arXiv:2303.04021v2 [math.OC] 5 Aug 2024

服务率区域多胞体

G. N. Alfarano*[1]、A. B. Kılıc†[2]、A. Ravagnani‡[3]和 E. Soljanin[4]

[1,2,3]荷兰埃因霍温理工大学[4]美国新泽西州罗格斯大学

摘要

我们研究了一类多胞体的性质，这类多胞体自然地与分布式数据存储中的一个问题相关，即服务率区域多胞体。分布式编码系统的服务速率区域描述了底层系统能够支持的数据访问请求。本文研究服务速率区域的多面体结构，主要旨在描述其几何形状和性质。为此，我们引入了服务速率区域的若干结构参数，并为这些参数建立上界和下界。本文采用的技术涵盖编码理论和优化。我们的一个主要结果表明，服务速率区域中的每个有理点都有一个所谓的有理分配，从而回答了该研究领域中的一个开放问题。

引言

分布式存储系统将数据拆分并分布到多个服务器上，以便为多个可能并发的用户提供访问服务。可靠处理并发请求的最简单方法是根据数据的受欢迎程度复制数据；例如参见[23, 28]。遗憾的是，这种方法在存储方面可能代价高昂。此外，预测人们对数据的兴趣如何变化并非总是易事。因此，纠删码作为一种冗余存储形式受到关注；例如参见 [10]及其中的参考文献。

近期研究确立了服务速率区域这一概念，将其作为衡量分布式编码系统效率的重要指标，并指出应在系统设计阶段予以考虑；参见 [1, 2, 4, 16–18]。为理解这一指标，考虑这样的分布式系统：其中存储着 k 个数据对象，并使用线性 $[n, k]_q$ 纠错码将数据编码后分布在 n 台服务器上，每台服务器的容量均为 $\mu \in \mathbb{R}$。该分布式编码系统的服务速率区域是系统能够同时处理的所有请求率（$\lambda_1, \ldots, \lambda_k$）$\in \mathbb{R}^k$ 的集合。这样的分布式系统由一个满秩 $k \times n$ 矩阵 G定义，其服务速率区域是 $\mathbb{R}^k$中的一个凸多胞体。

该领域以往的研究主要致力于刻画分布式编码系统的服务速率区域，并寻找将速率请求分配到各服务器上的最优策略，以最大化该区域；例如参见 [2]。目前，二元单纯形码以及两类最大距离可分（MDS）码的服务速率区域已得到刻画：1）当 $n \geq 2k$ 时的系统 MDS 码；2）长度和维数任意，且无法从少于 k 个存储对象中解码出任何数据对象的 MDS 码；另请参见 [2] 及其中的参考文献。

通过确立服务率问题与著名的分数

*G. N. A. 获得瑞士国家基金会通过编号为 210966 的资助。

†A. B. K. 获得荷兰研究理事会通过编号为 VI.Vidi.203.045 的资助。

‡A. R. 获得荷兰研究理事会通过编号为 VI.Vidi.203.045 和 OCENW.KLEIN.539 的资助，并获得荷兰皇家艺术与科学院的资助。

超图上的匹配问题之间的等价关系，人们提出了一种研究服务速率区域的组合方法。在同一项工作中，作者还精确地指出，服务率问题是以下问题的推广：批处理、*PIR*和交换编码；有关这些码类的详细信息和背景资料，请参见[11, 12, 15, 25, 29, 30]。在 [17], 中，人们利用几何方法确定了二元一阶Reed-Muller码和二元单纯形码的服务速率区域。在 [3], 中，人们利用编码理论工具找到了一个包含服务速率区域的多胞体，从而为其给出了一个外界。

贡献。与以往方法不同，本文着重描述服务速率区域的多胞体结构及其几何性质，例如其体积。为实现这一目标，我们采用的关键工具是服务速率区域多胞体的外界，这些外界是通过运用从编码理论到凸几何等领域的方法得到的。

我们还提出了服务率区域的一种离散化概念，它自然源自整数分配模型。我们证明，这种离散化概念恰好给出了最初提出的（连续）概念中的有理点。在研究分配与服务速率区域多胞体之间的联系时，我们表明，该区域中的每个有理点都对应一个有理分配。本文最后一部分聚焦于系统 MDS 矩阵的服务速率区域，并计算了其在维数为 2 和 3 时的体积等量。

本文其余部分安排如下：第 1 节介绍访问模型并提出服务率问题。第 2 节探讨服务率区域多胞体的各种表示。第 3节将服务率区域的概念离散化，并证明该区域中的有理点具有有理分配。第 4节介绍并研究r阶最大和容量及系统体积。第 5 节致力于证明服务率区域的不同外界。第 6 节聚焦于系统 MDS 码。附录中提供了关于多胞体和纠错码的基础知识。

1 系统模型与问题表述

我们考虑一个具有 n 个相同节点（服务器）的分布式服务系统。每个节点包含两个功能组件：一个用于数据存储，另一个用于处理系统用户提出的下载请求。本节将确立记号，定义分布式服务系统及其服务速率区域，并陈述本文所要解决的问题。

全文中，q 表示一个素数幂，$\mathbb{F}_q$ 表示含有 q 个元素的有限域。我们使用整数 $n > k \geq 2$。除非另有说明，以下所有向量均为行向量。

1.1 存储模型

我们考虑一个编码的分布式数据存储系统，其中 k 个对象（$\mathbb{F}_q$ 的元素）经过线性编码后存储在 n 台服务器上。每台服务器恰好存储 $\mathbb{F}_q$ 的一个元素。因此，该编码系统由一个秩为 k 的矩阵 $G \in \mathbb{F}_q^{k\times n}$ 指定，我们称之为系统的生成矩阵。若 $(x_1,\dots,x_k) \in \mathbb{F}_q^k$ 是待存储的 k 元对象组，则第 j 台服务器存储编码向量的第 j 个分量

$$(x_1, \dots, x_k) \cdot G \in \mathbb{F}_q^n.$$

请注意，根据定义， n 服务器存储的是数据对象的线性组合，而不仅仅是它们的副本。后一种存储策略称为（简单）复制。

按照编码理论的术语 [21], ，如果矩阵 G 的前 k 列构成单位 $k \times k$ 矩阵，我们就称其为系统的。如果 G 的第 ν列是标准基向量 e_i的非零倍数，那么我们就称 ν 为第 i个数据对象的系统节点。如果每个 $\nu \in \{1, \dots, n\}$ 都是一个系统节点，那么我们就称 G 为复制矩阵。请注意，复制矩阵描述的是这样一个系统：每个对象都以其原本的形式存储（允许乘以非零倍数）。

记号1.1。为简化全文中的表述，不失一般性，我们处理一个秩为 k的固定矩阵 $G \in \mathbb{F}_q^{k\times n}$。我们用 G^ν 表示 G 的第 ν列，并假设其所有列均不是零向量。

由于 $n > k$，我们或许能够从不同的服务器集合中恢复每个对象，这促使我们作出如下定义并引入相关术语。

定义 1.2。设 $R \subseteq \{1,\dots,n\}$，使得 $e_i \in \langle G^\nu \mid \nu \in R\rangle$，且 $\langle G^\nu \mid \nu \in R\rangle$ 是由 R 索引的 G 的列的 $\mathbb{F}_q$-张成空间。那么，R 称为第 i 个对象的一个恢复集。对于 $i \in \{1,\dots,k\}$，令

$$\mathcal{R}_i^{\text{all}}(G) := \{R \subseteq \{1,\dots,n\} \mid e_i \in \langle G^\nu \mid \nu \in R\rangle\},$$

其中 sup 上标“all”表示 $\mathcal{R}_i^{\text{all}}(G)$ 包含第 i 个对象的所有恢复集 t.

由于根据假设，G 的秩为 k，因此对于所有 $i \in \{1,\dots,k\}$，我们有 $\mathcal{R}_i^{\text{all}}(G) \neq \emptyset$。此外，对于所有 $i \in \{1,\dots,k\}$ 和 $R \in \mathcal{R}_i^{\text{all}}(G)$，都有 $R \neq \emptyset$。下面我们继续形式化恢复系统的概念。

定义 1.3。一个（恢复）G-系统是一个 k 元组 $\mathcal{R} = (\mathcal{R}_1,\dots,\mathcal{R}_k)$，其中 $\mathcal{R}_i \subseteq \mathcal{R}_i^{\text{all}}(G)$，并且对所有 $i \in \{1,\dots,k\}$均有$\mathcal{R}_i \neq \emptyset$。

1.2 服务与访问模型

我们采用两种服务模型，并将其称为排队模型和带宽模型。它们基于通过传入的数据访问请求实现资源共享的两种常见方式。在排队模型中，从某个节点下载的请求会被放入该节点的队列中。每个节点平均每单位时间可以服务 μ 个请求。为了保持每个队列的稳定性，每个节点的请求到达率总和不得超过其服务率 μ。

在带宽模型中，每个节点可以同时服务多个数据访问请求。当每个节点拥有一条具有有限访问带宽 W 比特/秒的 I/O 总线，且一次下载需要以固定带宽 b 比特/秒进行流媒体传输时，一个节点可以同时服务最多$\mu = \lfloor W/b \rfloor$ 个请求。在这两种情况下，我们将 μ 称为服务器容量（形式化定义将在后文给出）。在排队模型中，下载对象 i 的请求以 $\lambda_i \in \mathbb{R}_{\geq 0}$ 的速率到达。在带宽模型中，λ_i 是系统中同时存在的对象 i 请求数。在这两种模型中，$\lambda_{i,R}$ 是分配给恢复集 $R \in \mathcal{R}_i$服务的 λ_i 部分。我们将集合 $\{\lambda_{i,R} \mid R \in \mathcal{R}_i, i = 1,\dots,k\}$ 称为请求分配。

1.3 归一化与整数性

我们可以将所有请求分配值和速率除以节点服务容量 μ，从而对它们进行归一化。此时，所有归一化请求速率 λ_i 以及 $\{\lambda_{i,R} \mid R \in \mathcal{R}_i, i \in \{1,\dots,k\}\}$ 中的数值都是 $1/\mu$ 的倍数。由于在带宽模型中，这些数值表示请求数量，因此归一化后的数值是 $1/\mu$ 的整数倍。此外，在一些实际场景中，每个已服务请求都会占用其所访问服务器的全部带宽（例如，通过低带宽边缘设备进行流媒体传输）。在这种情况下，λ_i 是整数，而 $\lambda_{i,\,R}$ 是二进制数。在其他实际场景中，用户可以同时从多个节点下载数据，并从每个节点获取其部分带宽；因此，可以放宽 $\lambda_{i,\,R}$ 必须是 $1/\mu$ 的整数倍这一假设。

1.4 服务率区域与问题表述

我们关注如何刻画数据存储系统能够支持的速率请求 k-元组 $(\lambda_1,\dots,\lambda_k) \in \mathbb{R}^k$。这类元组的集合正式定义如下，由此引出分布式存储系统的服务速率区域这一概念。

定义 1.4. 设 $\mathcal{R}=(\mathcal{R}_1,\dots,\mathcal{R}_k)$ 是一个 G-系统。与 $\mathcal{R}$ 和 μ 相关联的服务速率区域，是满足以下条件的所有 $(\lambda_1,\dots,\lambda_k)\in\mathbb{R}^k$ 的集合：存在一组实数

$$\{\lambda_{i,R}\mid i\in\{1,\dots,k\},\ R\in\mathcal{R}_i\}$$

具有以下性质：

$$\begin{cases}\displaystyle\sum_{R\in\mathcal{R}_i}\lambda_{i,R}=\lambda_i \text{ for } 1\le i\le k, & (1)\\ \displaystyle\sum_{i=1}^{k}\sum_{\substack{R\in\mathcal{R}_i\\ \nu\in R}}\lambda_{i,R}\le\mu \text{ for } 1\le\nu\le n, & (2)\\ \lambda_{i,R}\ge 0 \text{ for } 1\le i\le k,\ R\in\mathcal{R}_i. & (3)\end{cases}$$

满足 $\{\lambda_{i,R}\}$ 上述性质（2）和（3）的集合称为一对（可行分配，$\mathcal{R}$，μ）。与 $\mathcal{R}$ 和 μ 相关联的服务速率区域记为

$$\Lambda(\mathcal{R},\mu)\subseteq\mathbb{R}^k.$$

注意，G-系统的服务速率区域与每个 $\mathcal{R}_i$ 中恢复集的排列顺序无关。

备注 1.5。事实证明，Λ（$\mathcal{R}$，μ）是一个向下单调多面体；定义见下文定理 2.3 和附录A 。我们将在第 2 节介绍分配多胞体时对此作进一步说明。

一个 G-系统的服务速率区域 $\mathcal{R}$ 在选取恢复集的适当子集后可能不会改变，这使我们能够减少定义 1.4中的变量和不等式数量。我们从以下观察开始，其证明很简单，因此略去。

命题 1.6。假设 $\mathcal{R}=(\mathcal{R}_1,\dots,\mathcal{R}_k)$ 和 $\mathcal{R}'=(\mathcal{R}'_1,\dots,\mathcal{R}'_k)$ 是 G-系统，并且对所有 $i\in\{1,\dots,k\}$ 都有$\mathcal{R}'_i\subseteq\mathcal{R}_i$。那么 $\Lambda(\mathcal{R},\mu)\supseteq\Lambda(\mathcal{R}',\mu)$。特别地，对于任意 G-系统 $\mathcal{R}$，$\Lambda(\mathcal{R},\mu)\subseteq\Lambda(\mathcal{R}^{\mathrm{all}}(G)\ \mu)$。

服务速率区域 $\Lambda(\mathcal{R}^{\mathrm{all}}(G)\ \mu)$ 不会因我们从 $\mathcal{R}^{\mathrm{all}}(G)$ 中选取相对于包含关系而言极小的恢复集而改变，具体来说：

