arXiv:2603.11040v1 [math.ST] 11 Mar 2026

相关矩阵的正定阈值化

SUJIT SAKHARAM DAMASE 和 JAMES ELDRED PASCOE

Abstract. 相关矩阵的标准阈值化技术往往会破坏半正定性。我们研究了正定函数的构造，这些函数在特定集合 $K \subseteq [-1, 1)$ 上取零，从而确保阈值化后的矩阵仍是有效的相关矩阵。我们确立了存在性结果，并类比编码理论中的 Delsarte 方法，基于归一化 Gegenbauer 展开的线性系数定义了忠实性判据，同时给出了在单点和点对上进行阈值化的界。

我们证明，对于秩为 n的相关矩阵，任何保持半正定性的软阈值化算子都必然导致特征空间发生几何塌缩，其程度可由忠实性常数的 $\mathcal{O}(1/n)$界定量化。这表明，几何上无偏的软阈值化会限制可恢复的信号。

1. Introduction

给定从某个样本得到的观测相关矩阵 $M = (m_{ij})_{ij}$，我们可能希望假设较小的元素实际上为零——也就是说，底层随机变量相互独立。在高维统计学中，阈值估计量是协方差和相关矩阵正则化的经典工具，尤其是在特征数量超过样本量时 [2, 12, 3]。

“硬阈值化”技术考虑 $f[M] = (f(m_{ij}))_{ij}$，其中当 $|x| \geq \varepsilon$ 时，$f(x) = x$；否则，$f(x) = 0$。**“软阈值化”**技术考虑 $f[M] = (f(m_{ij}))_{ij}$，其中 f 是在 $|x| \leq \varepsilon$ 上取零值的连续函数。

日期：2026年3月12日。2020 数学主题分类。 62H12, 43A35, 46E22, 15B48, 33C45

。关键词和短语。正定函数、软阈值化、Delsarte 估计。我们感谢 SERB 的 J C Bose Fellowship JCB/2021/000041（项目负责人：Tirthankar Bhattacharyya）对本研究的支持。1

该领域一个核心且广为人知的问题是，$f[M]$ 通常不是半正定的，因此不能作为有效的相关矩阵。虽然统计实践通常依赖事后特征值截断或假设严格的结构稀疏性来规避这一问题，但我们探究能够内在保持正定性的阈值函数所面临的严格代数限制。

定义 1.1. 我们称 $f\colon [-1,1] \to \mathbb{R}$ 在**上正定**S^{n-1}，如果对所有 N 以及所有 $x_1, \cdots, x_N \in S^{n-1}$，矩阵 $f[M]$ 是一个相关矩阵，其中 $M = (\langle x_i, x_j \rangle)_{i,j=1}^{N}$ 是向量 x_i. 的相关矩阵

这类映射在球面的情形下由 Schoenberg 分类，在紧致二点齐性空间的一般情形下由 Bochner 分类。为了在概率背景下恰当地分析这些函数，我们使用**归一化 Gegenbauer 多项式** $\tilde{C}_k^{(\alpha)}(t) = C_k^{(\alpha)}(t)/C_k^{(\alpha)}(1)$，其中 $\alpha = (n-2)/2$。这些多项式满足$\tilde{C}_k^{(\alpha)}(1) = 1$。Schoenberg 证明，连续函数 f 在 S^{n-1} 上正定，当且仅当它可以展开为：

$$f(t) = \sum_{k=0}^{\infty} a_k \tilde{C}_k^{(\alpha)}(t), \quad a_k \geq 0.$$

如果我们进一步要求 $f(1) = 1$（保持相关矩阵的对角线不变），那么 $\sum_{k\geq 0} a_k = 1$。

注 1.2. Schoenberg 的结果可以从相关矩阵正则化的角度来理解。注意，来自单位球面 S^{n-1} 上向量的 Gram 矩阵，恰好就是对角线元素为 1 的半正定矩阵，即相关矩阵。此外，无论选取的向量数量 x_i 是多少，球面的维数都会将相关矩阵的秩 限制在至多n。因此，定义 1.1 现在指出，函数 $f\colon [-1,1] \to \mathbb{R}$ 在 S^{n-1} 上正定，当且仅当对任意大小但秩为 $\leq n$的相关矩阵逐元素应用 $f[-]$ 后，得到的 $f(1)$乘以另一个相关矩阵。

因此，给定一个紧集 $K \subseteq [-1,1)$，我们希望将其阈值化为 0，并寻求在 K.上为零的正定函数。我们基于 [1] 为球面单位距离图的可测色数进行 Delsarte 分析所发展的方法，确立了它们的存在性——见定理 4.1。我们给出一个评估准则，其受 Delsarte 理论启发，并基于最大化 Gegenbauer 展开中线性项的相对权重，即**忠实性常数**。在存在性成立的前提下，根据 [6], 关于有限 K 的近期结果，可以

