

# 中国科学技术大学

# 本科毕业论文



## 加倍空间下的核心集

作者姓名：	沈俊杰
学    号：	PB19000182
专    业：	计算机科学与技术
导师姓名：	丁虎 教授    杨青 教授
完成时间：	2024 年 5 月 16 日



## 摘 要

摘要分中文和英文两种，中文在前，英文在后，博士论文中文摘要一般 800～1500 个汉字，硕士论文中文摘要一般 500～1000 个汉字。英文摘要的篇幅参照中文摘要。

关键词另起一行并隔行排列于摘要下方，左顶格，中文关键词间空一字或用分号“；”隔开，英文关键词之间用逗号“,”或分号“;”隔开。

中文摘要是论文内容的总结概括，应简要说明论文的研究目的、基本研究内容、研究方法或过程、结果和结论，突出论文的创新之处。摘要应具有独立性和自明性，即不用阅读全文，就能获得论文必要的信息。摘要中不宜使用公式、图表，不引用文献。

中文关键词是为了文献标引工作从论文中选取出来用以表示全文主题内容信息的单词和术语，一般 3～8 个词，要求能够准确概括论文的核心内容。

**关键词：**学位论文；摘要；关键词

## ABSTRACT

This is a sample document of USTC thesis  $\text{\LaTeX}$  template for bachelor, master and doctor. The template is created by zepinglee and seisman, which originate from the template created by ywg. The template meets the requirements of USTC thesis writing standards.

This document will show the usage of basic commands provided by  $\text{\LaTeX}$  and some features provided by the template. For more information, please refer to the template document ustcthesis.pdf.

**Key Words:** dissertation; abstract; keywords

## 目 录

第一章 绪论 .....	3
第一节 脚注 .....	3
第二章 浮动体 .....	4
第一节 三线表 .....	4
第二节 插图 .....	4
第三节 算法环境 .....	5
第三章 数学 .....	6
第一节 数学符号 .....	6
第二节 数学公式 .....	7
第三节 量和单位 .....	7
第四节 定理和证明 .....	7
第四章 引用文献的标注 .....	10
第一节 顺序编码制 .....	10
一、角标数字标注法 .....	10
二、数字标注法 .....	10
第二节 著者-出版年制标注法 .....	10
参考文献 .....	11
附录 A 补充材料 .....	14
第一节 补充章节 .....	14
致谢 .....	15
在读期间发表的学术论文与取得的研究成果 .....	16

## 符 号 说 明

$a$	The number of angels per unit area
$N$	The number of angels per needle point
$A$	The area of the needle point
$\sigma$	The total mass of angels per unit area
$m$	The mass of one angel
$\sum_{i=1}^n a_i$	The sum of $a_i$

## 第一章 绪论

在大数据时代，随着数据量的迅速增加，我们往往需要更低的时间复杂度算法来完成任务。此外，在一些实际场景中，我们的数据是动态的，即数据集合会随着时间的推移而发生变化。显然，如果数据量极其庞大，当训练数据动态变化时，重新训练整个模型的代价是不可行的。

为了解决这些问题，一个很自然的想法是构建一个小规模的训练集，以便我们可以在小规模集合而不是所有数据上运行现有的算法。Coreset，最初出现于计算几何研究中的概念，已成为许多大规模机器学习问题中广泛使用的数据压缩技术。作为一种简洁的数据压缩技术，Coreset 还具有一些优秀的性质。例如，Coreset 通常是可组合的，因此可以应用于分布式计算环境之中。此外，它往往能在流算法和包含数据增删的动态算法中取得较好的表现。

然而，现有的 Coreset 算法仍然有较大的改进空间。一个主要的瓶颈是，多数 Coreset 算法容易受到异常点（离群点）的影响而导致性能下降，而现实数据集中往往有许多噪声。许多现有研究也表明，只需在训练样本中加入少数攻击点，就可能导致原模型的效果显著下降。因此，如何设计一个鲁棒性更好的 Coreset 算法是一个重要的研究方向。