定义 1.7. 如果不存在满足 $R'\subsetneq R$ 的 $R'\in\mathcal{R}_i^{\mathrm{all}}(G)$，则称集合 $R\in\mathcal{R}_i^{\mathrm{all}}$ 为 i-极小。我们令 $\mathcal{R}^{\min}(G)$ 为如下定义的 G-系统，对所有 i，

$$\mathcal{R}_i^{\min}(G):=\{R\in\mathcal{R}_i^{\mathrm{all}}(G)\mid R\text{ is }i\text{-minimal}\}.$$

以下结果的证明并不困难，因此留给读者完成。

命题 1.8. 我们有 $\Lambda(\mathcal{R}^{\min}(G)\ \mu)=\Lambda(\mathcal{R}^{\mathrm{all}}(G)\ \mu)$。

注记 1.9. 根据定义可立即得出，对于任意 G-系统 $\mathcal{R}$，都有 $\Lambda(\mathcal{R},\mu)=\mu\Lambda(\mathcal{R},1)$，其中 $\mu\Lambda(\mathcal{R},1)=\{\mu\lambda\mid\lambda\in\Lambda(\mathcal{R},1)\}$。在下文中，我们常常不失一般性地假设 $\mu=1$。

下面的符号将进一步简化后续内容的表述。

记号1.10。对于 G-系统 $\mathcal{R}=(\mathcal{R}_1,\dots,\mathcal{R}_k)$，令 $\Lambda(\mathcal{R})=\Lambda(\mathcal{R},1)$。我们还记 $\Lambda(G,\mu)=\Lambda(\mathcal{R}^{\mathrm{all}}(G)\ \mu)=\Lambda(\mathcal{R}^{\min}(G)\ \mu)$，其中后一个等式由命题 1.8 得出。最后，令 $\Lambda(G)=\Lambda(G,1)$。

最后，我们通过一个例子来说明本节介绍的概念。

例 1.11。考虑以下矩阵

$$G_1 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \in \mathbb{F}_2^{2\times 4}, \qquad G_2 = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 2 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{F}_3^{3\times 6}.$$

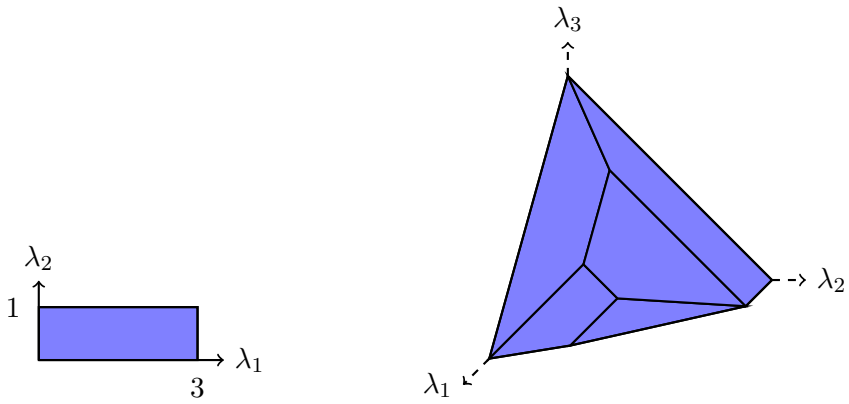
相应的服务速率区域如图1所示。对于矩阵 G_2，我们有

$$\begin{aligned} \mathcal{R}_1^{\min}(G_2) &= \{\{1\},\{5,6\},\{2,3,4\},\{2,4,5\},\{3,4,6\},\{2,3,6\},\{3,4,5\}\}, \\ \mathcal{R}_2^{\min}(G_2) &= \{\{2\},\{3,5\},\{4,6\},\{1,3,4\},\{1,4,5\},\{1,3,6\}\}, \\ \mathcal{R}_3^{\min}(G_2) &= \{\{3\},\{2,5\},\{1,2,4\},\{1,4,6\},\{1,2,6\},\{1,4,5\}\}. \end{aligned}$$

此外，对于所有的 $i \in \{1,2,3\}$，都有 $\mathcal{R}_i^{\text{all}} = \{R \subseteq \{1,\dots,6\} \mid S \subseteq R,$，其中 $S \in \mathcal{R}_i^{\min}(G_2)\}$。最后，例如，要说明点 $P = (3/2, 3/2, 1/2)$属于 $\Lambda(G_2)$，我们可以考虑如下可行分配：

$$\lambda_{1,R} = \begin{cases} 1 & \text{if } R = \{1\}, \\ 1/2 & \text{if } R = \{5,6\}, \\ 0 & \text{otherwise}, \end{cases} \qquad \lambda_{2,R} = \begin{cases} 1 & \text{if } R = \{2\}, \\ 1/2 & \text{if } R = \{3,5\}, \\ 0 & \text{otherwise}, \end{cases} \qquad \lambda_{3,R} = \begin{cases} 1/2 & \text{if } R = \{3\}, \\ 0 & \text{otherwise}. \end{cases}$$

不难看出集合 $\{\lambda_{i,R}\}$ 满足性质（1）–（3），对于（$\mathcal{R}^{\min}(G_2)$，1）。 ▲



(a) $\Lambda(G_1)$，例1.11。 (b) $\Lambda(G_2)$，例1.11。

图 1：示例 1.11 中系统的服务速率区域。

本文主要描述服务速率区域 $\Lambda(\mathcal{R}, \mu)$ 的几何性质。尽管我们的主要关注点是 $\Lambda(G)$，但我们的大多数结果都适用于任意的 G-系统 $\mathcal{R}$。

2 服务速率区域与分配多面体

本节描述与任意 G-系统 $\mathcal{R}$ 相关联的服务速率区域的多面体结构。我们还将说明这种几何结构对相应系统中的用户分配有何影响。为此，将服务速率区域视为高维多胞体在线性映射下的像通常很方便，我们称之为分配多胞体。

定义 2.1。设 $\mathcal{R} = (\mathcal{R}_1, \dots, \mathcal{R}_k)$ 为一个 G-系统，$m_i = |\mathcal{R}_i|$ 其中 $i \in \{1,\dots,k\}$，且 $m(\mathcal{R}) = m_1 + \ldots + m_k$。(分配多胞体)$\mathcal{R}, \mu$ 的分配多胞体是满足不等式 (2) 和 (3) 的 $(\lambda_{i,R} \mid i \in \{1,\dots,k\}, R \in \mathcal{R}_i)$ 的集合。我们将分配多胞体记为

$$\mathcal{A}(\mathcal{R}, \mu) \subseteq \mathbb{R}^{m(\mathcal{R})}.$$

我们还令 $\mathcal{A}(G) = \mathcal{A}(\mathcal{R}^{\text{all}}(G)) = \mathcal{A}(\mathcal{R}^{\min}(G))$，其中后一个恒等式可用与命题 1.8 类似的方法证明。

现在我们证明分配多胞体确实是一个多胞体，并说明它与服务速率区域之间的联系。我们将用到以下事实，它可由凸包的定义结合定理 A.1 及其后紧接的观察轻易推出。

引理 2.2。设 $\mathcal{P} \subseteq \mathbb{R}^m$ 是一个多胞体，$f\colon \mathbb{R}^m \to \mathbb{R}^k$ 是一个线性映射。那么 $f(\mathcal{P})$ 是一个多胞体，且 $\text{vert}(f(\mathcal{P})) \subseteq f(\text{vert}(\mathcal{P}))$。

下一条结果描述了分配多胞体与服务速率区域之间的联系，同时概述了这两个区域的一些性质。特别地，我们关注定理 2.3.

定理 2.3。设 $\mathcal{R}$ 是一个 G-系统，并设 $m(\mathcal{R})$ 如定义 2.1中所述。以下结论成立。

1. 有 $f(\mathcal{A}(\mathcal{R},\mu)) = \Lambda(\mathcal{R},\mu)$，其中 $f\colon \mathbb{R}^{m(\mathcal{R})} \to \mathbb{R}^k$ 是由下式定义的线性映射

$$f : \lambda = (\lambda_{i,R} \mid i \in \{1,\dots,k\},\ R \in \mathcal{R}_i) \mapsto \left(\sum_{R\in\mathcal{R}_1} \lambda_{1,R}, \dots, \sum_{R\in\mathcal{R}_k} \lambda_{k,R} \right).$$

2. $\mathcal{A}(\mathcal{R},\mu)$ 和 $\Lambda(\mathcal{R},\mu)$ 是向下单调多胞体。

证明。根据定义，$\Lambda(\mathcal{R},\mu)$ 是 $\mathcal{A}(\mathcal{R},\mu)$ 在 f 下的像，这一点很容易由定义 1.4和2.1得出。现在我们证明该陈述的第二部分。根据定义，集合 $\mathcal{A}(\mathcal{R},\mu)$ 是一个多面体。其有界性可如下证明。将（2）中的所有不等式相加，得到 $\sum_{i=1}^{k}\sum_{R\in\mathcal{R}_i} \lambda_{i,R} \le n\mu$。利用对于所有的配对 (i,R)，都有 $\lambda_{i,R} \ge 0$，可知任意 $\lambda \in \mathcal{A}(\mathcal{R},\mu)$ 满足

$$\|\lambda\|_2 \le \sum_{i=1}^{k}\sum_{R\in\mathcal{R}_i} |\lambda_{i,R}| = \sum_{i=1}^{k}\sum_{R\in\mathcal{R}_i} \lambda_{i,R} \le n\mu,$$

其中，$\|\lambda\|_2$ 是 λ的二范数。根据引理2.2 以及 $\mathcal{A}(\mathcal{R},\mu$的有界性，可知 $\Lambda(\mathcal{R},\mu)$ 有界。最后，不难直接验证多胞体 $\mathcal{A}(\mathcal{R},\mu)$ 和 $\Lambda(\mathcal{R},\mu)$ 都是向下单调的。

□

现在我们来探讨一个自然的问题：如何描述服务速率区域的顶点以及其中坐标为有理数的点。我们将利用服务速率区域与分配多胞体之间的联系来回答这些问题。

我们注意到，线性映射 $f\colon \mathbb{R}^m \to \mathbb{R}^k$ 称为有理的，如果其在典范基下的矩阵具有有理数项。下一个结果由命题A.2 和推论A.3推出。

引理 2.4. 设 $\mathcal{P} \subseteq \mathbb{R}^m$ 是一个有理多胞体，且 $f\colon \mathbb{R}^m \to \mathbb{R}^k$ 是一个有理线性映射。那么 $f(\mathcal{P})$是一个多胞体，其顶点的坐标均为有理数。

下面的结果对于定性描述分配多胞体与服务速率区域之间的联系至关重要。

引理 2.5. $\mathcal{P} \subseteq \mathbb{R}^m$ 是一个有理多胞体，且设 $f\colon \mathbb{R}^m \to \mathbb{R}^k$ 是一个有理线性映射。我们有 $f(\mathcal{P} \cap \mathbb{Q}^m) = f(\mathcal{P}) \cap \mathbb{Q}^k$。

证明。包含关系 $\subseteq$ 是显然的。为证明另一包含关系，令 $y \in f(\mathcal{P}) \cap \mathbb{Q}^k$。写作 $f = (f_1,\dots,f_k)$，其中 $f_i\colon \mathbb{R}^m \to \mathbb{R}$。我们需要证明存在 $x \in \mathcal{P} \cap \mathbb{Q}^m$，使得对于所有 $i \in \{1,\dots,k\}$，都有 $f_i(x) = y_i$。由于 $\mathcal{P}$是有理的，故对于某些$A \in \mathbb{Q}^{\ell\times m}$ 和 $b \in \mathbb{Q}^{\ell}$，有 $\mathcal{P} = \{x \in \mathbb{R}^m \mid Ax^\top \le b^\top\}$。我们在 A 和 b 中附加总共 $2k$ 行，其中 k 行对应不等式 $\{f_i(x) \le y_i \mid i = 1,\dots,k\}$，另外 k 行对应不等式 $\{-f_i(x) \le -y_i \mid i = 1,\dots,k\}$。

令 A' 和 b' 分别表示所得的矩阵和向量，其大小分别为 $(\ell + 2k) \times m$ 和长度 $\ell + 2k$。那么，根据定义，$\mathcal{P}' = \{x \in \mathbb{R}^m \mid A'x^\top \leq b'^\top\}$ 是一个多面体。注意，我们需要证明 $\mathcal{P}' \cap \mathbb{Q}^m \neq \emptyset$。首先注意到，由于 $\mathcal{P}' \subseteq \mathcal{P}$ 且 $\mathcal{P}$ 有界，$\mathcal{P}'$ 是有界的。此外，由于 A' 和 b' 具有有理数项，$\mathcal{P}'$ 是有理的（这里使用了以下事实：y 具有有理数项，且 f 是有理映射）。此外，由于根据假设 $y \in f(\mathcal{P})$，$\mathcal{P}'$ 非空，这意味着存在 $z \in \mathcal{P}$ 满足 $f(z) = y$，即存在 $z \in \mathcal{P}'$。于是，$\mathcal{P}'$ 至少有一个顶点 x，且根据推论 A.3，该顶点必具有有理数项。因此$\mathcal{P}' \cap \mathbb{Q}^m \neq \emptyset$，如所要求。 □

现在可以陈述本节的主要结果了（请注意，该陈述中的性质仅对 $\mu = 1$ 成立）。

定理 2.6. 设 $\mathcal{R}$ 是一个 G-系统。以下结论成立。

1. 的顶点 $\Lambda(\mathcal{R})$ 具有有理数项。2. $\Lambda(\mathcal{R}) \cap \mathbb{Q}^k = f(\mathcal{A}(\mathcal{R}) \cap \mathbb{Q}^k)$，
其中 f 的定义与定理 2.3 中相同。

证明。 的顶点 $\Lambda(\mathcal{R})$ 具有有理数项，这是定理 2.3 和引理2.4的结果。将定理2.3 与引理 2.5结合即可得到该陈述的第二部分。 □