将正定函数选为多项式。与 [9]中所考虑的无约束秩软阈值化情形不同——在该情形下可用的阈值化函数十分匮乏——我们的固定秩情形则提供了充裕的选择。然而，此类阈值化函数必然会在定理 4.3 与定理 5.3 所确立的几何意义上显著劣化信号；后者证明，软阈值化函数的忠实性为 $\mathcal{O}(1/n)$，其中 n 是数据所取自的球面的维数（根据注 1.2，这是可由该数据得到的相关矩阵的最大可能秩）。固定秩自然出现在低样本高特征数据的情形中。

软阈值化的一种常见变通方法是采用 Ledoit–Wolf 类型的收缩 [10]，并将硬阈值化（或分段线性软阈值化）的相关矩阵与单位矩阵进行适当的凸组合，使其进入正锥，同时保留诱导稀疏性，因为这实际上仅相当于对相关矩阵的非对角元素进行缩放。然而，Schoenberg 定理表明，对于大型通用秩 n 相关矩阵，如果阈值化函数不是正定函数，则该组合渐近地将全部权重置于单位矩阵上。

2. Background

2.1. 核嵌入。 正定函数的几何解释从根本上依赖于再生核希尔伯特空间（RKHS）的构造。在集合 X 上，集合 X 上的一个对称函数 $K\colon X\times X\to\mathbb{R}$ 被称为<款式 id='7'>正定核，如果对于任意有限点集 x_1，……，$x_N\in X$ 以及实数 c_1，……，c_N，相应的格拉姆矩阵都是半正定的：

$$\sum_{i=1}^{N}\sum_{j=1}^{N} c_i c_j K(x_i,x_j)\geq 0.$$

这类核与欧几里得几何之间的联系由阿龙扎扬定理确立，该定理保证每一个有效的相关结构都可以实现为潜在空间中的一组内积。

定理 2.1（Aronszajn 定理 [11]）。设 K 是集合 X上的正定核。那么存在本质上唯一的希尔伯特空间$\mathcal{H}$ 和特征映射 $\iota\colon\ X\to\mathcal{H}$，使得该核能够完美地恢复 $\mathcal{H}$中的内积：

$$\langle \iota(x),\iota(y)\rangle_{\mathcal{H}} = K(x,y)\quad \textit{for all } x,y\in X.$$

在相关矩阵的语境下，指标集 X 是单位球面 S^{n-1}。当且仅当复合函数 $K(x,\ y) = f(\langle x,\ y\rangle)$ 构成正定核时，连续函数 f：$[-1,1] \to \mathbb{R}$ 才称为在 S^{n-1} 上正定。

根据定理 2.1，将这样的函数 f 应用于相关矩阵，在几何上等价于通过映射 $\iota : S^{n-1} \to \mathcal{H}$ 将原始变量嵌入一个新的希尔伯特空间。因此，经过阈值处理的矩阵所受的约束完全由这种嵌入的几何性质决定。

2.2. **球谐函数与 Gegenbauer 多项式**。 关于球谐函数和正交多项式的全面论述，可参阅经典著作 [13]。令 $S^{n-1} \subset \mathbb{R}^n$ 表示配备旋转不变概率测度 σ 的单位球面。空间 $L^2(S^{n-1}, \sigma)$ 可分解为相互正交子空间 $\mathcal{H}_k$ 的直和，其中 $\mathcal{H}_k$ 由总次数为 k 的调和齐次多项式在球面上的限制构成：

$$L^2(S^{n-1}, \sigma) = \bigoplus_{k=0}^{\infty} \mathcal{H}_k.$$

$\mathcal{H}_k$的维数记为 $d_k(n)$，其表达式为：

$$d_k(n) = \binom{n+k-1}{k} - \binom{n+k-3}{k-2}.$$

Gegenbauer 多项式是与这一分解相关的带状球面函数。在经典文献中，它们通常记作 $C_k^{(\alpha)}(t)$，其中参数定义为 $\alpha = (n-2)/2$。

Gegenbauer 多项式也可以看作关于测度 $(1-t^2)^{\alpha-1/2}$ 正交的多项式族；该测度自然地（除缩放因子外）源自将 S^{n-1} 上的曲面测度正交投影到某个坐标轴上。

为便于在引用经典恒等式时进行数学表述，我们用 $\alpha = (n-2)/2$ 表示经典参数，并将我们的**归一化 Gegenbauer 多项式** 记为 $\tilde{C}_k^{(\alpha)}(t)$，其缩放方式满足 $\tilde{C}_k^{(\alpha)}(1) = 1$。具体而言，若 $C_k^{(\alpha)}(t)$ 是由生成函数 $(1 - 2rt + r^2)^{-\alpha} = \sum C_k^{(\alpha)}(t) r^k$定义的标准正交多项式，则：

$$\tilde{C}_k^{(\alpha)}(t) = \frac{C_k^{(\alpha)}(t)}{C_k^{(\alpha)}(1)}, \quad \text{where } C_k^{(\alpha)}(1) = \binom{k+2\alpha-1}{k}.$$