为了简要说明现有的 coreset 方法对离群值的敏感性，我们可以以一种基于采样的 coreset 算法为例。这个算法需要对每个数据点计算一个“敏感度”

本模板 `ustcthesis` 是中国科学技术大学本科生和研究生学位论文的  $\text{\LaTeX}$  模板，按照《中国科学技术大学研究生学位论文撰写手册》（最近在修订中，以下简称《撰写手册》）和《中国科学技术大学本科毕业论文（设计）格式》的要求编写。

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

## 第一节 脚注

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. <sup>①</sup>

---

<sup>①</sup>Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur.



## 第二章 浮动体

### 第一节 三线表

三线表是《撰写手册》推荐使用的格式，如表 2.1。

表 2.1 表号和表题在表的正上方

类型	描述
挂线表	挂线表也称系统表、组织表，用于表现系统结构
无线表	无线表一般用于设备配置单、技术参数列表等
卡线表	卡线表有完全表，不完全表和三线表三种

如果有表注，推荐使用 `threeparttable`。这样可以与表格对齐，满足部分评审老师的要求。

表 2.2 带表注的表格

类型	描述
挂线表	挂线表也称系统表、组织表，用于表现系统结构
无线表	无线表一般用于设备配置单、技术参数列表等
卡线表	卡线表有完全表，不完全表和三线表三种

注：表注分两种，第一种是对全表的注释，用不加阿拉伯数字排在表的下边，前面加“注：”；第二种是和表内的某处文字或数字相呼应的注，在表里面用带圈的阿拉伯数字在右上角标出，然后在表下面用同样的圈码注出来

编制表格应简单明了，表达一致，明晰易懂，表文呼应、内容一致。排版时表格字号略小，或变换字体，尽量不分页，尽量不跨节。表格太大需要转页时，需要在续表上方注明“续表”，表头页应重复排出。

### 第二节 插图

有的同学可能听说“ $\text{\LaTeX}$  只能使用 `eps` 格式的图片”，甚至把 `jpg` 格式转为 `eps`。事实上，这种做法已经过时。而且每次编译时都要调用外部工具解析 `eps`，导致降低编译速度。所以我们推荐矢量图直接使用 `pdf` 格式，位图使用 `jpeg` 或 `png` 格式。

关于图片的并排，推荐使用较新的 `subcaption` 宏包，不建议使用 `subfigure` 或 `subfig` 等宏包。



图 2.1 图号、图题置于图的下方

注：图注的内容不宜放到图题中。

### 第三节 算法环境

模板中使用 `algorithm2e` 宏包实现算法环境。关于该宏包的具体用法，请阅读宏包的官方文档。

---

#### 算法 2.1 算法示例 1

---

```
Data: this text
Result: how to write algorithm with LATEX2ε
1 initialization;
2 while not at end of this document do
3   read current;
4   if understand then
5     go to next section;
6     current section becomes this one;
7   else
8     go back to the beginning of current section;
9   end
10 end
```

---

注意，我们可以在论文中插入算法，但是插入大段的代码是愚蠢的。然而这并不妨碍有的同学选择这么做，对于这些同学，建议用 `listings` 宏包。

## 第三章 数学

### 第一节 数学符号

《撰写手册》要求数学符号遵循 GB/T 3102.11—1993《物理科学和技术中使用的数学符号》<sup>①</sup>。该标准参照采纳 ISO 31-11:1992<sup>②</sup>，但是与 T<sub>E</sub>X 默认的美国数学学会 (AMS) 的符号习惯有所区别。具体地来说主要有以下差异：