请注意，定理 2 的性质 2.6 意味着，服务速率区域中的每个有理点都有一个可行有理分配；参见定义 1.4。这一事实从表面上看并不明显，但从应用角度来看（见小节 1.3）非常重要。

3 整数分配模型

本节考虑第 1.3 节所述的实际场景：每台服务器都有特定的带宽，并且每个被服务的请求在服务期间都会占用全部带宽，即 $\lambda_{i,\ R}$ 必须为 0 或 1。我们定义该模型的服务速率区域，并说明它与上一节所考虑模型的关系。

假设我们使用系统 $s \in \mathbb{Z}_{\geq 1}$ 次，每次采用的分配可能不同。令 $\alpha_i(R)$ 表示在使用系统 s 次期间，用恢复集 $R \in \mathcal{R}_i$ 恢复第 i 个对象的次数。那么，第 i 个对象被恢复的次数等于 $\lambda_i = \sum_{R \in \mathcal{R}_i} \alpha_i(R)$。由此引出以下定义。

定义 3.1. 设 $\mathcal{R}$ 是一个 G-系统。一个 $\mathcal{R}$-分配是由函数 $\alpha = (\alpha_1, \ldots, \alpha_k)$ 组成的 k-元组，其中对所有 $i \in \{1, \ldots, k\}$，$\alpha_i \colon \mathcal{R}_i \to \mathbb{N}$。$\alpha$ 的服务率是向量 $\lambda(\alpha) = (\lambda_1, \ldots, \lambda_k) \in \mathbb{N}^k$，其中

$$\lambda_i = \sum_{R \in \mathcal{R}_i} \alpha_i(R) \qquad \text{for all } i \in \{1, \ldots, k\}.$$

对于 $\nu \in \{1, \ldots, n\}$，我们定义

$$\delta_\nu(\mathcal{R}, \alpha) = \sum_{i=1}^{k} \sum_{R \in \mathcal{R}_i} \delta_\nu(R)\, \alpha_i(R) \quad \text{where} \quad \delta_\nu(R) = \begin{cases} 1 & \text{if } \nu \in R, \\ 0 & \text{otherwise.} \end{cases}$$

在定义 3.1中，量 $\delta_\nu(\mathcal{R},\ \alpha)$ 表示服务器 ν 被联系的次数。

定义 3.2. 设 $\mathcal{R}$ 是一个 G-系统。$\mathcal{R}$ 的单次服务速率区域（容量为 $s \in \mathbb{Z}_{\geq 1}$）是 $\Lambda_1(\mathcal{R}, s) = \{\lambda(\alpha)/s \mid \alpha$ 一个 $\mathcal{R}$-分配 $\delta_\nu(\mathcal{R}, \alpha) \leq s$ 对于 $1 \leq \nu \leq n\}$。$\mathcal{R}$ 的有理服务速率区域是集合

$$\Lambda^{\mathbb{Q}}(\mathcal{R}) = \bigcup_{s \in \mathbb{Z}_{\geq 1}} \Lambda_1(\mathcal{R}, s).$$

本节将展示有理服务速率区域与第 1 节中定义的服务速率区域之间以下这种“拓扑的”联系。

定理 3.3。设 $\mathcal{R}$ 是一个 G-系统。以下结论成立。

1. $\Lambda^{\mathbb{Q}}(\mathcal{R}) = \underline{\Lambda(\mathcal{R})} \cap \mathbb{Q}^k$。2. $\Lambda(\mathcal{R}) = \Lambda^{\mathbb{Q}}(\mathcal{R})$，其中后者是 $\Lambda^{\mathbb{Q}}(\mathcal{R})$ 关于 $\mathbb{R}^k$ 中欧几里得拓扑的闭包。

注 3.4。 在证明定理 3.3之前，我们强调，这一定理与第 1.3 节所述的实际场景尤其相关，因为这些场景可能要求分配为整数或有理数。它表明：1）服务速率区域中的有理点可以通过有理分配实现；2）所有点都可以通过对系统多次使用取平均来实现。

我们将在定理 3.3 的证明中使用以下结果。

引理 3.5。设 $\mathcal{P} \subseteq \mathbb{R}^m$ 是一个向下单调多面体。那么 $\mathcal{P} = \overline{\mathcal{P} \cap \mathbb{Q}^m}$，其中后者是 $\mathcal{P} \cap \mathbb{Q}^m$ 关于欧几里得拓扑的闭包。

证明。包含关系 $\overline{\mathcal{P} \cap \mathbb{Q}^m} \subseteq \mathcal{P}$ 成立，因为 $\mathcal{P}$ 是闭集且 $\overline{\mathbb{Q}^m} = \mathbb{R}^m$，这意味着 $\overline{\mathcal{P} \cap \mathbb{Q}^m} \subseteq \overline{\mathcal{P}} \cap \overline{\mathbb{Q}^m} = \mathcal{P} \cap \mathbb{R}^m = \mathcal{P}$。对于另一个包含关系，我们将证明，对于所有 $x \in \mathcal{P}$ 和所有 $\varepsilon > 0$，都有 $B_\varepsilon(x) \cap \mathcal{P} \cap \mathbb{Q}^m \neq \emptyset$，其中 $B_\varepsilon(x)$ 是以 x为中心、半径为 ε 的球。这意味着 $\mathcal{P} \subseteq \overline{\mathcal{P} \cap \mathbb{Q}^m}$，例如可由 [22, 定理 17.5 和 20.3]得出。任取如上所述的 x 和 ε。记 $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$。由于 $\mathbb{Q}$ 在 $\mathbb{R}$中稠密，对于每个 $i \in \{1, \ldots, m\}$，都存在 $y_i \in \mathbb{Q}$使得 $x_i - \varepsilon/m \leq y_i \leq x_i$。由于 $\mathcal{P}$向下单调，我们有 $y = (y_1, \ldots, y_m) \in \mathcal{P}$。现在，$y \in B_\varepsilon(x) \cap \mathcal{P} \cap \mathbb{Q}^m$ 这一事实可由下式推出

$$\|x - y\|_2 \leq \|x - y\|_1 = \sum_{i=1}^{m} (x_i - y_i) \leq m \cdot \varepsilon/m = \varepsilon,$$

我们在 $\mathbb{R}^k$ 中使用了 p-范数的标准记号。 □

定理 *3.3* 的证明。 该陈述的第二部分由第一部分结合引理 3.5 得出。因此，只需证明第一部分。

设 $\lambda \in \Lambda^{\mathbb{Q}}(\mathcal{R})$。存在一个 $s \in \mathbb{Z}_{\geq 1}$ 和一个 $\mathcal{R}$-分配 α，使得对于所有 $\nu \in \{1, \ldots, n\}$，都有 $s\lambda = \lambda(\alpha) = (\lambda_1, \ldots, \lambda_k) \in s\Lambda_1(\mathcal{R}, s)$ 以及 $\delta_\nu(\mathcal{R}, \alpha) \leq s$。注意，$\lambda \in \mathbb{Q}^k$。我们将证明集合

$$\left\{ \frac{\alpha_i(R)}{s} \mid i \in \{1, \ldots, n\},\ R \in \mathcal{R}_i \right\}$$

满足性质（1）、（2）和（3）。根据定义，对于 $1 \leq i \leq k$，我们有 $\lambda_i/s = \sum_{R \in \mathcal{R}_i} \alpha_i(R)/s$。此外，条件 $\delta_\nu(\mathcal{R}, \alpha) \leq s$ 可改写为

$$\sum_{i=1}^{k} \sum_{\substack{R \in \mathcal{R}_i \\ \nu \in R}} \frac{\alpha_i(R)}{s} \leq 1 \quad \text{for } 1 \leq \nu \leq n.$$

最后，对于所有 $i \in \{1, \ldots, k\}$ 和所有 $R \in \mathcal{R}_i$，$\alpha_i(R)/s \geq 0$。因此，我们得出结论：$\lambda \in \Lambda(\mathcal{R}) \cap \mathbb{Q}^k$，故 $\Lambda^{\mathbb{Q}}(\mathcal{R}) \subseteq \Lambda(\mathcal{R}) \cap \mathbb{Q}^k$。

为证明另一包含关系，设 $\lambda \in \Lambda(\mathcal{R}) \cap \mathbb{Q}^k$。由定理 2.6，我们有$\Lambda(\mathcal{R}) \cap \mathbb{Q}^k = f(\mathcal{A}(\mathcal{R}) \cap \mathbb{Q}^k)$，其中 f 在定理 2.3 中定义。特别地，存在满足定义 1.4 的性质 (1)–(3) 的有理数$\{\lambda_{i,R} \in \mathbb{Q} \mid i \in \{1, \ldots, k\},\ R \in \mathcal{R}_i\}$。根据定义，对于所有 $i \in \{1, \ldots, k\}$ 和 $R \in \mathcal{R}_i$，有 $\lambda_{i,R} = u_{i,R}/v_{i,R}$, $u_{i,R}, v_{i,R} \in \mathbb{N}$，以及 $v_{i,R} > 0$。令$s := \operatorname{lcm}(v_{i,R} \mid i \in \{1, \ldots, k\},\ R \in \mathcal{R}_i)$，从而 $s\lambda = (s\lambda_1, \ldots, s\lambda_k) \in \mathbb{N}^k$。我们声称$s\lambda \in s\Lambda_1(\mathcal{R}, s)$。现在对于 $1 \leq i \leq k$，定义 $\alpha_i(R) = s\lambda_{i,R}$；注意到 α_i 总是通过 $\mathbb{N}$ 映射到

数 s的构造。于是对于 $1 \le i \le k$，我们有 $s\lambda_i = \sum_{R\in\mathcal{R}_i} \alpha_i(R)$，由此可得 $\alpha = (\alpha_1, \ldots, \alpha_k)$ 是一个 $\mathcal{R}$-分配。此外，对于 $\nu \in \{1, \ldots, n\}$，我们有

$$\delta_\nu(\mathcal{R}, \alpha) = \sum_{i=1}^{k} \sum_{R\in\mathcal{R}_i} \delta_\nu(R)\alpha_i(R) = \sum_{i=1}^{k} \sum_{\substack{R\in\mathcal{R}_i \\ \nu\in R}} \alpha_i(R) = \sum_{i=1}^{k} \sum_{\substack{R\in\mathcal{R}_i \\ \nu\in R}} s\lambda_{i,R} \le s,$$

其中第二个等式由 $\delta_\nu(R)$ 的定义得出，最后一个不等式由（2）得出。这表明 $s\lambda \in s\Lambda_1(\mathcal{R}, s)$，等价地 $\lambda \in \Lambda_1(\mathcal{R}, s)$，因此 $\lambda \in \Lambda^{\mathbb{Q}}(\mathcal{R})$，正如所要求的。

□

我们用一个说明定理 3.3 的例子结束本节。

例 3.6。设

$$G := \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix} \in \mathbb{F}_3^{2\times 4}.$$

考虑 G-系统 $\mathcal{R} = \mathcal{R}^{\min}(G) = (\mathcal{R}_1, \mathcal{R}_2)$，其中 $\mathcal{R}_1 := \{\{1\}, \{2,3\}, \{2,4\}, \{3,4\}\}$ 且 $\mathcal{R}_2 := \{\{2\}, \{1,3\}, \{1,4\}, \{3,4\}\}$。相应的服务速率区域如图 2 所示，同时标出了点 $P = (4/3, 2/3)$。

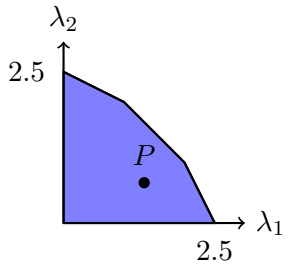


图 2：例 G-系统的服务速率区域，以及例 3.6 中的点 $P = (4/3, 2/3)$。

我们有 $P \in \Lambda(\mathcal{R}) \cap \mathbb{Q}^2$，并且 $P \in \Lambda_1(\mathcal{R}, 3)$，即 P可以通过使用系统三次来实现。根据定义 3.1，一个 $\mathcal{R}$-分配的示例如下：

$$\begin{array}{rl} \alpha_1 : \mathcal{R}_1 \to \mathbb{N} & \qquad \alpha_2 : \mathcal{R}_2 \to \mathbb{N} \\ \{1\} \mapsto 2 & \qquad \{2\} \mapsto 1 \\ \{2,3\} \mapsto 1 & \qquad \{1,3\} \mapsto 0 \\ \{2,4\} \mapsto 0 & \qquad \{1,4\} \mapsto 1 \\ \{3,4\} \mapsto 1 & \qquad \{3,4\} \mapsto 0 \end{array}$$

我们有 $\delta_1(\mathcal{R}, \alpha) = \alpha_1(\{1\}) + \alpha_2(\{1,4\}) = 3$，$\delta_2(\mathcal{R}, \alpha) = \alpha_1(\{2,3\}) + \alpha_2(\{2\}) = 2$，$\delta_3(\mathcal{R}, \alpha) = \alpha_1(\{2,3\}) + \alpha_1(\{3,4\}) = 2$，$\delta_4(\mathcal{R}, \alpha) = \alpha_1(\{3,4\}) + \alpha_2(\{1,4\}) = 2$。此外，$\sum_{R\in\mathcal{R}_1} \alpha_1(R) = 4$和 $\sum_{R\in\mathcal{R}_2} \alpha_2(R) = 2$，这表明 $(4,2) \in \Lambda_1(\mathcal{R}, 3)$。▲

4 服务速率区域的基本参数

本节介绍服务速率区域的一些基本参数，用以描述其“形状”。随后，我们运用各种技巧对这些参数进行研究。研究结果旨在说明，底层矩阵 G 的代数性质如何决定服务区域多胞体 $\Lambda(G, \mu)$的几何性质。为简化记号（且不失一般性），我们假设$\mu = 1$。首先回顾两类初等多胞体。

