arXiv:2303.10104v1 [math.ST] 17 Mar 2023

盲源分离的鲁棒性分析

ALEXANDER SCHELL

哥伦比亚大学统计系 *ity*

摘要。盲源分离旨在从混合信号 $X = f(S)$ 中恢复未观测到的信号 S，前提是产生该混合的变换 f 可逆但未知。由于这是一个具有诸多实际应用的基础问题，一个根本性问题是：当支撑该问题的统计先验假设遭到违背时，其解会如何表现。在经典的线性混合背景下，我们提出一个通用框架，用于分析此类违背并量化其对从 X 中盲恢复 S 的影响。我们将 S 建模为多维随机过程，为混合信号 X 背后的可能成因空间引入一种信息丰富的拓扑，并表明，对于定义盲源分离解的结构假设的一般偏离，可以通过相对于该拓扑的显式连续性保证，对一般盲源分离解的响应进行有效分析。这使得我们能够灵活、便捷地量化一般模型不确定性情形，并由此建立了首个全面的盲源分离鲁棒性框架。我们的方法完全具有构造性，并通过针对多种统计应用的新颖理论保证，展示了其效用。

1. 引言

The 盲源分离问题涉及对一个未知函数的（局部）求逆 on
give，但只能观察到该函数作用于某个同样未被观测到的自变量后所得的像：

$$\text{Provided only } X = f(S) \text{ for } f \text{ and } S \text{ unknown, invert } X \text{ for } S. \tag{1}$$

这一问题是众多应用领域的核心，应用范围从医学影像、神经科学和生物学 [6, 33, 61, 67] 到金融和天文学 [7, 72]再到物理学和工程学 [3, 30, 89]，这还只是其中的一小部分；为解决这一问题，人们提出了大量实用和理论方法，足见其重要性；例如，[28, 50, 68] 可参见综述。由于盲源分离问题的一般表述（1）高度欠定，所有这些方法都必须依赖于对 f 和 S施加的一些附加条件——即所谓的“可辨识性假设”——以确保盲反演 $X \mapsto S$ 确实有可能实现。

迄今为止，用于解释和求解（1）最为成功的方法，假设隐藏关系 f 是线性的[1]，并将源 S 建模为一个具有统计独立分量的多维随机变量。这一方法构建了一个通常称为独立成分分析（ICA）的框架，其依据是 Reiersøl、Skitovich 和 Darmois 分别获得的概率论见解 [78, 87, 32]，这些见解由 Comon 富有影响力地表述为一个优化问题 [27]。自其诞生以来 [50]，独立成分分析经历了多种改造和扩展，其中大多数涉及对其底层优化过程的修改与分析，或对源模型作出变化，即对其分量分布采用不同的（足以保证可识别性的）“非退化”假设；例如参见 [28, 50, 68] 及其中的参考文献。

电子邮箱地址： alexander.schell@columbia.edu。*MSC2020*主题分类：主要分类 62H25、62G35；次要分类 62H05、60L10、62M86。关键词：盲源分离、独立成分分析、统计稳健性、不确定性量化、无监督学习、逆问题、统计独立性、潜变量模型。[1] 尽管线性假设对于目前采用的绝大多数盲源分离方法至关重要（例如参见 [28, 50, 51], ），但这一曾经牢不可破的约束近来已被放宽，代价是对 S施加更强的结构假设；例如参见 [41, 52, 59] 以及我们近期的论文 [83]和其中的参考文献。

一个相对少有探讨、却显而易见且根本的问题是：尤其考虑到独立成分分析方法已广泛应用于众多领域，方程（1）的解在其关于 f 和 S 的底层可识别性假设遭到违反时，是否仍然成立，以及在多大程度上成立。下面这个信号处理领域的经典例子说明了这一问题的范围、重要性及其影响：

示例 1.1（更贴近现实的鸡尾酒会）。假设你正在监控一间拥挤的房间，里面有 $d \in \mathbb{N}$ 个人，比如（$S^1, \dots, S^d$）$=: S$，每个人都在交谈。再假设他们说的每句话都可能与你有关且包含有用信息，因此你的目标是获取一份关于每位说话者 S^i 说了什么的（经各方认可的）记录。为此，你在房间不同位置安装了 d 录音设备。然而，说话者们没有直接对着你的麦克风说话，而是全神贯注地交谈，以至于他们的谈话在声学上相互重叠。因此，每台录音设备录到的不是单一而清晰的声音，而是 $X^i = f_i$（$S^1, \dots, S^d$）个 d 不同声音的混合信号。你还能从录音中分离出每位说话者各自的声音吗？

至少乍看之下，鉴于声音传播的线性特性，将每段录音建模为说话者声音的未知线性叠加似乎合情合理，也就是说，假设 $X \equiv (X^1, \cdots X^d) = f(S)$ 其中某个 $f = (f_1, \cdots f_d)$：$\mathbb{R}^d \to \mathbb{R}^d$ 线性。此外，虽然在此情境下这一假设或许稍难论证，但进一步假设我们要寻找的声音[2] 统计独立，无疑在数学上相当方便。基于这些考虑，我们自然会尝试对记录的混合信号 X 执行 ICA，以此按照（1）恢复所需的各个语音信号。

虽然直觉上这无疑是一个合理的想法，但很快就会出现一些关于该方法及其底层结构假设有效性的局限性问题：

- 由于人类对话具有互动性，几乎可以肯定，部分或全部语音信号——例如参与共同讨论的说话者所发出的语音信号——彼此并非真正统计独立，至少并非完全独立；
- 由于麦克风之间存在空间距离，且声音传输和传播会受到声学色散或衰减等多种现象的影响，记录的混合信号与说话者各自声音之间的实际关系往往实质上是非线性的，至少存在轻微的非线性；
- 语音信号本身通常并非在时间上独立同分布，而是自然呈现出很强的跨期统计复杂性；此外，这些信号的物理传播及其录制过程也很少完全不受干扰，通常会受到各种噪声和技术干扰，例如环境噪声，或不同麦克风之间以非均匀采样率进行异步录音。

鉴于这些复杂因素会使真实混合构型偏离线性和独立性这一简化理想假设，对记录的混合信号进行独立成分分析（ICA），是否仍能重构出有意义的个体声音表达？这些表达即使可能并不完全准确，是否至少仍足够“好”，能够辨认出每位说话者传达的主要信息？抑或对基础模型假设的违反会严重影响 ICA 的表现，以至于返回的信号失真到无法辨认，或者预期的信号分离完全失败？这些结果之间的转变是否连续？能否根据导致真实情形偏离理想化可辨识性假设的“失真程度”，对返回信号的精度进行有意义的量化？

◆

2 ... 从一些描述声音的随机过程 $S^i = (S^i_t)_{t \geq 0}$ 中采样，这些过程定义在 $\mathbb{R}$ 中，而这些随机过程又是 ...2

严格来说，上述各项都代表了对源和混合的结构性先验假设的潜在违反，而这些假设是绝大多数 ICA 过程的基础，其中包括迄今为止最流行的所有此类方法，参见 [28, 50, 68]。将 $((S, f)$ 在 (1) 中称为 X的“成因”，那么核心问题便是：对观测成因的先验假设的此类违反，会以何种方式、在多大程度上影响 ICA 过程的算法一致性？此外，能否根据其支撑性可识别性假设所受到的一般性扰动，对这种一致性进行有意义的量化？

尽管人们一再呼吁对盲源分离和独立成分分析的解进行相应的鲁棒性研究， [49], 但文献中对这类问题的严谨关注却出人意料地少。据我们所知，目前尚无针对盲源分离或独立成分分析的鲁棒性理论能够全面分析上述问题；现有结果仅限于统计一致性的证明[3], ，例如 [19, 20, 80], ，或对数值子程序进行局部敏感性分析 [1, 16, 86], ，或仅适用于渐近情形 [19], ，或局限于高度特定的盲源分离模型 [8, 58] ，或仅针对非常特殊的（半）参数噪声模型或依赖模型 [9, 55, 77, 81], ；而且，即便是这少数结果，也大多仅限于“时不变”的特殊情形，即将 X 建模为由随机向量构成的独立同分布序列。

本文旨在弥补这些不足，为（独立同分布且）时变观测量的盲源分离提供一种灵活而全面的鲁棒性理论。将源建模为 S （1）中的离散时间或连续时间随机过程，取值于 $\mathbb{R}^d$ 并采用独立成分分析的一般范式，我们的主要贡献如下：

在对抽象的盲源分离问题进行简要分析（1）的基础上，我们提出一种通用且灵活的统计稳健性拓扑概念[4] ，适用于盲源分离和独立成分分析，并涵盖广泛的相关违规情形（第2节）。其核心思想是，将稳健性描述为：相对于一种信息丰富、面向独立成分分析的拓扑，在构成盲源分离观测量基础的隐藏原因空间上具有连续性。在第4节，我们从随机过程分布空间上一种自然且粗糙的预度量拓扑（第4.1节）构造出这样的独立成分分析拓扑；其中，预度量源自一族分级的路径空间泛函，这些泛函与粗糙路径理论中的签名映射相关（第3节）。结合我们对稳健性的通用形式化方法，这种独立成分分析拓扑提供了一种灵活且与模型无关的方式，用于刻画对理想化可辨识性假设的违反，并以有限（即“非渐近”）且便于量化的方式，将其与对独立成分分析逆变换精度所造成的影响联系起来。第5节通过展示一种通用盲反演映射对此进行了说明。该映射在符合独立成分分析典型特征的可辨识性假设下实现（1），同时首次针对其底层一致性假设的非常一般的偏离，提供统一的稳定性保证。从概念上说，这一反演映射具有相对于前述独立成分分析拓扑连续的性质；其鲁棒性保证则源于相对于该拓扑的信息丰富且完全显式的连续模（定理5.8 和5.12）。第6节的一系列统计应用案例展示了我们的方法对实际应用的意义，其中包括对示例1.1中实际违规情形的可证明鲁棒性的演示。

为使论述简洁明了，我们将大部分证明和一些技术性辅助结果放在附录中。

[3] 一致性结果可以视为“相对于经验近似的渐近（部分）鲁棒性”，因为它们确立了解沿着某些基于样本的估计量序列收敛（但不一定沿着所有收敛序列收敛；后者将刻画序列拓扑下的鲁棒性）。[4] 我们使用“鲁棒”一词时，采用的是 Huber 和 Ronchetti [47],在第2节中给出的定义。

我们假设读者熟悉第A 节及其符号约定，并采用以下约定。

约定 1.2（将信号与其概率律等同）。从第 2.1 节起，我们将默认把给定信号 $\boldsymbol{Y}$ 在 $\mathbb{R}^d$（定义在 $\mathcal{C}_d$上的随机变量）与其分布 $Y \equiv \mathbb{P}_{\boldsymbol{Y}} := \mathbb{P} \circ \boldsymbol{Y}^{-1}$（定义在 $\mathcal{C}_d$上的概率测度）等同。这使得常见的符号滥用 $\boldsymbol{Y}\,[\cong Y] \in \mathcal{M}_1(\mathcal{C}_d)$。

（这相当于将两个 $[\mathcal{C}_d$-值] 随机变量在分布相同的情况下视为“相等”。）分布的变换均理解为推前测度，即按（A.7）的含义理解。

2. The Problem of Blind Source Separation

要对盲源分离进行严谨的稳定性分析，必须对开篇提出的一般问题陈述（1）作出概念上精确的表述。将$\boldsymbol{S} = (\boldsymbol{S}_t)_{t\in\mathbb{I}}$ 视为 $\mathbb{R}^d$上的随机过程，$\mathbb{I} \subset \mathbb{R}$ 紧，本节将给出这样的表述：首先严格定义逆问题（1）及其相关的解的概念（第 2.1节），然后提出盲反演稳定性的一般概念，并据此给出 ICA 背景下鲁棒性的自然且精确的定义（第 2.2节）。

本节引入的概念乍读之下对一般读者而言可能显得有些抽象，图 1 对此作了说明，第 B节则进一步阐释了其动机。

2.1. 盲源分离。 盲源分离问题是一个逆问题，旨在从混合信号 $\boldsymbol{X} = f(\boldsymbol{S})$ 中恢复未观测到的信号 $\boldsymbol{S}$，前提是产生该混合信号的变换 f 可逆但未知。$\boldsymbol{X}$ 与 $\boldsymbol{S}$ 之间的基本关系由以下恒等式表示

$$\boldsymbol{X}_t = f(\boldsymbol{S}_t) \quad \text{for all } t \in \mathbb{I}, \tag{2}$$

当且仅当 f 的定义域包含 $\boldsymbol{S}$的空间支撑 $D_{\boldsymbol{S}}$ 时，该式才有定义，见（A.3）。假设 f 及其逆均连续，则有 $f(D_S) = D_X$（参见（A.6）），并将这样的

$$(X, S, f) \quad \text{a } \textit{BSS-triple} \tag{3}$$

其中，观测量 X、源 S 和混合变换 f。简要总结以上内容，注意到任何此类三元组的“因果部分”（S, f）都是空间中的一个元素

$$\mathfrak{C} := \left\{(\tilde{S}, \tilde{f}) \in \mathcal{M}_1 \times Z^Z \mid \tilde{f} \in C^{0,0}(D_{\tilde{S}})\right\} \tag{4}$$

where 我们令 $Z := \mathbb{R}^d$ 并采用约定 1.2。[5] 另外几个定义也会很有用 1.

2.1.1. 盲源分离问题。我们无法知晓式 (2) 的右侧，这使观测量 X 与其原因（S, f）之间存在不对称性：如果我们只知道 X，而没有关于 S或 f的任何信息，那么我们就无法区分产生~ ~ ~ ~X 的“原始”原因（S, f）与集合 $\{(S, f) \mid (X, S, f)$ 盲源分离三元组$\} \subseteq \mathfrak{C}$ 中的任何其他替代原因。在这种情况下，从 X 恢复原始 S，并不比找到一个元素

$$[S]_X := \left\{\tilde{S} \in \mathcal{M}_1 \;\middle|\; \exists\, \tilde{f} \in C^{0,0}(D_{\tilde{S}}) : (X, \tilde{S}, \tilde{f}) \text{ is BSS-triple}\right\} = \mathrm{ev}_X\left(C^{0,0}(D_X)\right).$$

The BSS 范式通过将我们掌握的信息从（仅有）X 提升至 :

$$X \quad \text{and} \quad \mathscr{I} \quad \text{and} \quad \text{the } \textit{inclusion } (S, f) \overset{!}{\in} \mathscr{I} \tag{5}$$

对于某个先验给定的 $\mathfrak{C}$ 的非空子集 $\mathscr{I}$，称为可辨识性假设 (IA)，它描述了我们对 X 的原始成因（S, f）的先验知识。集合 $\mathscr{I}$ 越大，我们对（S, f）了解得越少：如果 $\mathscr{I} = \{(S, f)\}$，那么我们就能精确知道原始成因；如果 $\mathscr{I} = \mathfrak{C}$，那么我们对原始成因就“完全一无所知”。

[5] 请回想，根据这一约定，我们将信号 $\tilde{S}$ 与其分布 $\mathbb{P}_{\tilde{S}} \in \mathcal{M}_1$等同起来。另请注意，在 (4) 中，我们稍微滥用了符号：$\tilde{f} \in C^{0,0}(D_{\tilde{S}}) :\Leftrightarrow \tilde{f}\big|_{D_{\tilde{S}}} \in C^{0,0}(D_{\tilde{S}})$，以免行文显得繁杂。

将数据从 X 升级为 $(X,\ \mathscr{I})$ 可以减小 X 与 $(S,\ f)$之间的不对称性：在（5）的条件下，现在可以将三元组 $(X,\ S,\ f)$ 的源恢复到更小的类中

$$\langle X\rangle_{\mathscr{I}} := \left\{\tilde{S} \equiv \tilde{g}(X) \;\middle|\; \tilde{g} \in C^{0,0}(D_X) \text{ and } \left(\tilde{g}(X), \tilde{g}^{-1}\right) \in \mathscr{I}\right\} \subseteq [S]_X. \tag{6}$$

换言之：中的每个元素$\langle X\rangle_{\mathscr{I}}$ 都是 S 在给定（5）时的最佳近似，而$\langle X\rangle_{\mathscr{I}}$中的任意两个元素彼此“无法区分”[6]，相对于 S给定的$(X,\mathscr{I})$。我们先约定术语。

定义 2.1（ $\mathscr{I}$上的 BSS 三元组）。设 $\mathscr{I} \subseteq \mathfrak{C}$。满足 $(S,f) \in \mathscr{I}$的 BSS 三元组 (X,S,f) 称为 $\mathscr{I}$上的 BSS 三元组，记作：$(X,S,f)_{\mathscr{I}}$。式（6）中的相关类 $\mathbf{h}X\mathbf{i}_{\mathscr{I}}$是由 X 在给定 $\mathscr{I}$时得到的极大解，其元素是 X 在给定 $\mathscr{I}$时的准源。

这些定义及后续定义的示意图见图1；而在示例B.1 中，第 2 页的鸡尾酒会情境将具体阐释本节的概念。

注 2.2。（i）BSS 三元组（6）的极大解$(X,S,f)_{\mathscr{I}}$按集合包含关系是极大的，因此由构造可知它是唯一的。

(ii) 若（5）中的可辨识性假设 $\mathscr{I}$in 是 $\mathcal{M}_1 \times C^{0,0}(\mathbb{R}^d)$ 中的一个笛卡尔积集合，记作 $\mathscr{I} = \mathscr{S} \times \mathfrak{T}$，则由 X 给定的极大解 $\langle X\rangle \equiv \langle X\rangle_{\mathscr{S}\times\mathfrak{T}}$ 为

$$\langle X\rangle = \mathrm{ev}_X(\mathfrak{T}) \cap \mathscr{S} \tag{7}$$

其中，$\mathrm{ev}_X(\mathfrak{T})$ 是求值映射$\mathrm{ev}_X:\ \mathfrak{T} \ni g \mapsto g(X)$的像，参见（A.7）。

(iii) 集合 $\langle X\rangle_{\mathscr{I}}$ 的 [大小取决于 $\mathscr{I}$。特别是：可辨识性假设越弱 $\mathscr{I}$（即集合 $\mathscr{I}$ 越大），类 $\langle X\rangle_{\mathscr{I}}$ 就越大，因此所有准源的集合与原始源 S $\cong\{S\}$ 的距离也就越远。相应地，任何包含 S 的子集 $[S]_\circ \subseteq \langle X\rangle_{\mathscr{I}}$ [$\subseteq [S]^\circ$，（分别地，]）都定义了从 X 恢复 S 时“精度”的上界（分别地，下界）。

盲反演任务 $X \mapsto S$ 的核心在于所能达到的精度 ——也就是说，所得准源] 与原始源 S 之间的偏离 [达到何种程度才算最优或最可接受？答案通常由盲源分离实践者或其应用情境决定，并转化为约束[7] $\langle X\rangle_{\mathscr{I}} \subseteq \lceil S\rceil$ ，其中$\lceil S\rceil$ 是先验给定的某个“最优”超集 S （即“精度界”）。因此，主要难点在于确定一个可辨识性假设 $\mathscr{I}$ ，使得这一包含关系成立；换言之，使集合$\langle X\rangle_{\mathscr{I}}$ 与最优超集 $[S]_\circ := \lceil S\rceil \cap \langle X\rangle_{\mathscr{I}}$ 关于 S相同。若非最优性成立，则可通过在选定 $\mathscr{I}$之后推导下界 $[S]^\circ$ ，来控制$\langle X\rangle_{\mathscr{I}}$与 $[S]_\circ$之间的偏离。

以上内容立即引出了以下概念。

定义 2.3（盲反演）。**给定一个映射** $\lceil\cdot\rceil:\ \mathcal{M}_1 \to 2^{\mathcal{M}_1}$，可辨识性假设 $\mathscr{I} \subseteq \mathfrak{C}$ 称为

$$\lceil\cdot\rceil\text{-}\textit{sufficient} \quad \text{if} \quad \langle X\rangle_{\mathscr{I}} \subseteq \lceil S\rceil \quad \text{for each BSS-triple } (X,S,f)_{\mathscr{I}}. \tag{8}$$

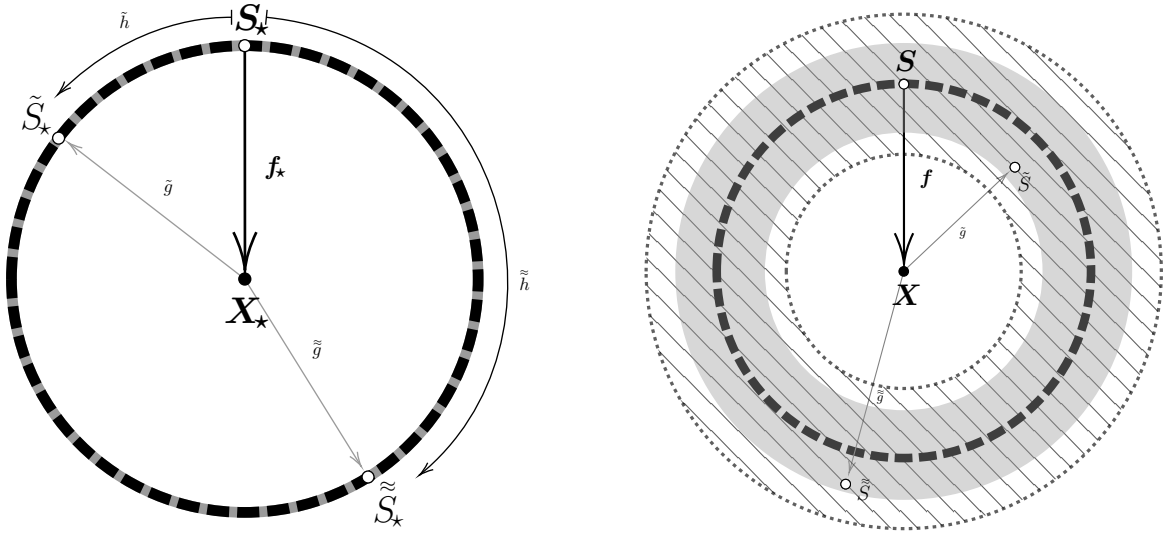
此外，对于任意 $\mathscr{I} \subseteq \mathfrak{C}$，我们将满足以下条件的任意函数称为定义在 $\mathscr{I}$上的反演映射 $\Phi:\ \mathcal{M}_1 \to 2^{\mathcal{M}_1}$

$$\emptyset \neq \Phi(X) \subseteq \langle X\rangle_{\mathscr{I}}, \quad \text{for each BSS-triple } (X,S,f)_{\mathscr{I}}. \tag{9}$$

最后，若映射 $\varphi:\ \mathcal{M}_1 \to 2^{\mathcal{M}_1}$ 满足：对于每个 $S \in \mathcal{M}_1$， $S \in \varphi(S)$，则称其为忠实的 。

盲反演的任务现在是实现以下目标：给定一个精度界$\lceil\cdot\rceil$，我们能否找到一个可辨识性假设 $\mathscr{I}$（越弱越好），但又足够强，从而保证对于任意原因$(S,\ f) \in \mathscr{I}$，源 S都能从其混合信号（2）中恢复出来，且精度达到$\lceil\cdot\rceil$，即误差不超过$\lceil S\rceil$？如果可以，是否存在一种可显式计算的算法来执行这种逆变换？

[6] 更准确地说，这可由等价关系 $\left[\tilde{S} \sim_{(X,\mathscr{I})} \check{S} :\Leftrightarrow\ \tilde{S},\ \check{S} \in \langle X\rangle_{\mathscr{I}}\right]$ 在 $[S]_X$ 上来刻画。[7] 因此，每个元素 $\tilde{S} \in \lceil S\rceil \cap \langle X\rangle_{\mathscr{I}}$都是这一约束下盲源分离问题的最优解。



a. $[S_\star]_\circ = \langle X_\star \rangle_{\mathscr{I}_*} = \Phi(X_\star)$

b. $[S]_\circ \subsetneq \langle X \rangle_{\mathscr{I}}, \quad \Phi(X) \neq [S]_\circ$

图 *1.* 两个盲源分离三元组 $(X_\star, S_\star, f_\star)_{\mathscr{I}_*}$、$(X, S, f)_{\mathscr{I}}$及其各自极大解的示意图。平面中的每个点代表一个信号，即 M_1，的一个元素；彩色区域表示 M_1 的子集，这些子集与可观测量 $X_\star$，X 或其源 $S_\star$，S相关。图中展示了对于某个精度界 $d \cdot e$：精确准源集合 $[S_\star]_\circ := hX_\star i_{\mathscr{I}_*} \cap dS_\star e$ 和 $[S]_\circ := hXi_{\mathscr{I}} \cap dSe$（黑色虚线圆），极大解 $hXi_{\mathscr{I}}$（带尖边界的条纹环形区域），以及某个定义在 $\mathscr{I}_*$ 上的反演映射 $\Phi \equiv \Phi_{\mathscr{I}_*}$ 的输出 $\Phi(X_\star)$ 和$\Phi(X)$（灰色环形区域）。支持 $X_\star$ 的可辨识性假设 $\mathscr{I}_*$ 是 $d \cdot e$-充分的，这意味着集合 $hX_\star i_{\mathscr{I}_*} = \Phi(X_\star)$（即给定 $X_\star$时，关于 $S_\star$ 的所有由 $\mathscr{I}_*$提供信息的最佳猜测集合）与 $[S_\star]_\circ$ 重合，见图 *a*。特别地，$hX_\star i_{\mathscr{I}_*}$ 中的每个准源 $\tilde{S}_\star \equiv \tilde{g}(X_\star)$都是对 $S_\star$ 的“精确”猜测，因为将 $S_\star$ 与 $\tilde{S}_\star$ 关联起来的残差 $h = \tilde{g} \circ f_\star$ 处于由 $d \cdot e$[8] 所确定意义下的 “ 到恒等映射的最小距离 ”。在图 *b* 中，作为 X 基础的可辨识性假设 $\mathscr{I}$过弱，未达到 $d \cdot e$-充分，因此极大解 $hXi_{\mathscr{I}}$是最优解 $[S]_\circ$的真超集。$(\mathscr{I}_*, \Phi)$ 的鲁棒性表明，映射 Φ 的反演性质能够抵御其输入信号背后的成因（属于 $\mathfrak{C}$）偏离 $\mathscr{I}_*$ 。这相当于“从 *a.* 到 *b.* 的连续过渡”，如式 *(10)* 所示：如果从某种一般且适当的意义上说，成因 $(S, f) \in \mathscr{I}$与 $(S_\star, f_\star) \in \mathscr{I}_*$“没有太大差异”，我们能否定量保证 $\Phi(X)$ 与 dSe 也不会相差太远?

形式上，（一般的）**盲源分离问题**表述如下：

$(\star)$	给定一个忠实映射 $\lceil\cdot\rceil: \mathcal{M}_1 \to 2^{\mathcal{M}_1}$，确定： (a) 一个 [弱] 可辨识性假设 $\mathscr{I} \subseteq \mathfrak{C}$，它是 $\lceil\cdot\rceil$-充分的； (b) 一个 Φ 关于 $\mathscr{I}$的可计算反演映射。

任何满足 (a) 和 (b) 的数对 $(\mathscr{I}, \Phi)$ 都称为 盲源分离问题的解，其精度为$\lceil\cdot\rceil$.

[8] 例如，广受欢迎的选择（B.8）是将 $\lceil S \rceil$ 视为单项式轨道 $\mathrm{M}_d \cdot S$ 要求 $\tilde{h} \in \mathrm{M}_d$。 6

注 2.4. 就目前的表述而言，上述问题（⋆）实际上包含了一整族逆问题，即使固定了 $\lceil\cdot\rceil$ 也是如此，因为 $\mathscr{I}$ 中限定词“弱”（⋆）$_{(a)}$ 并未明确[9]说明。若要求 $\mathscr{I}$ 是“最弱的”（即关于集合包含关系全局最大的）$\lceil\cdot\rceil$-充分可辨识性假设（IA），则可消除这种歧义；这会将（$\mathfrak{C}$）⋆转化为更强的任务，即刻画 $\lceil\cdot\rceil$-精确可逆性。此处我们不采用这一更强的表述，以便更好地反映当前文献中讨论和处理盲源分离问题的方式，参见例如 [28, 50, 68] 及其中的参考文献。

◆

注 2.5. 盲源分离问题（⋆）由一个抽象部分 (a) 和一个更具操作性的部分 (b) 构成。前者在理论上确保能够 [足够准确地] 将观测量盲逆变换回其源；后者则要求给出一种算法，在理论和统计实践中执行这一逆变换。文献中记载的（⋆）问题的解通常同时处理任务 (a) 和 (b)，典型做法是先构造性地证明某一可辨识性假设的 $\lceil\cdot\rceil$-充分性（8），然后将显式反演映射（9）作为所采用证明策略的直接结果推导出来，参见 [27, 28, 50, 68]。第 5 节中我们也遵循这一总体方案。

◆

In a除经典表述（⋆）外，我们还提出（暂且）将某个解称为 n

$$(\mathscr{I},\Phi)\text{ is }\textbf{robust}\quad\text{if,}\quad\text{in some appropriate sense,}\quad\text{the maps}$$
$$(\tilde{S},\tilde{f})\mapsto\Phi(\tilde{f}(\tilde{S}))\quad\text{and}\quad(\tilde{S},\tilde{f})\mapsto\operatorname{dist}_{\varrho}\big(\Phi(\tilde{f}(\tilde{S})),\lceil\tilde{S}\rceil\big)\quad\text{are continuous on }\mathscr{I}, \tag{10}$$

通常，若映射 Φ 在 $\mathscr{I}$ 上连续（或 $\mathscr{I}$-连续），则称其为连续映射，即 Φ 在 $\mathscr{I}$ 上逐点连续；关于我们在此采用的鲁棒性的一般概念，请参见 [47] 。

上述定义使用了“非对称距离”

$$\operatorname{dist}_{\varrho}(\mathcal{A},\mathcal{B}):=\sup_{a\in\mathcal{A}}\mathrm{d}_{\varrho}(a,\mathcal{B})\quad\text{with}\quad\mathrm{d}_{\varrho}(a,\mathcal{B}):=\inf_{b\in\mathcal{B}}\varrho(a,b) \tag{11}$$

其中 $\varrho\colon\mathcal{M}_1^{\times 2}\to\mathbb{R}_+$，且 $\varrho\circ(\mathsf{id}\times\mathsf{id})=0$，表示 Φ 的输出与 $\lceil\cdot\rceil$中的期望信号之间某种适当的距离函数； ϱ的具体选取取决于$\lceil\cdot\rceil$。

注 2.6. 从纯粹的概念角度看，（预期的）鲁棒性概念为问题（⋆）的传统分析增添了经典的(10)适定性（数值和分析方面）与稳定性，这些内容在纯数学和应用数学的诸多领域中都发挥着核心作用，参见例如 [40, 45, 47], [4, 第 139], 页或 [36,第 IV.12 节3]。

2.2. **独立成分分析**。如上所述，盲源分离鲁棒性（10）的初步描述仍只是非正式的期望，而非精确定义；要给出后者，还需要进一步明确 ϱ 以及合适的拓扑（“$\mathscr{I}$-连续性”）。本小节弥补了这一缺口，为最广为接受的（10）特例提供了相应的完整定义，该特例即独立成分分析（ICA），记为（⋆）。

为作准备，记 $\mathsf{d}\cdot\mathsf{e}_{\mathfrak{m}}\colon\mathsf{M}_1\to 2^{\mathcal{M}_1}$ 为将信号 S 映射到其在单项式群（作用）$\mathsf{M}_d:=\{\Lambda P\mid\Lambda\in\Delta_d,\ P\in\mathsf{P}_d\}$ 上的轨道$\tilde{}\mathsf{d}S\mathsf{e}_{\mathfrak{m}}:=\mathsf{M}_d\cdot S$ 的函数，其中 $_1$, 参见（B.8）。类似地，令 $\tilde{}\mathsf{M}_d:=\{h:u\mapsto Mu+b\mid M\in\mathsf{M}_d,\,b\in\mathbb{R}^d\}$ 为 C^1 上仿射单项式变换的子群 $[,\ \mathbf{1}(\mathbb{R}^d)]$，其中 $\mathsf{d}S\mathsf{e}_{\tilde{\mathfrak{m}}}:=\tilde{}\mathsf{M}_d\cdot S$。另外，记 $\mathfrak{R}:=\{g\colon\mathbb{R}^d\to\mathbb{R}^d\mid g\text{ 在 Borel 意义下可测}\}$ 为 $\mathbb{R}^d$ 上博雷尔可测变换的集合。

下文中，(14) 定义了对 (9) 的细化。注意，如果 $\mathscr{I}\subseteq\mathfrak{C}$ 是 $\lceil\cdot\rceil_{\mathfrak{m}}$-充分的，那么对于 Φ 上的任意反演映射 $\mathscr{I}$，都存在 $\hat{}\Phi:\mathcal{M}_1\to 2^{\mathfrak{R}}$, $X\mapsto\{B\in\mathfrak{R}\mid B\cdot X\in\Phi(X)\}$，使得

$$\Phi(X)\;=\;\hat{\Phi}(X)\cdot X,\quad\text{for each BSS-triple }(X,S,f)_{\mathscr{I}}. \tag{12}$$

在包含于 (⋆) 的逆问题中，迄今为止，下述变体是最成熟的一种，参见例如 [28, 49, 50, 68]。

[9] ~~例如，尽管选择~~ $(\lceil S\rceil\ \mathscr{I})=\big(\{S\},\ \{(S,f)\}\big)$ 很简单，但选择 $(\lceil S\rceil\ \mathscr{I})=\big([S]_X,\mathfrak{C}\big)$ 则毫无意义；具有实际或理论意义的 (⋆)$_{(a)}$ 的解将严格处于这两个极端之间。7

定义 2.7（独立成分分析）。若集合 $\mathscr{I}\subseteq\mathfrak{C}$ 满足以下条件，则称其为 *ICA* 条件：

$$\mathscr{I}=\mathscr{S}\times\mathrm{GL}_d\quad\text{for some}\quad\mathscr{S}\supseteq\hat{\mathscr{S}}_{\perp}\quad\text{and}\quad\mathscr{I}\ \text{ is }\ \lceil\cdot\rceil_{\mathfrak{m}}\text{-sufficient;}\tag{13}$$

一个 BSS 三元组 $(X,\ S,\ f)_{\mathscr{I}}$称为一个*ICA* 三元组（在 $\mathscr{I}$上），如果 $\mathscr{I}$是一个 ICA 条件。在这种情况下，我们将在 $\mathscr{I}$ 上的一个 ICA 逆变换称为任意函数

$$\hat{\Phi}:\mathcal{M}_1\to 2^{\mathfrak{N}}\quad\text{such that:}\quad\emptyset\neq\hat{\Phi}(f(S))\circ f\subseteq\tilde{\mathrm{M}}_d,\quad\text{for each }(S,f)\in\mathscr{I}.\tag{14}$$

独立成分分析（ICA）问题是（$\star$），针对 $\lceil\cdot\rceil\equiv\lceil\cdot\rceil_{\tilde{\mathfrak{m}}}$，并满足 $\mathscr{I}$且 Φ 具有形式（13）和（14）；称相应的一对（$\mathscr{I}$，）为 ICA 问题的解。

注 2.8。换言之，一个 ICA 三元组 $(X,S,f)_{\mathscr{I}}$是一个 BSS 三元组，其中源 S需要恢复到其原始分量经过某种置换和缩放（或许还包括某个常数偏移）后的形式；相应的先验信息 $\mathscr{I}\ni($ $S,f)$由一个关于源的条件（即 $\mathscr{S}$，它描述了各源分量的正则性及其统计独立性）和一个关于混合的线性条件（GL_d）构成，并由 (5) 分别施加。另请注意，如果 $\hat{}\Phi$ 是形式为 (14) 的 ICA 逆变换，那么根据 (9)$\tilde{}$和 (7)，并且在 $\mathrm{M}_d\cdot\mathscr{S}\subseteq\mathscr{S}$成立的情况下，$\hat{}\Phi(X)\cdot X$ 是 $\mathscr{I}$上的一个反演映射，符合定义 2.3。

◆

ICA 的特性（13）和（14）使得对一般鲁棒性概念（10）进行自然的解析性解释成为可能。为此，令 $(\mathscr{I},\hat{}\Phi)$ 为任意一个固定的 ICA 解，设 $\Phi\colon\boldsymbol{X}\mapsto\hat{}\Phi(X)\cdot\boldsymbol{X}$，并通过“偏差函数”量化准确性条件（14）的（偏离程度）

$$\begin{aligned}\partial_{\hat{\Phi}}:\mathfrak{C}\to\mathbb{R}_+\quad\text{given by}\quad\partial_{\hat{\Phi}}(S,f)&:=\sup_{B\in\hat{\Phi}(f(S))}\inf_{(M,v)\in\mathrm{M}_d\times\mathbb{R}^d}\varrho_B((S,f),(M,v))\\ \text{for}\quad\varrho_B((S,f),(M,v))&:=\sup_{u\in D_S^{(v)}}\tfrac{|Bof(u-v)-Mu|}{|Mu|}\mathbb{1}_{\times}(Mu),\end{aligned}\tag{15}$$

with $\mathbb{1}_{\times}:=\mathbb{1}_{\mathbb{R}^d\setminus\{0\}}$ 以及 $D_S^{(v)}:=v+D_S$。为理解下文中的不等式（16），请回顾符号约定（11）.