1. 大写希腊字母默认为斜体，如

$$\Gamma \Delta \Theta \Lambda \Xi \Pi \Sigma \Upsilon \Phi \Psi \Omega.$$

注意有限增量符号  $\Delta$  固定使用正体，模板提供了 `\increment` 命令。

2. 小于等于号和大于等于号使用倾斜的字形  $\leq$ 、 $\geq$ 。
3. 积分号使用正体，比如  $\int$ 、 $\oint$ 。
4. 偏微分符号  $\partial$  使用正体。
5. 省略号 `\dots` 按照中文的习惯固定居中，比如

$$1, 2, \dots, n \quad 1 + 2 + \dots + n.$$

6. 实部  $\operatorname{Re}$  和虚部  $\operatorname{Im}$  的字体使用罗马体。

以上数学符号样式的差异可以在模板中统一设置。但是还有一些需要用户在写作时进行处理：

1. 数学常数和特殊函数名用正体，如

$$\pi = 3.14 \dots; \quad i^2 = -1; \quad e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n.$$

2. 微分号使用正体，比如  $\mathrm{d}y/\mathrm{d}x$ 。
3. 向量、矩阵和张量用粗斜体 (`\symbf`)，如  $\mathbf{x}$ 、 $\mathbf{\Sigma}$ 、 $\mathbf{T}$ 。
4. 自然对数用  $\ln x$  不用  $\log x$ 。

模板中使用 `unicode-math` 宏包配置数学字体。该宏包与传统的 `amsfonts`、`amssymb`、`bm`、`mathrsfs`、`upgreek` 等宏包不兼容。本模板作了处理，用户可以直接使用 `\bm`、`\mathscr`、`\upGamma` 等命令。关于数学符号更多的用法，参见 `unicode-math` 宏包的使用说明和符号列表 `unimath-symbols`。

<sup>①</sup>原 GB 3102.11—1993，自 2017 年 3 月 23 日起，该标准转为推荐性标准。

<sup>②</sup>目前已更新为 ISO 80000-2:2019。

## 第二节 数学公式

数学公式可以使用 `equation` 和 `equation*` 环境。注意数学公式的引用应前后带括号，建议使用 `\eqref` 命令，比如式 (3.1)。

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi i x \xi} dx. \quad (3.1)$$

多行公式尽可能在“=”处对齐，推荐使用 `align` 环境，比如式 (3.3)。

$$a = b + c + d + e \quad (3.2)$$

$$= f + g. \quad (3.3)$$

## 第三节 量和单位

量和单位要求严格执行 GB 3100~3102—1993 有关量和单位的规定。宏包 `siunitx` 提供了更好的数字和单位支持：

- 为了阅读方便，四位以上的整数或小数推荐采用千分空的分节方式：55 235 367.346 23。四位以内的整数可以不加千分空：1256。
- 数值与单位符号间留适当空隙：25.4 mm， $5.97 \times 10^{24}$  kg， $-273.15^\circ\text{C}$ 。例外： $12.3^\circ$ ， $1^\circ 2' 3''$ 。
- 组合单位默认使用 APS 的格式，即相乘的单位之间留一定空隙： $\text{kg m s}^{-2}$ ，也可以使用居中的圆点： $\text{kg} \cdot \text{m} \cdot \text{s}^{-2}$ 。GB 3100—1993 对两者都允许，建议全文统一设置。
- 量值范围使用“~”： $10 \text{ mol/L} \sim 15 \text{ mol/L}$ 。
- 注意：词头  $\mu$  不能写为 u，如：umol 应为  $\mu\text{mol}$ 、 $\mu\text{mol}$ 。

## 第四节 定理和证明

示例文件中使用 `amsthm` 宏包配置了定理、引理和证明等环境。用户也可以使用 `ntheorem` 宏包。

**定义 3.1** If the integral of function  $f$  is measurable and non-negative, we define its (extended) **Lebesgue integral** by

$$\int f = \sup_g \int g, \quad (3.4)$$

where the supremum is taken over all measurable functions  $g$  such that  $0 \leq g \leq f$ , and where  $g$  is bounded and supported on a set of finite measure.