定义 4.1. 设 $h, \delta \in \mathbb{R}_{\geq 0}$。 $\mathbb{R}^k$ 中的 h-超立方体是集合 $\{x \in \mathbb{R}^k_{\geq 0} \mid x_i \in \{0, h\} \ (\ 1 \leq i \leq k\})$ 的凸包。我们称 h 为该超立方体的大小。 $\mathbb{R}^k$ 中的 δ-单纯形是集合 $\{\delta e_1, \ldots, \delta e_k\}$的凸包，其中 e_i 是 $\mathbb{F}_q^k$的第 i个标准基向量。同样，我们称 δ 为该单纯形的大小。

我们将介绍服务速率区域的第一组参数。其他参数将在后文介绍。

定义 4.2. 设 $\mathcal{R}$ 是一个 G-系统。我们定义：

$$\begin{aligned}
&\lambda^r(\mathcal{R}) = \max\{\textstyle\sum_{i=1}^k \lambda_i^r \mid \lambda \in \Lambda(\mathcal{R})\}, && [r\textbf{-th max-sum capacity}]\\
&\lambda(\mathcal{R}) = \lambda^1(\mathcal{R}), && [\textbf{max-sum capacity}]\\
&\lambda_i^*(\mathcal{R}) = \max\{x \in \mathbb{R} \mid xe_i \in \Lambda(\mathcal{R})\} \text{ for } 1 \leq i \leq k,\\
&\lambda^*(\mathcal{R}) = \max\{\lambda_i^*(\mathcal{R}) \mid 1 \leq i \leq k\}.
\end{aligned}$$

此外，我们将服务速率区域内所包含的超立方体和单纯形的最大尺寸分别记为：

$$\begin{aligned}
h(\mathcal{R}) &= \max\{x \in \mathbb{R} \mid (x, \ldots, x) \in \Lambda(\mathcal{R})\}\\
\delta(\mathcal{R}) &= \min\{\lambda_i^*(\mathcal{R}) \mid 1 \leq i \leq k\}.
\end{aligned}$$

当 $\mathcal{R} = \mathcal{R}^{\text{all}}(G)$ 或 $\mathcal{R} = \mathcal{R}^{\min}(G)$ 时，我们简记为 $\lambda^r(G)$、 $\lambda(G)$、$\lambda_i^*(G)$、 $\lambda^*(G)$、 $h(G)$ 和 $\delta(G)$。

下一个例子表明，一般而言，达到最大和容量的点并不能达到 $r > 1$ 的第 r 大和容量。

例 4.3. 考虑例 3.6中的服务速率区域，如图 2所示。可以证明 $\lambda(G)$ 由 (1,2) 和 $(2, 1)$ 达到。另一方面， $\lambda^2(G) = 6.25 > 5 = 1 + 4$ 由 (2.5,0) 和 (0,2.5) 达到。▲

我们先给出一个结果，说明参数 $h(\mathcal{R})$、 $\lambda(\mathcal{R})$ 和 $\delta(\mathcal{R})$ 之间的关系。

命题 4.4. 设 $\mathcal{R}$ 是一个 G-系统。我们有

$$h(\mathcal{R}) \leq \min\left\{\frac{\lambda(\mathcal{R})}{k}, \delta(\mathcal{R})\right\}.$$

证明。首先证明 $h(\mathcal{R}) \leq \lambda(\mathcal{R})/k$。假设 $h(\mathcal{R}) > \lambda(\mathcal{R})/k$。根据定义，我们有 $(h(\mathcal{R}) \ldots, h(\mathcal{R})) \in \Lambda(\mathcal{R})$。因此

$$\lambda(\mathcal{R}) \geq h(\mathcal{R})k > \frac{\lambda(\mathcal{R})}{k}k = \lambda(\mathcal{R}),$$

这就产生了矛盾。证明的第二部分，假设 $h(\mathcal{R}) > \delta(\mathcal{R})$。那么，根据 $\delta(\mathcal{R})$ 的定义，集合 $\{h(\mathcal{R})e_i \mid 1 \leq i \leq k\} \subseteq \mathbb{F}_q^k$ 中至少有一个元素不属于 $\Lambda(\mathcal{R})$。这与 $h(\mathcal{R})$ 的定义矛盾。 □

注 4.5. 命题 4.4 中的界在图 1a所示的服务速率区域中取等号。然而，这个界一般并不紧。举例来说，考虑服务速率区域 $\Lambda(G)$，其中

$$G = \begin{pmatrix} 1 & 1 & 0 & 1\\ 0 & 0 & 1 & 1\\ 0 & 0 & 0 & 1 \end{pmatrix} \in \mathbb{F}_2^{3\times 4}.$$

注意，(0,0,1)、(0,1,0)、(1,0,1)、$(2, 1, 0) \in \Lambda(G)$。可以证明， $h(G) = 0.5$， $\lambda(G) = 3$，且 $\delta(G) = 1$。

在下一个例子中，我们将说明这些值 $\delta(\mathcal{R})$ 和 $\lambda(\mathcal{R})/k$ 通常不可比较，这表明命题 4.4 界中的最小值确实是必要的。

例 4.6. 对于例 3.6中的服务速率区域，我们有 $\delta(G) = 2.5 > 1.5 = \lambda(G)/2$.然而，对于图 1b 中的服务速率区域，我们有 $\delta(G) = 1 < 2 = \lambda(G)/2$. ▲

量 $\lambda^2(\mathcal{R})$ 具有明确的几何意义；它给出了包含该服务速率区域的最小球形楔体。为说明 $\lambda^2(\mathcal{R})$ 与其他基本参数之间的关系，我们将采用一个基于统计学中的 Bhatia-Davis 不等式 [7] 的论证。请注意，以下界在下文例 G中的 4.18 -系统中是紧的。

定理 4.7. 设 $\mathcal{R}$ 是一个 G-系统。我们有

$$\lambda^2(\mathcal{R}) \leq \frac{k-1}{k}\lambda^*(\mathcal{R})\lambda(\mathcal{R}) + \frac{(\lambda(\mathcal{R}))^2}{k}.$$

证明。令 $\hat{\lambda} \in \Lambda(\mathcal{R})$ 达到 $\lambda^2(\mathcal{R})$。我们将 Bhatia-Davis 不等式 [7] 应用于 $\hat{\lambda}$的坐标，得到

$$\begin{aligned}
\frac{1}{k}\sum_{i=1}^{k}\hat{\lambda}_i^2 &\leq \left(\max_{1\leq i\leq k}\{\hat{\lambda}_i\} - \frac{1}{k}\sum_{i=1}^{k}\hat{\lambda}_i\right)\left(\frac{1}{k}\sum_{i=1}^{k}\hat{\lambda}_i - \min_{1\leq i\leq k}\{\hat{\lambda}_i\}\right) + \frac{1}{k^2}\left(\sum_{i=1}^{k}\hat{\lambda}_i\right)^2 \\
&\leq \left(\frac{k-1}{k}\max_{1\leq i\leq k}\{\hat{\lambda}_i\}\right)\left(\frac{1}{k}\sum_{i=1}^{k}\hat{\lambda}_i\right) + \frac{1}{k^2}\left(\sum_{i=1}^{k}\hat{\lambda}_i\right)^2 \\
&\leq \left(\frac{k-1}{k}\lambda^*(\mathcal{R})\right)\left(\frac{1}{k}\lambda(\mathcal{R})\right) + \frac{1}{k^2}\lambda(\mathcal{R})^2.
\end{aligned}$$

根据假设，$\hat{\lambda}$ 达到 $\lambda^2(\mathcal{R})$，因此我们可以将刚才得到的不等式改写如下：

$$\frac{1}{k}\lambda^2(\mathcal{R}) \leq \frac{k-1}{k^2}\lambda^*(\mathcal{R})\lambda(\mathcal{R}) + \frac{1}{k^2}(\lambda(\mathcal{R}))^2.$$

两边乘以 k 即可得到所需结果。 □

服务速率区域的另一个自然参数是其体积。回顾一下，凸多胞体 $\mathcal{P}$ 的体积是其内部的勒贝格测度 [20] ，记作 $\mathrm{vol}(\mathcal{P})$。分布式服务系统致力于支持同时用户的数据下载，而用户数量和兴趣会随时间变化。服务速率区域的体积越大，系统能够服务的不同用户数量配置就越多。

一般而言，计算多胞体的体积是一项困难的任务 [13]。不过，有些与我们目的相关的情形，只需简单观察即可得出服务速率区域体积的闭式公式。

命题 4.8。 假设 G 是一个复制矩阵；定义见第 2 页。我们有

$$\mathrm{vol}(\Lambda(G)) = \prod_{i=1}^{k} |\{1 \leq \nu \leq n \mid \text{the } \nu\text{-th column of } G \text{ is a nonzero multiple of } e_i\}|.$$

证明。容易看出，服务速率区域 $\Lambda(G)$ 是 $\mathbb{R}^k$ 中的一个超矩形，其中每条边的长度等于相应的标准基向量在 G 中作为列出现的次数。因此， $\Lambda(G)$ 的体积即为陈述中的量。

□

在定理6.5 中，我们将给出 $\Lambda(G)$ 的体积闭式公式，其中 G 生成一个长度至少为 6 的三维 MDS 码。该结果详见第6节，该节专门讨论系统 MDS 码的服务速率区域。

现在我们计算复制系统的分配多胞体的体积，并特别说明分配多胞体的体积并不能确定服务速率区域的体积。直观地说，这是因为分配多胞体的体积关于对应于同一对象的坐标是多重线性的，而服务速率多胞体的体积则关于这些坐标之和是线性的。我们将引入一类多胞体，稍后将在第5节中使用。

定义 4.9. 形如 $\mathcal{P} = \{x \in [0,1]^m \mid yx^\top \le n\} \subseteq \mathbb{R}^m$ 的多胞体，其中 n 和 m 为正整数，$y \in \mathbb{R}^m_{\ge 0}$ 为向量，称为 $\mathbb{R}^m$ 中的松弛背包多胞体。

定义 4.9 中的松弛背包多胞体的体积已知为

$$\operatorname{vol}(\mathcal{P}) = \frac{1}{m! \prod_{i=1}^m y_i} \sum_{x \in \{0,1\}^m \cap \mathcal{P}} (-1)^{\operatorname{wt}(x)} g(x)^m, \tag{4}$$

其中，$\operatorname{wt}(x)$ 是 x 的非零项数，并且对于所有 $x \in \mathbb{R}^m$，$g(x) = n - \sum_{i=1}^m y_i x_i$；参见例如 [6]。利用上述体积公式和一些初等生成函数理论，我们计算出复制矩阵的分配多胞体的体积。

命题4.10. 假设 G 是复制矩阵。我们有 $\operatorname{vol}(\mathcal{A}(G)) =$ 1.

证明。不难看出 $\mathcal{A}(G) = \{x \in [0,1]^n \mid x_1 + \cdots + x_n \le n\}$，这是在 $m = n$ 和 $y = 1^n = (1, \ldots, 1)$ 的条件下得到的松弛背包多胞体。因此，利用 (4) 并将 $[x^n]S(x)$ 记为幂级数 x^n 中的系数 $S(x)$，我们计算得

$$\begin{aligned}\operatorname{vol}(\mathcal{A}(G)) &= \frac{1}{n!} \sum_{x \in \{0,1\}^n} (-1)^{\operatorname{wt}(x)} (n - (x_1 + \ldots + x_n))^n = \frac{1}{n!} \sum_{x \in \{0,1\}^n} (-1)^{\operatorname{wt}(x)} (n - \operatorname{wt}(x))^n \\ &= \frac{1}{n!} \sum_{i=0}^n (-1)^i (n-i)^n \binom{n}{i} = \sum_{i=0}^n (-1)^i \binom{n}{i} [x^n] e^{(n-i)x} = [x^n](e^x - 1)^n = 1,\end{aligned}$$

其中所有推导都可以直接由二项式定理和指数函数的泰勒展开得到。

□

请注意，尽管对于每个复制矩阵，分配多胞体的体积都相同，但服务速率区域的体积并非常数；参见命题 4.8。

本节中，我们着重探讨 $\Lambda(G)$ 的参数与由 G生成的纠错码的参数之间的联系；有关编码理论的背景知识，请参见附录B 。

我们先回顾 [2], 中的以下结果，其表述依赖于将 G 的列解释为有限射影空间 $\mathrm{PG}(k-1, q)$ 中的点；一般性参考资料见 [31]。之所以可以这样做，是因为正如记号 1.1所述，G 的列均不是零向量。

命题 4.11. 设 $\lambda \in \Lambda(G)$，且 $I \subseteq \{1, \ldots, k\}$ 为一个指标集。设 H 为 $\mathrm{PG}(k-1, q)$ 的一个超平面，不包含任何标准基向量 e_i（$i \in I$），并设 S 表示 G 在 $\mathrm{PG}(k-1, q)$ 中的列所构成的多重集。则有

$$\sum_{i \in I} \lambda_i \le |S \setminus H|,$$

其中，$S \setminus H$ 是从 S 中移除所有包含在 H中的点后得到的多重集，并按其重数计数。

下面的引理广为人知，可以通过将 G 的列视为 $\mathrm{PG}(k-1, q)$ 中的点的多重集来证明。根据假设，G 的秩为 k，因此这些点并非全都包含在某个超平面中。

引理 4.12. 设 S 为 G 的列所构成的多重集，并将其视为 PG($k-1,q$) 中的射影点。设 d 为由 G 生成的码的最小距离。那么，PG($k-1,q$) 的每个超平面至多包含 S 中的 $n-d$ 个点，并且存在一个 PG($k-1,q$) 的超平面，恰好包含 S 中的 $n-d$ 个点。