这些多项式对所有 $t \in [-1,1]$ 都满足严格界 $|\tilde{C}_k^{(\alpha)}(t)| \leq 1$。

2.3. **Darboux 估计**。 Darboux 估计 [13] 有助于建立对 Gegenbauer 多项式的直观认识。将标准 Darboux 展开式应用归一化 $C_k^{(\alpha)}(1) \sim k^{2\alpha-1}/\Gamma(2\alpha)$ 后，我们在 $\theta \in (0,\pi)$范围内，将多项式取值于 $t = \cos\theta$处。由于 $k \to \infty$:

$$\tilde{C}_k^{(\alpha)}(\cos\theta) = \frac{2^{1-\alpha}\Gamma(2\alpha)}{\Gamma(\alpha)} k^{-\alpha} (\sin\theta)^{-\alpha} \cos\left((k+\alpha)\theta - \frac{\alpha\pi}{2}\right) + \mathcal{O}\left(k^{-\alpha-1}\right)$$

其中 $\alpha = \frac{n-2}{2}$。远离极点 $t = \pm 1$时，归一化多项式以 $\mathcal{O}(k^{\frac{2-n}{2}})$的速率一致有界地衰减。

2.4. **导数估计**。 为了分析区间上的阈值化（定理 5.3），我们需要估计 $\tilde{C}_k^{(\alpha)}(t)$在原点处的导数。

引理 2.2。 设 $K \subseteq (-1,1)$ 紧致。那么，$\sup_{t\in K} |(\tilde{C}_k^{(\alpha)})'(t)|$ 为$\mathcal{O}(k^{\frac{4-n}{2}})$。

证明。 Gegenbauer 多项式的经典求导恒等式为 $\frac{d}{dt}C_k^{(\alpha)}(t) = 2\alpha C_{k-1}^{(\alpha+1)}(t)$ [13]。 通过 $C_k^{(\alpha)}(1)$ 进行归一化，并应用经典维数参数 $\alpha = \frac{n-2}{2}$ 可得:

$$\frac{d}{dt}\tilde{C}_k^{(\alpha)}(t) = \frac{k(k+n-2)}{n-1} \frac{C_{k-1}^{(\alpha+1)}(t)}{C_{k-1}^{(\alpha+1)}(1)},$$

根据 Darboux 估计，该式为 $\mathcal{O}(k^{\frac{4-n}{2}})$。 □

2.5. **三项递推关系**。 为了提取阈值化函数系数的结构界，我们需要归一化 Gegenbauer 多项式的显式三项递推关系。

不同于标准正交多项式，归一化条件 $\tilde{C}_k^{(\alpha)}(1) = 1$ 导致次数之间出现不对称的转移。对于 $k \geq 0$，递推关系为:

$$t\tilde{C}_k^{(\alpha)}(t) = c_k \tilde{C}_{k+1}^{(\alpha)}(t) + b_k \tilde{C}_{k-1}^{(\alpha)}(t),$$

其中，向前和向后的转移权重分别为:

$$c_k = \frac{k+n-2}{2k+n-2}, \quad \text{and} \quad b_k = \frac{k}{2k+n-2}.$$

对于 $k = 0$，我们采用约定 $b_0 = 0$ and $\tilde{C}_{-1}^{(\alpha)}(t) = 0$，这会正确地得到 $c_0 = 1$。注意，对于所有 k, $c_k + b_k = 1$，都保留了多项式在极点 $t = 1$ 处取值为 1 的性质。

3. Delsarte's method and faithfulness

一种**球面码** $\mathcal{C}$ 是单位球面 S^{n-1}的有限子集。码的最小角是 $\mathcal{C}$ 中任意两个不同点之间的最小角距离。Delsarte 线性规划方法[7] 利用正定函数给出此类码基数的上界。

定理 3.1 (Delsarte 估计)。 设 $\theta \in (0, \pi)$ 且 $f\colon\ [-1,1] \to \mathbb{R}$ 是 S^{n-1} 上的正定函数，满足：

- $f(t) \leq 0$ 对所有 $t \in [-1, \text{余弦 } \theta]$;
- *Gegenbauer* 展开中的常数系数 a_0 为正。

那么，对于任意最小角为 θ 的球面码 $\mathcal{C}$，

$$|\mathcal{C}| \leq \frac{f(1)}{a_0}.$$

以下对 Delsarte 估计的证明，为我们研究软阈值化提供了路线图。

证明。 正定函数 f 定义了一个核嵌入 $\iota: S^{n-1} \to \mathcal{H}$，使得 $\langle \iota(x)\ \iota(y)\rangle_{\mathcal{H}} = f(\langle x, y\rangle)$。我们将希尔伯特空间分解为球谐函数所张成的正交子空间， $\mathcal{H} = \bigoplus_{k=0}^{\infty} \mathcal{H}_k$。 也就是说，我们将嵌入 ι 视为与核 $\tilde{C}_k^{(\alpha)}(\langle x, y\rangle)$ 对应的嵌入 ι_k 的加权直和。。