**假设 3.1** The communication graph is strongly connected.

**例 3.1** Simple examples of functions on  $\mathbb{R}^d$  that are integrable (or non-integrable) are given by

$$f_a(x) = \begin{cases} |x|^{-a} & \text{if } |x| \leq 1, \\ 0 & \text{if } |x| > 1. \end{cases} \quad (3.5)$$

$$F_a(x) = \frac{1}{1 + |x|^a}, \quad \text{all } x \in \mathbb{R}^d. \quad (3.6)$$

Then  $f_a$  is integrable exactly when  $a < d$ , while  $F_a$  is integrable exactly when  $a > d$ .

**引理 3.1** (Fatou) Suppose  $\{f_n\}$  is a sequence of measurable functions with  $f_n \geq 0$ . If  $\lim_{n \rightarrow \infty} f_n(x) = f(x)$  for a.e.  $x$ , then

$$\int f \leq \liminf_{n \rightarrow \infty} \int f_n. \quad (3.7)$$

**注** We do not exclude the cases  $\int f = \infty$ , or  $\liminf_{n \rightarrow \infty} \int f_n = \infty$ .

**推论 3.2** Suppose  $f$  is a non-negative measurable function, and  $\{f_n\}$  a sequence of non-negative measurable functions with  $f_n(x) \leq f(x)$  and  $f_n(x) \rightarrow f(x)$  for almost every  $x$ . Then

$$\lim_{n \rightarrow \infty} \int f_n = \int f. \quad (3.8)$$

**命题 3.3** Suppose  $f$  is integrable on  $\mathbb{R}^d$ . Then for every  $\epsilon > 0$ :

i. There exists a set of finite measure  $B$  (a ball, for example) such that

$$\int_{B^c} |f| < \epsilon. \quad (3.9)$$

ii. There is a  $\delta > 0$  such that

$$\int_E |f| < \epsilon \quad \text{whenever } m(E) < \delta. \quad (3.10)$$

**定理 3.4** Suppose  $\{f_n\}$  is a sequence of measurable functions such that  $f_n(x) \rightarrow f(x)$  a.e.  $x$ , as  $n$  tends to infinity. If  $|f_n(x)| \leq g(x)$ , where  $g$  is integrable, then

$$\int |f_n - f| \rightarrow 0 \quad \text{as } n \rightarrow \infty, \quad (3.11)$$

and consequently

$$\int f_n \rightarrow \int f \quad \text{as } n \rightarrow \infty. \quad (3.12)$$

**证明** Trivial. ■

**Axiom of choice** Suppose  $E$  is a set and  $E_\alpha$  is a collection of non-empty subsets of  $E$ . Then there is a function  $\alpha \mapsto x_\alpha$  (a “choice function”) such that

$$x_\alpha \in E_\alpha, \quad \text{for all } \alpha. \quad (3.13)$$

**Observation 1** Suppose a partially ordered set  $P$  has the property that every chain has an upper bound in  $P$ . Then the set  $P$  contains at least one maximal element.

**A concise proof** Obvious. ■

## 第四章 引用文献的标注

模板使用 **natbib** 宏包来设置参考文献引用的格式，更多引用方法可以参考该宏包的使用说明。

### 第一节 顺序编码制

#### 一、角标数字标注法

<code>\cite{knuth86a}</code>	⇒	[1]
<code>\citet{knuth86a}</code>	⇒	Knuth <sup>[1]</sup>
<code>\cite[42]{knuth86a}</code>	⇒	[1] <sup>42</sup>
<code>\cite{knuth86a,tlc2}</code>	⇒	[1-2]
<code>\cite{knuth86a, knuth84}</code>	⇒	[1,3]

#### 二、数字标注法

<code>\cite{knuth86a}</code>	⇒	[1]
<code>\citet{knuth86a}</code>	⇒	Knuth [1]
<code>\cite[42]{knuth86a}</code>	⇒	[1] <sup>42</sup>
<code>\cite{knuth86a,tlc2}</code>	⇒	[1-2]
<code>\cite{knuth86a, knuth84}</code>	⇒	[1, 3]