我们还可以应用命题 4.11 来说明由 G 生成的码的最小距离与服务速率区域中所包含的最大单纯形之间的联系。

推论 4.13。设 d 表示由 G生成的码的最小距离。我们有

$$\lceil \delta(G) \rceil \leq d.$$

证明。 固定 $i \in \{1,\dots,k\}$，并设 $H \subseteq \mathrm{PG}(k-1,q)$ 是一个不包含 e_i 的超平面。应用命题 4.11，取 $I=\{i\}$ 和 $\lambda = \delta(G)e_i \in \Lambda(G)$，可得

$$\delta(G) \leq |S \setminus H| = n - |S \cap H|,$$

其中， S 是 G的列所构成的多重集。对于每个 i，我们都可以采用相同的推理。由于每个超平面 H 都不包含某个 e_i,

$$d = n - \max\{|S \cap H| : H \subseteq \mathrm{PG}(k-1,q),\ H \text{ hyperplane}\}$$

并且 $\delta(G) \leq |S \setminus H|$，因此我们得出 $\lceil \delta(G) \rceil \leq d$。 □

推论 4.13 的界在某些矩阵 G中可以取等号，以下示例将对此加以说明。

例 4.14. 令

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \in \mathbb{F}_2^{3\times 4}$$

容易看出，对于 $i \in \{1,2,3\}$， $2e_i \in \Lambda(G)$，因此 $\delta(G) = 2$。 ▲

在本节的最后一部分中，受编码理论文献的启发，我们引入矩阵 G 的可用度概念。随后，我们将描述这一概念在塑造服务速率区域几何结构方面所发挥的作用。

定义 4.15. 假设 G 是系统的。如果对于所有 $i \in \{1,\dots,k\}$， G 的 $t \in \mathbb{Z}_{\geq 0}$ 可用度意味着 $\mathcal{R}_i^{\mathrm{all}}(G)$ 包含 $t+1$ 个两两不相交的集合。

以下结果可直接由定义推出。

命题 4.16. 假设 G 是系统的，并且具有 t 可用度。那么，对于所有 $i \in \{1,\dots,k\}$，$(t+1)e_i \in \Lambda(G)$。特别地， $\lfloor \delta(G) \rfloor \geq t+1$。

结合推论 4.13 与命题 4.16，我们得到以下结果。

推论 4.17。设 G 是系统的，并具有可用度 t。令 d 表示由 G 生成的码的最小距离。则有 $d \geq t+1$。

我们以一个命题 4.16 和推论 4.17 均为紧确的例子结束本节。

例 4.18（单纯形码）。设

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{F}_2^{3\times 7}.$$

那么 G 可用度为 3，因此命题 4.16 在这种情况下是紧的。注意， G 是一个最著名的纠错码——单纯形码的生成矩阵；例如参见 [21]。 ▲

5 外界

本节将推导服务速率区域 $\Lambda(G)$ 的外界，即界多胞体$\mathcal{P} \supseteq \Lambda(G)$。我们将运用编码理论和优化中的方法，并分别用一个小节介绍这两种方法。我们将通过示例说明如何应用这些界，并讨论它们的紧致性。

5.1 编码理论方法

我们先给出一个简单结果，只需将定义分配多胞体的各个不等式，即（2）中关于 $1 \le \nu \le n$ 的约束相加，即可轻松得到该结果。

引理 5.1 (总容量界)。设 $\mathcal{R}$ 是一个 G-系统，且 $\{\lambda_{i,R}\}$ 是（$\mathcal{R}$， μ）的可行分配；参见定义 1.4。我们有

$$\sum_{i=1}^{k} \sum_{R \in \mathcal{R}_i} |R| \lambda_{i,R} \le \mu n. \tag{5}$$

以下结果将 G-系统的恢复集大小与由 G 生成的纠错码（及其对偶码）的参数联系起来。

命题 5.2。假设 G 是系统的。令 $d^\perp$ 表示由 G 生成的码的对偶码的最小距离。对于所有 $i \in \{1, \dots, k\}$ 和 $R \in \mathcal{R}_i^{\text{all}}(G)$，我们有 $R = \{i\}$ 或 $|R| \ge d^\perp - 1$。

现在，我们给出本节的第一个外界。

定理 5.3（对偶距离界）。假设 G 是系统的。令 $d^\perp$ 表示由 G生成的码的对偶码的最小距离。若 $(\lambda_1, \dots, \lambda_k) \in \Lambda(G)$，则

$$\sum_{i=1}^{k} \Big(\min\{\lambda_i, 1\} + (d^\perp - 1) \max\{0, \lambda_i - 1\} \Big) \le n.$$

证明。令 $(\lambda_1, \dots, \lambda_k) \in \Lambda(G)$，并令 $\{\lambda_{i,R}\}$ 为 $(\mathcal{R}$,1) 的可行分配。根据命题 5.2，对于每个 $|R| \ge d^\perp - 1$ 以及满足 $i \in \{1, \dots, k\}$ 和 $R \in \mathcal{R}_i^{\text{all}}(G)$ 的 $R \ne \{i\}$，我们有。因而，可以将 (5) 的左侧改写如下：

$$\begin{aligned}
\sum_{i=1}^{k} \lambda_{i,\{i\}} + \sum_{i=1}^{k} \sum_{\substack{R \in \mathcal{R}_i^{\text{all}}(G) \\ R \ne \{i\}}} |R| \lambda_{i,R} &\ge \sum_{i=1}^{k} \lambda_{i,\{i\}} + (d^\perp - 1) \sum_{i=1}^{k} \sum_{\substack{R \in \mathcal{R}_i^{\text{all}}(G) \\ R \ne \{i\}}} \lambda_{i,R} \\
&= \sum_{i=1}^{k} \lambda_{i,\{i\}} + (d^\perp - 1) \sum_{i=1}^{k} \big(\lambda_i - \lambda_{i,\{i\}} \big) \\
&= (d^\perp - 1) \sum_{i=1}^{k} \lambda_i - (d^\perp - 2) \sum_{i=1}^{k} \lambda_{i,\{i\}}.
\end{aligned} \tag{6}$$

由于 G 没有全零列，因此有 $d^\perp \ge 2$。因此，利用 $\lambda_{i\text{这一事实，}\{i\}} \le \min\{\lambda_i, 1\}$对所有 i均成立，我们还可以得出，(6) 的右侧至少为

$$(d^\perp - 1) \sum_{i=1}^{k} \lambda_i - (d^\perp - 2) \sum_{i=1}^{k} \min\{\lambda_i, 1\} = \sum_{i=1}^{k} \Big(\min\{\lambda_i, 1\} + (d^\perp - 1) \max\{0, \lambda_i - 1\} \Big),$$

结合（5），即可得出该结论。 □

注 5.4。 由 [4] 可知，定理 5.3 的对偶距离界在 G 为系统 MDS 矩阵且 $n \geq 2k$时是紧的；MDS 矩阵的定义见附录 B 。系统矩阵 $G \in \mathbb{F}_q^{k\times n}$若生成 MDS 码且满足$n < 2k$，其界也可能是紧的。例如，例 G 中的矩阵 4.14就是这种情况。

结果表明，定理 5.3 对于主要实现复制的系统并不特别有效，也就是说，对于矩阵 G 与复制矩阵非常相似的情况。通过考察每个对象对应的系统节点数量，我们得到以下结果。由于证明与定理 5.3的证明类似，这里省略。

定理 5.5。假设 G 是系统的，并令 s_i 表示第 i个对象的系统节点数量，其中 $i \in \{1,\dots,k\}$。若 $(\lambda_1, \dots, \lambda_k) \in \Lambda(G)$，则

$$\sum_{i=1}^{k}\Big(\min\{\lambda_i, s_i\} + 2\max\{0, \lambda_i - s_i\}\Big) \leq n.$$

定理 5.3 和 5.5 给出的外界通常不可比较，以下示例对此作了说明。

例 5.6。 例 5.3 优于定理 5.5 的一个例子由例 4.14中的区域给出。根据备注 5.4，定理 5.3 给出的界正好是服务速率区域。因此，当 $(d^\perp - 1) = 3 > 2$时，它自然优于定理 5.5 给出的界。现在考虑例 1.11 中展示于图 1b的服务速率区域。定理 5.3 和 5.5 给出的界多胞体如图 3所示，这表明在这种情况下，定理 5.5 优于定理 5.3。

▲

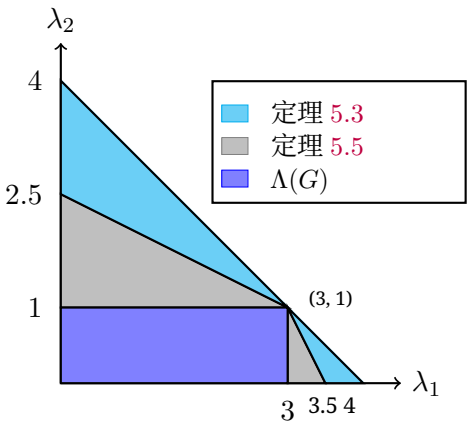


图 3：示例 1.11中服务速率区域的图示 1b，这是一个定理 5.5 给出的界比定理 5.3给出的界具有更好近似效果的示例。

下一个结果是定理5.3 和5.5之间的混合结果，因为它同时考虑了这两者：由 G 生成的码的对偶码的最小距离以及系统节点的数量。

定理 5.7. 设 $d^\perp$ 为由 G 生成的码的对偶码的最小距离，并令 s_i表示第 i 个对象的系统节点数，其中 $i \in \{1,\dots,k\}$。对于所有 $(\lambda_1, \dots, \lambda_k) \in \Lambda(G)$，我们有

$$\sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0}}\Big(\min\{s_i, \lambda_i\} + \max\{2, d^\perp - 1\}\max\{0, \lambda_i - s_i\}\Big) + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i = 0}} 2\lambda_i \leq n.$$

证明。 令 $(\lambda_1,\dots,\lambda_k) \in \Lambda(G)$，并令 $\{\lambda_{i,R}\}$ 为相应的可行分配。记 $\mathcal{R}$ 为$\mathcal{R}^{\min}(G)$。根据命题 5.2 以及以下事实：若 $d^\perp = 2$ 中存在两列线性相关，则 G 我们得到

$$\sum_{i\in\{1,\dots,k\}} \sum_{R\in\mathcal{R}} |R|\lambda_{i,R} \geq \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0}} \left(\sum_{\substack{R\in\mathcal{R}\\ |R|=1}} \lambda_{i,R} + \max\{2, d^\perp - 1\} \sum_{\substack{R\in\mathcal{R}\\ |R|\neq 1}} \lambda_{i,R} \right) + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i=0}} \sum_{R\in\mathcal{R}} 2\lambda_{i,R}.$$

因此，

$$\begin{aligned}
&\sum_{i\in\{1,\dots,k\}} \sum_{R\in\mathcal{R}} |R|\lambda_{i,R} \\
&\quad\geq \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0}} \sum_{\substack{R\in\mathcal{R}\\ |R|=1}} \lambda_{i,R} + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0}} \left(\max\{2, d^\perp - 1\} \left(\lambda_i - \sum_{\substack{R\in\mathcal{R}\\ |R|=1}} \lambda_{i,R} \right) \right) + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i=0}} 2\lambda_i \\
&\quad\geq \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0}} \max\{2, d^\perp - 1\}\lambda_i - \Big(\max\{2, d^\perp - 1\} - 1\Big) \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0}} \min\{s_i, \lambda_i\} + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i=0}} 2\lambda_i \\
&\quad= \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0}} \Big(\min\{s_i, \lambda_i\} + \max\{2, d^\perp - 1\} \left(\lambda_i - \min\{s_i, \lambda_i\}\right)\Big) + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i=0}} 2\lambda_i \\
&\quad= \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0}} \Big(\min\{s_i, \lambda_i\} + \max\{2, d^\perp - 1\} \max\{0, \lambda_i - s_i\}\Big) + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i=0}} 2\lambda_i,
\end{aligned}$$

其中，第一个不等式由定义 1.4 得出，第二个不等式则由以下不等式得出

$$\sum_{\substack{R\in\mathcal{R}\\ |R|=1}} \lambda_{i,R} \leq \min\{s_i, \lambda_i\}.$$

□

值得注意的是，如果 $d^\perp - 1 \geq 2$（因此 $s_i = 1$ 对所有 $i \in \{1,\dots,k\}$均成立），那么定理 5.7 就给出了定理 5.3。类似地，如果 $s_i \neq 0$ 对所有 $i \in \{1,\dots,k\}$ 均成立，且 $s_i \geq 2$ 对至少一个 i成立，那么定理 5.7 就变为定理 5.5。

引理5.1 表明，恢复集大小的多样性在塑造服务速率区域方面发挥着重要作用。考虑到满足恢复集大小均相同的索引 i，我们得到了以下结果。请注意，这恰好衡量了恢复集大小相同的索引所作出的贡献，因此改进了定理 5.7。证明是定理 5.7的简单扩展，此处从略。

定理 5.8. 设 $d^\perp$ 为由 G 生成的码的对偶码的最小距离，并设 s_i 表示第 i 个对象的系统节点数量，其中 $1 \in \{1,\dots,k\}$。设

$$\mu_i = \frac{1}{|\mathcal{R}_i^{\min}(G)| - s_i} \sum_{\substack{R\in\mathcal{R}_i^{\min}(G)\\ |R|\neq 1}} |R| \quad \text{for } i \in \{1,\dots,k\},$$

$$J = \{i \in \{1,\dots,k\} \mid \text{所有满足 } |R| \neq 1 \text{ 的 } R \in \mathcal{R}_i^{\min}(G) \text{ 都具有相同的基数}\}.$$

因此，对于所有 $\lambda \in \Lambda(G)$，我们有

$$n \geq \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0,\ i\notin J}} \Big(\min\{s_i,\lambda_i\} + \max\{2, d^{\perp}-1\}\max\{0,\lambda_i - s_i\}\Big) + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i=0,\ i\notin J}} 2\lambda_i$$
$$+ \sum_{\substack{i\in\{1,\dots,k\}\\ s_i=0,\ i\in J}} \mu_i\lambda_i + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0,\ |\mathcal{R}_i^{\min}(G)|=s_i}} s_i + \sum_{\substack{i\in\{1,\dots,k\}\\ s_i\neq 0,\ i\in J}} \left(\mu_i\lambda_i - (1-\mu_i)\min\{s_i,\lambda_i\}\right).$$