引理 2.9。映射 $\partial_{\hat{\Phi}}$ 在 $\mathscr{I}$上为零，即 $\partial_{\hat{\Phi}}\big|_{\mathscr{I}}=0$；对于任意函数 f 以及每个满足 $(S,f)\in\mathfrak{C}$ 的随机过程 $\boldsymbol{S}=(\boldsymbol{S}_t)$，几乎必然有：

$$\mathrm{dist}_{\tilde{\varrho}}(\Phi(f(\boldsymbol{S})),\lceil\boldsymbol{S}\rceil_{\tilde{\mathfrak{m}}})\ \leq\ \partial_{\hat{\Phi}}(S,f)\quad\textit{for}\quad\tilde{\varrho}(\boldsymbol{X},\boldsymbol{Y}):=\sup_{t\in\mathbb{I}}\tfrac{|\boldsymbol{X}_t-\boldsymbol{Y}_t|}{|\boldsymbol{Y}_t|}\mathbb{1}_{\times}(\boldsymbol{Y}_t).\tag{16}$$

由于 Φ 的输出与期望精度 $\lceil\cdot\rceil_{\tilde{\mathfrak{m}}}$ 之间的相对误差受 $\partial_{\hat{\Phi}}$控制，我们现在可以给出（10）在 ICA 特例下预期的形式化表述。请注意，以下内容自然结合了 BSS 特有的稳定性要求（10）与 [47, 第 1.2–1.4、2 节6]中阐述的经典统计稳健性概念。

Defi义 2.10（鲁棒独立成分分析）。ICA 问题的一个解（$\mathscr{I}$，$\hat{}\Phi$）称为鲁棒的 t if

$$\partial_{\hat{\Phi}}:\mathfrak{C}\to\mathbb{R}_+\quad\text{is continuous on }\mathscr{I}\tag{17}$$

相对于因果空间 $\mathfrak{C}$上的一个足够粗的拓扑而言。这里，$\tau_{\mathfrak{C}}$ 上的一个拓扑是指

足够粗的，如果（$\tau_{\mathfrak{C}}$，$\tilde{\pi}_1$）诱导的最终拓扑在$\mathcal{M}_1$ 上比 $\tau^{\star}$，$\tilde{}$ $\tilde{}$ $\tilde{}$更粗，其中 $\tilde{\pi}_1$:(18)
$\mathfrak{C}\to\mathcal{M}_1,(S,\ f)\mapsto S$是从原因到其信号的典范投影，而 $\tau^{\star}$是$\mathcal{M}_1(\mathcal{C}_d)$上的弱收敛拓扑[10] 。

除了定性论断（17）之外，通常还希望对这种鲁棒性进行明确的定量分析（例如参见 [47, 第 1.3f 节。]）。只要 $\mathfrak{C}$ 上的底层拓扑具有足够的信息量，就可以通过为 $\partial_{\hat{\Phi}}$ 在 $\mathscr{I}$处补充一个连续模来实现这一点，如下文第5 节所述。

[10] 例如参见 [12, 定义 8.1。1] 和 [10,第 7]节8

注释 2.11（“足够粗”）。直观地说， $\mathfrak{C}$ 上的任何一种使条件（17）成立的拓扑，都可以用来刻画 Φ 的 $\mathsf{d}\cdot\mathsf{e}_{\tilde{\mathfrak{m}}}$-精度对假设 $\mathscr{I}$遭到违背的反应有多“敏感”：这种拓扑 [越粗，也就是所包含的开集越少], ， Φ 在 $\mathscr{I}$发生一般扰动时的变化就越不敏感（“越鲁棒”）。如果相对于 $\mathfrak{C}$ 上的离散拓扑（即最细拓扑）来理解，连续性要求（17）便是多余的，因此，显然需要 $\mathfrak{C}$ 上的支持（17）的拓扑达到一定的粗糙程度。然而， $\mathfrak{C}$上的此类拓扑并没有显而易见或公认的选择，因此，为“粗到何种程度才算足够粗”设定阈值，一向取决于判断；与鲁棒性概念的其他用法一样 [43, 47], ，这通常取决于人们所关注的“ Φ 在 $\mathscr{I}$遭到违背时的稳定性”所处的具体（应用）情境。要找出一种能最佳涵盖所有可设想违背情形的通用拓扑，或普适的粗糙程度，通常并不容易。不过，以经典理论作为粗略指引（例如参见 [47, 第 1.3、2.6], 节），我们可以期望通过上述规格（18）涵盖“绝大多数”相关违背情形。但我们强调，（18）仅仅是一个参考依据，不应被视为（17）的必要或严格约束；定义 2.10 并不限于这类拓扑。事实上，在接下来的各节中，将证明（17）对 $\tau_{\mathfrak{C}}$ 上的一种拓扑成立。该拓扑在大多数应用场景下满足（18），但通常略微细于（18）；尽管如此，它仍足够粗，能够有效地刻画许多在理论和实践中都相关的 IA 违背（及其“趋于精确”），例如参见例 1.1 和第 6 节。

◆

有了盲源分离问题（$\star$）的可操作表述（1）——该问题通过 ICA（13）来处理——以及与之相关的鲁棒性概念（10），本文余下部分的总体目标便是让这一框架落地并充分发挥作用：在第 4 节中，我们将在 $\mathfrak{C}$ 上给出一种自然、易于处理且信息丰富的（预度量）粗拓扑，以支持一般情形下导出的 ICA 解 $(\mathscr{I},\ \hat{}\Phi)$ 的稳定性概念（10），并借助显式的连续模对由此得到的该解的盲源分离鲁棒性（17）进行全面量化，见定理 5.8 和 5.12。第 6 节中介绍的若干与应用相关的鲁棒性推论，展示了该理论的实际意义。

对于定义在 $\mathfrak{C}$上的上述拓扑，以及我们前述的整体鲁棒性方案而言，至关重要的是一组特殊的“坐标”，用于描述$\mathcal{M}_1$中的信号。这使得对$\mathcal{M}_1$ 的量化具有结构性且可显式计算；下一节将研究这种坐标化所刻画并赋予的拓扑。

3. 可坐标化信号及其收敛

我们在$\mathcal{M}_1$ 上引入坐标，其形式为矩型统计量。这些统计量通过反复将信号的样本路径与自身积分来定义。为使其定义良好，信号的实现必须具有一定的光滑性。为此，我们在$\mathcal{C}_d$ 中考虑若干“正则性类”，并假设可坐标化信号的样本属于这些类。一个表达力强且便于使用的[11]此类是子空间

$$\mathcal{C}^1 \equiv \mathcal{C}^1_d := \left\{x : [0,1] \to \mathbb{R}^d \;\middle|\; x \text{ is absolutely continuous}^{12}\right\} \tag{19}$$

$\mathcal{C}_d$的, 其样本属于该类的信号则是子空间的元素 ce

$$\mathfrak{D} := \{\mu \in \mathcal{M}_1(\mathcal{C}_d) \mid \mu(\mathcal{C}^1) = 1\} \tag{20}$$

的 $\mathcal{M}_1$, ，即由支撑在 $\mathcal{C}^1$ 上的、定义于 $(\mathcal{C}_d,\|\cdot\|_\infty)$ 上的所有博雷尔概率测度构成的集合。记$C^{1,1}(\mathbb{R}^d) =: C^{1,1}$ 为 C^1 上所有 $\mathbb{R}^d$-微分同胚构成的群，注意空间 $\mathfrak{D}$ 是左

[11] 。尽管它当然不是最大的可能正则性类，参见注 B.2 ...[12] 关于绝对连续性的定义，参见例如 [34, 定义 1。17]。

在 $C^{1,1}$-作用下不变（A.7），该作用于$\mathcal{M}_1(\mathcal{C}_d)$ （由链式法则可知，参见（21））。
在第 3.1 节中，我们将确立空间 $\mathcal{C}^1$, 的一些重要结构性质，并且还会看到，该空间包含所有离散时间观测到的信号。随后，我们将在第 $\mathfrak{D}$ 节 3.2介绍前面提到的信号坐标，并在第 3.3 节和3.4 节研究这些坐标在多大程度上能够刻画 $\mathcal{M}_1$ 上的一些自然拓扑。

3.1. **样本空间的性质**。 路径空间 $\mathcal{C}^1$ 具有某些拓扑正则性性质，因此适合作为信号的样本空间。不过需要指出，本文的主要思想很容易推广到样本路径正则性低得多的信号，参见本小节末尾的注 B.2 及其中的参考文献。

下述引理的内容众所周知。

引理 3.1。 上述空间 $\mathcal{C}^1$ 中的 $\mathbb{R}^d$ 绝对连续路径可写为

$$\mathcal{C}^1 = \left\{ x \in \mathcal{C}_d \;\middle|\; \exists!\, \dot{x} \in L^1([0,1];\mathbb{R}^d) \,:\, x = x_0 + \int_0^{\cdot} \dot{x}_s \,\mathrm{d}s \right\}; \tag{21}$$

该空间是$(\mathcal{C}_d, k\cdot k_\infty)$ 的一个博雷尔子集，并且关于 *1*-变差范数构成一个可分巴拿赫空间

$$\|x\|_{1\text{-var}} \;=\; |x_0| \;+\; \|\dot{x}\|_{L^1}. \tag{22}$$

证明。由 [34, 命题 1.31 和 1.32]，等式 (21) 成立。事实上， [34, 命题 1.31] 指出，对于 $\mathfrak{X} := \mathbb{R}^d \times L^1([0,1];\mathbb{R}^d)$ 和 $\mathfrak{Y} := \mathsf{C}_d$，映射 $f\colon \mathfrak{X} \to \mathfrak{Y}$ 由 $f(c,v) := c + \int_0^{\cdot} v_s \,\mathrm{d}s$ 给出，是一个巴拿赫空间同构（这也证明了 (22)）。由此， [56, 定理 15.1] 推出，像 $f(\mathfrak{X}) = \mathsf{C}^1$ 是 $(\mathsf{C}^1, \mathrm{k}\cdot\mathrm{k}_\infty)$ 的博雷尔子集，即 $\mathsf{C}^1 \in \mathsf{B}(\mathsf{C}_d)$。 [34, 推论 1.35]指出 $(\mathsf{C}^1, \mathrm{k}\cdot\mathrm{k}\text{-var})$ 是可分的巴拿赫空间。□

注3.2。 回顾一下，范数 $\|\cdot\|_{1\text{-var}}$实际上支配着一整族 $(\|\cdot\|_{p\text{-var}})_{p\in[1,\infty]}$ p-变差范数，它们定义在 $\mathcal{C}^1$ 上，是对 (22) 的推广（参见 [34, 第 5 节。1]），其形式如下：

$$\|x\|_{p\text{-var}} = |x_0| + \left[\sup_{(t_\nu)} \sum_{\nu} |x_{t_{\nu+1}} - x_{t_\nu}|^p \right]^{1/p} \qquad \text{for} \quad 1 \le p < \infty \tag{23}$$

以及 $\|\cdot\|_{\infty\text{-var}} := \|\cdot\|_\infty$，其中上述上确界取遍$[0,1]$的所有（有限）剖分 (t_ν)。由于这些范数会随着 p 增大而变弱，即对于每个$1 \le p \le q \le \infty$，都有$\|\cdot\|_{p\text{-var}} \ge \|\cdot\|_{q\text{-var}}$ （例如， [34, 命题 5.3] 和 [65, 引理 1.6.4.]），因此我们有以下子空间包含关系

$$C_b\big(\mathcal{C}^1; \|\cdot\|_{1\text{-var}}\big) \supseteq C_b\big(\mathcal{C}^1; \|\cdot\|_{p\text{-var}}\big) \supseteq C_b\big(\mathcal{C}^1; \|\cdot\|_{q\text{-var}}\big) \quad \text{for} \quad 1 \le p \le q \le \infty. \tag{24}$$

尽管 $\mathcal{C}^1$ 上的一致拓扑弱于其 1-变差拓扑，但值得注意的是，两者在 $\mathcal{C}^1$上诱导相同的可测结构。这基本上由 $23)|_{p=1}$ 中的定义，即 1-变差范数的定义，以及上述 $\mathcal{C}^1$ 关于 $\|\cdot\|_{1\text{-var}}$ 可分这一事实得出。

Le引理 3.3。 空间$(\mathcal{C}^1, k\cdot k_\infty)$ 和$(\mathcal{C}^1, k\cdot k_{1\text{-var}})$ 具有相同的 *Borel* σ-代数。

证明。 这基本上可由 23 式关于 1-变差范数的恒等式，以及 $\mathcal{C}^1$ 关于 $\|\cdot\|_{1\text{-var}}$ 可分这一事实（引理 3.1）推出，参见附录 C.2。□

有时，观测量 $\boldsymbol{X} = (\boldsymbol{X}_t)$ 在（2）中会预先建模为 $\mathbb{R}^d$ 中的随机向量或离散时间序列，而非连续时间随机过程。根据以下对应关系，这类离散时间模型仍是我们形式体系的特例。

注 3.4 （离散信号）。 对于 $Z := \mathbb{R}^d$ 和 $\hat{\mathcal{C}}_0 := \{x \in \mathcal{C}_{(0,1)} \mid x_0 = 0\}$，考虑映射

$$\hat{\iota}_1 \,:\, \mathbb{R}^d \ni u \overset{\sim}{\longmapsto} (t\cdot u)_{t\in[0,1]} \in \hat{\mathcal{C}}_0 \quad \text{and} \quad \hat{\iota}_n \,:\, Z^{\times n} \overset{\sim}{\longrightarrow} \hat{\mathcal{C}}_{\mathcal{E}_n} \quad (n \ge 2), \tag{25}$$

其中，对于 $[0,1]$ 的剖分 $\mathcal{I} = (t_\nu \mid 0 = t_0 < \ldots < t_{n-1} = 1)$，我们定义子空间（属于 $\mathcal{C}^1$）

$$\hat{\mathcal{C}}_{\mathcal{I}} := \left\{ x = (x_t)_{t\in[0,1]} \;\middle|\; x_t = x_{t_\nu} + \tfrac{t-t_\nu}{t_{\nu+1}-t_\nu}(x_{t_{\nu+1}} - x_{t_\nu}) \text{ for } t \in [t_\nu, t_{\nu+1}),\, \nu \in [n-1]_0 \right\} \quad (26)$$

（从 $[0,1]$ 到 $\mathbb{R}^d$的所有 I-分段线性路径所构成的空间），并将 $\hat{\iota}_n$ 记为映射，它将元组 $(z_\nu) \in Z^{\times n}$ 映射为沿 $\mathrm{E}_n := \big((\nu-1)/(n-1) \;\big|\; \nu \in [n]\big)$ 进行分段线性插值（ $[0,1]$的等距剖分；参见注 A.1）。因此，(25) 中的映射是其定义域与陪域之间的对应，因为它们是（关于 $n=1$ 的等距）巴拿赫同构，其逆矩阵 $\hat{}\pi_n := \hat{\iota}_n^{-1}$由 $\hat{}\pi_1 := \pi_1$ （分别为 $\hat{}\pi_n = \pi_{\mathcal{E}_n}$ ）给出，其中 $n \geq 2$。对于任意「离散信号」 $\upsilon \in \mathrm{M}_1(Z^{\times n})$，其连续时间对应是推前 $\hat{}\upsilon := (\hat{\iota}_1)_*\upsilon \in \widehat{\mathfrak{D}}_1 :\cong \{ \mu \in \mathrm{M}_1(\mathrm{C}^1) \mid \mu(\hat{}C_0) = 1\}$（若 $n=1$，则分别为）。

$$\hat{\upsilon} := (\hat{\iota}_n)_*\upsilon \in \widehat{\mathfrak{D}}_{\mathcal{E}_n} := \big\{ \mu \in \mathcal{M}_1(\mathcal{C}^1) \;\big|\; \mu(\hat{\mathcal{C}}_{\mathcal{E}_n}) = 1 \big\} \quad (\text{for } n \geq 2).$$

对于（“无时间”）情形 $n=1$，28式中的信号$\hat{}\upsilon$ 的签名矩与经典的[多元] 矩 υ 一致，二者相差一个因子 $1/m!$。 ◆

3.2. **签名矩**。 现在来定义我们前面提到的信号 μ 在 $\mathfrak{D}$ 中的坐标 .

这些坐标背后的关键思想是，信号 μ 的每个样本 $x \in \mathcal{C}^1$ 都可以用迭代积分统计量作简洁的非局部（关于时间的）描述

$$x = \big(x_t^1, \cdots, x_t^d\big)_{t\in[0,1]} \longmapsto \int_{0\leq t_1 \leq \cdots \leq t_m \leq 1} \mathrm{d}x_{t_1}^{i_1} \cdots \mathrm{d}x_{t_m}^{i_m} =: \mathfrak{sig}_{i_1\cdots i_m}(x) \quad (27)$$

对于 i_1、……、 $i_m \in [d]$、 $m \in \mathbb{N}$ 和 $\mathfrak{sig}_\emptyset(x) := 1$；由于每个 $x \in \mathrm{C}^1$ 都是有界变差的，积分（27）在勒贝格–斯蒂尔杰斯意义下存在。这些数称为 x 的签名系数，它们共同构成一个多重指标族 $\mathfrak{sig}(x) := \big(\mathfrak{sig}_w(x) \mid w \in [d]^\star\big)$，称为 x 的签名。（注意 $\mathfrak{sig}_{i_1\cdots i_m}$ 与矩映射 $x \mapsto x_0^{i_1}\cdots x_0^{i_m}$,1 在形式上相似，因此我们将（27）称为 x 的“非交换矩”。）

已有研究表明， [42]签名实际上能够刻画[13] 其路径，而且刻画方式颇具信息量，参见例如 [13, 21, 24, 22, 29, 60, 76] 以及其他许多文献。因此，可以将 $\mathfrak{sig}(\cdot)$ 视为$\mathcal{C}^1$, 上的（非线性）“坐标系”，其中每个“坐标向量” $\mathfrak{sig}(x)$ 都能唯一确定点 $x \in \mathcal{C}^1$ 的位置[13]。随后，可以通过对偶，将这种基于签名的路径描述（27）自然地从$\mathcal{C}^1$ 转移到 $\mathfrak{D}$ 中的信号上（通过用有序族非线性泛函（27）对 $\mathfrak{D}$ 中的测度进行检验）。

$\mathfrak{D}$ 上的“对偶坐标”将在下方的定义中引入。

引理 3.5。 映射 $(\mathcal{C}^1, \|\cdot\|_\infty) \ni x \mapsto \mathfrak{sig}_w(x) \in \mathbb{R}$ 对每个 $w \in [d]^\star$都是博雷尔可测的。证明。由引理 3.3 以及如下事实（例如， [66, 定理 3.1。3]）： $\mathfrak{sig}_w$关于 $\|\cdot\|_{1\text{-var}}$连续。 □

¯**定义 3.6** (签名矩)。 给定信号 μ 属于 $\mathfrak{D}$，这些数（属于 $\mathbb{R}$）

$$\langle\mu\rangle_{i_1\ldots i_m} := \int_{\mathcal{C}_d} \mathfrak{sig}_{i_1\ldots i_m}(x)\, \mu(\mathrm{d}x) \qquad (i_1, \ldots, i_m \in [d],\; m \in \mathbb{N}) \quad (28)$$

称为 μ的签名矩。我们将 μ 的核心坐标称为矩阵

$$[\mu]_0 := \big(\langle\mu\rangle_{ij}\big)_{i,j\in[d]} \quad \text{and} \quad [\mu]_\nu := \big(\langle\mu\rangle_{ij\nu}\big)_{i,j\in[d]} \quad (\nu = 1, \ldots, d), \quad (29)$$

并记 $[\mu]_{\mathfrak{c}} := ([\mu]_0,\ [\mu]_1、\ \cdots\ [\mu]_d)$ 为它们的有序集合。若 $[\mu]_{\mathfrak{c}}$ 存在，记作： $[\mu]_{\mathfrak{c}} < \infty$，表示（29）中的所有签名矩都存在于 $\mathbb{R}$中。

[13] ……直到在 $\mathcal{C}^1$ ……上的一个无关紧要的（类似于“模零测集相等”）等价关系。11

（“核心坐标”一词是“核心坐标”的混成词。）

由于基于（27）的$\mathcal{C}^1$, 坐标化具有上述特征性，签名矩（28）若包含于 $\mathbb{R}$，便能唯一地表征一个信号，参见 [23]。此外，由于族$(\mathfrak{sig}_w \,|\, w \in [d]^\star)$内存在代数相互关系，矩（28）还能高效编码由任意路径变换 $\mathfrak{D}$ （参见（A.7））所诱导的推前作用$f_*\colon\ \mathcal{C}^1 \to \mathcal{C}^1$ （例如 [22, 定理 5。1]）。事实上，核心坐标（29）已能捕捉线性作用

$$\mathbb{R}^{d\times d} \ni A \longmapsto A\cdot\mu \equiv A_*\mu := \mu\circ A^{-1} \tag{30}$$

足够：将三阶张量展平为矩阵 $\big(\langle\mu\rangle_{ijk}\big)$ ， $[\mu]_\nu$ 在（29）中得出一族仿射等变统计量（见下文）；这种等变性是 $[\mu]_\mathfrak{c}$的一种简单却极其有用的代数性质，并且结合对 μ 的温和假设，可用于研究（30）所示的作用，适用于一大类（非 IC）信号以及非线性扰动下的情形。

引理 3.7 (仿射等变性**)**。 令 $\mu \in \mathfrak{D}$ 并定义 $f \equiv f(u)\colon = A\cdot\, u + b$ 在 $\mathbb{R}^d$ 上，其中某个 $A = (a_{ij}) \in \mathbb{R}^{d\times d}$和 $b = (b_i) \in \mathbb{R}^d$。那么对于任意 $i_1, \dots, i_m \in [d]$ 和 $m \in \mathbb{N}$,

$$\big\langle f_*\mu\big\rangle_{i_1\dots i_m} = \sum_{j_1,\dots,j_m=1}^{d} a_{i_1 j_1}\cdots a_{i_m j_m}\langle\mu\rangle_{j_1\dots j_m}. \tag{31}$$

证明。 由迭代斯蒂尔杰斯积分的多重线性显然可得，但请参见附录 C.3。 □

签名统计量（29）与许多阶数相当的经典信号统计量具有相同的等变性（31），例如协方差矩阵、经典多元矩或累积量；这正是后者在许多经典独立成分分析方法中发挥作用的基础 [28, 50, 68]。因此，如有需要，也可以直接使用这些经典统计量来代替（28）；在这种情况下，只需将（29）替换为它们中任意一种（仿射等变的）经典对应项，本文提出的拓扑鲁棒性方法仍同样成立；另见注 3.4。由此引出以下注。

注 3.8 （为何使用签名？）。 与随机过程的经典（“可交换”）矩统计量截然不同，信号的核心坐标（29）是该信号关于时间的“全局” μ 统计量 []，也就是说，它们是能够穷尽整个信号（及其分布）的函数，而不像经典统计量那样，仅仅描述在少数预先选定的时间点上观测到的信号的有限维分布。信号统计量（28）的这种全局性，使其能够比经典（“固定t”）矩统计量更自然、更有效地捕捉非局部统计效应（例如时变噪声，或异步采样的通道数据导致的估计误差，参见第6节）。此外，推导得到的统计量（29）通常[14] 每个矩阵所承载的信息也多于同维数的经典统计量 [14], ，这一点体现为矩阵（29）通常是不对称的，因此包含的信息量大约是由可交换矩得到的同维数独立成分分析统计量的两倍；后者通常都是对称的。正是这些特点将在下文中加以阐释和利用，使信号坐标（28）及其推导出的统计量（29）成为一种十分自然且信息丰富的量化方式，用以刻画多维信号及其（非）线性变换内部及彼此之间的统计依赖关系。凭借这些优点，我们将看到，这些坐标为构建所需的、针对（17）的稳健性拓扑提供了极佳的基础。

◆

[14] 例外情况是均值平稳乘积信号的坐标矩阵及其线性变换（根据（69），两者与经典矩统计量一致）；参见引理4.8。12

3.3. **信号的收敛**。 我们通过引入一种自然的 p变差分级弱收敛概念，来形式化不同信号之间在拓扑意义上的关联方式$\mathfrak{D}$。由此得到的信号拓扑至多比 $\mathfrak{D}$ 上的经典弱收敛拓扑逐渐更强；事实上，在许多应用情境中，它与后者相同。

首先 ，我们回顾一下[15]，在 $\mathfrak{D}$ 中的信号网 (μ_α) 若满足以下条件，则称其弱收敛到 $\mu \in \mathfrak{D}$

$$\lim_\alpha \mu_\alpha(\tilde{f}) = \mu(\tilde{f}) \quad \text{for each } \tilde{f} \in C_{\mathrm{b}}(\mathcal{C}^1; \|\cdot\|_\infty), \quad \text{where} \quad \mu_\alpha(\tilde{f}) := \int_{\mathcal{C}^1} \tilde{f}\, \mathrm{d}\mu_\alpha \tag{32}$$

且 $C_b\big(\mathcal{C}^1;\ \|\cdot\|_\infty\big)$ 表示有界$\|\cdot\|_\infty$-连续函数 $\tilde{f}\colon \mathcal{C}^1 \to \mathbb{R}$ 的空间
a .

从上面的定义可以看出，弱收敛的概念本质上是一个拓扑概念：虽然 $\mathcal{C}^1$ 上不同的拓扑可能诱导出相同的博雷尔 σ-代数，但它们所提供的（有界）连续函数集合，进而它们在 $\mathfrak{D}$ 上诱导的“弱收敛模式”，可能大不相同。

引理 3.3 与此相关，因为它意味着信号空间 $\mathrm{M}_1(\mathrm{C}^1) \equiv \mathrm{M}_1\big((\mathrm{C}^1, \mathrm{k}\cdot\mathrm{k}_\infty)\big) \cong \mathfrak{D}$ 和 $\mathrm{M}_1\big((\mathrm{C}^1, \|\cdot\|_{1\text{-var}})\big)$ 相同。这使我们能够将每个测度 $\mu \in \mathrm{M}_1\big(\mathrm{C}^1\big)$ 与空间 $C_b\big(\mathrm{C}^1; \|\cdot\|_{1\text{-var}}\big)$ 中的任意元素进行积分；该空间由有界 $\|\cdot\|_{1\text{-var}}$-连续函数 $\tilde{f}\colon \mathrm{C}^1 \to \mathbb{R}$ 构成。

如前所述，参见 [12, 定义 8.1。1],由此可得出逐点（“弱”）收敛的概念

$$\mu_\alpha \xrightarrow{1\text{-var}} \mu \quad :\Longleftrightarrow \quad \Big[\lim_\alpha \mu_\alpha(\tilde{f}) = \mu(\tilde{f}),\ \ \forall\, \tilde{f} \in C_b\big(\mathcal{C}^1; \|\cdot\|_{1\text{-var}}\big)\Big] \tag{33}$$

对于网 $(\mu_\alpha)\ \subset \mathcal{M}_1(\mathcal{C}^1)$ 和 $\mu \in \mathcal{M}_1(\mathcal{C}^1)$。得益于包含关系 (24)，可以将这种收敛系统地推广为以下更弱的收敛概念。

定义 3.9 (p-弱拓扑)**.** 设 $p \in [1, \infty)$。网 $(\mu_\alpha) \subset \mathfrak{D}$ 称为收敛于信号 $\mu \in \mathfrak{D}$ 的 p-弱收敛，记作：$\mu_\alpha \xrightarrow{p\text{-var}} \mu$，若

$$\lim_\alpha \mu_\alpha(\tilde{f}) = \mu(\tilde{f}) \quad \text{for each } \tilde{f} \in C_b\big(\mathcal{C}^1; \|\cdot\|_{p\text{-var}}\big). \tag{34}$$

网的收敛 (34自然地[16] 定义了 $\mathfrak{D}$ 上的唯一拓扑，我们称之为p-弱拓扑；对于 q，其 $1 \le q < 2$ -弱拓扑称为弱'拓扑。

换言之，(34恰好是(μ_α) 关于 μ 相对于 p在 $\mathcal{C}^1$ 上的变差拓扑的弱收敛（参见 [12, 第 8 节1]）。以 p为指标的收敛 (34之间具有自然的层级关系：对于任意 $1 \le p \le q < \infty$， p-弱拓扑（根据 (24)）强于 q-弱拓扑，而后者又强于（即由 (32) 诱导的）经典弱拓扑$\mathcal{M}_1(\mathcal{C}^1)$ 但这种强弱差异是逐渐体现的，详见下面的注。

注 3.10。 范数 $\|\cdot\|_{p\text{-var}}$ $(1 < p < \infty)$ 和 $\|\cdot\|_\infty$ 可通过经典插值不等式（例如 [34, 命题 5.5]）联系起来；该不等式对每个 $x \in \mathcal{C}_d$都意味着

$$\|x\|_\infty \ \le\ \|x\|_{p\text{-var}} \ \le\ 2\|x\|_{1\text{-var}}^{\frac{1}{p}} \cdot \|x\|_\infty^{1-\frac{1}{p}} \quad \text{for each } p \in (1, \infty) \tag{35}$$

因此，范数 $\|\cdot\|_{p\text{-var}}$ 和 $\|\cdot\|_\infty$ 在极限 $p \to \infty$下渐近等价。特别是，测试函数的收敛定义集 $C_b\big(\mathcal{C}^1;\|\cdot\|_{p\text{-var}}\big)$ 和 $C_b\big(\mathcal{C}^1;\|\cdot\|_\infty\big)$在 $p \to \infty$时“渐近重合”，因此在此意义下，我们可以认为式（34）和式（32）分别定义的收敛拓扑在 p 较大时“实质上等价”。 ◆

[15] 例如参见 [12, 定义 8.1。1] 和 [10,第 7]节。[16] 例如，参见 [57, 第 2 章（第 73 页）]中关于收敛类的章节。另见该处，参见 [57, 定理 2。9], 收敛网的概念（按（33）诱导的 p-弱拓扑定义）与（33）所描述的收敛概念完全相同。

（强度上的）先验差距： p-弱拓扑与经典 [即（32）诱导的] 弱拓扑在 $\mathfrak{D}$ 之间，主要只是理论上的，对许多实际应用并无影响。这是因为对于许多具有实际意义的信号空间， p-弱拓扑与经典弱拓扑重合，如下例所示。

Pro位置 3.11. 对于任意 $p \in [1,\infty)$ 和 $K > 0$，考虑“紧化”的信号空间 s

$$\mathfrak{D}^p_K := \{\mu \in \mathcal{M}_1(\mathcal{C}^1) \mid \mu(\mathcal{C}^1_{p,K}) = 1\} \quad for \quad \mathcal{C}^1_{p,K} := \{x \in \mathcal{C}^1 \mid \|x\|_{p\text{-var}} \le K\}, \tag{36}$$

并且对于 $[0,1]$ 的任意剖分 I，考虑信号

$$\widehat{\mathfrak{D}}_{\mathcal{I}} := \{\mu \in \mathcal{M}_1(\mathcal{C}^1) \mid \mu(\hat{\mathcal{C}}_{\mathcal{I}}) = 1\} \quad supported\ on \quad \hat{\mathcal{C}}_{\mathcal{I}}, \tag{37}$$

从 $\mathcal{I}$ 到 (26) 的分段线性路径空间。在这些空间上，经典弱拓扑等价于：对任意 q，在 $\mathfrak{D}^p_K$ 上的 $q > p$-弱拓扑；以及在 1 上的弱拓扑 $\widehat{\mathfrak{D}}_{\mathcal{I}}$。

证明。 这基本上可由对（35）的推广得到，见附录 C.4。 □

上述信号空间（36）和（37）之所以具有实际意义，是因为现实世界中的信号处理系统总是受到容量限制，例如数据存储空间有限，以及记录的观测量分辨率有限；例如 [31, 第 5, 12] 节。前一种约束会导出（36），而后一种约束则要求对信号进行（有限时域）时间离散化，如（37）所示，参见例如 [73]。

3.4. 签名矩的收敛。 本节说明，在自然的增长条件下，信号的签名坐标关于上述信号拓扑是连续的。

为了将第 3.1 节和 3.3 节的考量结合起来，为（17）构造一个有意义的鲁棒性拓扑，我们需要将信号的收敛（34）与其矩的收敛（28）联系起来。与积分统计量的通常做法一样，这可以通过施加增长条件来实现。

D定义 3.12。 若子集 $\mathfrak{U}$ <style id='6}>属于 $\mathfrak{D}$，则称其为一致签名可积的，阶数为 m，如果

$$\inf_{a>0} \sup_{\mu\in\mathfrak{U}} \int_{\{|\mathfrak{sig}_w|>a\}} |\mathfrak{sig}_w| \, \mathrm{d}\mu \;=\; 0 \quad \text{for each } w \in [d]^\star_{\le m}. \tag{38}$$

如果 $\mathfrak{U}$ 是一致签名可积的，且 $\mathfrak{U}$ 是 3 阶一致签名可积的，我们就称其为一致签名可积。

简单来说，若一组信号（阶数为 m）中的元素以一致较低的概率取到那些签名系数（不超过 m阶）非常大的样本路径，则称该组信号是一致签名可积的。增长假设（38）确保在此类空间中，矩坐标（28）关于任意弱’拓扑（34）都是连续的|$_{p\in[1,2)}$。

引理 3.13. 设 $\mathfrak{U}$ 为 m阶一致签名可积，并且 $(\mu_\alpha) \subset \mathfrak{U}$ 是任意一个满足 $\mu_\alpha \to \mu \in \mathfrak{U}$ 在弱’拓扑下收敛的网。那么，对每个 $w \in [d]^\star_{\le m}$，$\boldsymbol{h}\mu_\alpha\boldsymbol{i}_w \to \boldsymbol{h}\mu\boldsymbol{i}_w$。

证明. 由于 $\langle\tilde{\mu}\rangle_w = \int_{\mathcal{C}^1} \mathfrak{sig}_w \mathrm{d}\tilde{\mu}$，且 $\mathfrak{sig}_w$ 关于 $\mathcal{C}^1$ 上的 p-变差拓扑连续（对任意 $p \in [1,2)$，例如参见 [66, 定理 3.1.3]，，由 [12, 引理 8.4.3]可得所述结论。□

作为一个直接的[17] 例子，任意阶的一致签名可积空间包括 $\mathfrak{D}$ 中所有元素具有一致 $\|\cdot\|$ $_{p\text{-var}}$-有界支撑的子集，参见命题 3.11。借鉴一致可积性的经典概念，这些一致签名可积空间的有界支撑示例可轻松推广如下。

[17] 回顾一下，对于任意 $x \in \mathcal{C}^1$ 和 $p \in [1,2$，$|\mathfrak{sig}_w(x)| \lesssim \|x\|^{p|w|}_{p\text{-var}}$，例如参见 [65, 定理 3.7 和第 1.2 节.2]；注意，所述不等式还蕴含（对所有 $\mu \in \mathcal{M}_1$）： $\int c_d \|x\|^3_{p\text{-var}} \, \mu(\mathrm{d}x) < \infty \implies [\mu]_c < \infty$。

命题 3.14。 设 $m \in \mathbb{N}$，并设 $\mathfrak{U}$ 是 $\mathfrak{D}$ 的一个子集，存在 $q > 1$ 使得

$$\sup_{\mu \in \mathfrak{U}} \int_{\mathcal{C}_d} \|x\|_{1\text{-var}}^{mq} \, \mu(\mathrm{d}x) < \infty. \tag{39}$$

那么， $\mathfrak{U}$ 是阶数为 m 的一致签名可积。

证明。 这可由若干标准不等式组合得出，参见附录 C.5。 □

3.4.1. 可积性。 对鲁棒性（17）进行分析，需要捕捉非线性变换如何通过（ $\mathbb{R}^d$ 影响邻近的源信号 [（A.7）]。为此，且仅为此（见注释4.14、E.1 (ii)），有必要通过对其强矩的一致可积性（ $\mathfrak{D}$ 39）加以要求，来强化局部性的概念，使信号（的核心坐标）具有可比性：**符号约定**。集合 $\mathfrak{U}$ 在 $\mathfrak{D}$ 中称为 核心可积，如果它是一致签名可积的，且

$$K_{\mathfrak{U}|\beta} := \sup_{\mu \in \mathfrak{U}} \int_{\mathcal{C}_d} \|x\|_{1\text{-var}}^{\beta} \, \mu(\mathrm{d}x) < \infty \quad \text{for some } \beta \in (5/2, 3). \tag{40}$$

对于（40）中的任意固定 β，记 $p_\beta := (\beta - 2)^{-1}$，并记 $\alpha = \alpha_\beta := 1 - 1/p_\beta$ [$\in (0, 1/2)$]。

若子集 $\mathfrak{U} \subseteq \mathfrak{D}$ 满足某个固定的 $\beta \in (5/2, 3)$ 及上文定义的相关常数 $p = p_\beta$ 和 $\alpha = \alpha_\beta$，使得（40）成立，则称其按照（$K_{\mathfrak{U}|\beta}, \beta, p, \alpha$）是核心可积的；对于任意这样的元组，我们还定义 $c_p := (1 - 2^{1-2/p})^{-1}$ 和 $C_p := 2^{\alpha+1} c_p (1 + K_{\mathfrak{U}|\beta})$。 ◆

