ \mathcal{H} = \left\{ \mathbf{z} \in \mathbb{R}^D \mid \sum_{j=1}^D z_j = 0 \right\}, \\ \mathbf{p} \in \mathcal{S}^D, \quad \mathcal{S}^D = \left\{ \mathbf{p} \in \mathbb{R}_{>0}^D \mid \sum_{j=1}^D p_j = 1 \right\}, \\ \mathbf{W}_g = \text{diag}(w_{g1}, \dots, w_{gD}), \quad w_{gj} \geq 0, \quad \forall j, \\ i \in S_g^{(w)} \Rightarrow g \text{ 是样本 } i \text{ 的玻璃类型标签.} \end{cases} \quad (4)$$

参考链接: legest.ufpr.br, econ-papers.upf.edu

这句话引用了文献 [?]。

这句话引用了文献^[2]。

5.4 模型求解

Step1:

Step2:

Step3:

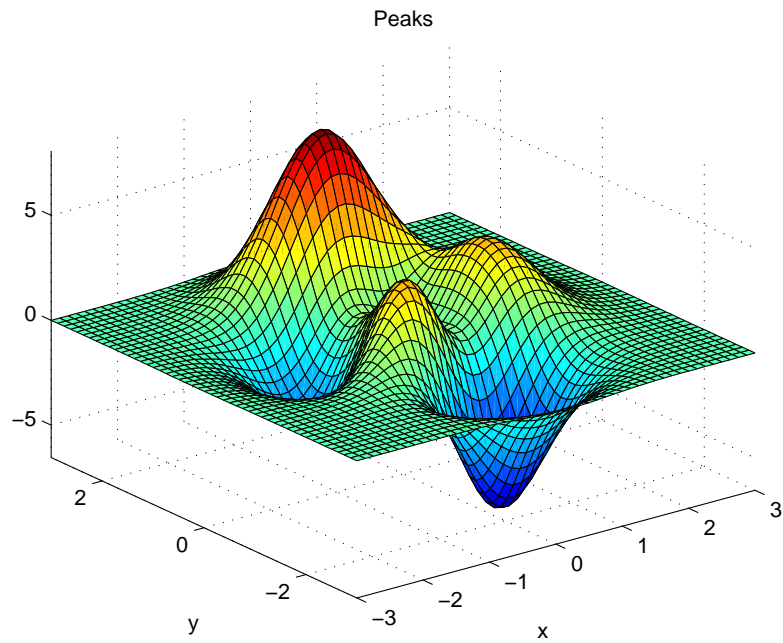


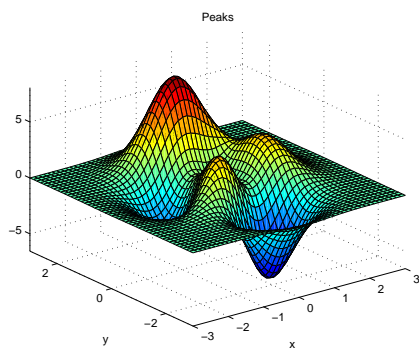
图 6 单图

5.5 求解结果

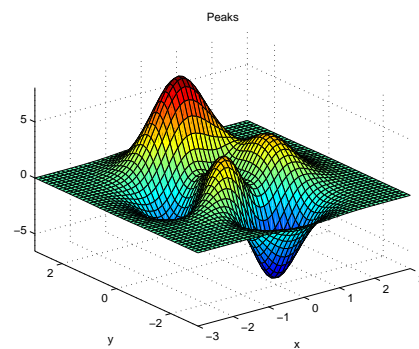
六、 问题二的模型的建立和求解

6.1 模型建立

引用??，引用??，引用??。



(a) 双图 a 子标题



(b) 双图 b 子标题

图 7 双图

6.2 模型求解

Step1:

Step2:

Step3:

6.3 求解结果

七、问题三的模型的建立和求解

7.1 模型建立

7.2 模型求解

Step1:

Step2:

Step3:

7.3 求解结果

八、问题四的模型的建立和求解

8.1 模型建立

8.2 模型求解

Step1:

Step2:

Step3:

8.3 求解结果

九、模型的分析与检验

9.1 灵敏度分析

9.2 误差分析

十、模型的评价

10.1 模型的优点

- 优点 1
- 优点 2
- 优点 3

10.2 模型的缺点

- 缺点 1
- 缺点 2

附录 A 文件列表

文件名	功能描述
q1.m	问题一程序代码
q2.py	问题二程序代码
q3.c	问题三程序代码
q4.cpp	问题四程序代码

附录 B 代码

q1.m

```
1 disp("Hello World!")
```

q2.py

```
1 print("Hello World!")
```

q3.c

```
1 #include <stdio.h>
2
3 int main()
4 {
5     printf("Hello World!");
6     return 0;
7 }
```

q4.cpp

```
1 #include <bits/stdc++.h>
2 using namespace std;
3
4 int main()
5 {
6     cout << "Hello World!" << endl;
7     return 0;
8 }
```