令 $T_0 : \mathcal{H} \to \mathbb{R}$ 为提取常数分量的线性算子，其作用定义为：对于任意单位向量 $x \in S^{n-1}$， $T_0\iota(x) = 1$ 。 $\iota(x)$ 在 $\mathcal{H}_0$ 中的分量的平方范数为 $a_0\tilde{C}_0^{(\alpha)}(\langle x, x\rangle) = a_0$。 由于 T_0 将一个范数为 $\sqrt{a_0}$ 的向量映射为标量 1， T_0 的算子范数恰为：

$$\|T_0\| = \frac{1}{\sqrt{a_0}}.$$

现在，设 $\mathcal{C} \subset S^{n-1}$ 是一个最小角为 θ 的球面码。考虑嵌入点的和：

$$V = \sum_{x \in \mathcal{C}} \iota(x) \in \mathcal{H}.$$

对该和应用算子 T_0，得到 $T_0(V) = \sum_{x\in\mathcal{C}} 1 = |\mathcal{C}|$。 两边平方，得到：

$$|\mathcal{C}|^2 = |T_0(V)|^2 \leq \|T_0\|\|V\| = \frac{1}{a_0}\|V\|^2. \tag{3.1}$$

我们直接利用核的性质计算 V 的平方范数：

$$\|V\|^2 = \sum_{x,y\in\mathcal{C}} \langle \iota(x), \iota(y)\rangle_{\mathcal{H}} = \sum_{x,y\in\mathcal{C}} f(\langle x, y\rangle) = |\mathcal{C}| f(1) + \sum_{x\neq y} f(\langle x, y\rangle).$$

由于码的最小角为 θ，且对于所有 $t \in [-1, \cos\theta]$，$f(t) \leq 0$，因此非对角项之和整体上非正。这就给出了该构形总能量的上界：

$$\|V\|^2 \leq |\mathcal{C}| f(1).$$

将此结果与 (3.1) 结合，即得所需的界，$|\mathcal{C}| \leq f(1)/a_0$。

□

备注 3.2。注意，f 的常数系数 a_0 是 $f(\langle x,\ y\rangle)$ 的期望，其中 x 和 y 从 S^{n-1}中独立且均匀地随机选取。

备注 3.3。传统上，Delsarte 估计的证明是通过对半正定格拉姆矩阵各项建立代数不等式来呈现的。然而，本文提出的再生核希尔伯特空间表述揭示了算子代数与概率所固有的、更深层的结构机制。

通过令非对角内积非正，我们在几何上强制嵌入的随机变量具有相互独立（或主动排斥）的先验。在这一框架中，算子 T_0 并非仅仅提取一个常数；它恰好充当规范的**条件期望**，将整个系统投影到平凡的球谐状态（常数子代数）上。我们通过优化，寻找具有最大常值函数条件期望的最佳无偏嵌入。从这个角度看，Delsarte 方法超越了几何球堆积；它给出了一个严格的算子范数容量界，限定了在独立性先验下经过条件期望后仍能保留的方差。

在经典的 Delsarte 方法中（以及 Cohn 和 Elkies [4], 将其应用于球面填充，并由 Viazovska [14, 5], 进一步发展；后者使用了针对 $\mathbb{R}^n$ 的类似理论），目标是最大化总质量 f（1）与常数项 a_0（平凡性）的比值，从而获得尽可能紧的码大小上界。

注释 3.4。在阈值化的语境中，我们将这一视角反转。我们不是在界定码大小，而是在构造一个核。我们的目标是最大化**线性系数** a_1（我们称之为忠实性），同时满足约束条件：f 在阈值化集合 K 上为零。正如 Delsarte 通过最大化平凡性来界定密度一样，

我们通过最大化忠实性来尽量减少信息损失。具体而言，我们有

$$a_1\langle x,y\rangle-(1-a_1)\le\langle\iota(x),\iota(y)\rangle\le a_1\langle x,y\rangle+1-a_1. \tag{3.2}$$

4. Existence and structural bounds

现在我们证明所需阈值函数的存在性。

定理 4.1. 设 $K\subseteq[-1,1)$ 是紧致的。存在一个非零正定函数 f，定义在 S^{n-1} 上，并且在 K.上消失

证明。 我们通过对球冠示性函数进行对称化来构造这样的函数。设 $C\subset S^{n-1}$ 是一个以北极 e_1 为中心、角半径为 θ的球冠，其闭包与阈值化区域不相交。具体而言，由于 $K\subseteq[-1,1)$ 是紧致的，存在某个 $r>0$，使得 $\sup K<\cos(2r)$。令 1_C 为半径为 r的球冠示性函数。考虑自相关核:

$$K(x,y)=\int_{O(n)}1_C(Ux)1_C(Uy)\,d\mu(U)$$

其中 μ 是正交群 $O(n)$上的Haar测度。根据构造，该核是正定的。值 $K(x,y)$ 与两个半径为 r 的球冠交集体积成正比，这两个球冠之间的角度由 x 和 y之间的角度确定。当且仅当 x和 y 之间的球面距离小于 $2r$时，这个交集体积非零。用内积表示，与该核对应的函数 $f(t)$的支撑集为 $[\cos(2r),1]$。由于 $\sup K<\cos(2r)$，函数 f在 K上恒为零。□