一个直接的例子是，子空间 $\mathfrak{U} \subseteq \mathfrak{D}$ 若满足条件（39），则为核心可积，其中 $m = 3$。注 C.1对条件（39）及其与 $\mathbb{R}^d$ 中随机向量的经典矩通常可积性条件之间的关系，提供了一些背景说明。

我们研究 ICA 鲁棒性（（10）、（17））的方法将用到信号的核心坐标（29），因此以下将重点讨论“可核心坐标化”的空间

$$\dot{\mathfrak{D}} := \{\mu \in \mathfrak{D} \mid [\mu]_{\mathfrak{c}} < \infty, \langle \mu \rangle_{ii} \neq 0 \text{ for } i = 1, \ldots, d\}. \tag{41}$$

请注意，这是 $\mathfrak{D}$ 的一个非常大的子空间：信号的核心坐标必须存在，这一要求包含了（例如）定义在 3-Wasserstein 空间 $(\mathcal{C}^1, \|\cdot\|_{1\text{-var}})$（脚注 17；参见 [74, 第 2 节。1]），而条件 $\langle \mu \rangle_{ii} \neq 0$ 非零 2nd 阶对角矩，只是一个温和的非退化假设，基本上可以在不造成任何损失的情况下施加。

现在，我们可以为信号空间 $\mathfrak{D}$ 赋予一种明确且结构良好的拓扑，这将有助于我们对鲁棒性进行有意义的量化（17）。

4. 因果空间上的面向 ICA 的拓扑

我们描述了因果空间 $\mathfrak{C}$上一种信息丰富且可显式计算的预度量拓扑，它与粗糙性要求（18）相关，同时又足够强，能够以便于量化的方式支持一般 ICA 解（$\mathscr{I}$，$\hat{}\Phi$）的鲁棒性（17），例如第5节推导出的解。为此，三个主要步骤中的第一步是，第4.1节展示了定义3.6中的签名矩如何对 $\mathfrak{D}$中的两个信号之间的（统计）“距离”进行系统化、类似散度的量化。由此编码的信息构成了信号内部及信号之间统计依赖的拓扑景观；这一景观在代数和分析方面具有灵活性，并且既对空间（非）线性变换敏感，也对内在的统计变化敏感，包括第

[18] 事实上：由于对于每个 $\mu = (\mu^1, \ldots\ldots, \mu^d) \in \mathfrak{D}$，我们有$\langle \mu \rangle_{ii} = \frac{1}{2}\mathbb{E}[(\mu^i_{0,\,1})^2]$，，条件$\langle \mu \rangle_{ii} \neq 0$要求增量 $\mu^i_{0,\,1} := \mu^i_1 - \mu^i_0$具有非零方差（符号约定见第A.3 节）。实际上，这并不构成限制，因为只需将A.2节所述的任意一个（“非退化”）信号在 $\mathfrak{D}$中截取到一个规则区间，使其端点间的增量具有正方差，再将截取后的信号重新缩放回 $[0, 1]$即可。

3.3。随后，第13节将这一拓扑依赖概况与 ICA 特有的可识别性结构联系起来，4.2节则将可准确识别的源信号描绘为（属于）这一拓扑景观全局极小值的集合。最后，第4.3 节落实了第2.2 节最后提出的主张，将我们的预度量信号拓扑扩展为整个因果空间 $\mathfrak{C}$上的粗拓扑，并将在第17节证明其鲁棒性（5）。

4.1. **信号空间上的预度量拓扑。** 由分层级签名矩提供的坐标具有可数无穷多个（28），低阶排列（29）捕捉了信号中最能体现作用于其上的任意线性作用（30）的那些方面。此外，与随机向量的协方差矩阵或累积量类似，坐标矩阵（29）描述了信号各分量之间的（二阶和三阶）统计依赖，具体体现为其非对角结构相对于对角结构的显著程度，参见引理4.8。结合第5.1节的逆变换理论，这将使我们能够对可由独立成分分析辨识的原因（图 1 a.）以及与其接近的原因（图 1 b.）进行显式的拓扑坐标化。令 $\mathfrak{D}$ 为（41）中的信号空间。

我们建议通过以下距离函数来实现这种图表化。回想一下，$\mu \equiv (\mu_t) \in \mathcal{M}_1$ 称为均值平稳，当且仅当 $\mathbb{E}\mu_t = \mathbb{E}\mu_0$ 对于每个 $t \in [0,1]$ (第 A.3节）。

Defi **定义 4.1** (ICA 预度量)。对于任意两个信号 μ，$\tilde{\mu} \in \mathfrak{D}$ 其核心坐标为 (29)，定义 ne

$$\delta(\mu,\tilde{\mu}) := \sqrt{\alpha_{\mu,\tilde{\mu}} + \beta_{\mu,\tilde{\mu}}} \qquad \text{for} \tag{42}$$

$$\alpha_{\mu,\tilde{\mu}} := \sum_{i,j=1}^{d} \left(\frac{\langle\tilde{\mu}\rangle_{ij} - \langle\mu\rangle_{ij}}{\sqrt{\langle\mu\rangle_{ii}\langle\mu\rangle_{jj}}} \right)^2 \quad \text{and} \quad \beta_{\mu,\tilde{\mu}} := \sum_{\nu=1}^{d}\sum_{i,j=1}^{d} \left(\frac{\langle\tilde{\mu}\rangle_{ij\nu} - \langle\mu\rangle_{ij\nu}}{\sqrt{\langle\mu\rangle_{ii}\langle\mu\rangle_{jj}\langle\mu\rangle_{\nu\nu}}} \right)^2 .$$

F对于均值平稳的 $\mu = (\mu^1, \dots, \mu^d) \in \dot{\mathfrak{D}}$，我们将 μ 的*IC* 缺陷 称为数目 r

$$\delta_{\perp}(\mu) := \delta(\mu^1 \otimes \cdots \otimes \mu^d, \mu)\,; \tag{43}$$

对于非均值平稳信号，该缺陷定义为[19] 下文公式 (49) 所示。

对于信号 μ，$\tilde{\mu} \in \mathfrak{D}$ 其核心坐标为 $[\mu]_{\mathfrak{c}}$，$[\tilde{\mu}]_{\mathfrak{c}}$ 如 (29) 所示，且对于标准化矩阵$N_\mu :=$ ddiag$\big(\langle\mu\rangle_{11}, \cdots, \langle\mu\rangle_{dd}\big)^{1/2}$，注意到 (42) 可以更紧凑地写为

$$\delta(\mu,\tilde{\mu}) = \left\| N_\mu^{-1} \cdot \left([\tilde{\mu}]_0 - [\mu]_0, \frac{[\tilde{\mu}]_1 - [\mu]_1}{\sqrt{\langle\mu\rangle_{11}}}, \dots, \frac{[\tilde{\mu}]_d - [\mu]_d}{\sqrt{\langle\mu\rangle_{dd}}} \right) \cdot N_\mu^{-1} \right\|_{\mathcal{V}} \tag{44}$$

其中 $\|(C_0, \cdots C_d)\|_{\mathcal{V}} := \sqrt{\sum_{\nu=0}^{d} \|C_\nu\|^2}$是空间 $\mathcal{V} := \big(\mathbb{R}^{d\times d}\big)^{\oplus(d+1)}$上的 ℓ_2-范数，由 $\mathbb{R}^{d\times d}$ 上的弗罗贝尼乌斯范数 $\|\cdot\|$ 导出。

映射 $(\mu,\tilde{\mu}) \mapsto \delta(\mu,\tilde{\mu})$ 在 $\mathfrak{D}$ 上定义了一个 预度量 （参见 [5, 定义 2.4.4]），也就是说

$$\delta(\mu,\tilde{\mu}) \geq 0 \quad \text{and} \quad \delta(\mu,\mu) = 0 \tag{45}$$

对于任意两个 $\mu,\tilde{\mu} \in \mathfrak{D}$。该预度量会以规范的方式（例如 [15, 第 2] 节）在信号空间上 τ_δ诱导出一个拓扑 $\mathfrak{D}$，以下列出该拓扑的一些基本事实。

引理 4.2. 对于由 τ_δ 在 $\mathfrak{D}$ 上由 (42)诱导的预度量拓扑，以下结论成立 *s.*

(i) 对于 $\mathfrak{D}$ 中的每个子集 O，若且唯若对每个 $\mu \in O$，都存在一个 $r > 0$ 使得

$$B_\delta(\mu, r) := \big\{ \tilde{\mu} \in \dot{\mathfrak{D}} \;\big|\; \delta(\mu,\tilde{\mu}) < r \big\}, \tag{46}$$

以 μ 为中心、半径为 r 的 δ 球包含于 O 中。

(ii) 拓扑空间 $(\mathfrak{D}, \tau_\delta)$ 是序列的。

[19] 引理 4.8 保证了定义（43）和（49）是一致的（参见注4.5）。

(iii) 对于任意度量空间 (W, d_W)，映射 $\phi:(\mathfrak{D}, \tau_\delta) \to W$ 在 $\mu \in \mathfrak{D}$ 处连续，当且仅当：

$$\forall \tilde{\varepsilon} > 0 : \exists \tilde{\delta} > 0 : \forall \tilde{\mu} \in \dot{\mathfrak{D}} : \ \delta(\mu, \tilde{\mu}) < \tilde{\delta} \implies d_W(\phi(\mu), \phi(\tilde{\mu})) < \tilde{\varepsilon}. \tag{47}$$

证明。 前两点可直接由（45）得出——供参考，条目(i) 见 [5,第 23] 页，而陈述 (ii) 见 [5, 命题 2.4。9]。条目 (iii) 可由陈述 (i)和 (ii) 以及 [5, 命题 3.1。3]推出；或者也可以由 [15, 引理 2.2 和 推论 2。3] 得出。□

注 4.3。 对于 ε-δ 特征（47），若 ϕ 是取值于度量空间的映射，定义在任意预度量空间上，则（拓扑）连续性的刻画仍更一般地成立，参见例如 [15,引理 2.2 和 推论 2。3]。

命题 4.4。 对于 $\mathfrak{D}$ 的任意一致签名可积子集 $\mathfrak{D}$， $\mathfrak{D}$ 上的（子空间）拓扑 τ_δ 比 $\mathfrak{D}$ 上的任意弱拓扑都更粗。

证明。 设 $\mathfrak{D} \subseteq (\mathfrak{D}, \tau_\delta)$ 是一致签名可积的，并回顾一下， τ_δ-子空间拓扑由限制后的预度量 $\tilde{\mathfrak{D}}$ 诱导。任取 $\delta\big|_{\tilde{\mathfrak{D}}\times\mathfrak{D}}$。若对任意网 $1 \le q < 2$，断言均成立，其中(μ_α) 属于 $\mathfrak{D}$ （回顾符号约定见定义3.9 ），则有：

$$\mu_\alpha \xrightarrow{q\text{-var}} \mu \in \tilde{\mathfrak{D}} \quad \text{implies} \quad \delta(\mu, \mu_\alpha) \to 0, \tag{48}$$

另见 [15, 引理 2.1]。根据 δ 的定义（42），对于上述任意 (μ_α) 和 μ，以及任意 i_1, ..., $i_m \in [d]$，若 $m \le 3$，则所需的蕴含式（48）显然成立： $\mu_\alpha \xrightarrow{q\text{-var}} \mu$意味着 $\mathrm{h}\mu_\alpha \mathbf{i}_{i_1 \cdots i_m} \to \mathrm{h}\mu \mathbf{i}_{i_1 \cdots i_m}$。 然而，引理 3.13 保证了这一蕴含式，从而完成证明。□

基于有限测试函数集合 44对 $\mathcal{M}_1$ 进行诱导拓扑化的思路，与最大均值差异（MMD）的自然概念密切相关；该概念针对测试函数集合 $\mathcal{H} := \{\mathfrak{sig}_w(\cdot) \ \ |w| \le 3\}$ 并结合适当的标准化，以吸收由 M_d引起的（“冗余”）自由度；这些自由度源于 $\lceil \cdot \rceil_{\tilde{\mathfrak{m}}}$所控制的真实源不可区分性（关于后者，请参见定义 2.7）。

虽然此处并不需要，但请注意，预度量（44）可以通过系统地加入高阶签名矩扩展为完整的度量，详见注 C.2。

我们在定义 4.1中补充如下内容。

注 4.5。 根据引理 4.8，均值平稳信号 $\mu \in \mathfrak{D}$ 的 IC 缺陷为

$$\delta_{\perp\!\!\!\perp}(\mu) = \left\| N_\mu^{-1} \left([\mu]_0 - \left(\delta_{ij}\langle\mu\rangle_{ij}\right), \frac{[\mu]_1 - \left(\delta_{ij1}\langle\mu\rangle_{ij1}\right)}{\sqrt{\langle\mu\rangle_{11}}}, \ldots, \frac{[\mu]_d - \left(\delta_{ijd}\langle\mu\rangle_{ijd}\right)}{\sqrt{\langle\mu\rangle_{dd}}} \right) N_\mu^{-1} \right\|_{\mathcal{V}}; \tag{49}$$

对于非均值平稳信号 $\mu \in \mathfrak{D}$，我们将其 IC 缺陷 $\delta_{\perp\!\!\!\perp}(\mu)$ 定义为式（49）右侧的表达式。显然，$\mu \in \mathfrak{D}$ 为正交当且仅当 $\delta_{\perp\!\!\!\perp}(\mu) = 0$，并且 $\delta_{\perp\!\!\!\perp}$ 具有不变性，因为由式（M_d在 $\mathfrak{D}$ 上的作用可直接从式（49）和引理 3.7）得出。 ◆

IC 缺陷 $\delta_{\perp\!\!\!\perp}$由 (42) 和 (49) 定义，是 $\mathfrak{D}$ 上的一个辅助函数，可对（非）线性变换信号的盲可辨识性（8）进行定量控制：该缺陷为零的源信号（第 4.2 节）将是完全可识别的，而随着源信号的 IC 缺陷增大，其盲反演将越来越不准确，甚至变得不可能，见第 5 节。鉴于此，下面的引理通过为 $\delta_{\perp\!\!\!\perp}$ 建立连续模，将为深入的鲁棒性分析（17）铺平道路；这些连续模分别关联于 上述预度量拓扑，以及空间混合变换空间上的一致拓扑。（由于证明主要涉及技术细节，故放在附录 $\mathfrak{D}$ C.8 中。）

此处及下文中，映射 $R \in C^1(\mathbb{R}^d; \mathbb{R}^d)$ 的雅可比矩阵 D_R 的范数定义为 $\|D_R\|_\infty := \sup_{x\in\mathbb{R}^d} \|D_R(x)\|_2$,，其中 $\|\cdot\|_2$ 是 $\mathbb{R}^{d\times d}$ 上的欧几里得算子范数（“谱范数”）。

（以下使用小节3.4.1 中的符号约定。）

引理 4.6。设 $\hat{\mu} \in \mathfrak{D}$，且 $B_r \equiv B_\delta(\hat{\mu}, r)$ 是以 $\hat{\mu}$ 为中心、半径为 $r > 0$ 的 δ-球（见 (46)）。假设 $r < \hat{r} := \rho_0 \wedge (\rho_0/\rho_1)$，其中 $\rho_0 := \min_{i\in[d]} h\hat{\mu}i_{ii}$，且 $\rho_1 := \max_{i\in[d]} h\hat{\mu}i_{ii}$。

(i) 存在常数 $\mathfrak{m}_{\hat{\mu}|r} = \mathfrak{m}_{\hat{\mu}|r}(\hat{\mu}, r)$ $K_{\hat{\mu}|r} = K_{\hat{\mu}|r}(\hat{\mu}, r) \geq 0$，它们关于 r 单调递增，使得对每个 $\mu \in B_r$，都有 $\max_{|w|=2,3} \left|h\mu i_w - h\hat{\mu}i_w\right| \leq \mathfrak{m}_{\hat{\mu}|r} \cdot r$，并且

$$\left|\delta_\perp(\mu) - \delta_\perp(\hat{\mu})\right| \leq K_{\hat{\mu}|r}\sqrt{r + r^2}\,; \tag{50}$$

(ii) 若 $B_\delta(\hat{\mu}, \rho_0)$ 按照 (K_0, β, p, α) 是核心可积的，则对每个 $R \in C^1(\mathbb{R}^d; \mathbb{R}^d)$，存在一个常数 $K_{\hat{\mu}|r} = K_{\hat{\mu}|r}(\hat{\mu}, r, K_0, R) \geq 0$，它关于 r 单调递增，使得：对每个 $\mu \in B_r$，对于引理 *C.3 (iv)* 中的辅助函数 $\phi_k(\cdot)$，

$$\left|\delta_\perp((I+R)_\star\mu) - \delta_\perp(\mu)\right| \leq \tilde{K}_{\hat{\mu}|r} \cdot \phi_3(\|D_R\|_\infty)^{\frac{1}{2}} \left(\|D_R\|_\infty \|R\|_\infty\right)^{\frac{\alpha}{2}} \tag{51}$$

provided that $\rho_R := C_p\phi_2(\|D_R\|_\infty)(\|D_R\|_\infty\|R\|_\infty)^\alpha < \rho_0 - \rho_1 r$.

对于任意给定的 c，$C > 0$ 满足 $c < \rho_0$，式 $K_{\hat{\mu}|r}$ 中的常数 (51) 可以选取得到一致适用的结果——也就是说，它只依赖于 $\hat{\mu}$、K_0 和 c，C 而与 R、μ 和 r 无关——在集合上

$$\mathcal{R}_c^C := \{(\mu, R) \in \dot{\mathfrak{D}} \times C^1(\mathbb{R}^d; \mathbb{R}^d) \mid \rho_1\delta(\hat{\mu}, \mu) + \rho_R \leq c \ \ and \ \ \rho_R \vee \|D_R\|_\infty \leq C\}. \tag{52}$$

以上所有常数均为显式给出的。

下一节将介绍 IC 缺陷 $\delta_\perp$ 最小的信号。这些信号将构成 $\mathscr{I}_*$ 中一大类可由独立成分分析辨识的信号 $\mathfrak{D}$，[参见（5）、（14）以及] 见第 5节。

4.2. 正交信号。众所周知，一般信号无法从其盲线性混合中辨识出来；但如果源的各个分量相互独立，或者在与时间相关的情况下至少两两不相关，则可以实现恢复。这里我们将展示，此类信号的核心坐标具有一种特别简单的代数形式。

定义 4.7（乘积信号）。若元素 $\mu \equiv (\mu^1, \cdots \ \mu^d) \in \mathfrak{D}$ 等于其各边缘分布的乘积，则称其为乘积信号，也就是说，若 $\mu = \mu^1 \otimes \cdots \otimes \mu^d$，则称其为乘积信号。

显然，过程 Y 属于 $\mathbb{R}^d$ 当且仅当其分量 $Y^1, \ldots, Y^d$ 相互独立时，它是乘积信号。均值平稳的乘积信号，其核心坐标（29）全都为对角的。

引理 4.8。对于 $\mu \in \mathfrak{D}$ 均值平稳乘积信号，有

$$\langle\mu\rangle_{ij} = \delta_{ij} \cdot \langle\mu\rangle_{ij} \quad \ and \quad \ \langle\mu\rangle_{ijk} = \delta_{ijk} \cdot \langle\mu\rangle_{ijk} \tag{53}$$

其中，δ_{ij} 和 δ_{ijk} 分别是二维和三维中的克罗内克δ。

证明。设 $\mu = \mu^1 \otimes \cdots \otimes \mu^d \in \mathfrak{D}$ 是路径空间有界变差上的均值平稳（符号）测度。那么，断言（53）可由以下更一般的陈述推出：

$$\langle\mu\rangle_{i_1\ldots i_m} = 0 \qquad (m \in \mathbb{N}) \tag{54}$$

对于每个多重指标 $(i_1, \ldots, i_m) \in [d]^{\times m}$，只要其至少有一个分量 i_ν 恰好出现一次；也就是说，对于包含在集合中的每个多重指标

$$J := \left\{(i_1, \ldots, i_m) \in [d]^{\times m} \;\middle|\; \exists\, \nu_0 \in [m] : i_{\nu_0} \neq i_\nu \text{ for each } \nu \neq \nu_0\right\}.$$

为证明（54），根据引理 3.1，每条路径 $x \equiv (x_t^1, \ldots, x_t^d)_{t\in\mathbb{I}} \in \mathcal{C}^1$ 几乎处处存在可积导数 $\dot{x} \equiv (\dot{x}^i)$（在其他所有地方将 $\dot{x}$ 设为零），且

$$x_t^i = x_s^i + \int_s^t \dot{x}_r^i \, \mathrm{d}r \qquad (0 \leq s \leq t \leq 1) \tag{55}$$

对所有 $i \in [d]$。现在根据 μ 的均值平稳性，对于每个 $i \in [d]$，我们有 $0 = \mathbb{E}[\mu_{s,t}^i] = \int c_d x_{s,t} \mu^i(\mathrm{d}x) = \int c_d(x_t^i - x_s^i)\,\mu(\mathrm{d}x)$ 对每个 $0 \leq s \leq t \leq 1$ (参见第 A.3 节中的符号约定),

由 (55) 和富比尼定理可知，$0 = \int_s^t \phi_\mu^i(r)\,\mathrm{d}r$ 对于 $\phi_\mu^i(r) := \int_{\mathcal{C}_d} \dot{x}_r^i\,\mu(\mathrm{d}x)$。由于$s,\ t \in [0,1]$ (其中 $s \le t$) 均为任意，因此可得

$$\text{现在，由于 } \phi_\mu^i(r) = \int_{\mathcal{C}_1} \dot{z}_r\,\mu^i(\mathrm{d}z),\ \text{由 (28) 可知，对每个 } \boldsymbol{i} \equiv (i_1,\dots,i_m) \in J,\ \text{都有} \qquad (56)$$

$$\langle \mu \rangle_{i_1\dots i_m} = \int_{\Delta_m}\int_{\mathcal{C}_d} \dot{x}_{t_1}^{i_1}\cdots\dot{x}_{t_m}^{i_m}\,\mu^1\otimes\cdots\otimes\mu^d(\mathrm{d}(x^i))\,\mathrm{d}^m\boldsymbol{t} = \int_{\Delta_m} \phi_\mu^{i_{\nu_0}}(t_{i_{\nu_0}})\cdot Q_\mu^{\boldsymbol{i}}(\boldsymbol{t})\,\mathrm{d}^m\boldsymbol{t} =: \beta$$

对于乘积 $Q_\mu^{\boldsymbol{i}}(\boldsymbol{t}) := \prod_{i\in[d]\setminus\{i_{\nu_0}\}} \int_{\mathcal{C}_1} z^{\boldsymbol{q}_i(\boldsymbol{t})}\,\mu^i(\mathrm{d}z)$，其中每个 $z^{\boldsymbol{q}_i(\boldsymbol{t})} := \prod_{\nu\in[m]:\ i_\nu = i} \dot{z}_{t_\nu}^{i_\nu} (z^i \in \mathcal{C}_1)$ 对应一个 $i \in [d]$ （若乘积为空，则令 $z^{\boldsymbol{q}_i(\boldsymbol{t})} := 1$）。因此，对于 $\boldsymbol{t}' := (t_2,\dots,t_m)$,

$$\beta = \int_{\mathbb{I}^{d-1}} Q_\mu^{\boldsymbol{i}}(\boldsymbol{t}')\left[\int_{\mathbb{I}} \phi_\mu^1(t_1)\cdot \mathbb{1}_{\Delta_m}(\boldsymbol{t})\,\mathrm{d}t_1\right]\mathrm{d}^{m-1}\boldsymbol{t}' = 0 \qquad (\text{by (56)}),$$

为方便起见，我们假设 $i_{\nu_0} = 1$ 不失一般性（否则相应地置换 $\Delta_m := \{(t_1,\cdots\ \ t_m)\ |\ 0 \le t_1 \le t_2 \le \dots \le t_m \le 1\}$）。由此证明了（54），证毕。□

Defi定义 4.9。满足（53）的信号 $\mu \in \mathfrak{D}$ 将称为（3rd阶）正交。

注 4.10。可通过直接计算验证，均值平稳乘积信号 μ 的签名矩（28）与其签名累积量 [14] 在阶数 $m = 3$以内相同，因此引理 4.8 的断言也可以由 [14, 定理 **1**。2]推导得出。

如第 2.2 节末尾所述，以下小节将引入一个可识别性空间 E，该空间赋有上述定义在 $\mathfrak{D}$上的签名拓扑的（粗）扩展。

4.3。 原因空间上的粗拓扑。本节将第 4.1 节的信号拓扑扩展为因果空间 $\mathfrak{C}$上的“可识别性”拓扑，并基于此对鲁棒性性质（17）进行适当分析。

利用预度量结构 τ_δ （参见第 4.1 节），出于技术上的便利，该拓扑将定义在‘正则’子空间 E 上，该子空间属于 $\mathfrak{C}$ 的乘积形式

$$E = \mathfrak{S}\times C^{1,1}(\mathbb{R}^d) \quad \text{with} \quad \mathfrak{S} \subseteq \dot{\mathfrak{D}} \text{ core-integrable} \qquad (57)$$

按照 $(K_{\mathfrak{S}},\ \beta,\ p,\ \alpha) =: c_{\mathfrak{S}}$；符号约定见小节 3.4.1。

因子 $\mathfrak{S}$ 赋予由 τ_δ诱导的子空间拓扑 τ'_δ，定义在 $\mathfrak{S}$上，即 $\tau'_\delta = \{\mathcal{O}\cap\mathfrak{S}\ |\mathcal{O}\in\tau_\delta\}$,[20] 而 $C^{1,1}(\mathbb{R}^d) =: C^{1,1}$ 赋予预度量 $\bar{d}\colon C^{1,1}\times C^{1,1}\to\mathbb{R}_+$，其定义如下:

$$\bar{d}(\theta_1,\theta_2) := \min\left(\left(\left\|\theta_1^{-1}\circ\theta_2 - \mathrm{id}\right\|_\infty + 1\right)\cdot\left\|D(\theta_1^{-1}\circ\theta_2 - \mathrm{id})\right\|_\infty, 1\right) \qquad (58)$$

它衡量两个变换之间的残差偏离，是对经典（伪）范数 $\min(\|\cdot\|_\infty + \|D(\cdot)\|_\infty, 1)$ 在 $C^1(\mathbb{R}^d;\mathbb{R}^d)$ 上的改编形式。两者 δ和 $\bar{d}$组合构成子空间 (57) 上的预度量 d： $E^{\times 2}\to\mathbb{R}_+$，其定义为

$$\mathrm{d}\big((\mu_1,\theta_1),(\mu_2,\theta_2)\big) := \delta(\mu_1,\mu_2) + \bar{d}(\theta_1,\theta_2). \qquad (59)$$

作为预度量（42）和（58）之和，预度量[21] d 将边缘拓扑（$\mathfrak{S},\delta$）和（$C^{1,1}$（$\mathbb{R}^d$） $\bar{d}$）结合为 E上的粗联合拓扑 τ_{d}。下面的引理说明了该拓扑与信号的“可识别性控制”缺陷 $\delta_\perp$ 之间的关系。

预度量空间 $(E,\ \mathrm{d})$ 是我们将展示并分析其鲁棒性（17）的定义域。（因此，相应的可辨识性假设 $\mathscr{I}$ 将是 E 的一个子集。）为与第 2.2 节的设定保持一致，我们可以（平凡地）将 τ_{d} 补充为 $\mathfrak{C}$ 上的一个拓扑，如下所示。

[20] 显然，τ'_δ是将 δ限制在子集 $\mathfrak{S}\times\mathfrak{S}$上所诱导的拓扑，而这又是 $\mathfrak{S}$上的一个预度量。（为方便起见，我们用同一符号表示该限制和原预度量（42）。）[21] 由于（E,d）是一个预度量空间，因此将（$\dot{\mathfrak{D}},\delta$）替换为（$E,\mathrm{d}$）后，引理4.2 仍按原样成立。

定义 4.11（因果拓扑）。给定 E 如（57）中所述，以及 $E^{\mathrm{c}} := \mathfrak{C} \setminus E$，我们称

$$\text{the disjoint union topology } \tau_{\mathfrak{C}} \text{ on } (E, \tau_{\mathfrak{d}}) \sqcup (E^{\mathrm{c}}, \tau_{\emptyset}) \tag{60}$$

该 I 支持因果的因果拓扑 于 $\mathfrak{C}$。（此处，$\tau_{\emptyset} = \{\emptyset,\ E^{\mathrm{c}}\}$ 是 E^{c} 上的平凡拓扑 .)

备注 4.12。根据定义，$\mathfrak{C}$ 上的拓扑 $\tau_{\mathfrak{C}}$ 就是不交并 $\tau_{\mathfrak{C}} = \tau_{\mathfrak{d}} \sqcup \tau_{\emptyset}$。因此，根据（59）的构造，拓扑 $\tau_{\mathfrak{C}}$ 通过命题 4.4 与指称粗糙度（18）相关联。由此，结合备注 4.3，映射 $\phi\colon (\mathfrak{C},\ \tau_{\mathfrak{C}}) \to (W,\ d_W)$ [（后者为一个度量空间]）在点 $\mathfrak{c}_\star \in E \subseteq \mathfrak{C}$ 处连续，当且仅当对于每个 $\varepsilon > 0$ ˜，都存在一个 $\delta > 0$，使得对于每个 $\mathfrak{c} \in \{\tilde{\mathfrak{c}} \in E \mid \mathfrak{d}(\mathfrak{c}_\star,\ \tilde{\mathfrak{c}}) < \delta\}$，均有 $d_W(\phi(\mathfrak{c})\ \phi(\mathfrak{c}_\star)) < \varepsilon$。 ◆

引理 4.13。对于固定的正交信号 $\mu_\star \in E$ 和 $A \in \mathrm{GL}_d$，考虑$\mathfrak{d}$-球

$$\mathbb{B}^\star_r \equiv \mathbb{B}_r(\mu_\star, A) := \big\{\mathfrak{p} \in E \;\big|\; \mathfrak{d}\big((\mu_\star, A), \mathfrak{p}\big) < r\big\} \tag{61}$$

对于 $r > 0$。进一步令 $\rho_0 := \min^{i\in[d]} h\mu_\star i_{ii}$，并令 $\rho_1 := \max^{i\in[d]} h\mu_\star i_{ii}$。若 $r < r_0 := (\rho_0/(\rho_1+2C_p))^{1/\alpha}$，则存在显式常数 $\hat{K} = \hat{K}(\mu_\star, C_p) \geq 1$ 和 $K_r = K_r(\mu_\star, r, C_p)$，它们随 r 增大；因此，对任意给定的 $(\mu, f) \in \mathbb{B}^\star_r$，存在唯一的 $\tilde{\mu} \in \mathfrak{D}$，使得：

$$f_*\mu = A_*\tilde{\mu} \quad \textit{with} \quad \delta_\perp(\tilde{\mu}) \leq K_r \cdot r^{\alpha/2} \quad \textit{and} \quad \big\|[\tilde{\mu}]_{\mathfrak{c}} - [\mu_\star]_{\mathfrak{c}}\big\|_{\mathcal{V}} \leq \hat{K} r^{\alpha}. \tag{62}$$

¯证明。任取 $(\mu, f) \in \mathbb{B}^\star_{r^\circ}$ 根据定义、(61) 和 (59)，有 $\delta(\mu_\star, \mu) + d(A, f) < r$。此外，对于 $R := A^{-1} \circ f - \mathrm{id}$，由 (58) 可得 $\bar{d}(A, f) \geq \|D_R\|_\infty \|R\|_\infty$（注意 $r_0 < (\rho_0/\rho_1)^{1/\alpha} \leq 1$）。简记 $r_\delta \equiv \delta(\mu_\star, \mu)$，并令 $\rho_R := C_p \phi_2(\|D_R\|_\infty)(\|D_R\|_\infty \|R\|_\infty)^\alpha$，于是可得

$$r_\delta + \|D_R\|_\infty \|R\|_\infty < r < r_0 \qquad \text{and hence} \qquad \rho_1 r_\delta + \rho_R < c_0 r^\alpha < \rho_0$$

对于 $c_0 := \rho_1 + 2C_p$（注意，$\|D_R\|_\infty \leq 1$ 蕴含 $\phi_2(\|D_R\|_\infty) \leq 2$ 且 $\phi_3(\|D_R\|_\infty) \leq 4$）。因此 $(\mu, R) \in \mathsf{R}^C_c$，参见 (52)，对于 $c := c_0 r^\alpha$ 以及 $C := \max\{2C_p, 1\}$。于是根据引理 4.6 (ii)˜ ˜，存在常数 $K_r = K_r(\mu_\star, r, K_{\mathfrak{D}}) \geq 0$，它与 μ、r_δ和 R无关，使得

$$\big|\delta_\perp((I+R)_*\mu) - \delta_\perp(\mu)\big| \leq \tilde{K}_r \cdot r_R^{\alpha/2} \tag{63}$$

对于 $r_R := \|D_R\|_\infty \|R\|_\infty$。记 $\tilde{\mu} := (I+R)_*\mu$，我们还可得

$$f_*\mu = [A(I+R)]_*\mu = A_*\tilde{\mu} \qquad (\text{by definition of } R)$$

a并且，结合 (63) 与引理 4.6 (i)，利用 $\delta_\perp(\mu_\star) = 0$（参见引理 4.8 ），

$$\big|\delta_\perp(\tilde{\mu})\big| \leq \big|\delta_\perp(\tilde{\mu}) - \delta_\perp(\mu)\big| + \big|\delta_\perp(\mu) - \delta_\perp(\mu_\star)\big| \leq K'_r\left(\sqrt{r_R^\alpha} + \sqrt{r_\delta}\right) \leq K_r \cdot r^{\alpha/2} \tag{64}$$

对于常数 $K'_r := \max\{\tilde{K}_r, \sqrt{2}K_{\mu_\star|r}\}$，¯对于 $K_{\mu_\star|r}$，如 (50) 中所示，以及对于 $K_r := 2K'_{r^\circ}$ 关于核心坐标之间的距离 $[\tilde{\mu}]_{\mathfrak{c}}$ 和 $[\mu_\star]_{\mathfrak{c}}$，注意，由 (C.19) 和 (C.18)，

$$\begin{aligned} \big\|[\tilde{\mu}]_{\mathfrak{c}} - [\mu_\star]_{\mathfrak{c}}\big\|_{\mathcal{V}} &\leq \big\|[\tilde{\mu}]_{\mathfrak{c}} - [\mu]_{\mathfrak{c}}\big\|_{\mathcal{V}} + \big\|[\mu]_{\mathfrak{c}} - [\mu_\star]_{\mathfrak{c}}\big\|_{\mathcal{V}} \\ = \sqrt{\textstyle\sum_{|w|=2,3} \Delta_w(\tilde{\mu}, \mu)^2} + \sqrt{\textstyle\sum_{|w|=2,3} \Delta_w(\mu, \mu_\star)^2} &\leq c_d\big[5C_p r_R^\alpha + \mathfrak{m}_{\mu_\star} r_\delta\big] \leq \hat{K} r^\alpha \end{aligned} \tag{65}$$

其中，我们记 $\Delta_w(\mu^{(1)},\ \mu^{(2)}) := |\langle\mu^{(1)}\rangle_w - \langle\mu^{(2)}\rangle_w|$（因此，根据 $\|\cdot\|_{\mathcal{V}}$ 的定义，该恒等式成立），并且记 $c_d := (\sum_{|w|=2,\ 3} 1)^{1/2} = d\sqrt{d+1}$ 和 $\hat{K} := c_d(5C_p + \mathfrak{m}_{\mu_\star})$。□

因果拓扑的定义（60）本身就包含一项选择：关于在（57）中选择 $\mathfrak{S}$。我们采用 $\mathfrak{D}$ 的一个可积子集 $\mathfrak{S}$，是为了确保在残差度量（58）下，不同的非线性变换信号具有一致的可比性，参见第3.4.1 节以及（63）、（64）；所包含的阈值（40）越大，$\tau_{\mathfrak{C}}$所支撑的（17）在 $\mathfrak{C}$上的变化就越“精细”。通过将（58）替换为定义在 $C^{1,1}$上的“非增量式”（例如商）[前]度量，或将后一个空间限制为变换的一个“受控”子空间（从而得到一个

μ（C.34）右侧的一致可积界。不需要可积性（即，$\mathfrak{S} = \mathfrak{D}$即可），只要考虑源中的变化，而不考虑其变换中的变化。

注 4.14。取 $\mathfrak{S} = \mathfrak{D}$，引理 4.13 的证明表明，在 A-截面上 s

$$E_{A,r} := \big(\dot{\mathfrak{D}} \times \{A\}\big) \cap \mathbb{B}_r(\mu_\star, A) = B_\delta(\mu_\star, r) \times \{A\},$$

即，当源与观测量之间的隐藏关系 A 是线性的，且对所有信号都相同，因此观测量到源信号的反演（17）中的任何不精确性都仅由源信号偏离（相对于 $\mu_\star$）造成时，则无需对 $\mathfrak{S}$ 作（40）形式的可积性假设，并且估计（62）可改进如下：

> **引理**。设 $\mu_\star \in \mathfrak{D}$ 为正交的，且 $A \in \mathrm{GL}_d$，并设 $0 < r < \rho_0/\rho_1$。则存在常数 $\hat{L} = \hat{L}(\mu_\star)$ 以及 $L_r = L_r(\mu_\star, r)$，后者随 r递增，且满足
>
> $$\delta_\perp(\mu) \le L_r\sqrt{r} \quad \textit{and} \quad \big\|[\mu]_{\mathfrak{C}} - [\mu_\star]_{\mathfrak{C}}\big\|_{\mathcal{V}} \le \hat{L}r \quad \textit{for each } \mu \in B_\delta(\mu_\star, r). \tag{66}$$
>
> 证明。重新审视引理 4.13 在 $E_{A,\,r}$上的证明，我们有 $R \equiv 0$，因此注意到（66）中的常数 $L_r := \sqrt{2}K_{\mu_\star|r}$ 由（64）得出，以及 $\hat{L} := \mathfrak{m}_{\mu_\star}$ 由（65）得出。 □

上述简化在许多应用中都很有用，因为源与观测量之间的关系通常可以建模为严格线性，但底层源信号可能偏离正交性；例如参见第6 节中的几个例子。 ◆

5. 鲁棒独立成分分析

我们提出了一种可量化地具有鲁棒性的统计程序，可根据矩阵对未观测到的 [非]正交源信号的作用来恢复该矩阵（的逆）。其总体策略基于经典思想：联合对角化一组导出的等变矩阵统计量；在我们的情形中，这些统计量由源信号的核心坐标（29）给出（第 5.1节）。通过将求逆过程与定义4.11中的因果拓扑联系起来，我们既得到了一个可辨识性定理，为恢复可能以任意方式偏离正交性的信号提供明确的稳定性界（定理5.8），也得到了作为主要成果的一种易于实现的新 ICA 解；根据定义2.10，该解已被证明具有鲁棒性。该解还具有强大且普适的鲁棒性保证，可用明确可计算且易于自然解释的连续模来表示（定理5.12）。

5.1. 基于核心坐标的独立成分分析反演。

在本节中，我们遵循经典的独立成分分析范式（13），并考虑两个线性相关的信号 ζ 和 χ，它们属于 $\mathfrak{D}$，即

$$\chi = A \cdot \zeta \quad \text{with} \quad A \equiv (a_{ij}) \equiv (a_1|\cdots|a_d) \in \mathbb{R}^{d\times d} \tag{67}$$

某个固定的可逆 $d \times d$ 矩阵，其列为 a_i。例如，我们总可以认为

$$\zeta = S \ \text{ and } \ \chi = X \quad \text{for a BSS-triple } (X, S, A) \text{ as in (3)}, \tag{68}$$

其中 $X = \mathbb{P}_{\boldsymbol{X}}$ 且 $S = \mathbb{P}_{\boldsymbol{S}}$ 对于（2）相关随机过程 $\boldsymbol{S} = (\boldsymbol{S}_t)$ 和 $\boldsymbol{X} = (\boldsymbol{X}_t)$ 属于 $\mathbb{R}^d$。独立成分分析问题（定义 2.7）就是从输入中求得逆 A^{-1}，即找到一个映射 $\chi \hat{}$在 $\dot{\mathfrak{D}}$ 上，使得 $\hat{\Phi}(\chi) \doteq A^{-1}$ 在某种最小歧义范围内成立，见（14）。

$\hat{}$原则上，这样的映射 Φ 可能需要作用于关于χ 的某些“信息片段”，这些信息片段能够有效关联 A 的作用与 ζ的相关统计性质，参见（5）。

以下内容保证了 χ 的核心坐标 $[\chi]_0, \ldots, [\chi]_d$可用于此目的 t.