### 第二节 著者-出版年制标注法

<code>\cite{knuth86a}</code>	⇒	Knuth (1986)
<code>\citep{knuth86a}</code>	⇒	(Knuth, 1986)
<code>\citet[42]{knuth86a}</code>	⇒	Knuth (1986) <sup>42</sup>
<code>\citep[42]{knuth86a}</code>	⇒	(Knuth, 1986) <sup>42</sup>
<code>\cite{knuth86a,tlc2}</code>	⇒	Knuth (1986); Mittelbach et al. (2004)
<code>\cite{knuth86a, knuth84}</code>	⇒	Knuth (1986, 1984)

## 参 考 文 献

- [1] KNUTH D E. Computers and typesetting: A the  $\text{\TeX}$ book[M]. Reading, MA, USA: Addison-Wesley, 1986.
- [2] MITTELBACH F, GOOSSENS M, BRAAMS J, et al. The  $\text{\LaTeX}$  companion[M]. 2nd ed. Reading, MA, USA: Addison-Wesley, 2004.
- [3] KNUTH D E. Literate programming[J]. The Computer Journal, 1984, 27(2): 97-111.
- [4] LAMPORT L.  $\text{\LaTeX}$ : a document preparation system[M]. 2nd ed. Reading, MA, USA: Addison-Wesley, 1994.
- [5] 孙立广. 极地科学前沿与热点: 顶级期刊论文摘要汇编(1999—2010) [M]. 合肥: 中国科学技术大学出版社, 2016: 222.
- [6] 李永池. 张量初步和近代连续介质力学概论[M]. 2 版. 合肥: 中国科学技术大学出版社, 2016: 61.
- [7] 刘景双. 湿地生态系统碳、氮、硫、磷生物地球化学过程[M]. 合肥: 中国科学技术大学出版社, 2014.
- [8] CRAWFORD W, GORMAN M. Future libraries: Dreams, madness, & reality [M]. Chicago: American Library Association, 1995.
- [9] International Federation of Library Association and Institutions. Names of persons: National usage for entry in catalogues[M]. 3rd ed. London: IFLA International Office for UBC, 1977.
- [10] 程根伟. 1998 年长江洪水的成因与减灾对策[M]//许厚泽, 赵其国. 长江流域洪涝灾害与科技对策. 北京: 科学出版社, 1999: 26-32.
- [11] 陈晋镛, 张惠民, 朱士兴, 等. 蓟县震旦亚界研究[M]//中国地质科学院天津地质矿产研究所. 中国震旦亚界. 天津: 天津科学技术出版社, 1980: 56-114.
- [12] BUSECK P R, NORD G L, Jr., VEBLEN D R. Subsolidus phenomena in pyroxenes[M]//PREWITT C T. Pyroxenes. Washington, D.C.: Mineralogical Society of America, 1980: 117-212.
- [13] FOURNEY M E. Advances in holographic photoelasticity[C]//American Society of Mechanical Engineers. Applied Mechanics Division. Symposium on Applications of Holography in Mechanics, August 23-25, 1971, University of Southern