注 4.2。 我们注意到，如有需要，可以得到一个恰好在 $K,$ 上为零的函数。具体而言，取由两个小球冠的并集所产生的核，并令它们之间的角度为 θ，即可得到一个支撑集位于 0 和 $\cos\theta$. 附近的函数。为了避免将函数在远离 $K.$ 处阈值化为 0，或许会有人想要寻找这样的函数。然而，由于我们正在扰动其余各处的值，这种定性改进可能微乎其微。

鉴于阈值函数的存在，一个自然的问题是如何衡量其表现。一个自然的度量是 Gegenbauer 展开中线性系数的权重。对于球面 n 维空间中的紧集 K，我们将其**忠实性常数**定义为：在 S^{n-1} 上消失的酉连续正定函数所能达到的最大线性系数；记作 $\tau_{K,\ n}$.

定理 4.3。设 f 是 S^{n-1} 上的正定函数，其 *Gegenbauer* 系数 a_k 已归一化。设 $\iota: S^{n-1} \to \mathcal{H}$ 为相应的核嵌入。线性映射 $T_1: \mathcal{H} \to \mathbb{R}^n$ 的范数满足 $T_1\iota(x) = x$，且等于 $1/\sqrt{a_1}$。

证明。将球谐函数空间写成一系列正交子空间的直和 $\mathcal{H}_k$，每个子空间都对应于总次数为 k 的球谐函数张成的空间，并赋予权重 a_k:

$$\mathcal{H}_f = \bigoplus_{k=0}^{\infty} \mathcal{H}_k.$$

嵌入 $\iota(x)$ 可写为 $\sum_{k=0}^{\infty} v_k(x)$，其中 $v_k(x) \in \mathcal{H}_k$。由于 Gegenbauer 展开的正交性（因此球谐函数也具有正交性），线性子空间 $\mathcal{H}_1$ 与所有满足 $k \neq 1$ 的 $\mathcal{H}_k$ 正交。因此，恢复映射 T_1 的范数恰好为 $1/\sqrt{a_1}$. □

鉴于最优阈值函数必须使这一线性系数最大化，我们可以对其行为确立严格的结构上限。优化器的线性系数受到一定的大小约束，例如 $a_1 \geq a_0$. 。将这些系数视为自然数上的概率测度后，我们可以直接利用正交多项式的三项递推关系导出这一界限。

定理 4.4. 设 $f(t) = \sum_{k=0}^{\infty} a_k \tilde{C}_k^{(\alpha)}(t)$ 是 S^{n-1} 上满足 $f(1) = 1$ 的连续正定函数，它在紧集 $K \subset [-1, 1)$上消失的约束下最大化忠实性常数 a_1 。则 f 的系数满足二阶差分不等式:

$$c_{k-1}a_{k-1} + b_{k+1}a_{k+1} \leq a_1 \frac{d_k}{n} \quad for\ all\ k \geq 0,$$

其中 $a_{-1} = c_{-1} = 0$、 d_k 是次数为 k的球谐函数空间的维数，而 c_k、b_k 是来自三项递推关系 $t\tilde{C}_k^{(\alpha)}(t) = c_k\tilde{C}_{k+1}^{(\alpha)}(t) + b_k\tilde{C}_{k-1}^{(\alpha)}(t)$的转移权重。

证明。令 $g(t) = \tilde{C}_m^{(\alpha)}(t)$。根据 Schoenberg 定理， g 是正定的。考虑逐点乘积 $h(t) = f(t)g(t)$。根据 Schur 乘积定理， h 是正定的。此外，由于对所有 $t \in K$， $f(t) = 0$， $h(t)$ 也在 K 上为零。由于 $h(1) = f(1)g(1) = 1$，函数 h 是一个可接受的阈值函数。因为 f 在所有此类函数中使线性系数最大，h 的线性系数（记作 $c_1(h)$）必须满足 $c_1(h) \leq a_1$。

为计算 $c_1(h)$，我们利用归一化 Gegenbauer 多项式的正交性，将 $h(t)$ 投影到 $\tilde{C}_1^{(\alpha)}(t) = t$ 上。在 S^{n-1} 上的旋转不变概率测度 σ 下，$\tilde{C}_k^{(\alpha)}$ 的平方 L^2 范数为 $\|\tilde{C}_k^{(\alpha)}\|^2 = 1/d_k$。

线性系数为：

$$c_1(h) = \frac{\langle f(t)\tilde{C}_m^{(\alpha)}(t), t\rangle}{\|\tilde{C}_1^{(\alpha)}\|^2}.$$

在 f 的级数中逐项应用非对称三项递推关系，并利用正交性：

$$\begin{aligned} c_1(h) &= n\left(c_m a_{m+1}\|\tilde{C}_{m+1}^{(\alpha)}\|^2 + b_m a_{m-1}\|\tilde{C}_{m-1}^{(\alpha)}\|^2\right) \\ &= n\left(\frac{c_m a_{m+1}}{d_{m+1}} + \frac{b_m a_{m-1}}{d_{m-1}}\right). \end{aligned}$$