命题 5.1. 设 χ 和 ζ 如(67) 所示，分别具有核心坐标 $[\chi]_{\mathfrak{c}}$ 和 $[\zeta]_{\mathfrak{c}}$。若 $[\zeta]_0$ 可逆，则其逆 $\theta := A^{-1}$ 满足

$$\theta \cdot [\chi]_0 \cdot \theta^{\intercal} = [\zeta]_0 \quad \textit{and} \quad \theta \cdot [\chi]_\nu \cdot \theta^{\intercal} = \sum_{\ell=1}^{d} a_{\nu\ell}[\zeta]_\ell \tag{69}$$

对于每个 $\nu \in [d]$，以及对于每个满足 $c \equiv (\tilde{c}_1, \ldots, \tilde{c}_d) \in \underline{\mathbb{R}}^{1+d}$ 的 $c \equiv (c_0, c)$，发现

$$\theta^{-1} \cdot [\chi]^{c}_{\odot} \cdot \theta \;=\; [\zeta]^{A|c}_{\odot} \qquad \textit{for the matrices} \tag{70}$$

$$[\chi]^{c}_{\odot} \;:=\; c_0[\chi]_0^{-1} \cdot \sum_{\nu=1}^{d} \tilde{c}_\nu [\chi]_\nu \quad \textit{and} \quad [\zeta]^{A|c}_{\odot} \;:=\; c_0[\zeta]_0^{-1} \cdot \sum_{\ell=1}^{d} \langle a_\ell, \underline{c} \rangle_2 [\zeta]_\ell. \tag{71}$$

如果 ζ 是 3rd 阶正交的，那么 $[\zeta]^{A|c}_{\odot}$ 是对角的。

Proof. 由引理 3.7、4.8 及定义（71）；参见附录 C.9 . □

命题 5.1的主要结论是，'输入'统计量与'目标'统计量 和 $[\zeta]_\nu$ 之间的合同关系（69）如何组合成 A所承载的共轭关系（70），它关联了（c参数化的）收缩统计量族 $[\chi]^{c}_{\odot}$ 和 $[\zeta]^{A|c}_{\odot}$。对核心坐标 $[\chi]_{\mathfrak{c}}$、$[\zeta]_{\mathfrak{c}}$ 的这种收缩带来了额外的自由度，由 c参数化；利用（70），可以方便地借此刻画 A^{-1}，使其在 $[\zeta]^{A|c}_{\odot}$ 的特征谱非退化时仅存在单项式歧义。而该谱又可以通过控制 ζ各分量之间统计依赖的程度来调节，参见引理 4.8 和（43），前提是这些分量满足某种温和的统计非退化条件，例如（86）。

上述策略构成了下面定理 5.8 证明的基础。其数学实现，即根据命题 14 构造一个 ICA 逆变换（5.1），如下所述。

5.1.1. **辅助统计量与核心统计量。** 给定 $\mu \in \mathfrak{D}$，定义 $C_\mu := \frac{1}{2}([\mu]_0 + [\mu]_0^{\intercal}) \in \mathbb{R}^{d\times d}$，它是对称半正定的。在（67）的背景下，假设

$$C_\chi \in \mathrm{GL}_d \quad \text{and} \quad R \equiv R_{(\chi)} := C_\chi^{-1/2} \text{ is the square root of the inverse } C_\chi^{-1}. \tag{72}$$

矩阵 R 存在且唯一，例如 [53, 推论 1。10]，并且根据其定义，$R^2 = C_\chi^{-1}$因此 $C_{R\cdot\chi} = RC_\chi R = \mathrm{I}$。这一恒等式体现了 R 的用途，而将其选为逆平方根（72）在文献中也称为马哈拉诺比斯白化（或 ZCA 白化）。

注 5.2。 根据谱定理，我们知道

$$R_{(\chi)} \equiv \mathcal{R}(C_\chi) \;=\; Q\Lambda_\chi^{-1/2}Q^{\intercal} \tag{73}$$

对于某些 $Q = Q(C_\chi) \in \mathsf{O}_d(\mathbb{R})$ 和 $\Lambda_\chi := \mathsf{ddiag}(\lambda_1^\chi, \ldots\ldots, \lambda_d^\chi)$，其中 $\lambda_1^\chi \geq \ldots \geq \lambda_d^\chi > 0$ 是 C_χ 的特征值（按降序排列，并计入重数）。另请注意

$$C_\chi \;=\; A\big[\tfrac{1}{2}\big([\zeta]_0 + [\zeta]_0^{\intercal}\big)\big]A^{\intercal} \;=\; AC_\zeta A^{\intercal}, \tag{74}$$

使得 C_χ 可逆当且仅当 C_ζ 可逆。仅在增量 C_ζ 退化情形下，$\zeta_{0,\,1}$ 支撑于 $\mathbb{R}^d$ 的一个超平面上时，D.1引理中的可逆性。 ◆

为便于下文使用，考虑预变换矩阵 $A_R := RA$ 以及

$$B_R \equiv (b_1|\cdots|b_d)^{\intercal} := (A_R)^{-1}, \;\text{ and }\; \bar{B}_R := \mathrm{ddiag}(|b_1|, \cdots, |b_d|)^{-1}B_R \tag{75}$$

对于按行归一化为单位行的矩阵 B_R。

我们求逆过程的核心分量是归一化统计量

$$\mathfrak{x}_0(\theta) := N_{\theta\cdot\chi}^{-1} \cdot [\theta\cdot\chi]_0 \cdot N_{\theta\cdot\chi}^{-1} \quad \text{and} \quad \mathfrak{x}_\nu(\theta) := N_{\theta\cdot\chi}^{-1} \cdot \frac{[\theta\cdot\chi]_\nu}{\sqrt{\langle\theta\cdot\chi\rangle_{\nu\nu}}} \cdot N_{\theta\cdot\chi}^{-1} \quad (\nu \in [d]) \tag{76}$$

对于 $\theta \in \mathrm{GL}_d =: \Theta$ 且对于每个 $\mu \in \mathfrak{D}$， $N_\mu = \mathrm{ddiag}\big(\langle\mu\rangle_{11}, \cdots, \langle\mu\rangle_{dd}\big)^{1/2}$，与之前一样。

注 5.3. (76) 中的统计量都有良好定义。事实上，由于 C_χ 根据假设可逆，因此 $C_\chi \in \mathrm{Sym}_d^+$，我们有 $C_{\theta\cdot\chi} = \theta C_\chi \theta^\intercal \in \mathrm{Sym}_d^+$，从而 $\langle\theta \cdot \chi\rangle_{ii} = e_i C_{\theta\cdot\chi} e_i > 0$ 对每个 $i \in [d]$ 以及所有 $\theta \in \Theta$均成立。

L引理 5.4. 统计量 $\mathfrak{r}_\nu(\theta)$ 在 (76) 中均具有 尺度不变性，对任意 $\chi \in \mathfrak{D}$，也就是说 s

$$\mathfrak{r}_\nu(\Lambda \cdot \theta) = \mathfrak{r}_\nu(\theta) \quad \textit{for all } \theta, \Lambda \in \mathrm{GL}_d \quad \textit{with } \Lambda \textit{ positive diagonal.} \tag{77}$$

证明。 由引理3.7可立即得出，但为完整起见，请参见附录 C.9。 □

5.1.2. 通过对比优化进行盲反演。 与经典独立成分分析方法的思路类似，参见 [28, 50] 和第 E.2节，我们旨在将隐藏的混合变换的逆作为某个特别构造的代价函数的极小化子来恢复：通过 θ 的 $\|\theta\|_\times$ 量化矩阵的“非对角性”$= \|\theta - \mathrm{ddiag}(\theta)\|$，其中 $\|\cdot\|$ 是 $\mathbb{R}^{d\times d}$上的弗罗贝尼乌斯范数，我们将统计量（76）组合成目标函数（或称“对比函数”）

$$\Theta \ni \theta \longmapsto \phi_\chi(\theta) := \sum_{\nu=0}^{d} \big\|\mathfrak{r}_\nu(\theta \cdot R)\big\|_\times^2 \tag{78}$$

它在矩阵（76）的（近似）“联合对角化矩阵”中取最小值，见下文。目标函数 ϕ_χ 具有单项式不变性，也就是说

$$\phi_\chi(M\theta) = \phi_\chi(\theta) \quad \text{for each } M \in \mathrm{M}_d,\ \theta \in \Theta. \tag{79}$$

事实上，由于每个 $M \in \mathrm{M}_d$ 都具有 $M = P\Lambda$ 的形式，其中 $P \in \mathrm{P}_d^\pm$ 且某个对角 $\Lambda > 0$，所述不变性可由引理 3.7、5.4 （参见（C.40））以及弗罗贝尼乌斯范数的定义直接推出。由于所需的解混矩阵将被确定为（78）的极小化者，因此适当地限制（“紧化”）原始定义域 Θ 是合适的。[22]

为此，我们提议考虑集合

$$\Xi_1^0 := \big\{\theta \equiv (\theta_1|\cdots|\theta_d)^\intercal \in \mathbb{R}^{d\times d} \,\big|\, \|\theta_i\|_2 = 1, \forall\, i \in [d]\big\} \tag{80}$$

由 $\mathbb{R}^{d\times d}$ 中所有行构成 $\mathbb{R}^d$的单位基的矩阵组成，并将 ϕ_χ 的超集 O_d 定义为优化的紧化子域（见引理 C.4），其定义如下：

$$\Xi_1 := \Xi_1^0 \cap \big\{\theta \in \mathrm{GL}_d \,\big|\, \kappa_2(\theta) \le \kappa_0\big\} \tag{81}$$

对于任意固定的条件数上界 $\kappa_0 \ge \kappa_2(\bar{B}_R)$ （作为先验假设[23]）。由于 Ξ_1 是紧的，且 ϕ_χ 关于 θ 连续（事实上连续可微，参见引理 3.7），因此有

$$\mathfrak{I}_\chi := \Big[\operatorname*{arg\,min}_{\theta\in\Xi_1} \phi_\chi(\theta)\Big] \cdot R_{(\chi)} \in \mathfrak{F} \tag{82}$$

对于子集 $\mathfrak{F} := \{\mathcal{M} \subseteq \mathrm{GL}_d \,|\, \mathcal{M}, \}$ 的 2^{GL_d}是非空且紧的。这里回想一下，

$$(\mathfrak{F},\ \mathfrak{d}) \text{ 关于 } \mathfrak{d}(\mathcal{A},\ \mathcal{B}) := \max\big\{\sup_{a\in\mathcal{A}} d(a,\ \mathcal{B}),\ \sup_{b\in\mathcal{B}} d(\mathcal{A}\ b)\big\} \text{是一个度量空间}, \tag{83}$$

在 $\mathfrak{F}$ 上的经典豪斯多夫距离，其中 $d(a,\ \mathcal{B}) := \inf_{b\in\mathcal{B}} \|b - a\|$。我们设定

$$\hat{\Phi} :\ \dot{\mathfrak{D}} \to \mathfrak{F},\quad \mu \to \mathfrak{I}_\mu, \tag{84}$$

[22] 在经典独立成分分析中，通常假设源 ζ 的分量相互独立（至少两两独立），并通过对 χ进行“白化”来实现定义域的紧化；事实上，这会将原始定义域 Θ 缩减为正交矩阵的紧子定义域 O_d。然而，这一白化步骤要求 ζ 的分量间协方差结构可对角化，而我们并不对自身的（可能非正交的）源作此假设。[23] 也就是说，假设观察者不会低估条件数 $\kappa_2(\bar{A}_R)$。

并声称，$\hat{\Phi}$ 不仅是某个尚待具体说明的可辨识性假设 $\mathscr{I}_\star$下的 ICA 逆变换（14），而且按照（10）和定义 2.10。

接下来的两个小节将论证这一结论。

注 5.5 (关于优化（82）)。 如第 E.2 节进一步详述的，在 Θ的受约束子集上优化形如（81）的对比函数，本身就是一个成熟的研究主题，相关研究仍在蓬勃开展。至于行约束（80），例如[86], 在优化（78）时施加额外的条件约束（81），已有相关研究，且不会引入任何新的数学难点，参见例如 [93]。 ◆

注 5.6。 由于可识别性映射（84）会为其输入可观测量返回可逆变换，因此也可以考虑在 $\mathfrak{F}$ 上采用由（83）给出的"残差"半度量作为另一种拓扑，不过这里使用的是另一种点集距离

$$\tilde{d}(a,\mathcal{B}) := \inf_{b\in\mathcal{B}} \tilde{d}(a,b) \quad \text{with} \quad \tilde{d}(a,b) := \big\|a\circ b^{-1} - \mathrm{id}\big\| \vee \big\|b\circ a^{-1} - \mathrm{id}\big\| \tag{85}$$

其中 id 表示GL_d中的单位矩阵。乘法设定（85）下的连续性可由[连续性推出，] （83），因为度量拓扑（83）比（85）诱导的拓扑更精细；这一点可由矩阵求逆在每一点处都是利普希茨连续的性质看出。

5.2. **正交与非正交源的盲反演**。 本节的主要目标是证明，在源受到一般扰动时（包括违反其正交性（"IC"）假设），信号的盲反演仍可具有鲁棒性——或稳定性。回想一下，后一类偏离程度由缺陷 $\delta_\perp(\zeta)$ （见 49）量化，下面对 $\hat{\Phi}(\chi)$ 与真实逆 A^{-1} （忽略M_d中的前置因子）之间关系的稳定性分析，是定量理解所需鲁棒性（17）的重要一步。这一步涉及若干常数和辅助函数，下面将予以介绍。

以下讨论针对位于 $\mathcal{M}_1$ 中的源信号，这些信号处于 se t

$$\mathscr{S}_0 := \big\{\mu\in\dot{\mathfrak{D}} \;\big|\; C_\mu\in\mathrm{GL}_d \text{ and } \sharp\{i\in[d]\,|\,\langle\mu\rangle_{iii}=0\}\le 1\big\} \tag{86}$$

of m度量 μ 使得 C_μ 可逆（引理 D.1），且 $\langle\mu\rangle_{iii}=0$ 至多对应一个 $i\in[d]$。[2] [4]

记号 5.7。对于一个 BSS 三元组 $(\chi,\zeta,A)\in\mathscr{S}_0\times\mathrm{GL}_d$，令 $\gamma := 1+\sqrt{5}$ 和 $k_d := \sqrt{\frac{1}{6}(d-1)d(2d-1)}$ 及 $R\equiv R_{(\chi)}$ 如（73）中所定义，并考虑依赖于 (ζ,A) 的 $\varepsilon_0 := q_0/(1+q_0) > 0$，其定义如下：

$$q_0 := (\gamma k_d r_0)^{-1} \quad\text{and}\quad r_0 := \kappa_0\varsigma_1\left[\frac{\|B_R\|}{\sqrt{d}}\big(1+\kappa_0+(1+\xi d)\kappa_0\kappa_2(B_R)\big)+\kappa_2(B_R)\varsigma\right]. \tag{87}$$

进一步令 $c_1 := 2dk_dr_0$ 并令 $c_2 := \sqrt{d}\kappa_2(B_R)\,c_1$，并注意到（87）包含以下项

$$\varsigma_1 := k_d^{-1}\sqrt{\sum_{i=1}^d \frac{(i-1)^2\langle\zeta\rangle_{ii}^3}{\langle\zeta\rangle_{iii}^2}},\quad \xi := \sqrt{\sum_{\nu=1}^d \big\|[A_R\cdot\zeta]_\nu\big\|^2},\quad\text{and}\quad \varsigma := \sqrt{\sum_{i=1}^d \langle\zeta\rangle_{iii}^2/\langle\zeta\rangle_{ii}^3},$$

参见（28）、（72）。在以下定理中，不应过于纠结上述常数是否最优，因为我们并未尝试优化不等式的尖锐性。

T以下定理是我们一般鲁棒性结果（定理5.12）的基石 .

定理 5.8（盲反演）。设 (χ,ζ,A) 是 $\mathscr{S}_0\times\mathrm{GL}_d$ 上的 *BSS* 三元组，且 $\hat{\Phi}$ 如（84）所示。如果 $\delta := \delta_\perp(\zeta)\le\varepsilon_0$，那么对于每个 $\theta_\star\in\hat{\Phi}(\chi)$，存在一个 $M\in\mathcal{M}_d$ 和 $E\in\mathbb{R}^{d\times d}$，使得

$$\theta_\star = M(\mathrm{I}+E)A^{-1} \quad \textit{with} \quad \|E\|\le c_1\tfrac{\delta}{1-\delta} \quad \textit{and} \quad \tfrac{\|\theta_\star - MA^{-1}\|}{\|MA^{-1}\|}\le c_2\tfrac{\delta}{1-\delta}. \tag{88}$$

[24] 注意，$\langle\mu\rangle_{iii}=\mathbb{E}[(\mu^i_{0,\,1})^3]$。此外，为方便起见，我们可以（不失一般性地重新编号分量后）假设可能为零的是$\langle\mu\rangle_{111}$；请注意，这一假设是 ς_1定义的基础。

此外，$\partial_{\hat{\Phi}}(\zeta,\ A) \le c_2 \frac{\delta}{1-\delta}$对于偏差（15），以及（$\mathscr{I}_\star$，$\hat{}\Phi$）是可辨识性假设下的独立成分分析解

$$\mathscr{I}_\star := \mathscr{S}_\star \times \mathrm{GL}_d \quad with \quad \mathscr{S}_\star := \{\mu \in \mathscr{S}_0 \mid \mu \ is\ orthogonal\}. \tag{89}$$

证明。（88）的证明结合了矩阵分析中的标准论证与针对（非对称）矩阵特征子空间的最新扰动界；证明篇幅较长，因此推迟至第 E.1 节。其余断言是该证明和（88）的直接推论：（88）中的陈述保证，对每个 $B \in \hat{}\Phi(\chi)$ 都存在 $M \in \mathrm{M}_d$ 和 $E \in \mathbb{R}^{d\times d}$，使得

$$\tfrac{|BAu-Mu|}{|Mu|} = \tfrac{|M\cdot E\cdot M^{-1}(Mu)|}{|Mu|} \le \|M\cdot E\cdot M^{-1}\|_2 \le c_2\tfrac{\delta}{1-\delta}, \quad \text{for each } u\in\mathbb{R}^d\setminus\{0\}, \tag{90}$$

其中，最后一个不等式由与（E.95）相同的论证得出。结合偏差函数 $\partial_{\hat{\Phi}}$ 的定义（15），即可推出所述的 $\partial_{\hat{\Phi}}(\zeta,A)$-不等式。

（89）的 d·em-充分性由（88）以及 $\delta_\perp(\mathscr{I}_\star) = 0$ 这一事实可得，参见（49）。~ ~事实上：对于任意 BSS 三元组（χ,ζ,A）$\mathscr{I}_\star$，按照（E.52）的符号约定，有（χ,ζ,A）$\mathscr{I}_\star$，因此 $\mathrm{I} = RC_\chi R = (R\tilde{A})C_{\tilde{\zeta}}(R\tilde{A}) = \tilde{A}_R\tilde{A}_R$，进而 $\tilde{B}_R := \tilde{A}_R^{-1} \in \mathrm{O}_d \subseteq \Xi_1$（参见（81））~ ~对每个（$\Xi_1$-定义的选择）$\kappa_0 \ge 1$均成立。因此，（88）中的陈述适用于 [三元组（$\chi,\zeta,A$）$\mathscr{I}_\star$，尤其适用于] 给定的映射 $\hat{}\Phi$，从而得到 $\hat{}\Phi(\chi) \subseteq \mathrm{M}_d\cdot\tilde{A}^{-1} = \mathrm{M}_d\cdot A^{-1}$，如所需。□

不出所料，$\mathrm{M}_d\cdot A^{-1}$ 与从 χ 恢复出的逆矩阵$\theta_\star \in \hat{}\Phi(\chi)$ 之间的偏离 (88)，以及通过 (16) 得到的精确准源与其估计之间的偏离 $\mathrm{dist}_{\tilde{\varrho}}(\hat{}\Phi(X)\cdot X, \mathrm{dSe_m})$，都以乘法尺度[（即相对误差]）表示，而非以加法尺度 [（即绝对误差],）表示，这是由于群 GL_d 及其在 M_1上的作用具有乘法性质。

我们还注意到，(88) 的证明（见第 E 节）可以量化盲反演过程 (84) 的崩溃点：当且仅当以下成因 (ζ, A) 属于 $\mathfrak{C}$ 且满足~时，其中一个特征值 (λ_i) 在 (E.87) 中的代数重数严格大于^一；我们可以将这些成因解释为 Φ 在因果空间 $\mathfrak{C}$中的“奇点”。

注 5.9. 让我们重新考察通常情形 (68)：ζ 和 χ 分别是某些$\mathbb{R}^d$值过程 $\boldsymbol{S} = (\boldsymbol{S}_t)_{t\in\mathbb{I}}$ 和 $\boldsymbol{X} = (\boldsymbol{X}_t)_{t\in\mathbb{I}}$ 的分布 S 和 X，参见 (68)。对于任意 $\theta_\star \in \hat{}\Phi(X)$，记 $\hat{}\boldsymbol{S} := \theta_\star\cdot\boldsymbol{X}$，则定理 5.8 和 (16) 得出[25]，即

$$\tilde{\varrho}(\hat{\boldsymbol{S}}, M\boldsymbol{S}) \equiv \sup_{t\in\mathbb{I}} \tfrac{|\hat{\boldsymbol{S}}_t - M\boldsymbol{S}_t|}{|M\boldsymbol{S}_t|} \le c_2\,\hat{\delta} \quad \text{for some } M = M(\theta_\star) \in \mathrm{M}_d \tag{91}$$

并且如上所述，~~取 $\hat{}\delta :=$~~ $\delta_\perp(s)$ $(1-\delta_\perp(s))$ 和 $c_2 \equiv c_2(S,A)$，前提是 $\delta_\perp(S) \le \varepsilon_0 \equiv \varepsilon_0(S,A)$。如果除了“全局”偏离 (91) 外，还关注源的单个分量 $\boldsymbol{S}^i = (\boldsymbol{S}^i_t)_{t\in\mathbb{I}}$ 的重构误差（任意 $i \in [d]$），则可以考虑比例因子 $p^i_t := |\hat{}\boldsymbol{S}_t|/|\hat{}\boldsymbol{S}^i_t|\cdot \mathbb{1}_\times(\hat{}\boldsymbol{S}^i_t)$，并注意到根据 (91)，令 $M =: (\beta_i\cdot\delta_{\sigma(i)j})_{ij}$，

$$\frac{\left|\hat{\boldsymbol{s}}^i_t - \beta_i\boldsymbol{s}^{\sigma(i)}_t\right|}{\left|\beta_i\boldsymbol{s}^{\sigma(i)}_t\right|}\cdot\imath \le \frac{p^i_t\cdot\delta'}{1-p^i_t\cdot\delta'} \qquad \text{for each } t\in\mathbb{I} \text{ with } (p^i_t\cdot\delta')\vee\tilde{\delta} < 1,$$

其中 $\imath := \mathbb{1}_\times(\hat{}\boldsymbol{S}^i_t)\mathbb{1}_\times(\boldsymbol{S}^{\sigma(i)}_t)$ 和 $\delta' := \tilde{\delta}/(1-\tilde{\delta})$，$\tilde{\delta} := c_2\hat{}\delta$; 例如参见 [38, 命题 1.12]。 ◆

如上所述，通过联合对角化观测信号的一族经过适当选择的等变张量统计量，对线性混合随机源进行盲反演的总体思路——如（82）中所做的那样——是当前盲源分离理论中的经典方法。第 5.8 定理的结果及其背后的论证如何融入这种联合对角化方法的既有研究，简要概述见第 E.2节。

[25] 注意，如果 $|M\cdot\boldsymbol{S}_t| = 0$ 那么 $\hat{}\boldsymbol{S}_t = \theta_\star A\boldsymbol{S}_t = 0$，而对于这样的 $t\in\mathbb{I}$，(91) 中的商定义为 $0/0 := 0$。25

5.3. **鲁棒 ICA 逆变换**。 本节将第 5.2 节的可识别性结果纳入第 2.2 节和第 4.1节的稳健性框架，从而将其推广到全局。更具体地说，我们证明，（84）中的反演映射 $\hat{\Phi}$ 可以直接扩展为一个 ICA 解 $(\mathscr{I}, \hat{\Phi})$，并且按照一般意义（即满足（10）和）定义 2.10）具有鲁棒性。

Defi义 5.10 $((\mathscr{I}_\star, \hat{\Phi}))$。 设 $\mathscr{I}_\star$ 如（89）中所述，并考虑集合值反演映 p

$$\hat{\Phi} : \dot{\mathfrak{D}} \longrightarrow \mathfrak{F}, \quad \mu \mapsto \hat{\Phi}(\mu) := \left[\operatorname*{arg\,min}_{\theta \in \tilde{\Xi}_1} \phi_\mu(\theta) \right] \cdot R_{(\mu)} \tag{92}$$

˜定义如（84）和（82）中所述，但将 Ξ_1 推广到形如（ Ξ_1 94）的紧域。（由于（84）与（92）之间的差异很小，因此我们用同一符号表示这两个映射。）

回忆定理 5.8 ， 上述一对 $(\mathscr{I}_\star, \hat{\Phi})$ 是一个 ICA 解，即

$$\hat{\Phi}(A \cdot \tilde{\mu}) \cdot A \subseteq \mathrm{M}_d \quad \text{for each } (\tilde{\mu}, A) \in \mathscr{I}_\star,$$

并回忆引理 4.8，另见注 3.4，可知族 $\mathscr{I}_\star$ 恰好可识别的原因并非空集，事实上其规模很大。

按照其蓝图（9），与 M 估计量（12）相关联的反演映射（92实质上对极大解（7）的（一个子集）进行了参数化。为了按照（10）对该反演映射进行有意义的鲁棒性分析，[26] 我们遵循（81），并用与原因条件数相关的正则次水平集（93）穷尽 $\mathscr{I}_\star$ 可识别原因的集合。

备注 5.11 (条件数界)。 回顾一下，优化域 $\Xi_1 = \Xi_1(\kappa_0)$ 在 (81) 中的定义取决于先验条件数界 $\kappa_0 < \infty$。为了将此定义“全局化”，以便用于 (92)，注意到“可识别性超集” $G := \mathfrak{D} \times \mathrm{GL}_d$ 可由以下集合穷尽：

$$G = \bigcup\nolimits_{b \geq 1} G_b \qquad \text{for} \qquad G_b := \left\{ (\mu, A) \in G \;\middle|\; \kappa_2\big(\bar{B}_{(A,\mu)}\big) \leq b \right\} \tag{93}$$

¯其中，对于任意给定的 $(\mu, A) \in G$，矩阵 $B_{(A,\mu)}$ 的定义与 (75) 类似，即： $B_{(A,\mu)} :=$ddiag$(|b_1|, \ldots, |b_d|)^{-1} B_{(A,\mu)} \in \Xi_1^0$，其中 $B_{(A,\mu)} \equiv (b_1 | \cdots | b_d) := (R_\mu A)^{-1}$ 表示 $R_\mu := R_{(A\mu)}$ 逆（半）协方差 $C_{A\mu}^{-1}$ 的平方根，参见 (72)。对于任意 $b \geq 1$ 和 $\Delta_\kappa > 0$，我们设

$$\tilde{\Xi}_1 := \Xi_1^0 \cap \left\{ \theta \in \mathrm{GL}_d \;\middle|\; \kappa_2(\theta) \leq b + \Delta_\kappa =: \tilde{\kappa}_0 \right\}. \tag{94}$$

对于任意预先选定的正则性参数 b 和容差 $\Delta_\kappa > 0$，集合 (94) 是我们在 (92) 中使用的紧集（见引理 C.4），也是相应的定义域。注意，对于每个 $(\mu, A) \in G_b$，条件数 $\kappa_2(\bar{B}_{(A,\mu)})$ 关于 $[\mu]_0$连续[27]，因此存在一个（可显式计算的） $r_* = r_*((\mu, A)\ \Delta_\kappa) > 0$，使得 $\sup\{\kappa_2(B_{(A,\tilde{\mu})}) \mid \tilde{\mu} \in \mathfrak{D} : \mathrm{k}[\tilde{\mu}]_0 - [\mu]_0 \mathrm{k} \leq r_*\} \leq b + \Delta_\kappa$。 ◆

我们固定任意一个 E，如（57）中所示，并将其关联的因果拓扑（ $\tau_{\mathfrak{C}}$ （60）记为。进一步令

$\mathscr{I} := \mathscr{I}_\star \cap G_b$ 对于任意固定的 $b \geq 1$，由此定义 Ξ_1，如（94）所示，并令 $\mathscr{I} := \mathscr{I} \cap E$。

下一个结果表明，ICA 解 $(\mathscr{I}, \hat{\Phi})$ 按照 (10) 和 (17) 的意义具有鲁棒性。它还通过给出底层连续性的显式模量，对这种鲁棒性进行了量化。

定理 *5.12*（鲁棒性）。*ICA* 解 $(\mathscr{I}, \hat{\Phi})$ 在 *(10)* 和 *(17)* 的意义下是鲁棒的。具体而言：映射 $\mathfrak{I} : (\mathfrak{C}, \tau_{\mathfrak{C}}) \to (\mathfrak{F}, \mathfrak{d})$，由 $(\mu, f) \mapsto \hat{\Phi}(f_*\mu)$ 给出；以及 $\hat{\Phi}$ 的偏差$\partial_{\hat{\Phi}} : (\mathfrak{C}, \tau_{\mathfrak{C}}) \to \mathbb{R}_+$，都在 $\mathscr{I}$ 上连续。事实上，对于每个原因 $(\mu_\star, A) \in \mathscr{I}$，存在

[26] 为与第 2.2 节保持形式上的一致，以下定理 5.12 考虑对 (92) 的一个“全定义域”扩展 $\hat{\Phi} : \mathrm{M}_1 \to \mathfrak{F}$，其中 $\hat{\Phi} : \dot{\mathfrak{D}} \to \mathfrak{F}$ 如 (92) 所述，而 $\hat{\Phi} : \mathrm{M}_1 \setminus \dot{\mathfrak{D}} \to \mathfrak{F}$ 则任意定义。更有意义的 M_1-扩展可以按照注 B.2 等方式定义。[27] 更确切地说，对于连续函数 $\tilde{\varphi}_A(\mathfrak{a}) := \kappa_2(\mathrm{R}(\mathrm{C}_\mathfrak{a}) A \cdot \mathrm{ddiag}_i(|(A^{-1}\mathrm{R}(\mathrm{C}_\mathfrak{a})^{-1}) \cdot e_i|) \in \bar{\mathbb{R}}$，在点 $[\mu]_0 \in \mathrm{GL}_d$ 处连续；见引理 D.7，我们有 $\kappa_2(\bar{B}_{(A,\mu)}) = \tilde{\varphi}_A([\mu]_0)$；此处若 $\mathfrak{b} \notin \mathrm{GL}_d$，则令 $\kappa_2(\mathfrak{b}) = +\infty$。

显式常数 $\tilde{c}_i = \tilde{c}_i(\mu_\star, A, c_{\mathfrak{S}}, \tilde{\kappa}_0) > 0$, $i = 0, 1, 2, 3, 4$, 具有以下性质：给定$\varepsilon > 0$令 $\tilde{\delta} := \tilde{c}_0 \wedge \big[\tilde{c}_1^{2/\alpha}\big(\varepsilon/(\tilde{c}_2 + \varepsilon)\big)^{2/\alpha}\big]$, 则对于每个原因 $\mathfrak{c} \equiv (\mu, f) \in \mathbb{B}_{\tilde{\delta}}(\mu_\star, A)$, 有：

$$\begin{aligned} &\textit{for every } \theta \in \mathfrak{I}(\mathfrak{c}) \textit{ there is } M = M(\theta) \in \mathrm{M}_d \textit{ with } \|M\| \le \tilde{c}_3 \textit{ and} \\ &E = E(\theta) \in \mathbb{R}^{d\times d} \textit{ with } \|E\| \le \varepsilon \textit{ such that } \theta = M(\mathrm{I} + E)A^{-1}; \end{aligned} \tag{95}$$

$$\textit{further,} \qquad \partial_{\hat{\Phi}}(\mathfrak{c}) \le \tilde{c}_4\big(\varepsilon + \bar{d}(A, f)(1 + \varepsilon)\big). \tag{96}$$

事实上，若 $\mu \in \mathfrak{D}$ 和 $f \equiv A$, 则上述结论仍按原样成立，但此时为 $\alpha = 1$, 且显式常数更简单： $\tilde{c}_0$、……、$\tilde{c}_4 > 0$, 每个都可以独立于 $c_{\mathfrak{S}}$ 选取。

证明。证明 $\mathfrak{I}$ 的 $\mathscr{I}$-连续性相当技术性，因此我们将其放在 E.3 节，并在此采用其中的符号约定。正如下文所示，所述常数 $\tilde{c}_0$、……、$\tilde{c}_4$ 可从该证明中提炼出来。具体而言：对于任意固定的 $\mathfrak{p}_\star \equiv (\mu_\star, A) \in \mathscr{I}$, 按照 E.3 节的推导，先令 $\tilde{c}_0 := r_3$, 其中 $r_3 = r_3(\mathfrak{p}_\star, c_{\mathfrak{S}}, \tilde{\kappa}_0) > 0$ 如 (E.106) 中所定义。进一步设 $\tilde{c}_1 := 1/K'$, 以及 $\tilde{c}_2 := K_1$, 其中 $K' (\equiv K'|_{r=r_3}) = K_{r_3}(\mu_\star, K_{\mathfrak{S}})$ 和 $K_1 = K_1(r_3)$ 均按紧接在公式 (E.106) 后的定义取值。于是，(95) 中与 θ 相关的第一个断言由 (E.107) 得出。事实上，(E.107) 给出了 (95) 中的 θ-恒等式，其中 $\|E\| \le \epsilon_r$; 而 $[0, C_0 \vee \hat{r}_\varepsilon] \ni \tilde{r} \mapsto \epsilon_{\tilde{r}}$ 的单调性意味着，对于 $\hat{r}_\varepsilon := \tilde{c}_1^{\alpha/2}(\varepsilon/(\tilde{c}_2 + \varepsilon))^{2/\alpha}$, 有 $\|E\| \le \epsilon_{\tilde{\delta}} \le \epsilon_{\hat{r}_\varepsilon} = \varepsilon$。由 ((E.93) 和) (E.90) 及 (E.101)，可得界 $\|M\| \le \sqrt{d}\|A_{R_{f_*\mu}}\|_2 \le \sqrt{d}L_{\mathfrak{p}_\star}\|A\|_2 =: \tilde{c}_3$。