在下一个例子中，我们将说明定理 5.8 对于某些服务速率区域而言，可能比定理 5.7 更为精确。

例 5.9。设 $k = 3$、$n = 6$、$q = 3$，以及

$$G := \begin{pmatrix} 0 & 1 & 1 & 2 & 1 & 2 \\ 1 & 2 & 2 & 2 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 & 2 \end{pmatrix} \in \mathbb{F}_3^{3\times 6}.$$

沿用定理 5.8 的记号，我们有 $d^{\perp} = 2$、$(\mu_1, \mu_2, \mu_3) = (2, 3, 11/4)$、$(s_1, s_2, s_3) = (0, 1, 0)$，以及 $J = \{1, 2\}$。图 4 描绘了服务速率区域 $\Lambda(G)$，以及定理 5.7 和 5.8 给出的外界。▲

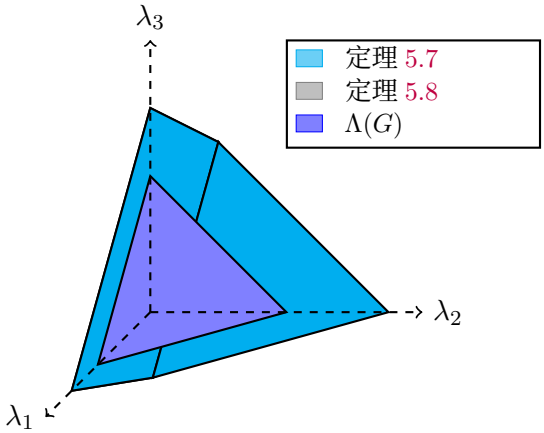


图 4：例 5.9的服务速率区域和外界。

5.2 优化方法

在本小节中，我们运用背包多面体理论（回顾定义 4.9），推导分配与服务速率区域多面体的外界及其界的推论。尤为重要的是，对于任意给定的指标集 $\sum_{i\in I}\lambda_i$，我们得到了量 $I \subseteq \{1,\dots,k\}$ 的上界。这些量在实践中具有重要意义，因为它们对应于对某些存储对象子集感兴趣的用户累计数量。请注意，当 $I = \{1,\dots,k\}$ 时，$\sum_{i\in I}\lambda_i$ 表示系统中的用户总数。本小节还通过一些示例说明如何应用这些界，并绘制输出结果。

记号 5.10。设 $\mathcal{R}$ 是一个 G-系统，且 $m(\mathcal{R}) = |\mathcal{R}_1| + \ldots + |\mathcal{R}_k|$。我们定义整数向量

$$y(\mathcal{R}) = (|R| : i \in \{1,\dots,k\},\ R \in \mathcal{R}_i) \in \mathbb{Z}_{\geq 0}^{m(\mathcal{R})},$$

其中我们采用与定义 2.1相同的顺序。注意，$y(\mathcal{R})$ 的所有分量均为正，因为根据定义，恢复集非空。

现在我们可以陈述本节的主要结果：它给出了无穷多个包含分配多胞体的半空间，每个向量 $c \in \mathbb{R}^{m(\mathcal{R})}$ 对应一个半空间。

定理 5.11。设 $\mathcal{R}$ 是一个 G-系统，$m = m(\mathcal{R})$，并设 $c \in \mathbb{R}^m$。定义 $y = y(\mathcal{R})$，并令 $\pi : \{1, \ldots, m\} \to \{1, \ldots, m\}$ 是任意一个满足下述条件的排列

$$\frac{c_{\pi(1)}}{y_{\pi(1)}} \geq \cdots \geq \frac{c_{\pi(m)}}{y_{\pi(m)}}.$$

定义 $J = \{j \mid y_{\pi(1)} + \ldots + y_{\pi(j)} > n\}$。若 $J = \emptyset$，则令 $r = m+1$、$\sigma = 0$ 和 $\pi(m+1) = 0$。若 $J \neq \emptyset$，则令 $r = \min(J)$ 和 $\sigma = \left(n - \sum_{j=1}^{r-1} y_{\pi(j)}\right) / y_{\pi(r)}$。那么，对于任意 $x \in \mathcal{A}(\mathcal{R})$，我们有

$$cx^\top \leq \left(\sum_{j=1}^{r-1} c_{\pi(j)} \right) + c_{\pi(r)}\, \sigma, \quad \text{where} \quad c_{\pi(m+1)} = 0. \tag{7}$$

在证明该定理之前，我们先给出服务速率区域的最大和容量的一个直接推论。取 $c = (1, \ldots, 1)$ 即可得到该结果，这使我们能够使用更简洁的记号。

推论 5.12. 设 $\mathcal{R}$ 是任意一个满足 $\Lambda(\mathcal{R}) = \Lambda(G)$ 的 G-系统。设 $y = y(\mathcal{R})$，并将其分量按非降序重新排列，得到向量 $\hat{y}$。假设 $\hat{y}_1 + \ldots + \hat{y}_m > n$，其中 $m = m(\mathcal{R})$，并令 $r = \min\{j \mid \hat{y}_1 + \ldots + \hat{y}_j > n\}$。我们有

$$\lambda(G) \leq r - 1 + \frac{n - \sum_{j=1}^{r-1} \hat{y}_j}{\hat{y}_r}.$$

我们举一个例子来说明如何应用推论 5.12。

例 5.13. 设 G 与例 4.14 中相同，其中 $n = 4$，且 $\mathcal{R} = \mathcal{R}^{\min}(G)$。我们有 $y = y(\mathcal{R}) = (1,3,1,3,1,3)$。与推论 5.12 中一样，我们构造 $\hat{y} = (1, 1, 1, 3, 3, 3)$，并得到 $\lambda(G) \leq 10/3$。可以验证 $\lambda(G) = 3$。▲

定理 *5.11* 的证明。令 $\mathcal{P} = \{x \in [0,1]^{m(\mathcal{R})} \mid y(\mathcal{R})x^\top \leq n\}$，它是一个松弛背包多胞体。令 $\beta = \max\{cx^\top \mid x \in \mathcal{P}\}$。我们有包含关系 $\mathcal{A}(\mathcal{R}) \subseteq \mathcal{P}$，因此 $\max\{cx^\top \mid x \in \mathcal{A}(\mathcal{R})\} \leq \beta$。我们应用 Dantzig 的一个经典结果 [9]（当 $J = \emptyset$ 时需要单独处理，但这里省略），该结果指出，达到最大值 $x \in \mathcal{P}$ 的点 β 由下式给出：

$$\hat{x}_j = \begin{cases} 1 & \text{if } 1 \leq j \leq r-1, \\ \dfrac{n - \sum_{j=1}^{r-1} y_{\pi(j)}}{y_{\pi(r)}} & \text{if } j = r, \\ 0 & \text{otherwise.} \end{cases}$$

通过计算 $\mu = c\hat{x}^\top$ 即可得出结果。 □

作为定理 5.11 的另一个推论，我们得到一个结果，它给出了无限多个半空间，服务速率区域 $\Lambda(\mathcal{R})$ 包含于其中。通过在陈述中选择不同的向量 b，即可得到每个半空间。

推论 5.14。设 $\mathcal{R}$ 是一个 G-系统，且设 $b \in \mathbb{R}^k$。令 $m = m(\mathcal{R})$，$m_0 = 1$，并且 $m_i = |\mathcal{R}_i|$ 对所有 $i \in \{1, \ldots, k\}$ 均成立。通过令 $c \in \mathbb{R}^{m(\mathcal{R})}$ 满足 $c_j = b_i$（当 $m_{i-1}+1 \leq j \leq m_i$ 时），构造一个排列 π，并按定理 r，σ，以及必要时的 $\pi(m+1)$，如定理 5.11 所示。于是，对所有 $\lambda \in \Lambda(\mathcal{R})$，均有

$$b\lambda^\top \leq \left(\sum_{j=1}^{r-1} c_{\pi(j)} \right) + c_{\pi(r)}\, \sigma.$$

将前述结果专门应用于向量 $b \in \{0,1\}^k$，可以得到形如 $\sum_{i\in I} \lambda_i$的部分和上界，其中 $I \subseteq \{1,\dots,k\}$ 且 $\lambda \in \Lambda(\mathcal{R})$。特别地，可以得到最大和容量 $\lambda(G$的上界。本节最后，我们将说明如何应用推论 5.14 以及它能得到何种结果。

例 5.15。 考虑例 1.11中的服务速率区域，该区域如图 1b所示。对所有 $b \in \{0,1\}^3$ 应用推论 5.14，我们得到服务速率区域的界多胞体，并将其绘于图 5 中，同时给出 $\Lambda(G)$. ▲

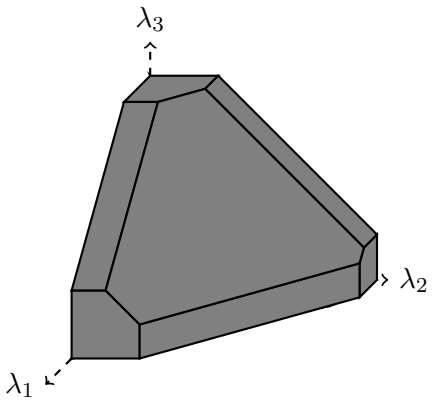


图 5：例 5.15的服务速率区域及界多胞体。

下面的例子表明，对非 0-1 向量应用推论 5.14 ，不仅可以得到更好的界，而且比仅对 0-1 向量应用该推论得到的界严格更优。

例 5.16。设

$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 2 \end{pmatrix} \in \mathbb{F}_3^{2\times 8}.$$

服务速率区域 $\Lambda(G)$ 是图 6中的紫色区域。对所有$b \in \{0,1\}^2$应用推论 5.14，可得到图 6中的浅蓝色区域。为了更好地近似服务速率区域，图 6中还绘出了相应结果，我们可以使用推论 5.14。例如，除 0-1 向量 $b \in \{0,1\}^2$外，对 $b = (3,2)$ 和 $b = (3,5)$ 应用上述推论，可得到图 6中的灰色区域。 ▲

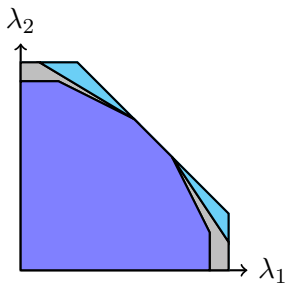


图 6：示例5.16的服务速率区域和界多胞体。

6 系统的 MDS 码

本节将全面讨论服务速率区域 $\Lambda(G)$，其中 G 是一个 MDS 矩阵；MDS 矩阵的定义见附录B 。我们重点关注这些服务速率区域在 $k \in \{2,3\}$时的体积，以及它们的最大和容量。

回顾一下，当 G 是一个系统 MDS 矩阵且 $n \geq 2k$时，服务速率区域 $\Lambda(G)$ 已知，其定义为集合

$$\left\{(\lambda_1, \ldots, \lambda_k) \in \mathbb{R}_{\geq 0}^k \;\middle|\; \sum_{i=1}^{k} \Big(\min\{\lambda_i, 1\} + k \cdot \max\{0, \lambda_i - 1\}\Big) \leq n \right\}; \tag{8}$$

参见备注 5.4。（8）中的描述不便于计算 $\Lambda(G)$ 的体积，而这正是本节的目标之一。因此，我们首先推导一种更便于使用的描述。

记号 6.1. 给定一个向量 $\lambda \in \mathbb{R}^k$，令 $\chi(\lambda) \in \mathbb{Z}^k$ 表示满足以下条件的向量：若 $\lambda_i < 1$，则为 $\chi(\lambda)_i = 1$；若 $\lambda_i \geq 1$，则为 $\chi(\lambda)_i = k$。此外，令

$$\lambda_{<1} = \{i \in \{1, \ldots, k\} \mid \lambda_i < 1\}.$$

下面的引理给出了系统 MDS 矩阵的服务速率区域的另一种表示 G，其中 $n \geq 2k$。在引理陈述中，我们使用记号 6.1。

引理 6.2. 假设 G 是一个系统 MDS 矩阵，且 $n \geq 2k$。则有

$$\Lambda(G) = \left\{\lambda \in \mathbb{R}_{\geq 0}^k \;\middle|\; \chi(\lambda)\, \lambda^\top \leq n + (k-1)(k - |\lambda_{<1}|)\right\}. \tag{9}$$

证明。令 $\lambda \in \Lambda(G)$。由（8），可得 $\sum_{i=1}^{k}\Big(\text{最小值}\{\lambda_i, 1\} + k \cdot \text{最大值}\{0, \lambda_i - 1\}\Big) \leq n$. 也就是说，

$$\sum_{i \in \lambda_{<1}} \lambda_i + (k - |\lambda_{<1}|) + \left(\sum_{i \in \{1,\ldots,k\} \setminus \lambda_{<1}} k\lambda_i\right) - k(k - |\lambda_{<1}|) \leq n.$$

后一个不等式可以改写为 $\chi(\lambda) \cdot \lambda^\top \leq n + (k-1)(k - |\lambda_{<1}|)$。这表明 $\subseteq$ 包含于（9）中。通过反向进行所有推导即可得到另一个包含关系，细节从略。

□

现在我们可以计算 $k \in \{2, 3\}$和 $n \geq 2k$时 MDS 矩阵的服务速率区域的体积。我们先处理 $k = 2$ 的情形。

定理 6.3。设 $G \in \mathbb{F}_q^{2\times n}$是一个系统 MDS 矩阵。假设 $n \geq 4$。则有 $\text{vol}(\Lambda(G)) = \frac{n^2+4n}{8}$。