球面的维数和转移权重满足详细平衡条件 $d_m c_m = d_{m+1} b_{m+1}$。应用这一对称性，我们代入 $c_m/d_{m+1} = b_{m+1}/d_m$ 和 $b_m/d_{m-1} = c_{m-1}/d_m$，得到：

$$c_1(h) = \frac{n}{d_m}\left(b_{m+1}a_{m+1} + c_{m-1}a_{m-1}\right).$$

施加变分界 $c_1(h) \leq a_1$，即可得到所需的不等式：

$$c_{m-1}a_{m-1} + b_{m+1}a_{m+1} \leq a_1\frac{d_m}{n}.$$

由于 $m \geq 0$ 的选择是任意的，该界对所有 $k \geq 0$均成立。 □

注4.5。不等式右侧出现维数比 d_k/n，在数学上具有深刻意义。由于 d_k 随 $\mathcal{O}(k^{n-2})$增长，结构界宽松得近乎荒谬。它在低频下仅仅充当局部反尖峰条件，而在高频下实际上变得几乎毫无约束力。它远远不足以单独迫使该序列逐渐衰减或截断。

尽管这一结构界极为宽松，球面的代数几何却对简单阈值集合施加了严格得多的限制。正如 [6], 中所确立的，当阈值集合 K 为有限集时，最优正定函数可以取为多项式。此外，如果我们假设 K 是对称的，那么优化器必然是一个多项式！

注4.6。经典 Delsarte 方法的类似界为 $a_k \leq d_k a_0$. 在 $\mathcal{H}_k \oplus \mathcal{H}_0$，至多有 d_k 个向量与 $\mathcal{H}_0$. 正交且投影为正。该界能否达到，还是一种 Delsarte幻觉 [6]，是一个困难的几何问题。

鉴于这种类比，定理 4.4 中的比值 d_k/n 衡量了 $\mathcal{H}_k$ 与 $\mathcal{H}_1$. 的相对大小。

5. Explicit faithfulness bounds and the two-point gap

现在我们讨论针对 K. 的特定取值进行阈值化。回顾定理 4.3 之前定义的量 $\tau_{K,\ n}$ 。

5.1. **单点阈值化**。

定理 5.1. 设 $K = \{\varepsilon\}$ ，且 $\varepsilon > 0$ 足够小。则：

$$\tau_{K,n} \geq \frac{|\tilde{C}_2^{(\alpha)}(\varepsilon)|}{\varepsilon + |\tilde{C}_2^{(\alpha)}(\varepsilon)|}.$$

具体而言，$\tau_{K,\ n} \to 1$ ，当 $\varepsilon \to 0$ 时。

证明。 我们仅使用第一和第二个标准化 Gegenbauer 多项式构造一个函数： $f(t) = a_1\tilde{C}_1^{(\alpha)}(t) + a_2\tilde{C}_2^{(\alpha)}(t)$。要使其成为有效的相关函数，我们要求 $f(1) = a_1 + a_2 = 1$，这意味着$a_2 = 1 - a_1$。条件 $f(\varepsilon) = 0$ 得出：

$$a_1\varepsilon + (1 - a_1)\tilde{C}_2^{(\alpha)}(\varepsilon) = 0$$

解出 a_1:

$$a_1(\varepsilon - \tilde{C}_2^{(\alpha)}(\varepsilon)) = -\tilde{C}_2^{(\alpha)}(\varepsilon) \implies a_1 = \frac{-\tilde{C}_2^{(\alpha)}(\varepsilon)}{\varepsilon - \tilde{C}_2^{(\alpha)}(\varepsilon)}$$

由于 $\tilde{C}_2^{(\alpha)}(0) = -1/(n-1) < 0$，当 ε 足够小时， $\tilde{C}_2^{(\alpha)}(\varepsilon)$ 为负，从而得到一个有效的 $a_1 \in (0,\ 1)$。□

5.2. **双点阈值化**。 对多个点进行阈值化的代价通常很高。

定理 5.2。 设 $K = \{-\varepsilon, \varepsilon\}$ 满足 $1 > \varepsilon > 0$。那么，

$$\tau_{K,n} = \frac{\Sigma}{\varepsilon + \Sigma}, \quad where\ \Sigma = \sup_k \tilde{C}_{2k-1}^{(\alpha)}(-\varepsilon).$$

证明。 设 $f(t) = \sum a_k\tilde{C}_k^{(\alpha)}(t)$ 是一个正定函数，使得 $f(x) = 0$ 当 $x \in K.$ 因此， $f(\pm\varepsilon) = 0.$ 不失一般性，我们可以令 f 为奇函数。取 k 为 $\tilde{C}_{2k-1}^{(\alpha)}(-\varepsilon)$ 的最大化元，并选择

$$f(x) = \frac{\tilde{C}_{2k-1}^{(\alpha)}(-\varepsilon)x + \varepsilon\tilde{C}_{2k-1}^{(\alpha)}(x)}{\varepsilon + \tilde{C}_{2k-1}^{(\alpha)}(-\varepsilon)}$$