现在我们证明 (96)，由此立即可得 $\partial_{\hat{\Phi}}$ 的 $\mathscr{I}$-连续性。为此，固定任意一个如前所述的 $\mathfrak{c} \equiv (\mu, f) \in \mathbb{B}_{\tilde{\delta}}(\mathfrak{p}_\star)$, 并令 $B \in \hat{\Phi}(f_*\mu)$ 任意。根据 (95)，存在 $M \in \mathrm{M}_d$, 事实上满足 $\kappa_2(M) \le \sqrt{d}\,\sigma^\star_{\tilde{c}_0} L_{\mathfrak{p}_\star}\kappa_2(A) =: \tilde{c}_4$ （参见 (E.91)，接着参见 (73)、(E.101)），使得

$$\eta := \sup_{u\in\mathbb{R}^d\setminus\{0\}} \tfrac{|BAu - Mu|}{|Mu|} \le \tilde{c}_4 \cdot \varepsilon, \tag{97}$$

这可由与 (90) 相同的论证得出。回顾 (61)、(59)、(58)，对于满足 $\|D_R\|_\infty \le \bar{d}(A, f) \le \tilde{\delta}$ 的 $R \in C^1(\mathbb{R}^d; \mathbb{R}^d)$, 有 $f = A(\mathrm{I} + R)$; 又由于 f 是双射，我们可以选择$v := -f^{-1}(0) \in \mathbb{R}^d$。此外，对于 $\tilde{R}(u) := R(u - v) - v$, 有 $(\mathrm{I} + R)(u - v) = (\mathrm{I} + \tilde{R})(u)$, 因此

$$\tfrac{|B\circ f(u-v) - Mu|}{|Mu|} = \tfrac{|BA(\mathrm{I}+R)(u-v) - Mu|}{|Mu|} = \tfrac{|BAu - Mu + BA\tilde{R}(u)|}{|Mu|} \le \eta + \tfrac{|BA\tilde{R}(u)|}{|Mu|}$$

对每个 $u \in \mathbb{R}^d \setminus \{0\}$。至于最后一项，请注意 $BA = M(\mathrm{I} + E)$ (根据 (95))，以及 $\tilde{R}(0) = R(f^{-1}(0)) + f^{-1}(0) = A^{-1}(f(-v)) = 0$, 因此对于 $\hat{R} := \tilde{R} \circ M^{-1}$ 以及每个 $u \in \mathbb{R}^d$,

$$|BA\tilde{R}(u)| \le \|M(\mathrm{I} + E)\|_2 \big|\hat{R}(Mu)\big| \quad \text{and} \quad \big|\hat{R}(Mu)\big| \le \|M^{-1}\|\|D_R\|_\infty |Mu|$$

其中，最后一个不等式用到了（链式法则和）中值定理。综上，

$$\sup_{u\in D_\mu^{(v)}} \tfrac{|B\circ f(u-v) - Mu|}{|Mu|} \mathbb{1}_\times(u) \le \tilde{c}_4 \cdot \varepsilon + \kappa_2(M)(1 + \varepsilon)\bar{d}(A, f) \le \tilde{c}_4\big(\varepsilon + \bar{d}(A, f)(1 + \varepsilon)\big)$$

由此，通过 (96) 可得所述 $\partial_{\hat{\Phi}}$-不等式 (, 15)。定理的最后一个断言可直接由注 E.1 (ii)应用于 (95) 和 (96) 得出。□

我们强调，定理 5.12 的证明完全是构造性的，并且上述每个常数 c_i 和 $\tilde{c}_i$虽然并未针对最优性进行精细调整，但都可以按照证明中所述以显式形式写出。

6. 应用

ICA 逆映射的鲁棒性（如定义2.10 所表述并由定理 5.12所确立）在任何精确的结构假设（ 13）被违反，或其他来源的（非）系统误差不可避免的情形下，都显然至关重要；这些误差包括对相关信号及其统计量的近似。这自然涵盖了几乎所有实际应用：信号的（精确）分布通常不可得，其相关统计量也只能根据经验数据进行估计，

而这些数据往往又受到噪声污染，或存在其他形式的不确定性。本节将介绍三个基本独立的应用示例（参见示例1.1），以说明我们的理论如何帮助人们便捷地量化各种与实践相关的鲁棒性情形。

在本节中，设 $\boldsymbol{S} = (\boldsymbol{S}_t)$ 是 $\mathbb{R}^d$ 中的一个随机过程（定义在概率空间 $(\Omega, \mathscr{F}, \mathbb{P})$ 上），并关联一个定义在 (89) 上的 BSS 三元组 $(X, S, A)_{\mathscr{I}_*}$，考虑如 (92) 中的 $\hat{\Phi}$，但其中 Ξ_1 对 $b := \kappa_2(B_{(A,S)})$ 和任意固定的 $\Delta_\kappa > 0$ 由 (94) 给出。

6.1. **统计估计**。在实践中，我们通常既无法获得观测量的分布 $X \equiv \mathbb{P}_{\boldsymbol{X}}$（计算统计量（28）所需），也无法获得其连续时间实现，只能获得 $\boldsymbol{X}$的时间离散化样本轨迹。时间离散化问题很容易解决，因为由（随机）向量组成的序列 $\boldsymbol{X}_{t_1}, \dots, \boldsymbol{X}_{t_n}$，属于 $\mathbb{R}^d$，可通过分段线性插值立即视为$\mathcal{C}_d$ 中的（随机）元素，见注 3.4。然而，分布无法获取的问题则更为微妙，必须通过利用现有数据估计 X 来解决。虽然 X 通常是一个无限维对象，但我们的求逆过程（92）只需要其低维投影（29）的信息，因此，任何可行的估计值 $\hat{\mu}$ 对于 X 只需在低阶层面足够准确即可

$$\langle \hat{\mu} \rangle_w \approx \langle X \rangle_w = \mathbb{E}\big[\mathfrak{sig}_w(\boldsymbol{X})\big] \quad \text{for} \quad |w| = 2, 3. \tag{98}$$

所有合理的估计量 $\hat{\mu} = (\hat{\mu}_n) \subset \mathfrak{D}$ 对 X 的估计都具有统计一致性这一性质；也就是说，在“无限数据极限”下 $n \to \infty$（详情参见例如 [83]）

$$\langle \hat{\mu}_n \rangle_w = \int_{\mathcal{C}_d} \mathfrak{sig}_w(x)\, \hat{\mu}_n(\mathrm{d}x) \xrightarrow{n\to\infty} \langle X \rangle_w \quad \text{for each} \quad |w| = 2, 3, \tag{99}$$

几乎必然收敛或依概率收敛；收敛（99）只是对非正式精度要求（98）的一种拓扑刻画。

注 6.1。要理解（99），需要注意，严格来说，任何基于样本的近似 $(\hat{\mu}_n)$ 对 X 实际上自然都是一列随机测度，即一列取值于 $\mathfrak{D}$的随机变量 $\hat{\mu}_n \equiv \hat{\mu}(\boldsymbol{X}, n) : \Omega \ni \omega \mapsto \hat{\mu}_n(\boldsymbol{X}(\omega)) \in \mathfrak{D}$。下文注 6.2 中的示例估计量也有这一情况，但为避免表述冗杂，我们在符号记法中省略了这种额外的复杂性。

注 6.2（签名估计量）。$\hat{\mu}$ 的“最优”选择通常取决于给定的数据，但在许多应用中，有用的估计可以很直接地得到。例如，若 $\boldsymbol{X}$ 描述某个典型患者的 d维健康轨迹的时间演变，而我们的数据由 n 个独立抽取的、以路径为值的[28] 观测值 $x_1, \dots, x_n \in \mathcal{C}_d$组成，分别对应于 n 位不同真实患者的轨迹，那么当 n 较大时，经验测度 $\hat{\mu} \equiv \hat{\mu}_n(x_1, \dots, x_n) = \frac{1}{n}\sum_{j=1}^n \delta_{x_j} \in \mathfrak{D}$ 就可能足够。对于其他应用，合适的 $\hat{\mu}$ 可能更难确定，并且可能需要对 $\boldsymbol{X}$ 作出额外的统计假设，以便根据其样本可行地近似其分布。一种非常常见的情形是，数据仅由 $\boldsymbol{X}$ 的一条轨迹 $x = (x_t)$组成；在这种情况下，对 $\boldsymbol{X}$ [或等价地对 $\boldsymbol{S}$] 作出常见的混合或遍历性假设，通常足以将常用的遍历插件估计量 $\hat{\mu} \equiv \hat{\mu}_n(x) = \frac{1}{n}\sum_{j=1}^n \delta_{x|_{((j-1)\tau, j\tau]}}$（当 $\tau > 0$足够大时）视为 X的合适近似；详情参见例如 [83, 第 8] 节。

◆

第 5 节中的稳定性结果给出了明确的界，将收敛（99）与逆的可控近似联系起来 A^{-1}。这使我们能够方便地推导出以下类型的有限样本结果。

[28] 例如，对离散测量数据点进行分段线性插值。28

记号 6.3（收敛速率）。与经典理论一致，若存在某个严格递减的零序列 $\tilde{\alpha}_n$，使得对于（$\tilde{\alpha}_n$）有：$\gamma_n := \max_{|w|=2,3} |\hat{\mu}_n i_w - X i_w| = O_P(\tilde{\alpha}_n)$，则称估计量 $\hat{\mu} \equiv \hat{\mu}_n$ of X 的收敛速率为 $\tilde{\alpha}_n$。也就是说，对于任意 $\epsilon > 0$，都存在数 $0 < m_\epsilon, M_\epsilon < \infty$，使得 $\sup_{n \geq m_\epsilon} \mathbb{P}(\gamma_n \geq M_\epsilon \tilde{\alpha}_n) \leq \epsilon$，参见例如第 2 节.2]。

命题 6.4. 设 $\hat{\mu} \equiv (\hat{\mu}_n) \subset \mathfrak{D}$ 是 X 的相合估计量，且 $\max_{|w|=2,3} |\hat{\mu}_n i_w - X i_w| = O_P(\tilde{\alpha}_n)$ 具有某个收敛速率 $\tilde{\alpha} \equiv (\tilde{\alpha}_n)$。那么，对于任意 $\varepsilon > 0$ 和任意 $0 \leq q < 1$，存在一个可显式计算的阈值 $n_0 = n_0(\varepsilon, q, \tilde{\alpha}, \hat{\mu}, A, S) \in \mathbb{N}$，使得对每个 $n \geq n_0$，以下结论以至少 q 的概率成立：

$$\text{对于每个 } \hat{\theta} \in \Phi(\hat{\mu}_n) \text{ 存在 } M \in \mathbb{M}_d \text{ 使得 } \sup_{t \in \mathbb{I}} \frac{|\theta X_t - M S_t|}{|M S_t|} \mathbb{1}_X(S_t) \leq \varepsilon; \tag{100}$$

具体而言，这样一个阈值 n_0 可由下式给出

$$n_0 := \min \left\{ n \geq m_{1-q} \;\middle|\; \tilde{\alpha}_n < \frac{\eta(\tilde{\delta}(\varepsilon/\tilde{c}_4))}{(1 + \|A^{-1}\|_\infty)^3 M_{1-q}} \right\} \tag{101}$$

对于 *Notation 6.3* 中的任意 m_{1-q} 和 M_{1-q}，$\tilde{c}_4$ 以及 *Theorem 5.12* 中的 $\delta \equiv \delta(\epsilon; \tilde{c}_i)$（针对 $\alpha \equiv 1$ 且与 $c_{\mathfrak{S}}$ 无关），并且 $\eta := \omega_\varphi^{-1}$: $\mathbb{R}_+ \to \mathbb{R}_+$ 在函数 $\varphi = \varphi([\tilde{\mu}]_\mathfrak{c})$: $V \to \mathbb{R}_+$ 于 $[S]_\mathfrak{c}$ 处的连续模 $\omega_\varphi(\tilde{\eta}) := \sup\{\varphi(v) \mid v \in V, \|v - [S]_\mathfrak{c}\|_\infty \leq \tilde{\eta}\}$ 的逆；该函数由式 *(44)*$|_{\mu=S}$ 的右端定义（固定 $\mu = S$）。

证明。令 $B := A^{-1}$，且 $v_n := B \cdot \hat{\mu}_n$。由引理 3.7 可得二阶和三阶张量的恒等式 $[v_n]^{(2)} := (h v_n i^{ij}) = B^{\otimes 2}(h \hat{\mu}_n i^{ij})$ 以及 $[v_n]^{(3)} := (h v_n i^{ijk}) = B^{\otimes 3}(h \hat{\mu}_n i^{ijk})$，其中 $B^{\otimes 2} = B \otimes B$ 和 $B^{\otimes 3} = B \otimes B \otimes B$ 是 B 的克罗内克幂，另见 [34, 命题 7.53]。由于 $B^{\otimes m}$ 在 $(\mathbb{R}^d)^{\otimes m}$ 上的作用是连续的，由收敛（99）依概率成立以及连续映射定理可得，对于 $m = 2$, 3，同样有 $[v_n]^{(m)} = B^{\otimes m}[\hat{\mu}_n]^{(m)} \longrightarrow_{n \to \infty} B^{\otimes m}[X]^{(m)} = [S]^{(m)}$ [依概率收敛]。事实上，我们有 $\tilde{\gamma}_n := \max_{|w|=2,3} |h v_n i_w - h S i_w| = \|[v_n]_\mathfrak{c} - [S]_\mathfrak{c}\|_\infty \leq \|B\|_\infty \|[\hat{\mu}_n]_\mathfrak{c} - [X]_\mathfrak{c}\|_\infty = C \gamma_n$，其中 $B := B^{\otimes 2} + B^{\otimes 3}$ 是 $(\mathbb{R}^d)^{\otimes 2} \oplus (\mathbb{R}^d)^{\otimes 3} \cong V$ 上的（直和）线性算子，其 ℓ_∞ 诱导的算子范数为 $C := \|B\|_\infty (\leq \max\{\|B^{\otimes 2}\|_\infty, \|B^{\otimes 3}\|_\infty\}) \leq \max\{\|B\|_\infty^2, \|B\|_\infty^3\}$。现在取 $\tilde{\varepsilon} > 0$，并令 $\varepsilon := \tilde{\varepsilon}/\tilde{c}_4$ for $\tilde{c}_4 \equiv \tilde{c}_4(S, A)$，其中 > 0 如定理 5.12 所述。后者进而给出一个显式半径 $\delta = \delta(\varepsilon, S, A, \tilde{\kappa}_0) > 0$，使得（96）对 $\mu_\star = S$ 成立，也就是说

$$\sup_{(v,A) \in \mathbb{B}_\delta(S,A)} \partial_{\hat{\Phi}}(v, A) \leq \tilde{\varepsilon}. \tag{102}$$

为将此与 (99) 联系起来，注意到根据 $\delta(\cdot, \cdot)$ 的定义 (42) 及连续性[29]，存在 $\eta > 0$，使得对每个满足 $\max_{|w|=2,3} |h v i_w - h S i_w| < \eta$ 的 $v \in \mathfrak{D}$，都有 $\delta(S, v) < \delta$。现在固定任意 $0 \leq q < 1$，并令 $q' := 1 - q$。由于 $\gamma_n = O_P(\tilde{\alpha}_n)$，我们可以找到 $m_q, M_q > 0$，使得 $\sup_{n \geq m_q} \mathbb{P}(\gamma_n \geq M_q \cdot \tilde{\alpha}_n) \leq q'$。由于 $\tilde{\alpha}_n \downarrow 0$，我们可以选取一个（最小的）$n_0 \geq m_q$，使得对每个 $n \geq n_0$，都有 $\tilde{\alpha}_n \leq \eta/(C M_q)$。对于这个 n_0，我们有 $\sup_{n \geq n_0} \mathbb{P}(\tilde{\gamma}_n \geq \eta) \leq \sup_{n \geq n_0} \mathbb{P}(C \gamma_n \geq \eta) \leq \sup_{n \geq n_0} \mathbb{P}(\gamma_n \geq M_q \tilde{\alpha}_n) \leq \sup_{n \geq m_q} \mathbb{P}(\gamma_n \geq M_q \tilde{\alpha}_n) \leq q'$，因此 $\inf_{n \geq n_0} \mathbb{P}(\tilde{\gamma}_n < \eta) \geq 1 - q' = q$。于是，对每个 $n \geq n_0$，以至少 q 的概率有 $\mathrm{d}((v_n, A), (\mu_S, A)) = \delta(S, v_n) < \delta$，也就是说 $(v_n, A) \in \mathbb{B}_\delta(S, A)$。由 (102) 和引理 2.9 可得断言 (100)（注意此处根据 (97)，(C.11) 中的空间支撑可以替换为整个 $\mathbb{R}^d$）。公式 (101) 只是对该证明的总结。□

6.2. **异步观测**。除了统计估计（99）之外，时变信号盲反演中的另一种系统误差来源，是偏离瞬时性假设（2的情况；这种偏离由分量的异步时间离散化引起 $X^1, \ldots, X^d$ 属于 X，参见例如 [2] 关于协方差估计的经典背景：

[29] 更确切地说，由于 $\delta(S, v) = \varphi([v]_\mathfrak{c})$ 对于 $\varphi: \mathcal{V} \to \mathbb{R}_+$ 连续，且 $\varphi([S]_\mathfrak{c}) = 0$。29

实际中，通道信号 $\boldsymbol{X}^i$ 通常分别由不同的传感器观测；这些传感器在某些离散时间点集合上进行测量，而这些集合可能因传感器而异。经验观察（例如 [63, 79]）表明，这种异步性会影响盲源分离性能，甚至使其失效。如下述命题所示，因果拓扑（60）足够粗略，可以捕捉这种“异步性扰动”，从而可通过定理5.12量化并分析其影响。

为此，我们可以将式 (2) 中的同步性违背形式化如下：记 $\boldsymbol{X}^i_{\mathcal{J}_i} := \big(\boldsymbol{X}^i_t \mid t \in \mathcal{J}_i := \{0 \le t^{(i)}_0 < t^{(i)}_1 < \ldots < t^{(i)}_{n_i-1} \le 1\}\big)$ 为通道 $\boldsymbol{X}^i$ 的传感器数据，并用 $\hat{}\,\boldsymbol{X}^i_{\mathcal{J}_i}$ 表示该数据的分段线性插值。形式上，对于 $d = 1$，$\hat{}\,\boldsymbol{X}^i_{\mathcal{J}_i} = \hat{\iota}_{n_i}(\boldsymbol{X}^i_{\mathcal{J}_i})$，其中 $\hat{\iota}_{n_i}$ 如式 (25) 所示，见注 3.4。若 $\mathcal{J}_1 = \mathcal{J}_2 = \ldots = \mathcal{J}_d$，则称通道 $\boldsymbol{X}^i$ 是同步观测的；否则称 $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^d$ 是异步观测的。

记 $\|\mathcal{J}_i\| := \max_{\nu \in [n_i-1]} |t^{(i)}_\nu - t^{(i)}_{\nu-1}|$ 为 $\mathcal{J}_i$ 的网格大小。根据定理 5.12，我们可以通过以下方式处理异步性：注意到，在依赖于通道的采样时间 $(\mathcal{J}_i) \equiv (\mathcal{J}1, \ldots, \mathcal{J}_d)$ 上，$\boldsymbol{X}$ 的插值观测 $\hat{}\,\boldsymbol{X}^*_{(\mathcal{J}_i)} := \big(\hat{}\,\boldsymbol{X}^1_{\mathcal{J}_1}, \cdots, \hat{}\,\boldsymbol{X}^d_{\mathcal{J}_d}\big)$ 在 $\|(\mathcal{J}_i)\| := \max_i \|\mathcal{J}_i\| \to 0$ 时收敛到式 (60) 中的 $\boldsymbol{X}$。下文中的符号 $\tilde{\varrho}$ 表示式 (16) 中的相对误差。

命题 *6.5.* 对于每个 $i \in [d]$，令 $(\mathcal{J}^n_i)_{n\in\mathbb{N}}$ 为 $[0,1]$ 的一列剖分，其网格大小 $\|\mathcal{J}^n_i\|$ 收敛到零。进一步假设 $\mathbb{E}\|\boldsymbol{X}\|^3_{1\text{-var}} < \infty$，并令 $\hat{}\,\mu_n := \mathbb{P} \circ \hat{}\,\boldsymbol{X}^{-1}_n \in \mathfrak{D}$ 为 $\hat{}\,\boldsymbol{X}^*_n := \hat{}\,\boldsymbol{X}^*_{(\mathcal{J}^n_i)}$ 的分布。那么对于每个 $\varepsilon > 0$，存在一个显式的 $n_0 = n_0(\varepsilon, A, S, (\hat{}\,\mu_n)) \in \mathbb{N}$，使得误差界 $\sup_{n \ge n_0} \mathrm{dist}_{\tilde{\varrho}}(\hat{}\,\Phi(\hat{}\,\mu_n) \cdot \boldsymbol{X}, \mathrm{dSe}_{\mathbb{m}}) \le \varepsilon$ 几乎必然成立。

证明. 这由定理 5.12 和一些标准的收敛论证得出。事实上：记 $\hat{}\,\pi^i_n : \mathrm{C}^1_1 \to \mathrm{C}^1_1$（参见 (19)）为线性算子，它将一个 C^1_1-路径 $y : [0,1] \to \mathbb{R}$ 映射为其 $\mathcal{J}^i_n$-分段线性插值 $\hat{}\,y_{\mathcal{J}^i_n} := \hat{}\,\iota_{|\mathcal{J}^i_n|}(\pi_{\mathcal{J}^i_n}(y))$（参见注 3.4）；注意，对于所有 $y \in \mathrm{C}^1_1$ 以及每个 $n \in \mathbb{N}$，都有 $\|\hat{}\,\pi^i_n(y)\|_{1\text{-var}} \le \|y\|$，并且 $\lim_{n\to\infty} \hat{}\,\pi^i_n = \mathrm{id}$ 逐点 [于 C^1_1] 上（相对于 $\|\cdot\|_\infty$），见 [34, 定理 5.23]。因此，对于每个 $x \in \mathrm{C}^1_d$ 和 $n \in \mathbb{N}$，

$$d^{-1/2}\|x - \hat{x}^*_n\|_\infty \le \max_{i\in[d]} \big\|x^i - \pi_i(\hat{x}^*_n)\big\|_\infty = \max_{i\in[d]} \big\|x^i - \hat{\pi}^i_n(x^i)\big\|_\infty \longrightarrow 0 \quad (\text{as } n \to \infty)$$

且 $\|\hat{}\,x^*_n\|_{1\text{-var}} \le \sum_{i=1}^d \|\hat{}\,\pi^i_n(x^i)\|_{1\text{-var}} \le \sum_{i=1}^d \|x^i\|_{1\text{-var}}$（回忆一下，第一个不等式由 (23) 得出）。由 [34, 引理 5.27 (i)], 可知，上述结果意味着，对于每个 $x \in \mathrm{C}^1_d$ 以及任意 $p > 1$，$\lim_{n\to\infty} \|x - \hat{}\,x^*_n\|_{p\text{-var}} = 0$。因此，根据 $\mathfrak{sig}(\cdot)$ 的 p-变差连续性（对 $p \in [1,2)$ 可得

$$\lim_{n\to\infty} \mathfrak{sig}_w(\hat{x}^*_n) = \mathfrak{sig}_w(x) \qquad (\text{each } |w| \le 3)$$

对于每个 $x \in \mathrm{C}^1_d$。由此以及 $L_1(\mathbb{P}_{\boldsymbol{X}})$-支配关系 $|\mathfrak{sig}_w(\hat{}\,x^*_n)| \lesssim \|\hat{}\,x^*_n\|^3_{1\text{-var}} \le d^3\|x\|^3_{1\text{-var}}$（该关系由 (C.13) 背后的签名估计保证），我们可以应用控制收敛定理得到 (99)，即 $[\hat{}\,\mu_n]_{\mathfrak{c}} \to [X]_{\mathfrak{c}}$，进而 $\delta(\hat{}\,\mu_n, X) \to 0$ 当 $n \to \infty$ 时。所述结论等价于 (100)，现可仿照命题 6.4 的证明得出。□

6.3. **加性噪声与乘性噪声**。尽管人们不断努力开发更精密的测量仪器和数据处理技术，经验观察（例如实验测量或计量经济学记录）却从来都不可能完全准确。相反，任何类型的基于观测的数据都容易受到不必要的偏差和污染；我们需要了解这些影响，否则它们可能会破坏特定统计方法的（算法）一致性。对此，一种应对方式是采用数学上的“噪声模型”。其中最为成熟的是影响信号 $\boldsymbol{Y} = (\boldsymbol{Y}_t)_{t\in\mathbb{I}}$ 的加性 噪声和乘性 噪声，

$$\boldsymbol{Y}_t = \boldsymbol{Y}_t + \boldsymbol{\eta}_t \qquad \text{or} \qquad \boldsymbol{Y}_t = \boldsymbol{\eta}_t \cdot \boldsymbol{Y}_t \qquad (t \in \mathbb{I}) \tag{103}$$

对于某个随机过程[30] $\boldsymbol{\eta} = (\boldsymbol{\eta}_t)_{t\in\mathbb{I}}$ 取值于 $\mathbb{R}^d$，称为“噪声”。对于含噪盲源分离模型 [69], ，噪声 $\boldsymbol{\eta}$ 通常还须满足额外的结构假设，这些假设往往是（半）参数化的，而且通常相当严格，参见例如 [9, 28, 55, 58]及其中的参考文献。[31]我们的拓扑框架具有一般性，因此可以将这类假设减至最低限度。

在经典的盲源分离语境（2）中，若 f 是线性的，那么加性噪声（103）是作用于 $\boldsymbol{Y} = \boldsymbol{X}$ 还是 $\boldsymbol{Y} = \boldsymbol{S}$ （或两者）本质上无关紧要，因为

$$\boldsymbol{X} + \boldsymbol{\eta} \;=\; A(\boldsymbol{S} + \boldsymbol{\eta}') \qquad \text{for} \quad \boldsymbol{\eta}' := A^{-1}\boldsymbol{\eta}; \tag{104}$$

对于乘性噪声造成的扰动，我们可以区分[32] 以下几种情况

$$\boldsymbol{X}_t^{\boldsymbol{\eta}} := \boldsymbol{\eta}_t \cdot \boldsymbol{X}_t \quad \text{or} \quad \boldsymbol{S}_t^{\boldsymbol{\eta}} := \boldsymbol{\eta}_t \cdot \boldsymbol{S}_t \qquad (t \in \mathbb{I}). \tag{105}$$

The定理 5.12 可以量化这些扰动对盲反演精度的影响 n.

P命题 6.6。 设 $\boldsymbol{\eta}$ 是一个分布为 $\eta \in \mathfrak{O}$的噪声过程，并考虑如下扰动 s

$$\tilde{\boldsymbol{X}}^{(1)} := A(\boldsymbol{S} + \boldsymbol{\eta}) \quad \textit{and} \quad \tilde{\boldsymbol{X}}^{(2)} := A\boldsymbol{S}^{\boldsymbol{\eta}} \quad \textit{and} \quad \tilde{\boldsymbol{X}}^{(3)} := \boldsymbol{X}^{\boldsymbol{\eta}}, \tag{106}$$

其中 $\tilde{X}^{(i)} : = \mathbb{P}_{\tilde{\boldsymbol{X}}^{(i)}}$ 分别表示它们的分布。于是存在显式常数 $c_j = c_j\,(S, A, \Delta_\kappa) \;>0$（ $j = 0, \ldots, 3$），具有如下性质：给定 $\varepsilon > 0$，令 $\beta_\varepsilon : = c_0 \wedge \big(c_1\ (\varepsilon/\ (c_2 + \varepsilon)\ )^{\,2}\big)$，则

$$\textit{for each}\ \ \hat{\theta} \in \hat{\Phi}(\tilde{X}^{(i)}) \ \ \textit{there is}\ \ M \in \mathrm{M}_d \ \ \textit{with}\ \ \tfrac{\|\hat{\theta} - MA^{-1}\|}{\|MA^{-1}\|} \le \varepsilon \ \ \textit{and} \tag{107}$$

$$\max\left\{\partial_{\hat{\Phi}}(\mathbb{P}_{\boldsymbol{S}+\boldsymbol{\eta}}, A), \partial_{\hat{\Phi}}(S^{\eta}, A), \partial_{\hat{\Phi}}(S, \boldsymbol{\eta}A)^*\right\} \;\le\; \varepsilon, \tag{108}$$

p假设 $\boldsymbol{S}$ 和噪声过程 $\boldsymbol{\eta}$ 满足以下依情况而定的假设：

- 对于情形 $i = 1$，假设 $\boldsymbol{\eta}$ 和 $\boldsymbol{S}$ 相互独立且均为均值平稳，并且 $\boldsymbol{\eta}$满足增长条件 $\sum_{\nu=0}^{d} \mathrm{hSi}_{\nu\nu}^{-1} \big\| N_S^{-1}[\eta]_\nu N_S^{-1} \big\|^2 \le \beta_\varepsilon^2$；
- 对于情形 $i = 2$，假设 $\boldsymbol{S}$ 是均值平稳且中心化的，并且 $\boldsymbol{\eta}$ 与 $\boldsymbol{S}$相互独立，且满足增长条件 $\beta_2(S, \eta) \le \beta_\varepsilon^2$;
- 对于情形 $i = 3$，假设 $\boldsymbol{\eta}$与 $\boldsymbol{S}$相互独立，且均值平稳，具有 $\mathbb{E}[\boldsymbol{\eta}] \equiv \mathrm{I}$ 且满足增长条件 $\beta_3(X, \eta) \le \gamma_\varepsilon$，其中 β_3 来自 (F.121)，并且对于 $\gamma_\varepsilon = \gamma_\varepsilon(\beta_\varepsilon, A) > 0$，其定义紧接着给出；

这里，$\mathrm{hSi}_{00} := 1$ 和 $\beta_2(S, \eta) := \sum_{i=1}^{d} \mathrm{hSi}_{ii}^{-2}\,(\tilde{\alpha}_i^2 + \mathrm{hSi}_{ii}^{-1}\check{}\alpha_i^2)$，其中 $\tilde{\alpha}_i := $(F.118)，$\check{}\alpha_i := $(F.119)。证明。定理 5.12 给出常数 $\tilde{c}_j = \tilde{c}_j(S, A) > 0\,(j = 0, 1, 2)$ 和 $\tilde{c}_4 = \tilde{c}_4(S, A, \Delta_\kappa) > 0$，使得对任意 $\varepsilon > 0$ 和 $r_{\tilde{\varepsilon}} := \tilde{c}_0 \wedge \big[\tilde{c}_1^2\big(\tilde{\varepsilon}/(\tilde{c}_2 + \tilde{\varepsilon})\big)^2\big]$，若满足 $\tilde{\varepsilon} := \varepsilon/\tilde{c}_4$，则有

若 $\delta(S,\ \tilde{\mu}) \le r_{\tilde{\varepsilon}}$ 则：对于每个 $\hat{}\ \theta \in \Phi(A\tilde{\mu})$，都存在 $M \in \mathrm{M}_d$，使得 $\frac{\|\hat{}\theta - MA^{-1}\|}{\|MA^{-1}\|} \le \varepsilon,$

˙上述蕴含关系适用于所有 $\tilde{\mu} \in \mathfrak{O}$。要证明这一点，只需将（95）与（E.94）背后的论证、((E.95) 以及 $\tilde{c}_4$ 的定义结合起来。令 $c_j : = \tilde{c}_j$ 对于

---

[30] 乘法 $\boldsymbol{\eta}_t \cdot \boldsymbol{Y}_t$ 按分量理解，即对于每个 $t \in \mathbb{I}$，$\boldsymbol{\eta}_t \cdot \boldsymbol{Y}_t \equiv (\boldsymbol{\eta}_t^1 \boldsymbol{Y}_t^1, \cdots\ \boldsymbol{\eta}_t^d \boldsymbol{Y}_t^d)$。[31] 为了完全避免处理任何可识别性问题，一些作者对 $\boldsymbol{S}$或 $\boldsymbol{\eta}$施加相当严格的（“对比性”）结构假设，以确保加性噪声模型 (104) 保持精确可逆。此类假设的典型例子包括：$\boldsymbol{S}$的时间依赖性与 $\boldsymbol{\eta}$的时间独立性 [25], ; $\boldsymbol{S}$的非高斯性与 $\boldsymbol{\eta}$的高斯性 [48], ; 或 [自]协方差的时变性 $(\boldsymbol{S})$ 与组内平稳混杂 $(\boldsymbol{\eta})$ [77]。当然，这些情形都是特意挑选的例外；要处理一般的无模型情形 (104)，则需要更全面的分析框架，见命题 6.6。[32] 由于 A 是一个同胚，这些情形本质上也等价，但请参见命题 6.6 的证明。* 记 $\boldsymbol{\eta}A$：$[0,1] \times \mathbb{R}^d \;\to\; \mathbb{R}^d, (t, u) \;\mapsto\; \boldsymbol{\eta}_t Au := \mathrm{ddiag}(\boldsymbol{\eta}_t^1, \cdots\ \boldsymbol{\eta}_t^d) \cdot\ Au$（一个随机时变变换），则 (108) 中最后一个不等式（“$\partial_{\hat{\Phi}}(S, \boldsymbol{\eta}A) \;\le\; \varepsilon$”）应理解为：$\partial_{\hat{\Phi}}(S, \boldsymbol{\eta}A) \;\equiv \sup^{B \in \hat{\Phi}(\mathbb{P}_{(\boldsymbol{\eta}_t A S_t \,|\, t\in[0,1])})} \inf_{(M,v)\in \mathrm{M}_d \times \mathbb{R}^d} \sup_{t\in[0,1]} \sup_{u \in D_{S_t}^{(v)}} {}^{|B} \circ (\boldsymbol{\eta}_t A)\big(\tfrac{u-v}{|Mu|} - Mu|\ \mathbb{1}_{\times}\big)\,(Mu) \;\le \varepsilon$ 以概率 1 成立。

$j=0$，1 和 $c_2 := \tilde{c}_2\tilde{c}_4$ 可得 $r_{\tilde{\varepsilon}} = \beta_\varepsilon$，对于 β_ε（如上所定义）。因此，只要满足以下条件，（107）便成立

$$\delta(S, A^{-1}\tilde{X}^{(i)}) \le \beta_\varepsilon \quad \text{for each of the corruptions in (106).} \tag{109}$$

此外，由定理 5.12 (96) 还可进一步得到，(109) 也意味着 $\partial_{\hat{\Phi}}(\upsilon^{(i)}, A) \le \varepsilon$对于 $\upsilon^{(i)} := A^{-1}\tilde{X}^{(i)}$ 以及每个 $i = 1, 2, 3$ 均成立，从而证明了情形 $i = 1, 2$ 下的 (108)。为说明它也意味着情形 $\partial_{\hat{\Phi}}(S, \eta A) \le \varepsilon$，因此整体上成立 (108)，请注意（参见脚注 $*$）

$$\sup_{t\in[0,1]} \sup_{u\in D_{S_t}^{(\upsilon)}} \frac{|B\circ(\eta_t A)(u-v)-Mu|}{|Mu|}\mathbb{1}_\times \le \sup_{u\in D_{\upsilon^{(3)}}^{(\upsilon)}} \frac{|BA(u-v)-Mu|}{|Mu|}\mathbb{1}_\times \quad \text{almost surely}$$

（对于任意 $B \in \hat{\Phi}(\tilde{X}^{(3)})$ 以及每个 $(M, v) \in \mathbb{M}_d \times \mathbb{R}^d$），因此 $\partial_{\hat{\Phi}}(S, \boldsymbol{\eta} A) \le \partial_{\hat{\Phi}}(\upsilon^{(3)}, A)$ 几乎处处成立。首先证明加性噪声情形 $i = 1$ 下的 (109)；三元组 $(\tilde{X}^{(1)}, A, \tilde{S})$ 连同 $\boldsymbol{S} := \boldsymbol{S} + \boldsymbol{\eta}$，可对该情形作完全一般的表示（参见 (104)）。对于任意 $ijk \in [d]^\star_3$，根据多重线性，

$$\langle\tilde{S}\rangle_{ij} = \mathbb{E}\left[\int_0^1\int_0^t \mathrm{d}\tilde{\boldsymbol{S}}_s^i \,\mathrm{d}\tilde{\boldsymbol{S}}_t^j\right] = \langle S\rangle_{ij} + \langle S^i, \eta^j\rangle + \langle\eta^i, S^j\rangle + \langle\eta\rangle_{ij}, \quad \text{and si}\text{似地} \tag{110}$$

$$\begin{aligned}\langle\tilde{S}\rangle_{ijk} = {} & \langle S\rangle_{ijk} + \langle S^i, S^j, \eta^k\rangle + \langle S^i, \eta^j, S^k\rangle + \langle S^i, \eta^j, \eta^k\rangle + \langle\eta^i, S^j, S^k\rangle \\ & + \langle\eta^i, S^j, \eta^k\rangle + \langle\eta^i, \eta^j, S^k\rangle + \langle\eta\rangle_{ijk},\end{aligned} \tag{111}$$