证明。根据引理 6.2，$\Lambda(G)$ 由以下五个方程定义：

$$\lambda_1 + \lambda_2 \leq \frac{n+2}{2}, \qquad 2\lambda_1 + \lambda_2 \leq n+1, \qquad \lambda_1 + 2\lambda_2 \leq n+1, \qquad \lambda_1 \geq 0, \qquad \lambda_2 \geq 0.$$

不难验证，$\Lambda(G)$的顶点为以下各点

$$(0,0), \qquad \left(1, \frac{n}{2}\right), \qquad \left(\frac{n}{2}, 1\right), \qquad \left(\frac{n+1}{2}, 0\right), \qquad \left(0, \frac{n+1}{2}\right).$$

现在可以使用初等方法计算体积（即面积）。 □

我们可以将定理 6.3 与由具有相同参数的矩阵生成的复制系统进行比较。

命题 6.4。假设 $G \in \mathbb{F}_q^{k\times n}$ 是一个复制矩阵。我们有

$$n - k + 1 \leq \text{vol}(\Lambda(G)) \leq \left\lfloor \left(\frac{n}{k}\right)^k \right\rfloor.$$

下界可由某个 G 达到，而上界可由某个 G 达到（若 $k=2$）。

证明。令 j_i 表示 G 中作为标准基向量 e_i倍数的列数，其中$i \in \{1,\dots,k\}$. 每个 j_i 都是正整数，因为 G 是满秩的。由命题4.8可知，$\Lambda(G)$ 的体积为 $\prod_{i=1}^{k} j_i$. 。由算术-几何平均不等式可得所需的上界。令 $j_1 = n-k+1$ 且对 $j_i = 1$ 的 $i \in \{2,\dots,k\}$，即可达到下界。当 $k=2$时，令 $j_1 \in \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$ 且令 $j_2 = n - j_1$，即可达到上界。 □

注意，命题6.4 表明，可以为 $G \in \mathbb{F}_q^{2\times n}$找到一个复制矩阵 $n \geq 5$，其服务速率区域的体积严格大于同等大小的 MDS 矩阵的体积。

现在转而讨论 $k=3$ 和 $n \geq 6$ 的情形，计算对应于 MDS 矩阵 $G \in \mathbb{F}_q^{3\times n}$ 的服务速率区域的体积。该计算比二维情形更为复杂。

定理 6.5。设 $G \in \mathbb{F}_q^{3\times n}$ 是一个系统 MDS 矩阵，并假设 $n \geq 6$。则有

$$\mathrm{vol}(\Lambda(G)) = \frac{n^3 + 18n^2 + 54n - 18}{162}.$$

证明。定义函数 $f(z)$ =最小值$\{z,1\}$ + 3 最大值$\{0, z-1\}$。利用引理 6.2 可以看出，$\Lambda(G)$ 是满足以下不等式的三元组 $(\lambda_1, \lambda_2, z$的集合：

$$\begin{cases} \lambda_1 + \lambda_2 \leq n - f(z) & (*) \\ 3\lambda_1 + \lambda_2 \leq n - f(z) + 2 & \\ \lambda_1 + 3\lambda_2 \leq n - f(z) + 2 & \\ 3\lambda_1 + 3\lambda_2 \leq n - f(z) + 4 & (**) \\ \lambda_1, \lambda_2, z \geq 0 & \end{cases}$$

此外请注意，z 可以取到的最大值为 $(n+2)/3$。在上述系统中取 $\lambda_1 = \lambda_2 = 0$ 即可达到该值。于是有

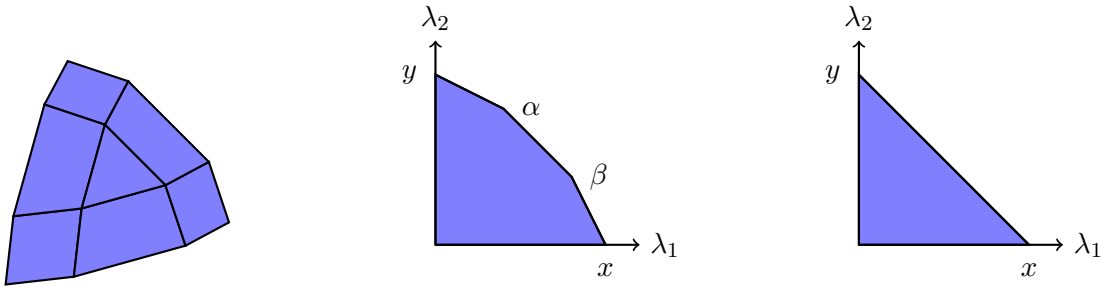
$$f(z) = \begin{cases} z & \text{if } 0 \leq z \leq 1, \\ 3z-2 & \text{if } 1 \leq z. \end{cases}$$

可以验证，对于 $z \leq n/3$，(∗∗) 比 (∗) 更严格；而在其他情况下，(∗) 比 (∗∗) 更严格。此外，当 $z \geq (n+1)/3$ 时，除了 (∗) 以及 λ_1, λ_2 和 z 的非负性之外，所有不等式都可以忽略。这说明了给定 z值时 $\Lambda(G)$ 的“切片”形状。我们在下表和图 7 中总结这一讨论。

表 1

z	图	y	x	α	β
$0 \leq z \leq 1$	7b	$(0, \frac{n-z+2}{3})$	$(0, \frac{n-z+2}{3})$	$(1, \frac{n-z+1}{3})$	$(\frac{n-z+1}{3}, 1)$
$1 \leq z \leq \frac{n}{3}$	7b	$(0, \frac{n-3z+4}{3})$	$(0, \frac{n-3z+4}{3})$	$(1, \frac{n-3z+3}{3})$	$(\frac{n-3z+3}{3}, 1)$
$\frac{n}{3} \leq z \leq \frac{n+1}{3}$	7b	$(0, \frac{n-3z+4}{3})$	$(0, \frac{n-3z+4}{3})$	$(n-3z+1, 1)$	$(1, n-3z+1)$
$\frac{n+1}{3} \leq z \leq \frac{n+2}{3}$	7c	$(0, n-3z+2)$	$(n-3z+2, 0)$	—	—

切片的面积很容易计算，因此可以通过对 z 积分来计算 $\Lambda(G)$ 的体积。例如，[5, 可从数学上证明这一方法的合理性。



(a) 系统的典型服务速率区域，属于系统的 MDS 矩阵 $G \in \mathbb{F}_q^{3\times n}$ 适用于 $n \geq 6$。

(b) 服务速率区域的一个截面属于系统的 MDS 矩阵矩阵 $G \in \mathbb{F}_q^{3\times n}$ $n \geq 6$，对于 $z \leq (n+1)/3$。

（c）系统的 MDS 矩阵 $G \in \mathbb{F}_q^{3\times n}$ 的服务速率区域切片 $n \geq 6$，对于 $(n+1)/3 \leq z \leq (n+2)/3$。

图 7：用于证明定理 6.5 的服务速率区域及其切片。

定理 2.7]。所需公式由此得出

$$\begin{aligned}\mathrm{vol}(\Lambda(G)) = \int_0^1 \Big(\frac{1}{18}z^2 - \frac{n+4}{9}z + \frac{n^2+8n+4}{18}\Big)\,dz + \int_1^{n/3} \Big(\frac{1}{2}z^2 - \frac{n+6}{3}z + \frac{n^2+12n+24}{18}\Big)\,dz \\ + \int_{n/3}^{(n+1)/3} \Big(-\frac{3}{2}z^2 + (n-2)z + \frac{n^2-4n-8}{6}\Big)\,dz \\ + \int_{(n+1)/3}^{(n+2)/3} \Big(\frac{9}{2}z^2 - (3n+6)z + \frac{n^2+4n+4}{2}\Big)\,dz\end{aligned}$$

以及繁琐但直接的计算。 □

出于好奇，我们指出，定理 6.5 也可以通过著名的三角剖分方法推导得出，该方法利用单形的体积计算多胞体的体积；参见 [8]。对于定理 6.5 中的多胞体，这种方法的形式化过程相当复杂，因此我们采用了积分方法。

我们还注意到，当 G 是一个 MDS 矩阵时，命题 B.3可以给出 $\mathrm{vol}(\Lambda(G))$ 的一般下界。任意 k 列都可用于恢复任意数据对象。因此，$(n/k)e_i \in \Lambda(G)$，这意味着以这些顶点构成的单形包含在服务速率区域中。因此，$\mathrm{vol}(\Lambda(G)) \geq (n/k)^k/k!$ 。

在本节的第二部分，我们研究系统 MDS 矩阵的服务速率区域的其他参数。首先注意到，推论 5.12 蕴含如下结论。

C推论 6.6. 设 $G \in \mathbb{F}_q^{k\times n}$ 是一个系统 MDS 矩阵。我们有 $\lambda(G) \leq k + {}^{n-}\ \frac{k}{k}$.

证明。 按照推论 5.12中的记号，我们有 $m = k\left(\binom{n-1}{k} + 1\right)$ 且 $\hat{y} = (v_1, v_2)$，其中 $v_1 = (1, \ldots, 1) \in \mathbb{R}^k$ 且 $v_2 = (k, \ldots, k) \in \mathbb{R}^{m-k}$。此外，

$$r = \min\{j \mid \hat{y}_1 + \ldots + \hat{y}_j > n,\ 1 \leq j \leq m\} = k + \lfloor (n-k)/k \rfloor + 1.$$

然后，应用推论 5.12，即可得到所需结果：

$$\lambda(G) \leq k + \lfloor (n-k)/k \rfloor + \frac{n - \sum_{i=1}^{r-1} \hat{y}_i}{\hat{y}_r} = k + \frac{n-k}{k}.$$

□

现在我们将证明，在 6.6 的情形下，系统 MDS 矩阵以等号达到推论 $n \geq 2k$给出的界。首先，我们引入一些用于证明该结果的对象。

记号 6.7。设 $a,b\in\mathbb{Z}$ 满足 $2\le a<b$。假设 $q>b$，并令 α 为 $\mathbb{F}_q$ 的一个本原元。定义矩阵

$$G_\alpha^{a,b}=\begin{pmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{b-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & \alpha^{(a-1)} & \alpha^{2(a-1)} & \cdots & \alpha^{(b-1)(a-1)}\end{pmatrix}\in\mathbb{F}_q^{a\times b}.$$

注意，$q>b$ 是 G^{a,α^b}的各列两两不同的充要条件。充分性可由考察 G^{a,α^b}的第二行，以及 α 的乘法阶为 $q-1$这一事实得出。必要性则源于：若 $q\le b$，那么 G^{a,α^b}中至少有一列由 $\{2,\dots,b\}$ 索引，且该列等于第一列。此外，矩阵G^{a,α^b}是一个 Reed-Solomon码 [24], 的生成矩阵，而 Reed-Solomon 码是一种 MDS 码；参见 [21]。

引理 6.8。按照记号 6.7，令 $\mathcal{R}=\mathcal{R}^{\min}(G^{a,\alpha^b})$。以下结论成立。

1. 设 $i\in\{1,\dots,a\}$ 和 $R\subseteq\{1,\dots,b\}$。那么，$R\in\mathcal{R}_i$ 当且仅当 $|R|=a$。
2. 设 $\nu\in\{1,\dots,b\}$。我们有

$$|\{R\in\mathcal{R}_i\mid i\in\{1,\dots,a\},\,\nu\in R\}|=\binom{b-1}{a-1}.$$

证明。我们首 先注意到，引理的第二部分由第一部分以及以下事实可得

$$\{S\subseteq\{1,\dots,b\}\mid |S|=a,\ \nu\in S\}=\binom{b-1}{a-1}\ \text{for all } 1\le\nu\le b.$$

为证明第一部分，令 $G=G^{a,\alpha^b}$。假设 $R\in\mathcal{R}_i$ 并证明 $|R|=a$。我们首先证明 $|R|\le a$。反设 $|R|>a$. 根据命题 B.3，存在 $R'\subseteq R$ 使得 $|R|=a$ 且 $R'\in\mathcal{R}_i$，这与 R 是 i-极小的事实矛盾。现在我们证明 $|R|\ge a$。反设 $|R|=c<a$. 由于 G的结构，不失一般性可设 $i=a$ 且 $R=\{1,\dots,c\}$。我们将证明$R\notin\mathcal{R}_a$，这就产生了矛盾。也就是说，只需证明 $e_a\notin\langle G^\nu\mid\nu\in\{1,\dots,c\}\rangle$，这可由矩阵

$$\begin{pmatrix}1 & 1 & 1 & \cdots & 1\\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{c-1}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & \alpha^{(a-1)} & \alpha^{2(a-1)} & \cdots & \alpha^{(c-1)(a-1)}\end{pmatrix}\ \text{and}\ \begin{pmatrix}1 & 1 & 1 & \cdots & 1 & 0\\ 1 & \alpha & \alpha^2 & \cdots & \alpha^{c-1} & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 1 & \alpha^{(a-1)} & \alpha^{2(a-1)} & \cdots & \alpha^{(c-1)(a-1)} & 1\end{pmatrix}$$

具有不同的秩。反方向，假设 $|R|=a$。由于 G 是一个 MDS 矩阵，我们由命题$R\in\mathcal{R}_i^{\text{all}}$（$G$）得到B.3。为了说明 $R\in\mathcal{R}_i$，只需证明 $\mathcal{R}_i$ 的所有元素的基数均为 a，这一点已在证明的第一部分中得到证明。

□

注意，前一引理的第二个性质表明，G^{a,α^b}的每个列索引都参与系统 $\mathcal{R}$的 $\binom{b-1}{a-1}$ 个恢复集 G^{a,α^b}。

定理 6.9。设 $(\lambda_1,\dots,\lambda_a)\in\mathbb{R}^a_{\ge0}$。按照记号 6.7，若 $\lambda_1+\cdots+\lambda_a\le b/a$，则 $(\lambda_1,\dots,\lambda_a)\in\Lambda(G^{a,\alpha^b})$。