给出适当的 f 来证明忠实性。 □

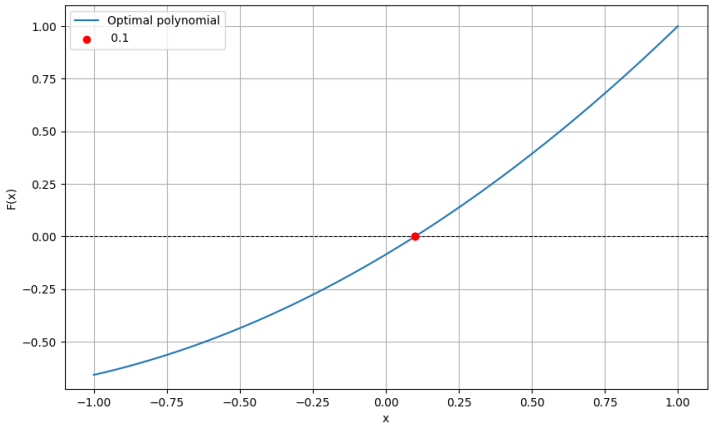


FIGURE 1. $K = \{0.1\}$, $n = 3$的阈值函数

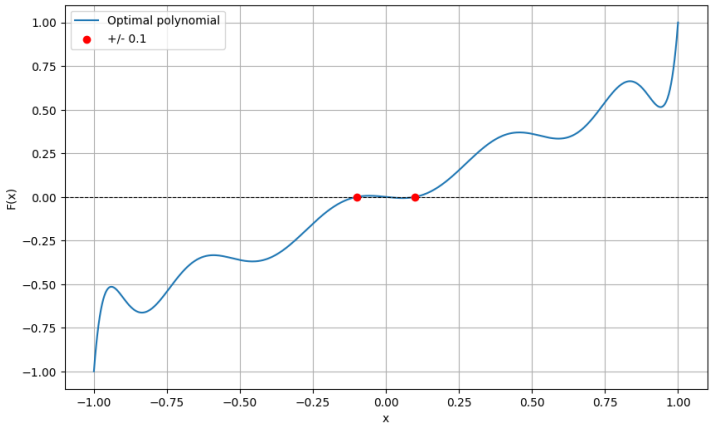


FIGURE 2. $K = \{\pm 0.1\}$, $n = 3$的阈值函数

若 $n = 2$, 3, 则当 ε 趋于 0 时，我们的界趋近于 1。若 $n \geq 4$，则可得 $\frac{3}{n+2}$. （对于足够大的 n，当 ε.较小时，最优函数近似为 x^3 ）

上述结果凸显了鲜明的二分性。对零附近的单个值进行阈值化，可以使 $a_1 \approx 1$。然而，对一个区间（甚至仅仅两个点 $\pm\varepsilon$）进行阈值化，会迫使 a_1 变小（其上界约与 $1/n$成正比）。

5.3. **区间阈值化**。

定理 5.3。 设 $K=[-\varepsilon,\varepsilon]$ 满足 $1>\varepsilon>0$. 那么，

$$\lim_{\varepsilon\to 0}\tau_{K,n}\le\frac{\Sigma}{1+\Sigma},\quad where\ \Sigma=\sup_{k\ odd,k\ge 3}\left|\frac{d}{dt}\tilde{C}_k^{(\alpha)}(0)\right|.$$

证明。 若 $n=2$,3 则上确界为无穷，因此该断言是平凡的。我们将假设 $n\ge 4$. 设 $f(t)=\sum a_k\tilde{C}_k^{(\alpha)}(t)$ 为正定函数，使得当 $x\in K$. 时 $f(x)=0$注意 $f'(0)=0$. 此外，由于$f(0)$ 必须为零，我们假设 f 为奇数。对展开式逐项求导:

$$f'(0)=\sum_{k=0}^{\infty}a_k\frac{d}{dt}\tilde{C}_k^{(\alpha)}(0)=0.$$

（在求和号下求导是合理的，因为 Darboux 估计表明导数之和在 K. 上一致收敛于一个连续函数）我们分离出线性项 $k=1$，并注意到 $\frac{d}{dt}\tilde{C}_1^{(\alpha)}(0)=1$:

$$a_1\cdot 1+\sum_{k\ne 1}a_k\frac{d}{dt}\tilde{C}_k^{(\alpha)}(0)=0.$$

由于当 k 为偶数时 $\tilde{C}_k^{(\alpha)}$ 为偶函数，它们在 0 处的导数为零。该和式仅限于奇数的 $k\ge 3$。

$$a_1=-\sum_{k\ \mathrm{odd},k\ge 3}a_k\frac{d}{dt}\tilde{C}_k^{(\alpha)}(0)=\sum_{k\ \mathrm{odd},k\ge 3}a_k\left|\frac{d}{dt}\tilde{C}_k^{(\alpha)}(0)\right|.$$

我们在 $\sum a_k=1$ 的约束下最大化 a_1。令 $S_k=|\frac{d}{dt}\tilde{C}_k^{(\alpha)}(0)|$。于是 $a_1=\sum_{k\ge 3}a_kS_k$。利用 $\sum_{k\ge 3}a_k=1-a_1$，我们有 $a_1\le(1-a_1)\sup S_k$。整理可得 $a_1\le\sup 1+\overline{\sup S_k}S_k$。□