其中 $\mathsf{h}Y^i, \tilde{Y}^j\mathsf{i} := \mathbb{E}\big[\int_0^1\int_0^t \mathrm{d}\boldsymbol{Y}_s^i \mathrm{d}\,\tilde{\boldsymbol{Y}}_t^j\big]$ 且 $\mathsf{h}Y^i, \tilde{Y}^j, \hat{Y}^k\mathsf{i} := \mathbb{E}\big[\int_0^1\int_0^t\int_0^s \mathrm{d}\boldsymbol{Y}_r^i \mathrm{d}\,\tilde{\boldsymbol{Y}}_s^j \mathrm{d}\,\hat{\boldsymbol{Y}}_t^k\big]$ 对于 $Y, \tilde{Y}, \hat{Y} \in \{S, \eta\}$。由于 $\boldsymbol{\eta}, \boldsymbol{S} \in \mathsf{C}_d^1$，我们可以利用富比尼定理计算上述统计量 (110) 和 (111)，得到

$$\langle S^i, \eta^j\rangle = \int_0^1\int_0^t \mathbb{E}\big[\dot{\boldsymbol{S}}_s^i\dot{\boldsymbol{\eta}}_t^j\big]\mathrm{d}s\mathrm{d}t \quad \text{and} \quad \langle\eta^i, S^j, \eta^k\rangle = \int_0^1\int_0^t\int_0^s \mathbb{E}\big[\dot{\boldsymbol{\eta}}_r^i\dot{\boldsymbol{S}}_s^j\dot{\boldsymbol{\eta}}_t^k\big]\mathrm{d}r\mathrm{d}s\mathrm{d}t \quad \text{etc.} \tag{112}$$

根据以下假设：$\boldsymbol{\eta}$ 以及 $\boldsymbol{S}$ 彼此独立且均为均值平稳过程，我们得到$\mathbb{E}[\dot{\boldsymbol{S}}_s^i\dot{\boldsymbol{\eta}}_t^j] = \mathbb{E}[\dot{\boldsymbol{S}}_s^i]\mathbb{E}[\dot{\boldsymbol{\eta}}_t^j] = 0$ 以及 $\mathbb{E}[\dot{\boldsymbol{\eta}}_r^i\dot{\boldsymbol{S}}_s^j\dot{\boldsymbol{\eta}}_t^k] = \mathbb{E}[\dot{\boldsymbol{S}}_s^j]\mathbb{E}[\dot{\boldsymbol{\eta}}_r^i\dot{\boldsymbol{\eta}}_t^k] = 0$ 等等。为此，我们利用了$\mathbb{E}[\dot{\boldsymbol{S}}_t^i] = 0 = \mathbb{E}[\dot{\boldsymbol{\eta}}_t^j]$ 对每个 i, $j \in [d]$ 以及 $t \in [0, 1]$，这可通过交换期望与求导的次序得到（根据导数 $\dot{\boldsymbol{S}}$ 和 $\dot{\boldsymbol{\eta}}$ 是 $L^1(\mathbb{P})$-支配的这一假设，这是允许的）。这意味着混合统计量（112）均为零，因此剩下

$$[\tilde{S}]_\nu = [S]_\nu + [\eta]_\nu \qquad \text{for each} \qquad \nu = 0, 1, \ldots, d.$$

因此，由（44），我们得到

$$\delta(S, \tilde{S})^2 = \sum_{\nu=0}^d \langle S\rangle_{\nu\nu}^{-1} \big\|N_S^{-1}[\eta]_\nu N_S^{-1}\big\|^2 \le \beta_\varepsilon^2,$$

这证明了加性噪声情形下的（109） $i = 1$，如所需。乘性噪声情形下的（109）的证明 $i = 2$、3 类似，但技术上略复杂，参见附录 F。□

请注意，偏差界（108）简洁地量化了在模型偏差（$\hat{\Phi}$ 下，106）的反演稳定性。事实上，该命题表明，加性噪声情形和乘性噪声情形分别是“无噪声”理想化 $\eta \equiv 0$ 和 $\eta \equiv 1$的连续推广。

6.4. **扩展与展望**。本文的鲁棒性分析还可拓展并应用于若干其他方向，包括推导强非线性混合信号盲反演的稳定性保证，以及研究迭代函数系统的随机动力学。举例而言，我们简要考察如何调整我们的方法，以适应另一种基础假设：源与观测量之间的隐藏混合关系（1）是（强）非线性的。为此，核心更新在于推广线性等变性的代数性质（31），这也是签名的一项核心性质，我们的盲反演算法正是基于该签名构建的，参见第 5.1 节：是否存在一种结构类似的方式，使信号的非线性函数也能

编码在未变换信号的签名表示（28）中？对于一般多项式变换，答案是肯定的，而且这实际上是对（31）的直接推广，参见例如[26]。签名的这种“非线性等变性”使得我们可以将第 4 节中的低阶鲁棒性拓扑直接扩展到更高阶的签名矩；这些高阶矩的相关阶数会随源与观测量之间任意多项式（混合）关系的次数而可控地增长。利用源的恒等式（54）在任意阶上都成立这一事实（至少对均值平稳独立成分源如此），可以从源的这些高阶签名矩推导出相应的一系列可适当对角化的张量缩并（类似于（29）），参见例如 [82, 第 5.9 节。4], 继而，将其推广命题 5.1 和第5.2节中的方法加以推广，便可将这些张量缩并安排成一种形式，通过上述多项式等变关系，明确刻画观测量的精确可逆非线性重变换（模置换和尺度不计，参见 [83, 定义 5] ，这是对（14）的推广）；这些关系描述了相应变换对信号多元签名表示（28）所产生的作用。由此可得一种可明确计算的多项式混合精确盲反演方法，并提供与定理 5.8 和 5.12中的结果类似的明确鲁棒性保证。潜在应用包括对多项式函数迭代所诱导的随机动力系统进行稳定性分析 [37], 等。对这些及其他扩展的进一步探索留待未来研究。

致谢。 作者衷心感谢 Harald Oberhauser 对本研究的指导，以及他提出的许多有益讨论和意见；同时感谢 Aapo Hyvärinen、Zhongmin Qian 和 Markus Reiß 提供的有益讨论和建议。在本项目的主要研究期间，作者获得了牛津大学提供的 Oxford-Cocker 研究生奖学金和数学研究所奖学金的资助。

参考文献

[1] B. Afsari. 矩阵联合对角化问题的敏感性分析。《*SIAM J. Matrix Anal. Appl.*》，**30**.3:1148–1171，2008年。[2] Y. Aït-Sahalia、J. Fan 和 D. Xiu. 基于含噪和异步金融数据的高频协方差估计。《*J. Amer. Statist. Assoc.*》，**105**.492:1504–1517，2010年。[3] T Akutsu 等。独立成分分析在 iKAGRA 数据中的应用。《*Progress of Theoretical and Experimental Physics*》，**2020年**.5:053F01，2020年。[4] W. Arendt 和 K. Urban.《偏微分方程》（德文）。施普林格，2018年。[5] A.V. Arkhangel'skii 和 V.V. Fedorchuk.《一般拓扑学 *I*：基本概念与构造、维数理论》。数学科学百科全书，第 17 卷，施普林格，1990年。[6] R. Aziz、C. Verma 和 N. Srivastava. 基于模糊方法从独立成分子空间中进行特征选择，用于微阵列数据的机器学习分类。《*Genomics Data*》，**8**:4–15，2016年。[7] A. Back 和 A. Weigend. 独立成分分析在从股票收益中提取结构方面的首次应用。《*Int. J. Neural Syst.*》，**8**.4:473–484，1997年。[8] M. Behr、C. Holmes 和 A. Munk. 多尺度盲源分离。《*Ann. Statist.*》，**46**.2:711–744，2018年。[9] M. Belkin、L. Rademacher 和 J. Voss. 高斯噪声下的盲信号分离。第 270–287 页，2013年。[10] P. Billingsley.《概率测度的收敛》。第 2 版，约翰·威利父子出版公司，1999年。[11] V. Bogachev. 测度的平均近似与矩。《*J. Complexity*》，16:390–410，2000年。[12] V. I. Bogachev.《测度论》。第 I、II 卷。施普林格，2007年。[13] P. Bonnier、C. Liu 和 H. Oberhauser. 适配拓扑与高秩签名。《*Ann. Appl. Prob.*》， 2023+。

[14] P. Bonnier 和 H. Oberhauser. 签名累积量、有序划分与随机过程的独立性。《Bernoulli》，26.4:2727–2757，2020年。[15] J. Bruno 和 P. Szeptycki. 量子格、广义预度量与自由局部格。《Appl. Categor. Struct.》，25:1045–1058，2017年。[16] Y. Cai 和 R.-C. Li. 矩阵联合块对角化的扰动分析。《线性代数及其应用》，581:163–197，2019年。[17] Y. Cai 和 C. Liu. 非正交一般联合块对角化的代数方法。《SIAM J. Matrix Anal. Appl.》，38.1:50–71，2017年。[18] G. Chabriel、M. Kleinsteuber、E. Moreau、H. Shen、P. Tichavsky 和 A. Yeredor. 矩阵联合分解与盲源分离：方法综述。《IEEE 信号处理杂志》，31.3:34–43，2014年。[19] A. Chen 和 P. J. Bickel. 一致的独立成分分析与预白化。《IEEE Trans. Signal Process.》，53.10:3625–3632，2005年。[20] A. Chen 和 P-J. Bickel. 高效独立成分分析。《Ann. Statist.》，34.6:2825–2855，2006年。[21] I. Chevyrev 和 T. Lyons. 几何粗糙路径上测度的特征函数。《Ann. Probab.》，44.6:4049–4082，2016年。[22] I. Chevyrev、V. Nanda 和 H. Oberhauser. 拓扑数据分析中的持久性路径与签名特征。《IEEE 模式分析与机器智能汇刊》，42(1):192–202，2018年。[23] I. Chevyrev 和 H. Oberhauser. 用签名矩刻画随机过程的分布。预印本，第 arXiv:1810.1097 页，2018年。[24] I. Chevyrev 和 H. Oberhauser. 用签名矩刻画随机过程的分布。《J. Mach. Learn. Res.》，23(176):1–42，2022年。[25] S. Choi 和 A. Cichocki. 含噪混合信号中非平稳源的盲分离。《电子快报》，36.9:848–849，2000年。[26] L. Colmenarejo 和 R. Preiß. 经多项式映射变换的路径签名。预印本，第 arXiv:1812.05962v1 页，2018年。[27]P. Comon. 独立成分分析：一个新概念？《Signal Processing》，36.3:287–314，1994年。[28] P. Comon 和 C. Jutten（编）。《盲源分离手册：独立成分分析及其应用》。学术出版社，2010年。[29] D. Crisan、J. Diehl、P.K. Friz 和 H. Oberhauser. 鲁棒滤波：相关噪声与多维观测。《Ann. Appl. Prob.》，23(5)，2013年。[30] N. Cvejic、D. Bull 和 N. Canagarajah. 使用独立成分分析改进传感器网络中的监控图像融合。《IEEE 消费电子汇刊》，53.3:1029–1035，2007年。[31] S. B. Damelin 和 W. Miller Jr.《信号处理数学》。剑桥应用数学教材48，剑桥大学出版社，2012年。[32] G. Darmois. 随机联系的一般分析。《国际统计学会评论》，21:2–8，1953年。[33] A. Delorme 和 S. Makeig. EEGLAB：一个用于分析单试次脑电动力学（包括独立成分分析）的开源工具箱。《神经科学方法杂志》，134.1:9–21，2004年。[34] P. K. Friz 和 N. B. Victoir.《作为粗糙路径的多维随机过程：理论与应用》。剑桥高等数学研究 120，剑桥大学出版社，2010年。[35] D. J. H. Garling. Radon 化映射与函数型中心极限定理。《Z. Wahrscheinlichkeitstheorie verw. Geb.》，26:297–308，1973年。

[36] T. Gowers、J. Barrow-Green 和 I. Leader（编）。《普林斯顿数学指南》。普林斯顿大学出版社，2008年。[37] P Grassberger、H Kantz 和 U Moenig。关于 Hénon映射的符号动力系统。《物理学杂志A：数学与一般》，22(24):5217，1989年。[38] P. Green。误差估计札记。讲义，马里兰大学，网址：http://www.math.umd.edu/~psg/401/errorestimation.pdf，2003年。[39] H. W Guggenheimer、A. S Edelman 和 C. R. Johnson。线性方程组条件数的一个简单估计。《大学数学杂志》，26:2–5，1995年。[40] J. Hadamard。关于偏微分问题及其物理意义。《普林斯顿大学公报》，第49–52页，1902年。[41] H. Hälvä、S. Le Corff、L. Lehéricy、J. So、Y. Zhu、E. Gassiat 和 A. Hyvarinen。利用结构化非线性独立成分分析，从含噪数据中解耦可识别特征。第35届神经信息处理系统会议（NeurIPS），34:1624–1633，2021年。[42] B. Hambly 和 T. Lyons。有界变差路径的签名及约化路径群的唯一性。《Ann. of Math.》，171.1:109–167，2010年。[43] F. R. Hampel。鲁棒性的一般定性定义。《Ann. Math. Stat.》，42.6:1887–1896，1971年。[44] N. J. Higham。《矩阵函数：理论与计算》。工业与应用数学学会，2008年。[45] N. J. Higham。《数值算法的精度与稳定性》。工业与应用数学学会，2002年。[46] R. A. Horn 和 C. R. Johnson。《矩阵分析》。剑桥大学出版社，2012年。[47] P. Huber 和 E. Ronchetti。《稳健统计学》。威利出版社，2009年。[48] A. Hyvärinen。利用高斯矩对含噪数据进行快速独立成分分析。载于 ISCAS'99。1999年 IEEE 国际电路与系统研讨会论文集（VLSI），第57–61页。IEEE，1999年。[49] A. Hyvärinen。独立成分分析：最新进展。《英国皇家学会哲学汇刊A辑》，371.1984:20110534，2013年。[50] A. Hyvärinen、J. Karhunen 和 E. Oja。《独立成分分析》。约翰·威利父子出版公司，2001年。[51] A. Hyvärinen 和 P. Pajunen。非线性独立成分分析：存在唯一性结果。《神经网络》，12.3:429–439，1999年。[52] A. Hyvärinen、H. Sasaki 和 R. Turner。利用辅助变量与广义对比学习进行非线性独立成分分析。人工智能与统计学会议，2019年。[53]C. Kanzow。《线性方程组数值计算：直接与迭代方法》。施普林格出版社，2007年。[54] M. Karow 和 D. Kressner。关于矩阵不变子空间的扰动界。《SIAM J. Matrix Anal. Appl.》，35.2:599–618，2014年。[55] M. Kawanabe 和 K.-R. Müller。当源具有方差依赖时，盲分离的估计函数。《J. Mach. Learn. Res.》，6.4，2005年。[56]A Kechris。《经典描述集合论》。《数学研究生教材》第156卷，施普林格，1995年。[57] J. L. Kelley。《一般拓扑学》。D. van Nostrand，1961年。[58] C. Kervazo、N. Gillis 和 N. Dobigeon。线性二次近可分混合的可证明鲁棒盲源分离。《SIAM J. Imaging Sci.》，14.4:1848–1889，2021年。[59] I. Khemakhem、D. P. Kingma、R. P. Monti 和 A. Hyvärinen。变分自编码器与非线性独立成分分析：统一框架。人工智能与统计学会议论文集（AISTATS2020），2020年。[60] F. Kiraly 和 H. Oberhauser。序列有序数据的核。《J. Mach. Learn. Res.》，19:1–45，2018年。[61] S.-I. Lee 和 S. Batzoglou。独立成分分析在微阵列中的应用。《基因组生物学》，4.11:1–21，2003年。

[62] B. Li、J. Lu 和 Z. Yu。近似可交换矩阵的向量式联合对角化。arXiv预印本 arXiv:2205.15519，2022年。[63] H. Li。用于时间序列数据挖掘的基于异步性的主成分分析。《Expert Systems with Applications》，41.6:2842–2850，2014年。[64] W. Linde。巴拿赫空间上测度的矩。《Math. Ann.》，258:277–287，1982年。[65] T. J. Lyons、M. Caruana 和 T. Lévy。《粗糙路径驱动的微分方程》。施普林格，2007年。[66] T. J. Lyons 和 Z. Qian。《系统控制与粗糙路径》。牛津数学专著，牛津大学出版社，2002年。[67] S. Makeig 和 J. Onton。事件相关电位特征与脑电动力学：独立成分分析视角。载于《牛津事件相关电位成分手册》。牛津大学出版社，2011年。[68] J. Miettinen、K. Nordhausen 和 S. Taskinen。基于联合对角化的实数集上的盲源分离：JADE 和 BSSasymp 软件包。《J. Stat. Softw.》，76.2，2017年。[69] E. Moulines、J. F. Cardoso 和 E. Gassiat。利用混合模型对含噪信号进行盲分离和反卷积的最大似然方法。IEEE 国际声学、语音与信号处理会议论文集（ICASSP'97），第3617–3620页，1997年。[70] K. Nordhausen。关于对非平稳时间序列的若干二阶盲源分离方法进行稳健化。《Statistical Papers》，55:141–156，2014年。[71] K. Nordhausen 和 A. Ruiz-Gazen。关于联合对角化在多元统计学中的应用。《J. Multivariate Anal.》，188:104844，2022年。[72] D. Nuzillard 和 A. Bijaoui。多光谱天文图像的盲源分离与分析。《Astron. Astrophys. Suppl. Ser.》，147.1:129–138，2000年。[73]A. Olevskii 和 A. Ulanovskii。《信号重建：唯一性与稳定采样》。载于《采样：理论与应用》，第9–49页。施普林格，2020年。[74] V. M. Panaretos 和 Y. Zemel。《瓦瑟斯坦空间中的统计学导论》。Springer概率与数理统计简明丛书。施普林格·自然，2020年。[75] A Papavasiliou 和 C. Ladroue。《粗糙微分方程的参数估计》。《Ann. Statist.》，39.4:2047–2073，2011年。[76] I. Perez Arribas、G. M. Goodwin、J-R. Geddes、T. Lyons 和 K. Saunders。基于签名的机器学习模型：区分双相障碍与边缘型人格障碍。《Translational Psychiatry》，8(1):274，2018年。[77] N. Pfister、S. Weichwald、P. Bühlmann 和 B. Schölkopf。通过校正组间平稳噪声实现独立成分分析的稳健化。《J. Mach. Learn. Res.》，20.147:1–50，2019年。[78] O. Reiersøl。受误差影响的变量之间线性关系的可识别性。《Econometrica: Journal of the Econometric Society》，第375–389页，1950年。[79]rmcerafl（ https://dsp.stackexchange.com/users/59697/rmcerafl）和 Gillespie（ https://dsp.stackexchange.com/users/55647/gillespie）。异步观测的混合通道的盲源分离。dsp.stackexchange。https://dsp.stackexchange.com/questions/78796（版本：2022年9月16日）。[80]A. Samarov 和 A. Tsybakov。《非参数独立成分分析》。《Bernoulli》，10.4:565–582，2004年。[81] R.J. Samworth 和 M. Yuan。通过非参数最大似然估计进行独立成分分析。《Ann. Statist.》，40.6:2973–3002，2012年。[82] A. Schell。《随机过程的非线性与鲁棒独立成分分析》。博士论文，牛津大学，2022年。[83] A. Schell 和 H. Oberhauser。离散时间与连续时间信号的非线性独立成分分析。（参考文献指向 arXiv 版本 arXiv:2102.02876）。《Ann. Statist.》，2023+。

[84] L. Schwartz. Radon 化映射的应用。巴黎综合理工学院研讨班，巴黎，1969- 1970年。[85] L. Schwartz. 柱状概率与 Radon 化映射。《东京大学理学部纪要，*IA* 分册》，**18**:139–286，1971年。[86] D. Shi、Y. Cai 和 S. Xu. 归一化非正交联合对角化问题的一些扰动结果。《线性代数及其应用》，**484**:457–476，2015年。[87] V. P. Skitovich. 关于正态分布的一个性质。《苏联科学院院刊》，**89**:217–219（俄文），1953年。[88] Y. Takahashi. 巴拿赫空间上 Radon 化映射与概率测度的核之间的关系。《北海道数学杂志》，14(1):97–106，1985年。[89] A. Takahata、E. Nadalin、R. Ferrari、L. Duarte、R. Suyama、R. Lopes、J. Romano 和 M. Tygel. 地球物理信号的无监督处理：盲反卷积和盲源分离若干关键方面的综述。《*IEEE* 信号处理杂志》，**29**.4:27–35，2012年。[90] F. J. Theis 和 Y. Inouye. 联合对角化在盲信号处理中的应用。*2006*年 *IEEE* 国际电路与系统研讨会，第 4 页，2006年。[91] K. Usevich、J. Li 和 P. Comon. 通过酉变换近似矩阵和张量对角化：Jacobi 型算法的收敛性。《*SIAM* 优化杂志》，**30**.4:2998–3028，2020年。[92] A. W. van der Vaart.《渐近统计学》。剑桥统计与概率数学丛书，1998年。[93] G. Zhou、S. Xie、Z. Yang 和 J. Zhang. 使用良态对角化矩阵的非正交近似联合对角化。《*IEEE* 神经网络与学习系统汇刊》，**20**.11:1810–1819，2009年。

附录

A 记号与预备知识 .38

A.1 符号列表.38A.2 离散时间与连续时间信号.39A.3 测度运算.40**B 第2 节的补充动机.** 41B.1 定义2.7 的补充.42B.2 更粗糙路径的扩展.42**C 技术性证明与引理（第2 至5节）.** 42

C.1 引理 2.9 的证明.42 C.2 引理 3.3 的证明.43 C.3 引理 3.7 的证明.43 C.4 命题 3.11 的证明.44 C.5 命题 3.14 的证明.44 C.6 关于定义 4.1 的注记.45 C.7 一些技术估计.45 C.8 引理 4.6 的证明.49 C.9 命题 5.1 和引理 5.4 的证明.50 C.10 Ξ_1 和 Ξ_1 的紧致性. 50

D 第 5.2 节和第 5.3 节的技术引理 5.2 和 5.3 .51

D.1 矩阵 C_μ 是泛型可逆的.51D.2 超平面、逆矩阵与缩并张量.51D.3 特征值间隙. 52D.4 非退化特征子空间上的扰动界 . . . . . 52D.5 若干矩阵函数的连续性与界. 53D.6 极值函数的连续性.54

E 定理证明 5.8 和 5.12.54

E.1 定理证明 5.8 .54E.2 相关工作.60E.3 定理证明 5.12：的连续性 $\mathfrak{I}$.60

F 命题的证明 6.6 (情形 $i = 2,\ 3$) .64

A. 记号与预备知识

A.1. 符号列表。 以下是我们使用的一些符号和术语。

符号	含义	Page
$\mathcal{C}_d$	$:= \{x : [0,1] \to \mathbb{R}^d \mid x$连续$\}$； $\mathbb{R}^d$中的连续路径空间，通常赋予一致范数 $\|x\|_\infty := \sup_{t\in[0,1]} \|x_t\|$。	4
$\mathcal{M}_1(E)$	$:= \{\mu : \mathcal{B}(E) \to [0,1]\}$；空间 E 上所有博雷尔概率测度的集合。	4
$\mathcal{M}_1$	$:= \mathcal{M}_1(\mathcal{C}_d)$；在 $(\mathcal{C}_d, \|\cdot\|_\infty)$ 上的博雷尔概率测度集合。	4
D_Y	参见（A.3）；信号的空间支撑 $Y \in \mathcal{M}_1$。	4
$C^{0,\ 0}(D)$	$: = \{\tilde{f} :\ D \to \mathbb{R}^d \mid \tilde{f}$ 到 $\tilde{f}(D)$ 的同胚映射$\}$； D 上所有同胚映射的集合。	4
$\mathrm{ev}_X(g)$	$\equiv \mathrm{ev}(X,\ g) : = g(X)$，其中 $g :\ D_X \to \mathbb{R}^d$；过程 X 处的求值映射。	4
M_d	$: = \{M \in \mathrm{GL}_d \mid M = D \cdot P$ 对于 $D \in \mathrm{GL}_d$ 对角的且 $P \in \mathrm{P}_d\}$；群的次数为 d 的（实）单项矩阵。	6
$\tilde{\Delta}_d$	$:= \{\Lambda = (\lambda_i \cdot \delta_{ij}) \in \mathrm{GL}_d \mid \lambda_1, \ldots, \lambda_d \in \mathbb{R}\backslash\{0\}\}$；可逆对角矩阵构成的群。	7
$C^{1,1}$	$\equiv C^{1,1}(\mathbb{R}^d) := C^{1,1}(\mathbb{R}^d; \mathbb{R}^d)$；所有 $\mathbb{R}^d$ 上的 C^1-微分同胚构成的群。	7
$\varphi\big\vert_{\tilde{A}}$	映射 $\varphi :\ A \to B$ 到子域 $\tilde{A} \subseteq A$ 的限制。	8

$\mathrm{GL}_d := \{A \in \mathbb{R}^{d\times d} \mid \det(A) \neq 0\}$；域 $\mathbb{R}$ 上 d 次一般线性群。8$|\cdot|$：$\mathbb{R}^n \to \mathbb{R}_+$，$\mathbb{R}^n$ 上的欧几里得范数；特别地，$|\cdot|$ 是 $\mathbb{R} = \mathbb{R}^1$上的绝对值。8$\mathfrak{D}$ [$\hat{\mathfrak{D}}$] 参见 (20)/(41)；由 $\mathtt{sig}$ 个坐标可表示的 [可预度量化信号空间。9/15$[k] := \{1, \dots, k\}$，以及 $[k]_0 := [k] \cup \{0\}$ ($k \in \mathbb{N}$)。11$[d]^\star := \bigcup_{m\geq 0}[d]^{\times m} = \{\emptyset, 1, 2, \dots, 11, 12, 112, \dots\}$；其元素取自 $[d]$ 的所有多重指标的集合，其中 $[d]^0 := \{\emptyset\}$，且 $i_1 \cdots i_m \equiv (i_1, \dots, i_m)$。11$\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$；实数的仿射闭包。11$[d]^\star_{\leq m} := \bigcup_{0\leq j\leq m}[d]^{\times j}$；次数至多为 m 的所有多重指标的集合（$m \in \mathbb{N}$）。14 $\mathrm{ddiag}_{i\in[d]}[a_i]$：$= a_i\ \delta_{ij}\ _{ij}$ $a_1 \quad a_d$ $\theta = \theta_{ij} \in (\cdot)$；主对角线为 $(, \dots,)$ $\mathbb{R}^{d\times d}$ 的对角矩阵；对于 $(\,)$，我们写作 $\mathrm{ddiag}(\,)$ $\mathrm{ddiag}_{i\in[d]}[\theta_{ii}]$。16$D_\varphi := \left(\frac{\partial}{\partial x_j}\varphi_i\right)_{ij}$；映射 $\varphi \equiv (\varphi_1, \cdots\ \varphi_d) \in C^1(G; \mathbb{R}^d)$ 的雅可比矩阵。18$\Delta_m(\mathbb{I}) := \{(t_1, \dots, t_m) \in \mathbb{I}^{\times m} \mid t_1 < \dots < t_m\}$；$\mathbb{I}^{\times m}$ 上相对开 m 单纯形；记作 $\Delta_m \equiv \Delta_m([0,1])$。19 $\mathrm{h}u, v\mathrm{i}_2 := u_1v_1 + \dots + u_dv_d$；$\mathbb{R}^d$. 22 中两个向量 $u = (u_i)$ $v = (v_i)$ 的点积。$\mathrm{P}^\pm_d$：$= \{(\epsilon \cdot \delta_{\sigma(i),j})_{i,j\in[d]} \in \mathrm{GL}_d \mid \epsilon \in \{0, 1\}\ \ \sigma \in S_d\}$；带符号置换矩阵。23$S_d := \{\sigma : [d] \to [d] \mid \sigma$ 为双射$\}$；$[d]$ 的所有置换构成的群。38$f_1 \curlywedge f_2$：$U_1 \times U_2 \to V_1 \times V_2, (u_1, u_2) \mapsto (f_1(u_1)\ f_2(u_2))$；两个映射的笛卡尔积 $f_i : U_i \to V_i$ ($i = 1, 2$)。40 wlog/wrt./s.t. “不失一般性”/“关于”/“使得”。i.o.w. “换言之”

A.2. **离散时间与连续时间信号**。本文中，信号 一词指 $\mathbb{R}^d$中的随机过程，也就是说，指一个有序族（以粗体印刷）

$$\boldsymbol{Y} = (\boldsymbol{Y}_t \mid t \in \mathbb{I}) \quad \text{of (Borel) random vectors} \quad \boldsymbol{Y}_t = (\boldsymbol{Y}_t^1, \cdots, \boldsymbol{Y}_t^d) \text{ in } \mathbb{R}^d, \tag{A.1}$$

对于 $\mathbb{I}$，它是 $\mathbb{R}$的一个有序子集，通常为 $\mathbb{I} = \mathbb{N}$ （“离散时间”）或 $\mathbb{I} = [0, 1]$ （“连续时间”）。标量过程 $\boldsymbol{Y}^i = (\boldsymbol{Y}_t^i \mid t \in \mathbb{I}) \equiv (\boldsymbol{Y}_t^i)$，$i \in [d]$称为 $\boldsymbol{Y}$的分量。

通过缩放 $\mathbb{I}$ 并借助注 A.1 A.1，不失一般性，我们可以假设 $\mathbb{I} = [0, 1]$和[33] 任意信号的向量（A.1）均连续有序，即以概率 1

$$\mathbb{I} \ni t \longmapsto \boldsymbol{Y}_t \quad \text{is continuous.} \tag{A.2}$$

后者是指（以概率 1）样本 $\boldsymbol{Y}(\omega) = (\boldsymbol{Y}_t(\omega))_{t\in\mathbb{I}}$ 属于 $\boldsymbol{Y}$ 中的$\mathcal{C}_d$，即从 $[0, 1]$ 到 $\mathbb{R}^d$的连续路径空间；默认情况下，空间 $\mathcal{C}_d$ 赋有一致范数拓扑。连续性假设（A.2）使我们能够

- 除非另有说明，我们都将“随机过程”与“连续时间随机过程”视为同义词，并将
- 将信号视为随机路径（A.1），也就是说，将其视为某个概率空间上的 $\mathcal{C}_d$值（博雷尔）随机变量 $(\Omega\ \mathscr{F}, \mathbb{P})$。

注 A.1。任何信号 $\boldsymbol{Y} = (\boldsymbol{Y}_t)_{t\in\mathbb{I}}$ 只要具有离散 $\mathbb{I}$，就可以通过对其各点进行分段线性插值，将其视为连续时间过程 $\hat{\boldsymbol{Y}}$，$(\boldsymbol{Y}_t \mid t \in \mathbb{I}) \equiv (\boldsymbol{Y}_{t_j} \mid j \in \mathbb{N})$，即

$$\hat{\boldsymbol{Y}} = \left(\hat{\boldsymbol{Y}}_t := \boldsymbol{Y}_{t_j} + \tfrac{t-t_j}{t_{j+1}-t_j}(\boldsymbol{Y}_{t_{j+1}} - \boldsymbol{Y}_{t_j}) \text{ for } t \in [t_j, t_{j+1})\right).$$

（请注意，该插值与线性变换可交换：$A \cdot \hat{\boldsymbol{Y}} \equiv (A\,\hat{\boldsymbol{Y}}_t) = \widehat{\ \ }(A\boldsymbol{Y}_{t_j})$，对每个 $A \in \mathbb{R}^{d\times d}$ 均成立。）缩放 $\mathbb{I}$ 后，我们还可以假设 $\hat{\boldsymbol{Y}}$ 取值于 $\mathcal{C}_d$。 ◆

[33] 另请参见注 3.4。

通过将信号与其分布等同（约定 1.2），我们可以将[34] 所有信号组成的类嵌入$\mathcal{M}_1 \equiv \mathcal{M}_1(\mathcal{C}_d)$，即定义在$\mathcal{C}_d$上的所有（博雷尔）概率测度构成的空间。这在数学上十分便利，因为这样我们便能用单一的预先赋予结构的定义域（$\mathcal{M}_1$）上的映射来描述盲源分离问题，进而可从一般拓扑的角度灵活地表述并分析鲁棒性概念（见第 2节）。

约定 1.2（将信号与其概率律等同）。从第 2.1 节开始，我们将默认把给定的信号 $\boldsymbol{Y}$（属于 $\mathbb{R}^d$ 的随机变量，定义在 $\mathcal{C}_d$上）与其分布 $Y \equiv \mathbb{P}_{\boldsymbol{Y}} := \mathbb{P}\circ\boldsymbol{Y}^{-1}$（定义在 $\mathcal{C}_d$上的概率测度）等同起来。这也解释了常见的符号滥用 $\boldsymbol{Y}\,[\cong Y] \in \mathcal{M}_1(\mathcal{C}_d)$。

除非另有说明，约定1.2 在本文中始终适用 r.

有时我们会用到信号的空间支撑（A.1），其定义为集合

$$D_{\boldsymbol{Y}} = \overline{\bigcup_{t\in\mathbb{I}} \operatorname{supp}(\boldsymbol{Y}_t)}, \quad \text{where} \quad \operatorname{supp}(\boldsymbol{Y}_t) \equiv \operatorname{supp}(\mathbb{P}_{\boldsymbol{Y}_t}) \tag{A.3}$$

是分布 $\mathbb{P}_{\boldsymbol{Y}_t}$ 的支撑，其中 $\boldsymbol{Y}_t$ 的闭包是相对于$\mathbb{R}^d$上的欧几里得拓扑而言的。集合 $D_{\boldsymbol{Y}}$ 是 $\mathbb{R}^d$ 的最小闭子集，其中包含 $\mathbb{P}$-几乎每条 $\boldsymbol{Y}$的样本路径（的迹）；详情请参见 [83, 第 4.2] 节。若信号 $\boldsymbol{Y}$ 的空间支撑 $D_{\boldsymbol{Y}}$ 包含于 $\mathbb{R}^d$的某个超平面内，我们称其为“退化”信号。

（若读者不习惯（A.3），为求简便，可假设 $D_{\boldsymbol{Y}} = \mathbb{R}^d$ 始终成立。）

支撑（A.3）以及信号的其他概念和运算（A.1）都可以自然地推广到$\mathcal{M}_1$, 中的测度，如下一小节所述。

A.3. **测度运算**。盲源分离问题自然涉及对信号施加的一系列（非）线性变换 φ 。由于 $\mathcal{M}_1(\mathcal{C}_d) \equiv \mathcal{M}_1$ 上的（博雷尔）概率测度空间是凸的，但不是向量空间，因此需要在 $\mathcal{M}_1(\mathcal{C}_d)$ 上“诱导”这些变换，才能使其定义良好。

给定某个目标向量空间 V，一种规范的实现方式是通过推前

$$\varphi_* : \mathcal{M}_1 \ni \mu \longmapsto \varphi_*(\mu) \equiv \mu\circ\varphi^{-1} \in \mathcal{M}_1(V) \quad \text{of a (Borel) map } \varphi : \mathcal{C}_d \to V. \tag{A.4}$$

O由于某些特殊信号作用极为常见（A.4），因此将为它们单独命名 :

对于 $x = (x_t^1, \cdots x_t^d)_{t\in[0,1]} \in \mathcal{C}_d$ 中任意固定路径，考虑映射

$\pi_i : x \mapsto x^i \equiv (x_t^i)_{t\in[0,1]}$ 和 $\pi_t : x \mapsto x_t$ 以及 $\pi_t^i : x \mapsto x_t^i$ $(i \in [d], t \in [0,1])$，并将它们(A.5)

别由信号 $\mu \in \mathcal{M}_1$ 诱导的推前测度记为$\mu^i := \mu\circ\pi_i^{-1}$和 $\mu_t := \mu\circ\pi_t^{-1}$以及 $\mu_t^i := \mu\circ(\pi_t^i)^{-1}$; 于是，我们也可以写作 $\mu = (\mu^1, \cdots \mu^d)$ 和 $\mu = (\mu_t)$，以表示 μ 具有空间边缘测度 μ^i 和固定时刻边缘测度 μ_t。

按照（A.3）， $\mu \in \mathcal{M}_1$ 的空间支撑定义为

$$D_\mu := \overline{\bigcup_{t\in\mathbb{I}} \operatorname{supp}(\mu_t)} \tag{A.6}$$

任意 $g : \mathbb{R}^d \to \mathbb{R}^d$ 都会诱导一个映射 $\tilde{g} : \mathcal{C}_d \to \mathcal{C}_d$，其定义为 $\tilde{g}(x) := (g(x_t))_{t\in[0,1]}$，并令

$$g(\mu) \equiv g_*\mu := \mu\circ\tilde{g}^{-1}. \tag{A.7}$$

扩展（A.5），我们还可以定义映射 $\mu_{s,t} \equiv \mu_t - \mu_s := (\pi_t - \pi_s)_*(\mu)$ 的时间增量 $\mu_{s,t} \equiv \mu_t - \mu_s$，以及乘积 $\mu_{s,t}^i \cdot \mu_{s,t}^j := \left[\mathfrak{m}\circ(\pi_{s,t}^i \times \pi_{s,t}^j)\right]_*(\mu)$，其中映射为 $\pi_{s,t}^k := \pi_t^k - \pi_s^k$ 和 $\mathfrak{m}(a,b) := a\cdot b$ 以及 $s,t \in \mathbb{I}$。此外，$\mathbb{E}\nu := \int_{\mathbb{R}^n} u\,\nu(\mathrm{d}u)$ 是定义在 $\mathbb{R}^n$ 上的（可能带符号的）博雷尔测度ν 的期望，而测度 $\mu = (\mu_t) \in \mathbf{M}_1$ 若满足对每个 $t \in \mathbb{I}$ 都有 $\mathbb{E}[\mu_t] = \mathbb{E}[\mu_0]$，则称为均值平稳。

[34] .....直至分布相等，这对于我们的目的而言完全足够.....