证明。令 $\mathcal{R}=\mathcal{R}^{\min}(G^{a,\alpha^b})$。对于所有 $i\in\{1,\dots,a\}$ 和 $R\in\mathcal{R}_i$，令

$$\lambda_{i,R}=\frac{\lambda_i}{\binom{b}{a}}=\lambda_i\,\frac{a}{b}\,\frac{1}{\binom{b-1}{a-1}}.$$

现在我们证明定义 1.4 中的约束 (3) 按定义成立。约束 (1) 得以满足，因为

$$\sum_{R\in\mathcal{R}_i}\lambda_{i,R}=\lambda_i\frac{a}{b}\frac{1}{\binom{b-1}{a-1}}|\mathcal{R}_i|=\lambda_i\frac{a}{b}\frac{1}{\binom{b-1}{a-1}}\binom{b}{a}=\lambda_i$$

对于所有 $i\in\{1,\ldots,a\}$，其中 $|\mathcal{R}_i|=\binom{b}{a}$ 这一事实由引理 6.8的第一部分可得。根据引理6.8的第二部分，约束条件 (2) 可写为

$$\binom{b-1}{a-1}\sum_{i=1}^{a}\lambda_i\frac{a}{b}\frac{1}{\binom{b-1}{a-1}}\leq 1, \tag{10}$$

根据定理的假设 $\lambda_1+\ldots+\lambda_a\leq b/a$，该结论成立，因为

$$\binom{b-1}{a-1}\sum_{i=1}^{a}\lambda_i\frac{a}{b}\frac{1}{\binom{b-1}{a-1}}=\frac{a}{b}(\lambda_1+\ldots+\lambda_a)\leq\frac{a}{b}\frac{b}{a}=1.$$

□

现在我们可以证明，满足 $n\geq 2k$ 的系统 MDS 矩阵达到了推论 6.6(参见 [2])。

定理 6.10. 假设 $n\geq 2k$。若 $G\in\mathbb{F}_q^{k\times n}$是一个系统 MDS 矩阵，则 $\lambda(G)=k+\frac{n-k}{k}$。

证明。 所有满足 $n\geq 2k$ 的系统 MDS 矩阵具有相同的服务速率区域；参见注5.4。因此，只需证明 $G=\left[\mathrm{Id}_k\mid G^{k,n-k}_{\alpha}\right]\in\mathbb{F}_q^{k\times n}$，其中 Id_k 是 $k\times k$ 上的单位矩阵， $\mathbb{F}_q$ 且 $G^{k,n-k}_{\alpha}$如记号 6.7中所述。

事实 $\lambda_1+\cdots+\lambda_k\leq k+(n-k)/k$ 由推论 6.6 得出。令 $\lambda_{i,\{i\}}=1$ 对所有 $i\in\{1,\ldots,k\}$ 均成立，并注意到，根据定理 6.9 取 $a=k$ 和 $b=n-k$，可得 $((n-k)/k,0,\ldots,0)\in\Lambda(G^{k,n-k}_{\alpha})$。那么，根据 G 的定义，$((n-k)/k+1,1,\ldots,1)\in\Lambda(G)$。 □

我们通过指出推论 6.6 在以下情况下未必取等号，来结束本节： $n<2k$，或者 G 不是系统的。对于 $n<2k$的情况，例如参见例 5.13。对于 $n\geq 2k$ 且 G 是非系统 MDS 矩阵的情况，考虑

$$G=\begin{pmatrix}2&1&3&4\\1&2&3&5\end{pmatrix}\in\mathbb{F}_7^{2\times 4}.$$

于是 $\lambda(G)=2<3$。

致谢

作者感谢 Laura Sanit`a 就组合优化和向下单调多胞体展开富有成效的讨论，也感谢匿名审稿人提出的意见和建议。

参考文献

[1] M. Akta¸s、S. E. Anderson、A. Johnston、G. Joshi、S. Kadhe、G. L. Matthews、C. Mayer 和 E. Soljanin。论访问擦除编码内容的服务容量区域。载于第55届 Allerton 通信、控制与计算年会，2017年。

[2] M. Aktas、G. Joshi、S. Kadhe、F. Kazemi 和 E. Soljanin。服务速率区域：编码分布式系统设计的一个新方面。《IEEE信息论汇刊》，67(12):7940– 7963，2022年。[3] G. N. Alfarano、A. Ravagnani 和 E. Soljanin。关于多个并发（局部）数据恢复的对偶码界。载于IEEE国际信息论研讨会，2022年。[4] S. E. Anderson、A. Johnston、G. Joshi、G. L. Matthews、C. Mayer 和 E. Soljanin。从擦除编码存储中访问内容的服务容量区域。载于IEEE信息论研讨会，2018年。[5] T. M. Apostol。《微积分》，第1卷。John Wiley & Sons，1991年。[6] D. L. Barrow 和 P. W. Smith。将样条记法应用于体积问题。《美国数学月刊》，86(1):50–51，1979年。[7] R. Bhatia 和 C. Davis。一个更优的方差界。《美国数学月刊》，107(4):353–357，2000年。[8] J. Cohen 和 T. Hickey。确定凸多面体体积的两种算法。《ACM学报》，26(3):401–414，1979年。[9] G. B. Dantzig。离散变量极值问题。《运筹学》，5(2):266–288，1957年。[10] A. G. Dimakis、K. Ramchandran、Y. Wu 和 C. Suh。分布式存储网络码综述。《IEEE会刊》，99(3):476–489，2011年。[11] A. Fazeli、A. Vardy 和 E. Yaakobi。低存储开销的分布式PIR编码。载于IEEE国际信息论研讨会，第2852–2856页，2015年。[12] A. Fazeli、A. Vardy 和 E. Yaakobi。低存储开销的PIR：以编码取代复制。arXiv预印本，1505.06241，2015年。[13] Z. F¨uredi 和 I. B´ar´any。计算体积很困难。载于第18届 ACM 年度计算理论研讨会论文集，1986年。[14] B. Gr¨unbaum、V. Klee、M. A. Perles 和 G. C. Shephard。《凸多胞体》，第16卷。Springer，1967年。[15] Y. Ishai、E. Kushilevitz、R. Ostrovsky 和 A. Sahai。批处理码及其应用。载于第36届 ACM 年度计算理论研讨会论文集，2004年。[16] F. Kazemi、E. Karimi、E. Soljanin 和 A. Sprintson。编码的服务速率问题的组合视角、其与分数匹配的等价性及其与批处理码的联系。载于IEEE国际信息论研讨会，2020年。[17] F. Kazemi、S. Kurz 和 E. Soljanin。编码的服务速率问题的几何视角及其在一阶Reed–Muller码服务速率中的应用。载于IEEE国际信息论研讨会，2020年。[18] F. Kazemi、S. Kurz 和 E. Soljanin。实现目标服务速率区域的高效存储方案。载于IEEE信息论研讨会，2021年。[19] B. Korte 和 J. Vygen。《组合优化》，第1卷。Springer，2011年。[20] R. Lang。关于凸集可测性的一则注记。《数学档案》，47:90–92，1986年。[21] J. MacWilliams 和 N. Sloane。《纠错码理论》。Elsevier，1977年。[22] J. R. Munkres。《拓扑学》。Prentice Hall，第2版，2000年。

[23] D. A. Patterson、G. Gibson 和 R. H. Katz。关于廉价磁盘冗余阵列（RAID）的论述。载于 ACM SIGMOD 会议论文集，1988 年。[24] I. S. Reed 和 G. Solomon。某些有限域上的多项式码。《工业与应用数学学会杂志》，8(2):300–304，1960年。[25] A.-E. Riet、V. Skachek 和 E. K. Thomas。由超图构造的异步批量码和 PIR 码。载于 IEEE信息论研讨会，2018年。[26] A. Schrijver。线性规划与整数规划理论。John Wiley & Sons，1998年。[27] B. Segre。有限空间中的有理正规曲线与 k-弧。《纯粹与应用数学年刊》，39(1):357–379，1955年。[28] K. Shvachko、H. Kuang、S. Radia 和 R. Chansler。Hadoop分布式文件系统。载于 IEEE大容量存储系统与技术研讨会，2010年。[29] V. Skachek。批量码和 PIR 码及其与局部可修复码的联系。《网络编码与子空间设计》，第427–442页，2018年。[30] I. Tamo 和 A. Barg。一类最优局部可恢复码。《IEEE信息论汇刊》，60(8):4661–4676，2014年。[31] M. A. Tsfasman 和 S. G. Vl˘adut。代数几何码。克鲁维尔学术出版社，1991年。[32] G. M. Ziegler。《多胞体讲义》，第152卷。Springer，2012年。

A 多胞体

在本附录中，我们汇集了一些有关多胞体及其性质的背景材料。更多详情可参阅标准参考资料，例如 [19, 26, 32]。

我们首先回顾一下，多面体是形如 $\mathcal{P} = \{x \in \mathbb{R}^m \mid Ax^\top \leq b^\top\}$ 的集合，其中 $A \in \mathbb{R}^{\ell \times m}$、$\ell$、$m \geq 1$、$b \in \mathbb{R}^\ell$ 和 $\leq$ 是逐分量应用的。如果该集合有界，则称其为多胞体。关于多面体的一个基本结果指出，每个多胞体都是有限个点的凸包。对于一个（可能是无限的）集合 $S \subseteq \mathbb{R}^m$，我们用 conv(S) 表示其凸包，其中 conv($\emptyset$) $= \emptyset$。

定理 A.1（例如参见 [14]）。设 $\mathcal{P} \subseteq \mathbb{R}^m$ 是一个多胞体。那么存在一个有限集合 $S \subseteq \mathbb{R}^m$，使得 $\mathcal{P}$ =conv(S)。

多胞体 $\mathcal{P} \subseteq \mathbb{R}^m$ 的一个顶点是满足 $v \notin$ conv($\mathcal{P} \setminus \{v\}$) 的元素 $v \in \mathcal{P}$。$\mathcal{P}$ 的顶点集合记为 vert($\mathcal{P}$)。注意，如果 $\mathcal{P} =$ conv(S) 是一个多胞体，那么 vert($\mathcal{P}$) $\subseteq S$。因此 $\mathcal{P}$ =conv(vert($\mathcal{P}$))。此外，非空多胞体至少有一个顶点。

我们回顾顶点的以下关键性质。

命题 A.2. 设 $\mathcal{P} = \{x \in \mathbb{R}^m \mid Ax^\top \leq b^\top\}$ 是一个多胞体，且 $A \in \mathbb{R}^{\ell \times m}$。设 v 是 $\mathcal{P}$的一个顶点。则存在 $I \subseteq \{1, \ldots, \ell\}$，使得 rank($A[I]$) $= m$ 且 $\{v\} = \{x \in \mathbb{R}^m \mid A[I]x^\top = b[I]^\top\}$，其中 $A[I]$ 和 $b[I]$ 由 A 和 b 分别删除未由 I索引的行和分量后得到。

前一个结果令人瞩目地表明，有理多胞体具有有理顶点（即顶点的项均为有理数）。回顾一下，形如 $\{x \in \mathbb{R}^m \mid Ax \leq b^\top\}$ 且满足 $A \in \mathbb{Q}^{\ell \times m}$和 $b \in \mathbb{Q}^\ell$ 的多面体称为有理多面体。接着，命题 A.2 结合高斯消元可得以下结果。

推论 A.3。有理多胞体的顶点均为有理数。

我们通过回顾多胞体 $\mathcal{P} \subseteq \mathbb{R}^m$ 是向下单调的这一性质来结束本附录：如果 $x \geq 0$对所有 $x \in \mathcal{P}$ 以及所有满足 $0 \leq y \leq x$ 的 $y \in \mathbb{R}^m$ 和 $x \in \mathcal{P}$ 都有 $y \in \mathcal{P}$，则称其为向下单调。本文中的所有多胞体都是向下单调的。

B 纠错码

定义 B.1。 $[n,k]_q$ （纠错）码是一个 k维的 $\mathbb{F}_q$线性子空间 $\mathcal{C} \leq \mathbb{F}_q^n$。我们称 n 为 $\mathcal{C}$的长度。若矩阵 $G \in \mathbb{F}_q^{k\times n}$的行张成 $\mathcal{C}$ ，则称其为 $\mathcal{C}$的生成矩阵。对于所有 $y \in \mathcal{C}\} \leq \mathbb{F}_q^n$， $[n,n-k]_q$码 $\mathcal{C}^\perp = \{x \in \mathbb{F}_q^n \mid xy^\top = 0$是 $\mathcal{C}$的对偶码。

$\mathcal{C}$ 的纠错能力由一个基本参数衡量，定义如下。

定义 B.2。向量 $x \in \mathbb{F}_q^n$的汉明重量是整数 $\mathrm{wt}^{\mathrm{H}}(x) = |\{i \mid x_i \neq 0\}|$。码 $\mathcal{C} \leq \mathbb{F}_q^n$的最小（汉明）距离为 $d^{\mathrm{H}}(\mathcal{C}) = \min\left\{\mathrm{wt}^{\mathrm{H}}(x) \mid x \in \mathcal{C}, x \neq 0\right\}$。

本文主要关注 $[n,k]_q$ 满足 $k+d-1=n$的码。这类码称为**MDS**（最大距离可分码）。生成 MDS 码的满秩矩阵称为MDS 矩阵。已知这类矩阵仅在规模足够大的有限域上存在（ $q \geq n-1$即可满足条件）。自 1955 年以来，确定 MDS 矩阵在哪些域大小下存在一直是一个未解问题；参见[27]。最后，我们给出 MDS 矩阵的一个实用刻画。证明见 [21, 第 318]页。

命题B.3. 设 $G \in \mathbb{F}_q^{k\times n}$为一个矩阵。那么， G 是一个 MDS 矩阵，当且仅当 G的每 k列都 $\mathbb{F}_q$-线性无关。