也可以利用洛必达法则，将上述结果看作定理 5.2 的推论。

6. Conjectures and remarks

以下猜想在 [6] 中以齐性空间的 I 类表示的语言表述。

猜想 6.1。 任意定义在 S^{n-1} 上的正定函数都是连续函数的逐点极限。

当然，使用不连续的正定函数来运行 Delsarte 方法或我们用于软阈值化的线性德萨特方法，与使用连续函数一样容易。由此会产生如下悖论：假如该猜想不成立，我们很可能会针对某些特定的

有限集 $K.$。迪克斯米耶通过卷积论证证明，任何可测的正定函数都是连续函数的逐点极限 [8], ，因此任何此类病态函数本质上都是不可测的，而且或许依赖于公理。也就是说，对于 K 中具有某种特定条目排列的情形，人们会预期不连续表示所导致的 Delsarte 型界会出现大幅且无法解释的下降。

6.1. **聚类与稀疏性**。 根据定理 5.3 和 5.2，我们可以看出，对包含两个或更多点的集合进行软阈值化，必然会大幅削弱协方差矩阵的非对角元素。换言之，为了施加一种先验，假定随机变量以几何无偏的方式保持一致（即由正定函数产生的核嵌入），所付出的代价高得令人咋舌。

在统计学文献中，阈值估计量通常以假设真实总体矩阵高度稀疏或呈带状为依据[2]。我们的结果从严格的几何角度解释了为何这类假设不可避免：若不存在固有的聚类或稀疏结构，那么在阈值化过程中保持正定性就会使信号急剧衰减。

然而，分布嵌入（例如 $\sum_x \omega_x \iota(x)$，其中 $\omega_x \geq 0$和 $1 = \sum_x \omega_x$）可能具有显著较大的相关性。这很自然，因为我们的方法天然适用于样本量小、特征维度高的数据（此时协方差矩阵为低秩），而这类数据本质上需要进行聚类、选择代表性特征（例如 LASSO 类型的方法），或采取其他处理，才能达到足够严谨的程度。寻找恰当且有意义的分布似乎属于某种聚类问题。

致谢。 我们感谢 Apoorva Khare 提出的诸多意见和建议。

References

[1] Christine Bachoc、Gabriele Nebe、Fernando M´ario de Oliveira Filho 和 Frank Vallentin。可测色数的下界。《几何与泛函分析》，19(3):645–661，2009。[2] Peter J Bickel 和 Elizaveta Levina。通过阈值化进行协方差正则化。《统计年刊》，第2577–2604页，2008。[3] Tony Cai 和 Weidong Liu。稀疏协方差矩阵估计的自适应阈值化。《美国统计学会杂志》，106(494):672–684，2011。[4] Henry Cohn 和 Noam Elkies。球堆积的新上界。I。《数学年刊》（第*2*辑），157(2):689–714，2003。

[5]Henry Cohn、Abhinav Kumar、Stephen Miller、Danylo Radchenko 和 Maryna S Viazovska。24维球堆积问题。《数学年刊》，185(3):1017–1033，2017。[6]Sujit Sakharam Damase 和 James Eldred Pascoe。齐性空间上的完备离散 Schoenberg–Delsarte 理论。*arXiv* 预印本 *arXiv:2602.09010*，2026。[7] Phillippe Delsarte。通过线性规划得到的无限制码界。《飞利浦研究报告》，27:272–289，1972。[8] Jacques Dixmier。 C^*-代数，北荷兰数学文库第15卷。北荷兰出版公司，阿姆斯特丹–纽约–牛津，1977。由 Francis Jellett 从法文译出。[9] Dominique Guillot 和 Bala Rajaratnam。使阈值化矩阵保持正定性。《线性代数及其应用》，436(11):4143–4160，2012。[10]Olivier Ledoit 和 Michael Wolf。亲爱的，我把协方差矩阵缩小了——均值–方差优化中的问题。《投资组合管理杂志》，30(4):110–119，2004。[11] VernIPaulsen 和 Mrinal Raghupathi。再生核希尔伯特空间理论导论，第152卷。剑桥大学出版社，2016。[12] Adam J Rothman、Elizaveta Levina 和 Ji Zhu。大型协方差矩阵的广义阈值化。《美国统计学会杂志》，104(485):177–186，2009。[13] G´abor Szeg˝o。正交多项式，美国数学会学术会议出版物第XXIII卷。美国数学会，罗得岛州普罗维登斯，第四版，1975。[14] Maryna S Viazovska。8维球堆积问题。《数学年刊》，第991–1015页，2017。

(S.S. Damase) Department of Mathematics, Indian Institute of Science, Bangalore 560012, India;电子邮件地址: sujits@iisc.ac.in

(J.E.Pascoe) Department of Mathematics, Drexel University, Philadelphia, PA 19104, United States;电子邮件地址: jep362@drexel.edu