- California, Los Angeles, California. New York: ASME, 1971: 17-38.
- [14] 孔庆勇, 郭红健, 孔庆和. 我国科技期刊的金字塔分层模型及发展路径初探[J]. 中国科技期刊研究, 2015, 26(10): 1100-1103.
- [15] 杨洪升. 四库馆私家抄校书考略[J]. 文献, 2013(1): 56-75.
- [16] 于潇, 刘义, 柴跃廷, 等. 互联网药品可信交易环境中主体资质审核备案模式[J]. 清华大学学报(自然科学版), 2012, 52(11): 1518-1521.
- [17] DES MARAIS D J, STRAUSS H, SUMMONS R E, et al. Carbon isotope evidence for the stepwise oxidation of the proterozoic environment[J]. Nature, 1992, 359: 605-609.
- [18] HEWITT J A. Technical services in 1983[J]. Library Resource Services, 1984.
- [19] 丁文详. 数字革命与竞争国际化[N]. 中国青年报, 2000-11-20(15).
- [20] 姜锡洲. 一种温热外敷药制备方案: 中国, 88105607.3[P]. 1989-07-26.
- [21] 万锦坤. 中国大学学报论文文摘(1983-1993)(英文版)[DB/CD]. 北京: 中国大百科全书出版社, 1996.
- [22] MLOT C. Plant physiology: Plant biology in the Genome Era[J]. Science, 1998, 281: 331-332.
- [23] 孙玉文. 汉语变调构词研究[D]. 北京: 北京大学, 2000.
- [24] CAIRNS B R. Infrared spectroscopic studies of solid oxygen[D]. Berkeley: Univ. of California, 1965.
- [25] 中国力学学会. 第3届全国实验流体力学学术会议论文集[C]. 天津, 1990.
- [26] ROSENTHALL E M. Proceedings of the Fifth Canadian Mathematical Congress, University of Montreal, 1961[C]. Toronto: University of Toronto Press, 1963.
- [27] BAKER S K, JACKSON M E. The future of resource sharing[M]. New York: The Haworth Press, 1995.
- [28] 尼葛洛庞帝. 数字化生存[M]. 胡泳, 范海燕, 译. 海口: 海南出版社, 1996.
- [29] 杨宗英. 电子图书馆的现实模型[J]. 中国图书馆学报, 1996(2): 24-29.
- [30] 刘斌. 力学[M]. 合肥, 2014: 24-29.
- [31] 刘文富, 顾丽梅. 网络时代经济发展战略特征[J/OL]. 学术研究, 2000, 21(4): 35-40. <https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C44YLTlOAI TRKgchrJ08w1e79zTD32bjb4xSTlZqRyr7rTlf7ok1AFpatLUPx8UT1OWFBnk w65sK7Uwj b q66SQOt&uniplatform=NZKPT>.
- [32] 肖渡, 沈群红, 张芸, 等. 知识时代的企业合作经营[M/OL]. 北京: 北京大学出

- 版社, 2000: 67-69. <https://book.douban.com/subject/1632549/>.
- [33] The White House. Technology for economic growth[R/OL]. Washington, 1993. <https://clintonwhitehouse6.archives.gov/1993/11/1993-11-04-technology-for-economic-growth-table-of-contents.html>.
- [34] HUTSON J M. Vibrational dependence of the anisotropic intermolecular potential of argon-hydrogen chloride[J/OL]. J Phys Chem, 1992, 96(11): 4237-4247[2023-05-31]. <https://doi.org/10.1021/j100190a026>.

## 附录 A 补充材料

### 第一节 补充章节

补充内容。

## 致 谢

在研究学习期间，我有幸得到了三位老师的教导，他们是：我的导师，中国科大 XXX 研究员，中科院 X 昆明动物所马老师以及美国犹他大学的 XXX 老师。三位深厚的学术功底，严谨的工作态度和敏锐的科学洞察力使我受益良多。衷心感谢他们多年来给予我的悉心教导和热情帮助。

感谢 XXX 老师在实验方面的指导以及教授的帮助。科大的 XXX 同学和 XXX 同学参与了部分试验工作，在此深表谢意。

2024 年 5 月

## 在读期间发表的学术论文与取得的研究成果

### 已发表论文

1. A A A A A A A A A
2. A A A A A A A A A
3. A A A A A A A A A

### 待发表论文

1. A A A A A A A A A
2. A A A A A A A A A
3. A A A A A A A A A

### 研究报告

1. A A A A A A A A A
2. A A A A A A A A A
3. A A A A A A A A A