最后，若信号 $\mu \in \mathcal{M}_1$ 具有（相互）独立分量，或称为IC，则 $\mu = \mu^1 \otimes \cdots \otimes \mu^d$（在识别 $\mathcal{M}_1(\mathcal{C}_d) \cong \mathcal{M}_1(\mathcal{C}_1^{\times d})$ 的意义下，参见例如 [83, 附录 A。1]）。所有属于 $\mathcal{M}_1$ 的 IC 信号构成的集合记为 $\mathscr{S}_\perp$。

例 A.2。为作说明，假设 $\mu = \mathbb{P}_{\boldsymbol{Y}} \in \mathcal{M}_1$ 对于某个信号 $\boldsymbol{Y} \equiv (\boldsymbol{Y}^i) = (\boldsymbol{Y}_t)$ 属于 $\mathbb{R}^d$。则 $\mu^i = Y^i$ 且 $\mu_t = Y_t$ 且 $g(\mu) = g(Y) \equiv \mathbb{P}_{g(\boldsymbol{Y})}$ 对于任意可测的 $g\colon \mathbb{R}^d \to \mathbb{R}^d$。此外 $\mathbb{E}\,\mu_{s,\,t} = \mathbb{E}\boldsymbol{Y}_t - \mathbb{E}\boldsymbol{Y}_s$ 且 $\mathbb{E}[\mu^i_{s,t}\mu^j_{s,t}] = \mathbb{E}[\boldsymbol{Y}^i_{s,t}\boldsymbol{Y}^j_{s,t}]$ 对于任意 $i,\ j \in [d]$ 以及 $s,\ t \in \mathbb{I}$，并且μ 是均值平稳的，当且仅当过程 $\boldsymbol{Y}$ 是均值平稳的；μ 是独立同分布且中心化的，当且仅当 $\boldsymbol{Y}$的分量$\boldsymbol{Y}^1$，……，$\boldsymbol{Y}^d$（作为随机过程）彼此统计独立。

B. 第 2 节的补充动机

以下内容基于例 1.1，旨在阐释并说明第 2 节中引入的定义。尽管下文关于盲源分离鲁棒性的观点和思想具有相关性且直观，但现有文献似乎并未提供灵活且信息丰富的理论框架来对其进行形式化并加以利用。本文是朝这一方向迈出的首次尝试。

例 B.1（重访鸡尾酒会）。例 1.1 中说话者所发出的原始语音信号，可以建模为从随机过程的分量 S^i 中抽取的样本 [drawn from]，$S = (S^1, \dots, S^d) \in \mathrm{M}_1$。这些信号的麦克风录音随后形成如下混合信号 $X^i = f_i(S^1, \dots, S^d)$，经组合后得到观测量 $X = (X^1, \dots, X^d) = f(S)$，其中 $f = (f_1, \dots, f_d)\colon D_S \to \mathbb{R}^d$ 是某个可逆映射。因此，例 1.1 中的逆问题可转化为针对三元组 (X, S, f) 的盲源分离问题 $(\star)$，其求解精度为 $\mathsf{d} \cdot \mathsf{e}$，而例中并未明确给出该精度。不过，例中指出，目标是“从 X 中恢复每位说话者说过的内容”。因此，目标并非从 X 中精确恢复 S，而只是从 X 中恢复消息 —— 也就是他们各自讲话的信息内容；究竟是哪位说话者说了什么，或他们原本说话时的音量有多大，都无关紧要。

这忽略了原始源 S 各分量的顺序 [即语音信号归属于其原始说话者的情况（通过枚举确定）] 以及各分量的尺度[即原始语音的音量]。因此，任何两个信号（$b_1 S^{j_1}, \dots, b_d S^{j_d}$）和（$\hat{}b_1 S^{k_1}, \dots, \hat{}b_d S^{k_d}$），只要满足 b_i、$\hat{}b_i \neq 0$ 以及 $\{j_i\} = \{k_i\} = [d]$，都是盲反演同样理想的结果。换言之，例 1.1 中的任务若能从 X 得到（不一定是 S，而是）任意一个形如以下形式的信号，就算是“精确地”完成了

$$\tilde{S} \in \left\{ \left(\alpha_1 S^{\tau(1)}, \dots, \alpha_d S^{\tau(d)}\right) \,\middle|\, \alpha_1, \dots, \alpha_d \in \mathbb{R}_\times,\, \tau \in S_d \right\} = \mathrm{M}_d \cdot S, \tag{B.8}$$

~其中 $\mathrm{M}_d := \{\Lambda P \mid \Lambda \in \Delta_d, P \in \mathrm{P}_d\}$ 是 [的子群，属于 GL_d] 中的单项矩阵。由此可得单项式轨道 $\lceil S \rceil := \mathrm{M}_d \cdot S$，可将其作为例 8中任务特定准确度的合适定义，用于例 1.1。
由于我们对说话者的先验知识 S以及对其录音情境中“物理学”的了解f都非常有限，因此通常只能将原始配置$(S,\ f)$定位为 $\mathscr{I}$ 中的某个子集元素 $\mathfrak{C}$，而非该空间中的一个单点——这就是我们的“盲性”。因此，在这一先验层面，任何符合 $\mathscr{I}$ 的信号（即任何$\hat{}S := g(X)$ 使得$(\hat{}S,\ g^{-1}) \in \mathscr{I}$）在我们看来都与我们所要寻找的原始讲话无法区分。因此，我们将这类信号称为“准源”，并将其归入集合（6）。

现阶段我们主要关心的问题是：关于说话者及其混合方式的有限知识 $\mathscr{I}$ 是否至少在原理上足以恢复信号 S，使其 X 达到所需的精度（B.8）。这意味着必须满足以下要求：$\langle X \rangle_{\mathscr{I}} \subseteq \lceil S \rceil$ on $\mathscr{I}$；若满足这一要求，我们便希望相应地进行逆变换 $X \mapsto S$。后者要求

35 … 也就是说，要从 X 获得的不仅是所说内容，还包括是哪位说话者（由其编号 i标识）说的，以及其原始音量；这通常是不可能的。

求出一个可显式计算的求逆过程 $\Phi:\ \mathcal{M}_1 \to 2^{\mathcal{M}_1}$。由于它应当充分利用数据（$X$，$\mathscr{I}$），我们自然期望它满足 $\Phi(X) \neq \emptyset$ 和 $\Phi(X) \subseteq \langle X \rangle_{\mathscr{I}}$。

现在，由于说话者 $S^1, \dots, S^d$ 之间存在复杂的相互依赖关系，或因为录音情境 f 有些复杂，亦或两者兼有，我们可能无法针对 (S, f) 提出一个假设 $\mathscr{I}$，使准确性条件 (8) 得到精确满足。不过，我们往往仍能发现，(S, f) 在某种程度上可以由一个理想化配置 $(S_\star, f_\star)$「良好近似」，而该配置符合 $\mathfrak{C}$ 中的某个假设 $\mathscr{I}_*$，且满足准确性条件。在我们的例子中，对 (S, f) 一种可行的近似方式是：令 $S_\star$ 具有与 S 相同的分量信号，但它们之间不存在统计依赖；并令 $f_\star$ 为 f 在某点 $x_0 \in D_S$ 处的线性化，符号表示为：$(S_\star, f_\star) \equiv (\mathbb{P}_{S^1} \otimes \cdots \otimes \mathbb{P}_{S^d}, D_f(x_0))$（参见第 2.2 节）。

这种理想化由此带来如下益处。$\mathscr{I}_*$ 的 $\mathsf{d} \cdot \mathsf{e}$ 充分性允许存在一个形式如 (9) 的映射 $\Phi \equiv \Phi_{\mathscr{I}_*}$，使其能够对由 $\mathscr{I}_*$ 引起的可观测量实现精确的盲反演。若该映射的反演性质 $\Phi(f(S)) \subseteq \mathsf{d}S\mathsf{e}$ 在理想化 $\tilde{}$ $\tilde{}$ $(S, f) \equiv (S_\star, f_\star) \in \mathscr{I}_*$ 附近稳定（见 (10)），那么邻近关系 $(S, f) \approx (S_\star, f_\star)$ 就能保证 $\Phi(X)$ 近似为 $\mathsf{d}S\mathsf{e}$ 的子集，正如所期望的那样。更具体地说，给定 (S, f)（原始成因）与 $(S_\star, f_\star)$（该成因的 $\mathscr{I}_*$ 理想化）之间“偏离” Δ 的估计，我们可以给出相应结果，以精确且有意义的误差界和完全显式的常数表明：$\Phi(X)$ 中的每个元素，都与 $\mathsf{d}S\mathsf{e}$ 中的某个元素相等，误差至多为 $\mathsf{O}(\Delta)$ 量级的相对误差。这对我们的示例应用显然大有裨益：尽管 X 的实际成因 (S, f) 可能过于复杂，无法通过盲反演 S 精确恢复，但我们仍可以找到 (S, f) 的一种“结构更简单的近似”，例如 $\mathscr{I}_* \ni (S_\star, f_\star)$，并在如上所述的 $\mathscr{I}_*$ 上进行逆变换 Φ；而且可以发现，估计语音 $\tilde{S}^\Delta \equiv \Phi(X)$ 仍然达到 $\mathsf{d} \cdot \mathsf{e}$ 精度，与 (B.8) 的相对距离也受到控制。只要理想化误差 Δ 足够小，[36]，所得信号 $\tilde{S}^\Delta$ 或许会出现声学失真，但仍会足够接近最优结果 (B.8)（清晰可听的语音），从而能够辨明说话者传达的原始信息。

◆

B.1. **定义的补充 2.7**。为在（13）中使用，定义如下集合（“光滑性类别”）

$$\hat{\mathscr{S}}_\perp := \left\{ \mu \in \mathscr{S}_\perp \;\middle|\; \sharp\{i \in [d] \mid \mathbb{E}\big[(\mu^i_{0,1})^3\big] = 0\} \leq 1 \text{ and } \sharp\sigma(\mathrm{Cov}(\mu_0, \mu_1)) = d \right\} \tag{B.9}$$

所有增量协方差具有两两不同特征值，且其三阶增量矩至多一个为零的 IC 信号。按照常规正则性的含义，这些信号具有常规正则性，具体而言，它们是非高斯的，因此可由经典范式进行（线性）辨识 [27], ，而且具有足够丰富的时间结构，可通过 [70]中的方法恢复。换言之，（B.9）是一类正则（“光滑”）程度足够高的源，盲反演过程应当能够恢复它们；只有这样，才能称其为 ICA 方法。

B.2. **对更粗糙路径的扩展**。尽管样本空间（19）已经相当一般，尤其涵盖了统计实践中会遇到的所有信号（参见注 3.4、命题 3.11），但请注意，利用粗糙路径理论，可以轻松地将本文的讨论扩展到更不规则的路径空间 [34, 65, 66]。由于这主要是一项技术性工作，为避免超出本文的范围，我们不在此实现这些扩展，而是请感兴趣的读者参阅上述文献。

C. Technical Proofs and Lemmas (Sections 2 to 5)

C.1. **引理2.9的证明**。

[36] 可以尝试通过调整麦克风位置或衰减外部噪声来实现。42

证明。由（14）可知，$\partial_{\hat{\Phi}}(\mathscr{I}) \equiv 0$ 显然成立。令 $\varphi(\boldsymbol{S}, f)$ 表示不等式（16）左侧，并注意这里稍微滥用了符号，参见约定 1.2：此处，$\varphi(\boldsymbol{S}, f)$ 是过程 $\boldsymbol{S}$ 在 $\mathbb{R}^d$ 中的函数，而 $\partial_{\hat{\Phi}}(S, f)$ 是其分布 $S \equiv \mathbb{P}_S \in \mathsf{M}_1$, 的函数。由（11）得，

$$\varphi(\boldsymbol{S}, f) = \sup_{\tilde{\boldsymbol{X}} \in \Phi(\boldsymbol{X})} \inf_{h \in \tilde{\mathsf{M}}_d} \tilde{\varrho}(\tilde{\boldsymbol{X}}, h(\boldsymbol{S})) \quad \text{for} \quad \boldsymbol{X} = f(\boldsymbol{S}), \tag{C.10}$$

其中，根据定义，$\Phi(\boldsymbol{X}) = (\hat{\Phi}(X) \circ f)(\boldsymbol{S}) \equiv \{(B \circ f)(\boldsymbol{S}) \mid B \in \hat{\Phi}(X)\}$。因此，对于任意 $\tilde{\boldsymbol{X}} \in \Phi(\boldsymbol{X})$, 以及所有 $h \equiv M + v \in \tilde{\mathsf{M}}_d$ (某个 $M \in \mathsf{M}_d, v \in \mathbb{R}^d$)，存在 $B \equiv B(\tilde{\boldsymbol{X}}) \in \hat{\Phi}(X)$，使得

$$\tilde{\varrho}(\tilde{\boldsymbol{X}}, h(\boldsymbol{S})) = \sup_{t \in \mathbb{I}} \tfrac{|(B \circ f)(\boldsymbol{S}_t) - (M\boldsymbol{S}_t + v)|}{|M\boldsymbol{S}_t + v|} \mathbb{1}_{\times(h(\boldsymbol{S}_t))} = \sup_{t \in \mathbb{I}} \tfrac{|(B \circ f)(\tilde{\boldsymbol{S}}_t - \tilde{v}) - M\tilde{\boldsymbol{S}}_t|}{|M\tilde{\boldsymbol{S}}_t|} \mathbb{1}_{\times(M\tilde{\boldsymbol{S}}_t)}$$

其中 $\tilde{\boldsymbol{S}} = (\tilde{\boldsymbol{S}}_t)$ 由 $\tilde{\boldsymbol{S}}_t := \boldsymbol{S}_t + \tilde{v}$ 给出，且 $\tilde{v} := M^{-1}v$。因此，由于一个过程的轨迹几乎必然包含于其空间支撑中，我们有，以概率 1，

$$\tilde{\varrho}(\tilde{\boldsymbol{X}}, h(\boldsymbol{S})) \leq \sup_{u \in D_{\tilde{S}}} \tfrac{|(B \circ f)(u - \tilde{v}) - Mu|}{|Mu|} \mathbb{1}_{\times(Mu)} = \varrho_B((S, f), (M, \tilde{v})) \tag{C.11}$$

其中 $D_{\tilde{S}} (= D_{\tilde{\boldsymbol{S}}}) = D_S^{(\tilde{v})}$。现在由 (16)、(C.11) 以及定义 (C.10) 和 (15)，可得结论 (16)。 □

C.2. **引理3.3的证明**。

证明。首先注意到，由于 $\|\cdot\|_{1\text{-var}} \geq \mathsf{k} \cdot \mathsf{k}_\infty$（这一点很容易看出），我们发现，$\mathsf{C}^1$ 上的 1-变差拓扑比 C^1, 上的一致拓扑更精细，这当然意味着 $\mathsf{B}_{1\text{-var}} := \sigma(\mathsf{C}^1, \|\cdot\|_{1\text{-var}}) \supseteq \sigma(\mathsf{C}^1, \mathsf{k} \cdot \mathsf{k}_\infty) =: \mathsf{B}_\infty$。由于 $(\mathsf{C}^1, \|\cdot\|_{1\text{-var}})$ 的可分性保证 σ- 代数 $\mathsf{B}_{1\text{-var}}$ 由闭的 $\|\cdot\|_{1\text{-var}}$- 球生成，因此只要证明下式成立，反向包含关系 $\mathsf{B}_{1\text{-var}} \subseteq \mathsf{B}_\infty$ 便可得出：

$$B_r^1(x) := \{y \in \mathcal{C}^1 \mid \|y - x\|_{1\text{-var}} \leq r\} \in \mathcal{B}_\infty \quad \text{for every} \quad x \in \mathcal{C}^1 \text{ and any } r \geq 0. \tag{C.12}$$

为证明这一点，任取 $x \in \mathcal{C}^1$ 以及 $r \geq 0$，并回顾定义（参见（23））：

$$\|z\|_{1\text{-var}} = \sup_{\mathcal{I} \in \mathfrak{I}} V_{\mathcal{I}}(z) \quad \text{with} \quad V_{(t_\nu)}(z) := |z_0| + \sum_\nu \left|z_{t_{\nu+1}} - z_{t_\nu}\right|$$

且其中 $\mathfrak{I} := \{\mathcal{I} = (t_\nu) \mid \mathcal{I} \text{ 是 } [0,1]\}$ 的一个（有限）剖分。任取 $\mathcal{I} \in \mathfrak{I}$，显然函数 $Q_{\mathcal{I}} : \mathcal{C}^1 \ni y \mapsto V_{\mathcal{I}}(y - x)$ 关于 $\|\cdot\|_\infty$ 连续，因此水平集 $C_{\mathcal{I}} := \{y \in \mathcal{C}^1 \mid Q_{\mathcal{I}}(y) \leq r\}$ 是 $\|\cdot\|_\infty$-闭的。结合这一点，立即可得恒等式

$$B_r^1(x) = \bigcap\nolimits_{\mathcal{I} \in \mathfrak{I}} C_{\mathcal{I}} \quad \text{implies that} \quad B_r^1(x) \text{ is closed wrt. } \|\cdot\|_\infty,$$

这表明（C.12）如所需成立。 □

C.3. **引理3.7的证明**。

证明。任取一个多重指标 $(i_1, \ldots\ldots, i_m) \in [d]^{\times m}$。于是，根据迭代（斯蒂尔切斯）积分的多重线性，对每个 $x \in$ 有界变差，我们得到

$$\begin{aligned}
\mathfrak{sig}_{i_1 \cdots i_m}(f(x)) &= \int_{\Delta_m} \mathrm{d}\big[f(x)\big]^{i_1} \cdots \mathrm{d}\big[f(x)\big]^{i_m} \\
&= \int_{\Delta_m} \mathrm{d}\big[\textstyle\sum_{j_1=1}^d a_{i_1 j_1} x^{j_1}\big] \cdots \mathrm{d}\big[\sum_{j_m=1}^d a_{i_m j_m} x^{j_m}\big] \\
&= \sum_{j_1, \ldots, j_m = 1}^d a_{i_1 j_1} \cdots a_{i_m j_m} \mathfrak{sig}_{j_1 \cdots j_m}(x)
\end{aligned}$$

对于 m单形 $\Delta_m := \{t \in [0,1]^m \mid t_1 \le \cdots \le t_m\}$。因此，根据定义（28）和（A.7），

$$\begin{aligned}\langle f \cdot \mu \rangle_{i_1 \dots i_m} &= \int_{\mathcal{C}_d} \mathfrak{sig}_{i_1 \cdots i_m}(x)\, f_\star \mu(\mathrm{d}x) = \int_{\mathcal{BV}} \mathfrak{sig}_{i_1 \cdots i_m}(f(x))\, \mu(\mathrm{d}x) \\ &= \sum_{j_1,\dots,j_m=1}^{d} a_{i_1 j_1} \cdots a_{i_m j_m} \cdot \langle \mu \rangle_{j_1 \dots j_m},\end{aligned}$$

其中第二个等式是推前测度的变量替换公式 s. □

C.4. **命题 3.11 的证明。**

证明。首先注意，子空间 $\mathfrak{D}_K^p$ 和 $\widehat{\mathfrak{D}}_{\mathcal{I}}$ 定义良好，因为 $\mathsf{C}^1_{p,K}$ 和 $\hat{}\mathsf{C}_{\mathcal{I}}$ 关于 $\mathsf{k}\cdot\mathsf{k}_\infty$均为博雷尔可测。事实上：由 (24) 可知，$\mathsf{C}^1_{p,K}$ 是 $\|\cdot\|_{1\text{-var}}$-闭的 [，因为它是一个次水平集]（当然，即使去掉 b，这些包含关系仍然成立）；根据引理 3.3，$\mathsf{B}(\mathsf{C}^1; \|\cdot\|_{1\text{-var}}) = \mathsf{B}(\mathsf{C}^1; \mathsf{k}\cdot\mathsf{k}_\infty)$；而 C^1-子空间 $\hat{}\mathsf{C}_{\mathcal{I}}$ 关于 $\mathsf{k}\cdot\mathsf{k}_\infty$是闭的，因为它是有限维的。

接下来利用 (35) 可推广为 $\|\cdot\|_\infty \le \|\cdot\|_{q\text{-var}} \le 2\|\cdot\|_\infty^{\frac{p}{q}}\, {}_{p\text{-var}}\mathsf{k}\cdot\mathsf{k}^{1-\frac{p}{q}}$ 对任意有限的 $q > p$，参见 [34, 命题 5.5 (i)]，，可得范数 $\mathsf{k}\cdot\mathsf{k}_\infty$ 和 $\mathsf{k}\cdot\mathsf{k}_{q\text{-var}}$[在 $\mathsf{C}^1_{p,K}$ 上具有相同的收敛序列，因此]它们在 $\mathsf{C}^1_{p,K}$上等价。特别地$C_b(\mathsf{C}^1_{p,K}; \mathsf{k}\cdot\mathsf{k}_\infty) = C_b(\mathsf{C}^1_{p,K}; \mathsf{k}\cdot\mathsf{k}_{q\text{-var}})$，于是对于任意网 $(\mu_\alpha) \subset \mathfrak{D}_K^p \cong \mathsf{M}_1(\mathsf{C}^1_{p,K})$，有：$\mu_\alpha \to \mu \in \mathfrak{D}_K^p$ 弱收敛当且仅当 $\mu_\alpha \overset{q\text{-变差}}{\longrightarrow} \mu$（参见 (33) 和 (34)，其中以 $\mathsf{C}^1_{p,K}$替换 C^1）。引理的断言现在可由以下事实推出：两个拓扑相同当且仅当它们具有相同的收敛网；并注意到 $C_b(\hat{}\mathsf{C}_{\mathcal{I}}; \mathsf{k}\cdot\mathsf{k}_\infty) = C_b(\hat{}\mathsf{C}_{\mathcal{I}}; \mathsf{k}\cdot\mathsf{k}_{1\text{-var}})$，这是因为 $\mathsf{k}\cdot\mathsf{k}_\infty$ 和 $\mathsf{k}\cdot\mathsf{k}_{1\text{-var}}$ 在 $\hat{}\mathsf{C}_{\mathcal{I}}$上等价。已知 $\mathsf{k}\cdot\mathsf{k}_{1\text{-var}} \ge \mathsf{k}\cdot\mathsf{k}_\infty$ 后，后者可由 $\mathsf{k}x\mathsf{k}_{1\text{-var}} = |x_0| + \sum_{\nu=1}^{n-1} |x_{t_\nu} - x_{t_{\nu-1}}| \le (2|\mathrm{I}| - 1)\mathsf{k}x\mathsf{k}_\infty$ 推出（任意 $x \in \hat{}\mathsf{C}_{\mathcal{I}}$）。□

C.5. **命题 3.14 的证明。**

证明。设 $\mathfrak{U} \subseteq \mathcal{M}_1(\mathcal{C}^1)$ 对某个 $q > 1$满足 (39)。那么对任意 $\mu \in \mathfrak{U}$，并令 r 为 q的共轭指数，

$$\int_{\{|\mathfrak{sig}_w| > a\}} |\mathfrak{sig}_w|\, \mathrm{d}\mu \le \frac{c}{m!} \int_{\mathcal{C}_d} \|x\|_{1\text{-var}}^m \cdot \mathbb{1}_{\{|\mathfrak{sig}_w| > a\}}\, \mu(\mathrm{d}x) \tag{C.13}$$

$$\le \frac{c}{m!} \left(\int_{\mathcal{C}_d} \|x\|_{1\text{-var}}^{mq}\, \mu(\mathrm{d}x) \right)^{\frac{1}{q}} \cdot \left(\mu\big(\{|\mathfrak{sig}_w| > a\}\big) \right)^{\frac{1}{r}} \overset{a\to\infty}{\longrightarrow} 0 \tag{C.14}$$

对所有 $w \in [d]^{\le m}$ 均成立，其中第一个不等式源于经典签名界 [65,定理 3.7]（其中给出了常数 c），第二个不等式源于霍尔德不等式，而最后的收敛到零则由切比雪夫不等式推出。事实上，后者蕴含

$$\sup_{\mu\in\mathfrak{U}} \mu\big(\{|\mathfrak{sig}_w| > a\}\big) \le \frac{1}{a} \sup_{\mu\in\mathfrak{U}} \int_{\mathcal{C}_d} |\mathfrak{sig}_w(x)|\, \mu(\mathrm{d}x) \lesssim \frac{1}{a} \sup_{\mu\in\mathfrak{U}} \int_{\mathcal{C}_d} \|x\|_{1\text{-var}}^m\, \mu(\mathrm{d}x) \to 0$$

当 $a \to \infty$时，最后一个不等式再次源于所引述的签名界，而最后一个上确界由假设 (39) 可知有限（回忆 $L^q(\mu) \subseteq L^1(\mu)$，因为测度 μ 均有限）。(C.14) 与上述 $\mathfrak{U}$-一致收敛性共同推出 (38)。□

注 C.1. 满足 $\mathfrak{U}$ 中 $\mathfrak{D}$ 式（39）的集合，恰好是mq上 $(\mathcal{C}^1, \|\cdot\|_{1\text{-var}})$-Wasserstein 空间中的有界子集；关于这些集合的讨论，参见例如 [74, 第 2.1、2 节2]。此外，满足式（39）的信号也称为阶数为 mq 的 Radon 测度，它们是 Laurent Schwartz [84, 85]等人发展起来的 Radon 化映射理论的核心内容。许多文献都描述了这类信号及其一致有界族，例如参见 [11, 35, 64, 74, 88] 及其相关参考文献。对于特殊情形 $\mu \equiv \hat{}\nu \in \mathfrak{D}_1$ 的

实数空间中的随机向量 $\mathbb{R}^d$ （参见注 3.4），式（38）和（40）中的积分可化为经典统计量

$$\frac{1}{|w|!}\int_{\{|u^w|>|w|!a\}} |u^w|\,\nu(\mathrm{d}u) \quad \text{and} \quad \int_{\hat{\mathcal{C}}_0} \|x\|_{1\text{-var}}^{\tilde{q}}\,\hat{\nu}(\mathrm{d}x) = \int_{\mathbb{R}^d} |u|^{\tilde{q}}\,\nu(\mathrm{d}u),$$

分别满足 $u^{i_1\dots i_m} := u^{\sharp\{j|i_j=1\}}{}_1 \cdots u^{\sharp\{j|i_j=d\}}{}_d$ 对于 $u=(u_i)\in\mathbb{R}^d$。因此，(38) 和 (40) 都可以看作多元矩经典可积性条件（从 $\mathbb{R}^d$ 到 $\mathcal{C}_d$）的自然推广，参见例如 [10,第 30]页。

◆

C.6. **关于定义的注4.1**。

注 C.2. 预度量 (44) 不满足不可分辨者同一性（“$\delta(\mu,\tilde{\mu})=0$当且仅当 $\mu=\tilde{\mu}$”），而这正是（拟半伪度量或）定义在 $\mathfrak{D}$上的统计散度所要求的。然而，这种“缺失”的同一性可以恢复（例如，证明见 [23, 第 5 节.4] ）：不只考虑信号的 2nd 阶和 3^{rd}阶矩，而是纳入信号的所有签名矩 (28)，并计算它们在相应的（希尔伯特）坐标空间 $\mathrm{V}^\infty := \bigoplus_{m\geq 1}(\mathbb{R}^d)^{\otimes m}$中的 ℓ_2距离。该空间是上述坐标空间 $\mathrm{V} \cong (\mathbb{R}^d)^{\otimes 2}\oplus(\mathbb{R}^d)^{\otimes 3}$ 的等距扩张，后者包含核心坐标 (29)。尽管还可以构造更高阶及其他类型的扩张（参见第 6.4 节），但以下各节将表明，对于当前经典的 [即线性]盲源分离/独立成分分析情形，通过 (42) 与三阶上限 (29) 相联系的预度量拓扑 τ_δ 已经足够 []。

◆

C.7. **一些技术估计**。

引理 C.3. 设 $p\in(1,2)$和 $\eta>0$任意，并记 $\alpha:=1-1/p$和 $\gamma:=1/p-\alpha$。

(i) 设 $\{Y,\tilde{Y}\}\subset \mathcal{C}_{d|\mathbb{P}}\cap\mathfrak{D}$ 为核心可积，且 $\|Y-\tilde{Y}\|_\infty\leq\eta$几乎必然成立。那么对于每个满足 $2\leq|w|\leq 3$ 的 $w\in[d]^\star$，我们有

$$\left|\langle\tilde{Y}\rangle_w - \langle Y\rangle_w\right| \leq \kappa_{p,|w|}\cdot\eta^\alpha \tag{C.15}$$

其中 $\kappa_{p,k}:=(2+k)c_p(2^{\beta_p}K_p+1)$，常数为 $\beta_p:=2+1/p$ 和 $c_p:=(1-2^{1-2/p})^{-1}$，且 $K_p:=\max_{V\in\{\tilde{Y},Y\}}\mathbb{E}\big[\|V\|_1^{\beta_p\text{-var}}\big]$。

(ii) 对于任意满足 $\max_{|w|=2,3}|\tilde{\mu}_w-\mu_w|\leq\eta$ 的 $\tilde{\mu},\mu\in\mathfrak{D}$，其中某个 $\kappa\geq 0$，我们有

$$\left|\delta_\perp(\tilde{\mu})-\delta_\perp(\mu)\right| \leq \varphi_\kappa(\mu,\tilde{\mu})\sqrt{\eta+\eta^2} \tag{C.16}$$

对于辅助函数 $\varphi_\kappa(\mu,\tilde{\mu}):=\sqrt{(3\vartheta_{\mu,\tilde{\mu}}+\kappa)^\kappa\sum_{k=0}^d\sum_{i,j=1}^d\varphi_{ijk}(\mu,\tilde{\mu})}$ 其中

$$\varphi_{ijk}(\mu,\tilde{\mu}) := \frac{2\big(\langle\mu\rangle_{ijk}\vee\sqrt{\langle\tilde{\mu}\rangle_{ii}\langle\tilde{\mu}\rangle_{jj}\langle\mu\rangle_{kk}}\big)^2}{\big(\sqrt{\langle\mu\rangle_{ii}\langle\mu\rangle_{jj}\langle\tilde{\mu}\rangle_{kk}}\wedge\sqrt{\langle\tilde{\mu}\rangle_{ii}\langle\tilde{\mu}\rangle_{jj}\langle\mu\rangle_{kk}}\big)^4} \tag{C.17}$$

以及 $\vartheta_{\mu,\tilde{\mu}}:=$最大值$\{|\mu_{ii}-\tilde{\mu}_{ii}|,|\mu_{ii}-\tilde{\mu}_{jj}|\mid i,j\in[d]\}$。

(iii) 对于任意 $\mu,\tilde{\mu}\in\mathfrak{D}$，若 $\delta(\mu,\tilde{\mu})\leq\eta$，则有

$$\max_{|w|=2,3}\left|\langle\tilde{\mu}\rangle_w-\langle\mu\rangle_w\right| \leq \mathfrak{m}_\mu\eta \tag{C.18}$$

对于 $\mathfrak{m}_\mu:=\max\{\sqrt{|\mu_{ii}||\mu_{ij}||\mu_{jj}||\mu_{ik}|}\mid i,j\in[d],k\in[d]_0\}$；若进一步满足 μ 正交，则 $\delta_\perp(\tilde{\mu})\leq\hat{\varphi}(\mu,\tilde{\mu})\sqrt{\eta+\eta^2}$适用于 $\hat{\varphi}(\mu,\tilde{\mu})^2:=\max\{3\vartheta_{\mu,\tilde{\mu}}\mathfrak{m}_\mu,\mathfrak{m}_\mu^2\}\sum_{k=0}^d\sum_{i,j=1}^d\varphi_{ijk}(\mu,\tilde{\mu})$。

(iv) 令 $R \in C^1(\mathbb{R}^d;\mathbb{R}^d)$，其中[37] $\|D_R\|_\infty < \infty$，并令 $\mu \in \mathfrak{D}$ 为核心可积的，且满足 $\tilde{\mu} := (\mathrm{I} + R)_*\mu$。那么，对于每个满足 $|w| = 2$ 的 $w \in [d]^\star$，3：

$$\big|\langle\tilde{\mu}\rangle_w - \langle\mu\rangle_w\big| \le \tilde{K}_{p,|w|} \cdot \phi_{|w|}(\|D_R\|_\infty)\big(\|D_R\|_\infty\|R\|_\infty\big)^\alpha, \quad \textit{and} \tag{C.19}$$

$$\big|\delta_\perp(\tilde{\mu}) - \delta_\perp(\mu)\big| \le \varphi_{\tilde{K}_{p,3}}(\mu,\tilde{\mu})\,\hat{\phi}_p(R)\cdot\phi_3(\|D_R\|_\infty)^{\frac{1}{2}}\big(\|D_R\|_\infty\|R\|_\infty\big)^{\frac{\alpha}{2}} \tag{C.20}$$

对于 $\tilde{K}_{p,k} := 2^\alpha c_p(1 + k/2)(1 + \mathbb{E}_\mu[\|x\|_1^\beta$ -方差$])$ 和 $\phi_k(u) := (1 + u)^{k-1}u^\gamma$，且 $\varphi_{\tilde{K}_p,3}(\mu,\tilde{\mu})$ 如 *(ii)* 中所述，但将 κ 改为 $= \tilde{K}_p$:3，以及 $\hat{}\,\phi_p(R) := \sqrt{1 + \phi_3(\|D_R\|_\infty)(\|D_R\|_\infty\|R\|_\infty)^\alpha}$。

˜ ˜证明。(i)：设 $\{Y,\ Y\} \subset \mathfrak{D}$ 为平方可积，且几乎处处满足 $\|Y - Y\|_\infty \le \eta$，并设 $w \in [d]^{\le m}$ 为长度为 $|w| \le m$ 的任意多重指标。取任意 $p \in (1,\ 2)$。经典的˜插值不等式 [34, 命题 5.5 (i)] 于是推出，对于 $Z := Y - Y$ 有

$$\|Z\|_{p\text{-var}} \le 2^\alpha\|Z\|_{1\text{-var}}^{1/p}\|Z\|_\infty^{(1-1/p)} \le 2^\alpha\|Z\|_{1\text{-var}}^{1/p}\eta^\alpha \le 2\Xi^{1/p}\cdot\eta^\alpha \tag{C.21}$$

其中 $\alpha := 1 - 1/p$ 且 $\Xi := \max\big\{\|Y\|_{1\text{-var}},\ \|\tilde{Y}\|_{1\text{-var}}\big\}$。签名变换的局部利普希茨连续性（参见 [34, 命题 7.63]）意味着，对于每个 $2 \le |w| \le m$，

$$\big|\langle\tilde{Y}\rangle_w - \langle Y\rangle_w\big| \le \mathbb{E}\big|\mathfrak{sig}_w(\tilde{Y}) - \mathfrak{sig}_w(Y)\big| \le (2 + |w|)c_p\,\mathbb{E}\big[\Xi^{|w|-1+1/p}\big]\cdot\eta^\alpha \tag{C.22}$$

对于 $c_p := (1-2^{1-2/p})^{-1}$，这可由 [34, 命题 7.63] 的证明推出，只需改用杨-勒夫估计 $|\tilde{Y}_{s,\ r} - Y_{s,\ r}| = |Z_{s,\ r}| \le \|Z\|_{p\text{-var};[s,t]}$ 和 $\big|\int_s^t \tilde{Y}_{s,\ r}\,\mathrm{d}Z_r\big| \le c_p\|\tilde{Y}\|_{p\text{-var};[s,t]}\|Z\|_{p\text{-var};[s,t]}$。事实上：对于每个 $|w| \ge 2$，我们有

$$\begin{aligned}\big|\mathfrak{sig}_w(\tilde{Y}) - \mathfrak{sig}_w(Y)\big| &= \big|\langle\pi_{|w|}\big[\mathfrak{sig}_{0,1}(\tilde{Y}) - \mathfrak{sig}_{0,1}(Y)\big], w\rangle\big| \le \big\|\pi_{|w|}\big[\mathfrak{sig}_{0,1}(\tilde{Y}) - \mathfrak{sig}_{0,1}(Y)\big]\big\| \\ &= \left\|\int_0^1 \pi_{|w|-1}\big[\mathfrak{sig}_{0,r}(\tilde{Y}) - \mathfrak{sig}_{0,r}(Y)\big]\mathrm{d}\tilde{Y}_r + \int_0^1 \pi_{|w|-1}\big[\mathfrak{sig}_{0,r}(Y)\big]\mathrm{d}Z_r\right\|\end{aligned} \tag{C.23}$$

其中 $\|\cdot\| := |\cdot|_1$ 是 $\mathbb{R}^d$的张量幂上的射影张量范数，参见 [66, 第 5.6]节。利用对任意 $x \in \mathcal{C}^1([0,1];\mathbb{R}^d)$ 和 $y \in \mathcal{C}^1([0,1];\mathbb{R}^{e\times d})$，我们有

$$\int_s^{\cdot} y_{s,r}\,\mathrm{d}x_r \in \mathcal{C}^1([s,t];\mathbb{R}^e) \quad \text{and} \quad \left|\int_s^t y_{s,r}\,\mathrm{d}x_r\right| \le c_p\|y\|_{p\text{-var};[s,t]}\|x\|_{p\text{-var};[s,t]}$$

对每个 $0 \le s \le t \le 1$，例如参见 [65, 第 1.3] 节和 [34, 命题 6.4]，，由 (C.23) 可得$|\mathfrak{sig}_w(\tilde{Y})$ $- \mathfrak{sig}_w(Y)| \le c_p(\|Y\|_{p\text{-var}} + \|\tilde{Y}\|_{p\text{-var}})\|Z\|_{p\text{-var}} \le 2c_p\Xi\|Z\|_{p\text{-var}}$ 若 $|w| = 2$，以及$|\mathfrak{sig}_w(\tilde{Y})$ $- \mathfrak{sig}_w(Y)| \le \big(2c_p\Xi\|Y\|_{1\text{-var}} + \frac{c_p}{2!}\|\tilde{Y}\|_{1\text{-var}}^2\big)\|Z\|_{p\text{-var}} \le \frac{5}{2}c_p\Xi^2\|Z\|_{p\text{-var}}$ 若 $|w| = 3$，其中，在倒数第二个不等式中我们使用了标准估计 $\|\pi_2[\mathfrak{sig}_{s,t}(Y)]\|_{1\text{-var}} \le 1\ 2!\|\tilde{Y}\|_{1\text{-var}}^2$（例如参见 [65, 命题 2.2]）。结合 (C.21)，我们得到对 $|w| = 2,\ 3$，

$$\big|\langle\tilde{Y}\rangle_w - \langle Y\rangle_w\big| \le (1 + |w|/2)c_p\mathbb{E}\Big[\Xi^{|w|-1}\|Z\|_{p\text{-var}}\Big] \le (2 + |w|)c_p\mathbb{E}\Big[\Xi^{|w|-1+1/p}\Big]\cdot\eta^\alpha,$$

如（C.22）所述。最后，为了得到（C.15），注意到 $\Xi^{|w|-1+1/p} \le 1 + \Xi^{2+1/p} \le 1 + \big(\|\tilde{Y}\|_{1\text{-var}} + \|Y\|_{1\text{-var}}\big)^\beta$ 对每个 $|w| = 2,\ 3$ 以及 $\beta := 2 + 1/p$ 因此，由闵可夫斯基不等式可得，

$$\mathbb{E}\Big[\Xi^{|w|-1+1/p}\Big] - 1 \le \Big[\mathbb{E}\big[(\tilde{U} + U)^\beta\big]^{1/\beta}\Big]^\beta \le \Big[\mathbb{E}\big[\tilde{U}^\beta\big]^{1/\beta} + \mathbb{E}\big[U^\beta\big]^{1/\beta}\Big]^\beta \le 2^\beta K$$

˜ ˜对于 $U := \|Y\|_{1\text{-var}}$ 以及 $U := \|Y\|_{1\text{-var}}$，如所述。

(ii)：令 $\mu,\ \tilde{\mu} \in \mathfrak{D}$ 并假设（如所要求的）存在 $\kappa \ge 0$ 和 $\eta > 0$ 使得

$$\max_{|w|=2,3}\big|\langle\tilde{\mu}\rangle_w - \langle\mu\rangle_w\big| \le \kappa\eta. \tag{C.24}$$

[37] 这里以及下文中，我们将映射 D_R 的雅可比矩阵 $R \in C^1(\mathbb{R}^d;\ \mathbb{R}^d)$ 按如下方式定义范数：$\|D_R\|_\infty := \sup_{x\in\mathbb{R}^d}\|D_R(x)\|_2$ 其中 $\|\cdot\|_2$ 是 $\mathbb{R}^{d\times d}$上的欧几里得算子范数。

由 (C.24) 以及定义 (42) 和 (43) 可得界 (C.16)，如下所示。为方便起见，我们采用 δ 的表示法 (44)。按照直和的惯例，我们将任意 $C \equiv (C_0, \cdots C_d) \in \mathbb{V}$（粗体“+”）记作 $C \equiv C_0 + \ldots + C_d$，并相应地对任意 $N_1, N_2 \in \mathbb{R}^{d\times d}$ 设定 $N_1 \cdot C \cdot N_2 \equiv \sum_{\nu=0}^{d} N_1 C_\nu N_2 \equiv (N_1 C_\nu N_2)_{\nu=0}^{d} \in \mathbb{V}$。于是

$$\delta(\mu,\tilde{\mu}) \;=\; \Big\| N_\mu^{-1}\big(C_0 + \textstyle\sum_{k=1}^{d} C_k\big) N_\mu^{-1} \Big\|_{\mathcal{V}} \qquad \text{with} \qquad C_\nu = \left(\frac{\langle\tilde{\mu}\rangle_{ij\nu} - \langle\mu\rangle_{ij\nu}}{\sqrt{\langle\mu\rangle_{\nu\nu}}} \right)_{ij}$$

其中 $\langle\mu\rangle_{ij0} := \langle\mu\rangle_{ij}$，$\langle\mu\rangle_{00} := 1$ 以及 N_μ 如（44）中所示；我们进一步令 $C_{\mu\tilde{\mu}} := C_0 + \sum_{k=1}^{d} C_k$。假设 $\{\mu,\ \tilde{\mu}\} \subset \mathfrak{D}$ 具有核心可积性，便可确保所有这些量均存在。

界（C.16）基于简单的三角不等式

$$\begin{aligned} \big|\delta_{\perp}(\tilde{\mu}) - \delta_{\perp}(\mu)\big| \;&=\; \Big| \big\| N_{\tilde{\mu}}^{-1} C_{\tilde{\mu}_\star\tilde{\mu}} N_{\tilde{\mu}}^{-1} \big\|_{\mathcal{V}} - \big\| N_\mu^{-1} C_{\mu_\star\mu} N_\mu^{-1} \big\|_{\mathcal{V}} \Big| \\ &\le\; \Big\| N_{\tilde{\mu}}^{-1} \Big[C_{\tilde{\mu}_\star\tilde{\mu}} R - R^{-1} C_{\mu_\star\mu} \Big] N_\mu^{-1} \Big\|_{\mathcal{V}} \qquad \text{for} \quad R := N_{\tilde{\mu}}^{-1} N_\mu, \end{aligned} \tag{C.25}$$

其中，对于信号 $\mu = (\mu^i) \in \mathfrak{D}$，我们将其对应的乘积信号记作 $\mu_\star := \mu^1 \otimes \cdots \otimes \mu^d$。根据定义，矩阵 $R \equiv (R_{ij})$ 是对角的，且具有 $R_{ii} = \sqrt{\mathfrak{h}\mu\mathfrak{i}_{ii}/\mathfrak{h}\tilde{\mu}\mathfrak{i}_{ii}}$。此外，若 $\tilde{\mu}$、μ 均为均值平稳信号，则它们各自的乘积信号 $\tilde{\mu}_\star$、$\mu_\star$ 也均为均值平稳信号，因此引理 4.8 得出，（展平后的）张量 $C_{\mu_\star\mu} =: (C_{ijk})$ 和 $C_{\tilde{\mu}_\star\tilde{\mu}} =: (C_{ijk})$ 为

$$C_{ij0} = (1-\delta_{ij})\langle\mu\rangle_{ij} \qquad \text{and} \qquad C_{ijk} = \frac{(1-\delta_{ijk})\langle\mu\rangle_{ijk}}{\sqrt{\langle\mu\rangle_{kk}}} \qquad (i,j,k \in [d]), \tag{C.26}$$

并且对 (C_{ijk}) 也同样成立。若 $\tilde{\mu}$ 或 μ 不是均值平稳的，则上述所有结论仍按原样成立，但 (C.26) 根据 $\delta_{\perp}(\cdot)$ 的定义成立，参见注 4.5。沿其 [维度展平的 3rd，即嵌入 $\mathcal{V}$] 张量 $M \equiv (m_{ijk}) := N_{\tilde{\mu}}^{-1}(C_{\tilde{\mu}_\star\tilde{\mu}} R - R^{-1} C_{\mu_\star\mu}) N_\mu^{-1}$ 于是可写为

$$m_{ij0} = \frac{C_{ij0}R_{jj} - \tilde{C}_{ij0}R_{ii}^{-1}}{\sqrt{\langle\tilde{\mu}\rangle_{ii}\langle\mu\rangle_{jj}}} = \left[\frac{\langle\mu\rangle_{ij}}{\sqrt{\langle\tilde{\mu}\rangle_{ii}\langle\tilde{\mu}\rangle_{jj}}} - \frac{\langle\tilde{\mu}\rangle_{ij}}{\sqrt{\langle\mu\rangle_{ii}\langle\mu\rangle_{jj}}} \right](1-\delta_{ij}), \qquad \text{and}$$

$$m_{ijk} = \frac{C_{ijk}R_{jj} - \tilde{C}_{ijk}R_{ii}^{-1}}{\sqrt{\langle\tilde{\mu}\rangle_{ii}\langle\mu\rangle_{jj}}} = \left[\frac{\langle\mu\rangle_{ijk}}{\sqrt{\langle\tilde{\mu}\rangle_{ii}\langle\tilde{\mu}\rangle_{jj}\langle\mu\rangle_{kk}}} - \frac{\langle\tilde{\mu}\rangle_{ijk}}{\sqrt{\langle\mu\rangle_{ii}\langle\mu\rangle_{jj}\langle\tilde{\mu}\rangle_{kk}}} \right](1-\delta_{ijk})$$

记 $\mathfrak{h}\mu\mathfrak{i}_{00} := 1$ 和 $\mathfrak{h}\mu\mathfrak{i}_{ij0} := \mathfrak{h}\mu\mathfrak{i}_{ij}$，以及 $\beta_{ijk} := \sqrt{\mathfrak{h}\mu\mathfrak{i}_{ii}\mathfrak{h}\mu\mathfrak{i}_{jj}\mathfrak{h}\tilde{\mu}\mathfrak{i}_{kk}}$ 和 $\tilde{\beta}_{ijk} := \sqrt{\mathfrak{h}\tilde{\mu}\mathfrak{i}_{ii}\mathfrak{h}\tilde{\mu}\mathfrak{i}_{jj}\mathfrak{h}\mu\mathfrak{i}_{kk}}$ 以及 $\gamma_{ijk} := \min\{\beta_{ijk}, \tilde{\beta}_{ijk}\}$ 和 $\varsigma_{ijk} := \max\{|\mathfrak{h}\mu\mathfrak{i}_{ijk}|, |\tilde{\beta}_{ijk}|\}$，可得

$$|m_{ijk}| = \left| \frac{\langle\mu\rangle_{ijk}\beta_{ijk} - \langle\tilde{\mu}\rangle_{ijk}\tilde{\beta}_{ijk}}{\beta_{ijk}\tilde{\beta}_{ijk}} \right| \;\le\; \frac{\varsigma_{ijk}}{\gamma_{ijk}^2}\Big(|\beta_{ijk} - \tilde{\beta}_{ijk}| + |\langle\mu\rangle_{ijk} - \langle\tilde{\mu}\rangle_{ijk}| \Big). \tag{C.27}$$

Writ对于第一个上述有界求和项，我们得到 $\nu_j := \langle\mu\rangle_{jj}$ 以及 $\tilde{\nu}_j := \langle\tilde{\mu}\rangle_{jj}$，[8]

$$\begin{aligned} \big|\beta_{ijk} - \tilde{\beta}_{ijk}\big|^2 \;&\le\; \big|\nu_i\nu_j\nu_k - \tilde{\nu}_i\tilde{\nu}_j\nu_k\big| \;\le\; \nu_i\nu_j\big|\tilde{\nu}_k - \nu_k\big| + \nu_k\big|\nu_i\nu_j - \tilde{\nu}_i\tilde{\nu}_j\big| \\ &\le\; \nu_i\nu_j\big|\tilde{\nu}_k - \nu_k\big| + \nu_k\big(\nu_i|\nu_j - \tilde{\nu}_j| + \tilde{\nu}_j|\nu_i - \tilde{\nu}_i|\big). \end{aligned}$$

因此，记 $\vartheta :=$ 最大值$\{\nu_i\nu_j,\ \nu_i\tilde{\nu}_j \mid i,\ j \in [d]\}$ 并根据（C.24），我们得到

$$\big|\beta_{ijk} - \tilde{\beta}_{ijk}\big| \;\le\; \sqrt{3\vartheta\kappa}\cdot\eta^{\frac{1}{2}}. \tag{C.28}$$

定义 $\xi_{ijk} := \sqrt{2\varsigma_{ijk}/\gamma_{ijk}^2}$，现在我们可以将 (C.28) 和 (C.24) 与 (C.27) 结合，得到

$$m_{ijk}^2 \;\le\; \xi_{ijk}^2\big(|\beta_{ijk} - \tilde{\beta}_{ijk}|^2 + |\langle\mu\rangle_{ijk} - \langle\tilde{\mu}\rangle_{ijk}|^2\big) \;\le\; K_1\xi_{ijk}^2\cdot(\eta + \eta^2)$$

[38] 回想一下，对于任意 $a,\ b \ge 0$。

对于 $K_1 := (3\vartheta + \kappa)\kappa_\circ$。利用 (C.25)，由此可得

$$\big|\delta_{\perp\!\!\!\perp}(\tilde{\mu}) - \delta_{\perp\!\!\!\perp}(\mu)\big|^2 \;\le\; \sum_{k=0}^{d}\sum_{i,j=1}^{d} m_{ijk}^2 \;\le\; K_2 \cdot (\eta + \eta^2)$$

对于 $K_2 := K_1 \sum_{k=0}^{d}\sum_{i,\,j=1}^{d} \xi_{ijk}^2$，由此得到所要求的（C.16）。(ii) 给定 $\delta(\mu,\ \tilde{\mu}) \le \eta$，根据定义 4.1 可知，对于每个 $w \equiv ijk \in [d]^\star_{|\cdot|=2,\ 3}$，我们有

$$\big|\langle\tilde{\mu}\rangle_w - \langle\mu\rangle_w\big| \le \sqrt{\langle\mu\rangle_{ii}\langle\mu\rangle_{jj}\langle\mu\rangle_{kk}}\cdot\eta \;\le\; \mathfrak{m}_\mu\eta \tag{C.29}$$

对于 $\mathfrak{m}_\mu = \max\{\sqrt{\langle\mu\rangle_{ii}\langle\mu\rangle_{jj}\langle\mu\rangle_{kk}} \mid i,\ \ j \in [d],\ \ k \in [d]_0\}$。因此，按照第 (ii) 点中的相同计算（将假设（C.24）替换为（C.29））， 可得

$$\big|\delta_{\perp\!\!\!\perp}(\tilde{\mu}) - \delta_{\perp\!\!\!\perp}(\mu)\big|^2 \;\le\; \max\{3\vartheta_{\mu,\tilde{\mu}}\mathfrak{m}_\mu, \mathfrak{m}_\mu^2\}\sum_{k=0}^{d}\sum_{i,j=1}^{d}\xi_{ijk}^2\cdot(\eta+\eta^2) \tag{C.30}$$

其中 $\vartheta_{\mu,\ \tilde{\mu}}$ 以及 ξ_{ijk} 如前所定义。若 $\delta_{\perp\!\!\!\perp}(\mu) = 0$ （参见引理 4.8），则结论成立。(iv)：令 $w \in [d]^\star$ 为任意多重指标。则根据变量替换公式，

$$\begin{aligned}\big|\langle\tilde{\mu}\rangle_w - \langle\mu\rangle_w\big| \;&=\; \left|\int_{\mathcal{C}_d} \mathfrak{sig}_w\big(x + R(x)\big)\,\mu(\mathrm{d}x) - \int_{\mathcal{C}_d}\mathfrak{sig}_w(x)\,\mu(\mathrm{d}x)\right| \\ &\le \int_{\mathcal{C}_d}\big|\mathfrak{sig}_w\big(x+R(x)\big) - \mathfrak{sig}_w(x)\big|\,\mu(\mathrm{d}x).\end{aligned} \tag{C.31}$$

Now 根据 $\|\cdot\|_{1\text{-var}}$ 的定义，我们得到（上确界取遍 $[0,1]$ 的所有剖分 (t_ν)）

$$\|R(x)\|_{1\text{-var}} = \sup_{(t_\nu)}\sum_{\nu}\big|R(x_{t_{\nu+1}}) - R(x_{t_\nu})\big| \;\le\; \|D_R\|_\infty\|x\|_{1\text{-var}} \tag{C.32}$$

因此——再次根据 [34, 命题 5.5 (i)]，正如 (C.21) 中一样——

$$\|R(x)\|_{p\text{-var}} \;\le\; 2^\alpha\|D_R\|_\infty^{1/p}\|x\|_{1\text{-var}}^{1/p}\|R\|_\infty^\alpha \;=\; 2^\alpha\|x\|_{1\text{-var}}^{1/p}\|D_R\|_\infty^\gamma\cdot\big(\|D_R\|_\infty\|R\|_\infty\big)^\alpha. \tag{C.33}$$

对于 $\gamma := 1/p - \alpha > 0$。根据第 (i) 点的估计——将 $Y(\omega)$ 和 $Y(\omega)$ 分别替换为$x + R(x)$ 和 x——结合 (C.32)，对于 $|w| = 2,\ 3$ 可得

$$\big|\mathfrak{sig}_w\big(x+R(x)\big) - \mathfrak{sig}_w(x)\big| \;\le\; (1+|w|/2)c_p(1+\|D_R\|_\infty)^{|w|-1}\|x\|_{1\text{-var}}^{|w|-1}\|R(x)\|_{p\text{-var}}. \tag{C.34}$$

结合 (C.31) 和 (C.33)，上述结果合并后得到

$$\big|\langle\tilde{\mu}\rangle_w - \langle\mu\rangle_w\big| \;\le\; \tilde{c}_{p,|w|}\mathbb{E}_\mu\big[\|x\|_{1\text{-var}}^{|w|-1+1/p}\big]\cdot\phi_{|w|}(\|D_R\|_\infty)\big(\|D_R\|_\infty\|R\|_\infty\big)^\alpha$$

对于 $\tilde{c}_{p,\ |w|} := 2^\alpha c_p(1+|w|/2)$ 和 $\phi_k(u) := (1+u)^{k-1}u^\gamma$。由于 $\mathbb{E}_\mu\big[\|x\|_{1\text{-var}}^{|w|-1+1/p}\big] \le 1 + \mathbb{E}_\mu\big[\|x\|_{1\text{-var}}^{\beta}\big] =: 1 + \tilde{K}_p$ 对于 $\beta = 2 + 1/p$，因此我们对每个满足 $|w| = 2,\ 3$ 的 $w \in [d]^\star$ 都得到

$$\big|\langle\tilde{\mu}\rangle_w - \langle\mu\rangle_w\big| \;\le\; \tilde{c}_{p,|w|}(1+\tilde{K}_p)\cdot\phi_{|w|}(\|D_R\|_\infty)\big(\|D_R\|_\infty\|R\|_\infty\big)^\alpha \tag{C.35}$$

如所述。根据（C.35），由(ii) 可得，最后一个不等式（C.20）对 $\kappa := \tilde{c}_{p,\ 3}(1+K_p)$ 和 $\eta := \phi_3(\|D_R\|_\infty)(\|D_R\|_\infty\|R\|_\infty)^\alpha$成立。□

C.8. **引理4.6的证明。**

证明。注意，任何通过信号的核心坐标（29）作用于 $\mathfrak{D}$ 中信号的泛函 φ （例如引理 C.3 中的 φ_κ、$\mathfrak{m}_\mu$ 等），都可分解为 $\varphi = \phi_\varphi \circ [\cdot]^{\times k}$，其中 $[\cdot]^{\times k}\colon \dot{\mathfrak{D}}^{\times k} \to \mathbf{V}^{\times k}$，且$\phi_\varphi\colon \mathbf{V}^{\times k} \to \mathbb{R}$（对于 $(\mathbf{V}, \mathbf{k}\cdot\mathbf{k}_\mathcal{V})$，如（44）所示，且 k为 φ 的参数个数），其中$[\cdot]\colon \mu \mapsto ([\mu]_\nu) \in \mathbf{V}$ 是核心坐标映射（见（28））， 而 ϕ_φ 是 φ 的坐标作用。

进一步注意，根据 46 中 δ-球的定义以及限制 $r < \hat{r}$，像集

$$[B_r] \equiv [\cdot](B_r) \quad \text{is contained in} \quad \hat{\mathfrak{B}} := \bar{B}_{q\rho_0}([\hat{\mu}]_0) \times \bar{B}_{q\rho_0\sqrt{\rho_1}}([\hat{\mu}]_{\nu\geq 1}) \equiv \hat{\mathfrak{B}}_0 \times \hat{\mathfrak{B}}_1 \tag{C.36}$$

对某个 $q \equiv q_r \in (0,1)$，其中我们定义了 $\hat{\mathfrak{B}}_0 \equiv \bar{B}_{q\rho_0}([\hat{\mu}]_0) := \{C \in \mathbb{R}^{d\times d} \mid \|C - [\hat{\mu}]_0\| \leq q\rho_0\}$ 以及 $\hat{\mathfrak{B}}_1 \equiv \bar{B}_{q\rho_0\sqrt{\rho_1}}([\hat{\mu}]_{\nu\geq 1}) := \{C' \in (\mathbb{R}^{d\times d})^{\oplus d} \mid \|(0, C') - (0, [\hat{\mu}]_1, \ldots, [\hat{\mu}]_d)\|_\mathcal{V} \leq q\rho_0\sqrt{\rho_1}\}$。

显然 $\hat{\mathfrak{B}} \subset \mathcal{V}$ 是紧集，因此（由极值定理）我们有

$$K_{\varphi|r} := \sup_{\mu,\tilde{\mu}\in B_r} |\varphi(\mu,\tilde{\mu})| \leq \sup_{(\mathfrak{q},\tilde{\mathfrak{q}})\in\hat{\mathfrak{B}}^{\times 2}} |\phi_\varphi(\mathfrak{q},\tilde{\mathfrak{q}})| < \infty \quad \text{if} \quad \phi_\varphi\big|_{\hat{\mathfrak{B}}^{\times 2}} \text{ is continuous.} \tag{C.37}$$

这使我们能够从引理 C.3 推导出所述不等式，如下所示。

(i)：任取 $\tilde{\mu} \in B_r$，此时（C.29）和（C.30）（对于 $\mu := \hat{\mu}$ 和 $\eta := r$）成立。因此

$$\sup_{\mu\in B_r}\Big[\max_{|w|=2,3}\big|\langle\mu\rangle_w - \langle\hat{\mu}\rangle_w\big|\Big] \leq \mathfrak{m}_{\hat{\mu}|r} r \quad \text{and} \quad \sup_{\mu\in B_r}\Big[\big|\delta_\perp(\mu) - \delta_\perp(\hat{\mu})\big|\Big] \leq K_{\hat{\mu}|r}\sqrt{r + r^2}$$

对于 $\mathfrak{m}_{\hat{\mu}|r} := \sup^{\mu\in B_r} \mathfrak{m}_\mu$ 和 $K_{\hat{\mu}|r} \equiv K_{\hat{\varphi}|r} := \sup_{\mu\in B_r} |\hat{\varphi}(\mu,\hat{\mu})|$，其中 $\mathfrak{m}_\mu$ 和 $\hat{\varphi}$ 如引理 C.3 (iii) 中所述。现在，由于——正如本证明开头所指出的—— $\mathfrak{m}_\mu$ 和 $\hat{\varphi}$ 都通过映射 $[\cdot\,]$（分别为 $[\cdot\,]^{\times 2}$）分解，且其坐标作用分别为 $\phi_\mathfrak{m}$ 和 $\phi_{\hat{\varphi}}$，根据（C.37），$\mathfrak{m}_{\hat{\mu}|r}$ 和 $K_{\hat{\mu}|r}$ 的有限性成立；该结论适用是因为受限作用 $\phi_{\hat{\varphi}}\big|_{\hat{\mathfrak{B}}^{\times 2}}$ 和 $\phi_\mathfrak{m}\big|_{\hat{\mathfrak{B}}^{\times 2}}$ 都是连续的；后者分别由 [39]（C.36）和 $\mathfrak{m}_\mu$ 的定义得出。

(ii)：此断言可由引理 C.3 (iv)同理推出。事实上，任取 $\mu \in B_r$。则 $\|[\mu]_0 - [\hat{\mu}]_0\| \leq \rho_1 r$，因此 $\langle\mu\rangle_{ii} \geq \langle\hat{\mu}\rangle_{ii} - |\langle\hat{\mu}\rangle_{ii} - \langle\mu\rangle_{ii}| \geq \rho_0 - \rho_1 r$，从而，

$$\langle\tilde{\mu}\rangle_{ii} \geq \langle\mu\rangle_{ii} - |\langle\tilde{\mu}\rangle_{ii} - \langle\mu\rangle_{ii}| \geq \rho_0 - \rho_1 r - \rho_R > 0 \quad \text{for} \quad \tilde{\mu} := (I+R)_\star\mu \tag{C.38}$$

以及任意 $i \in [d]$（请注意，倒数第二个不等式由 (C.19) 得出）。$\mathfrak{h}_{\mu\mathfrak{i}_{ii}}$ 和 $\mathfrak{h}_{\tilde{\mu}\mathfrak{i}_{ii}}$ 的下界阻止泛函 $\varphi_R\colon \mu \mapsto \varphi_{\tilde{\kappa}}(\mu, (I+R)_\star\mu)\,\hat{\phi}_p(R)$（其中 $\hat{\phi}_p$ 和 $\varphi_{\tilde{\kappa}}$ 如 (C.20) 中所述，但将 $\tilde{\kappa} := 2^{\alpha+1}c_p(1+K_0) \geq \tilde{K}_p$ 替换为 3），以及其坐标作用 ϕ_{φ_R}，分别在定义域 B_r 和 $\tilde{\mathfrak{B}} \equiv \bar{B}_{\rho_1 r}([\hat{\mu}]_0) \times \hat{\mathfrak{B}}_1 \supseteq [B_r]$ 上趋于无穷。与点 (i) 一样，这保证作用 $\phi_{\varphi_R}\big|_{\tilde{\mathfrak{B}}}$ 连续，因此由 (C.20) 可如愿得到

$$\sup_{\mu\in B_r}\big|\delta_\perp((I+R)_\star\mu) - \delta_\perp(\mu)\big| \leq \tilde{K}_{\hat{\mu}|r}\cdot\phi_3(\|D_R\|_\infty)^{\frac{1}{2}}\big(\|D_R\|_\infty\|R\|_\infty\big)^{\frac{\alpha}{2}}$$

~ ~对于有限的常数（由C.37给出）$\tilde{K}_{\hat{\mu}|r} \equiv K_{\hat{\mu}|r,\ R} := \sup_{\mu\in B_r}|\varphi_R(\mu)| < \infty$。

对于任意 C，$c > 0$ 满足 $c < \rho_0$ 且 $\mathcal{R}_c^C$如（52）中所示，引理的最后一点可通过将~ ~上述 $K_{\hat{\mu}|r}$ 替换为 $K_{\hat{\mu}|c,\ C} := \sup_{(\mu,R)\in\mathcal{R}_c^C}|\varphi_R(\mu)|$来得到；根据 φ_R的定义、界（下界和上界）（C.38）[及其前一行] 和（C.19），显然它是有限的。□

---

[39] Indeed: Clearly $\phi_{\hat{\varphi}} = \sqrt{\max\{3\phi_\vartheta, \phi_\mathfrak{m}, \phi_\mathfrak{m}^2\}\sum_{k,i,j}\phi_{\varphi_{ijk}}}$ with $\phi_{\varphi_{ijk}} \circ [\cdot]^{\times 2} = \varphi_{ijk}$, (cf. (C.17)), and the inclusion (C.36), specifically $\{[\mu]_0 \mid \mu \in B_r\} \subseteq \hat{\mathfrak{B}}_0$, implies that the functions $\phi_{\varphi_{ijk}}\big|_{\hat{\mathfrak{B}}^{\times 2}}$ are each continuous since their denominators are bounded away from zero. (To be explicit, the latter is seen from

$$\langle\mu\rangle_{ii} \geq \langle\hat{\mu}\rangle_{ii} - |\langle\hat{\mu}\rangle_{ii} - \langle\mu\rangle_{ii}| \geq \langle\hat{\mu}\rangle_{ii} - \|[\hat{\mu}]_0 - [\mu]_0\| \geq \langle\hat{\mu}\rangle_{ii} - q\rho_0 \geq (1-q)\rho_0 > 0$$

该结论对每个 $\mu \in B_r$ 以及所有 $i \in [d]$均成立。）其余 $\phi_{\hat{\varphi}}$ 的组成部分，即 ϕ_ϑ 和 $\phi_\mathfrak{m}$，根据 ϑ 和 $\mathfrak{m}$的定义，分别显然是连续的（参见引理 C.3，其中给出了上述定义）。

C.9. **命题 5.1和引理 5.4的证明。**

命题 *5.1* 的证明。将其应用于恒等式 $\chi = A\cdot\zeta$，并展开至 $m=2$阶，3，上述签名矩的等变性可写为

$$\begin{aligned}\langle\chi\rangle_{ij} &= \textstyle\sum_{\alpha,\beta=1}^{d} a_{i\alpha}a_{j\beta}\cdot\langle\zeta\rangle_{\alpha\beta} \qquad \big((i,j)\in[d]^2\big) \qquad \text{and}\\ \langle\chi\rangle_{ijk} &= \textstyle\sum_{\alpha,\beta,\gamma=1}^{d} a_{i\alpha}a_{j\beta}a_{k\gamma}\cdot\langle\zeta\rangle_{\alpha\beta\gamma} \quad \big((i,j,k)\in[d]^3\big)\end{aligned} \tag{C.39}$$

回忆一下（29），这是一个等价于以下矩阵合同关系的方程组

$$[\chi]_0 = A[\zeta]_0A^{\intercal} \quad \text{and} \quad [\chi]_k = A\Big[\textstyle\sum_{\gamma=1}^{d} a_{k\gamma}[\zeta]_\gamma\Big]A^{\intercal} \tag{C.40}$$

对每个 $k\in[d]$。由恒等式（69）可得所述结论。至于共轭关系（70），由（C.40）可知，注意到对于线性组合，有 $[\chi]^{c}_{\odot} = c_0[\chi]_0^{-1}\cdot[\chi]_c$，

$$[\chi]_{\underline{c}} \;\equiv\; \textstyle\sum_{\nu=1}^{d}\tilde{c}_\nu[\chi]_\nu \;=\; A\Big[\sum_{\gamma=1}^{d}\big(\sum_{\nu=1}^{d}a_{\nu\gamma}c_\nu\big)[\zeta]_\gamma\Big]A^{\intercal} \;=\; A\Big[\sum_{\gamma=1}^{d}(a_\gamma\cdot\underline{c})[\zeta]_\gamma\Big]A^{\intercal}$$

因为这意味着

$$[\chi]^{c}_{\odot} \;=\; A^{-\intercal}c_0[\zeta]_0^{-1}A^{-1}\cdot A\Big[\textstyle\sum_{\gamma=1}^{d}(a_\gamma\cdot\underline{c})[\zeta]_\gamma\Big]A^{\intercal} \;=\; A^{-\intercal}\cdot[\zeta]^{A|c}_{\odot}\cdot A^{\intercal}.$$

由引理 $[\zeta]^{A|c}_{\odot}$ 的对角性显然成立4.8。 □

引理 *5.4* 的证明。由于 χ 是任意的，且GL_d-作用（A.7）保持 $\mathcal{D}$ 不变，因此只需证明（77）对 $\theta:=\mathrm{id}$ 以及任意满足 $\Lambda\equiv$对角$[\lambda_1,\dots,\lambda_d]$的 $\lambda_i>0$。那么由引理 3.7，参见（C.39），

$$\begin{aligned}\big\langle\Lambda\cdot\chi\big\rangle_{ij} &= \lambda_i\lambda_j\textstyle\sum_{\alpha,\beta=1}^{d}\delta_{i\alpha}\delta_{j\beta}\langle\chi\rangle_{\alpha\beta} = \lambda_i\lambda_j\langle\chi\rangle_{ij}\\ \big\langle\Lambda\cdot\chi\big\rangle_{ijk} &= \lambda_i\lambda_j\lambda_k\textstyle\sum_{\alpha,\beta,\gamma=1}^{d}\delta^{ijk}_{\alpha\beta\gamma}\langle\chi\rangle_{\alpha\beta\gamma} = \lambda_i\lambda_j\lambda_k\langle\chi\rangle_{ijk}\end{aligned}$$

对于 $i,\ j,\ k\in[d]$。特别是，$N_{\Lambda\cdot\chi}=\Lambda\cdot N_\chi$ 以及

$$\big[\Lambda\cdot\chi\big]_0 = \Lambda[\chi]_0\Lambda \quad\text{and}\quad \big[\Lambda\cdot\chi\big]_\nu = \lambda_\nu\Lambda[\chi]_\nu\Lambda \quad \text{for each } \nu\in[d],$$

这就通过定义（77）推出了（76）。 □

˜C.10. Ξ_1 **和** Ξ_1**的紧致性。**

引理 C.4. 对于 Ξ_1^0 如(80) 中所示的 $\gamma\geq1$，集合

$$\Xi_1^{\gamma} := \Xi_1^0\cap\{\theta\in\Theta\mid\kappa_2(\theta)\leq\gamma\} \quad \textit{is compact in} \quad \mathbb{R}^{d\times d}.$$

证明。集合 $\Xi_1^\gamma\subset\mathbb{R}^{d\times d}$ 有界（根据 Ξ_1^0的定义），因此只需证明它也是闭的。为此，取 $(\theta_n)\subset\Xi_1^\gamma$ 和 $\theta\in\mathbb{R}^{d\times d}$，使得 $\theta_n\to\theta$ 属于 $\mathbb{R}^{d\times d}$。那么必然也有 $\theta\in\Xi_1^0$，并且 $\lim_{n\to\infty}\|\theta_n\|_2=\|\theta\|_2\neq0$。后者意味着，对于某个 $c>0$ 以及几乎所有 $n\in\mathbb{N}$（例如对所有 $n\geq n_0$，其中某个 $n_0\in\mathbb{N}$），都有 $\|\theta_n\|_2\geq c$。现在假设θ 是奇异的，即 $\theta\notin\Theta$。那么，由于$\|\cdot\|_2$是次乘性的（由此可得下面的最后一个不等式，参见例如 [46, 习题 5.6.P47（第 369 页）]），我们必有

$$\gamma \;\geq\; \kappa_2(\theta_n) = \|\theta_n\|_2\|\theta_n^{-1}\|_2 \;\geq\; c\|\theta_n^{-1}\|_2 \;\geq\; c/\|\theta_n-\theta\|_2 \quad \text{for all } n\geq n_0,$$

这显然是一个矛盾。这意味着 $\theta\in\Theta$，因此，由于 $\kappa_2(\cdot)$ 在Θ上连续，也有 $\kappa_2(\theta)=\lim_{n\to\infty}\kappa_2(\theta_n)\leq\gamma$，从而得到 $\theta\in\Xi_1^\gamma$，如所求。 □

D. 第 5.2 和5.3 节的技术引理

D.1. **矩阵 C_μ 是泛型可逆的。**

引理 D.1. 设 $C_\mu := \frac{1}{2}([\mu]_0 + [\mu]_0)$ 为一个具有 0th 阶坐标 $[\mu]_0$ 的信号 $\mu \in \mathfrak{D}$。那么$C_\mu \notin \mathrm{GL}_d$ 当且仅当 $\mu_{0,1}(H) = 1$ 对于某个属于 $\mathbb{R}^d$ 的超平面 H 成立。

证明。假设 $C_\zeta \notin \mathrm{GL}_d$，这意味着存在 $v = (v_i) \in \mathbb{R}^d \setminus \{0\}$，使得 $C_\zeta v = 0$。回顾 $[\mu]_0$ 的定义 (29)，并快速计算 C_ζ定义的积分和 (28)，可知 $C_\zeta = \frac{1}{2}\big(\mathbb{E}[\zeta^i_{0,1} \cdot \zeta^j_{0,1}]\big)_{ij}$。因此，对于 $Z_v := \sum_{i=1}^d \zeta^i_{0,1} v_i = \zeta_{0,1} \cdot v (:= \varphi_*(\zeta)$，有 $0 = v^\intercal C_\zeta v = \sum_{i,j=1}^d v_i C^{ij}_\zeta v_j = \mathbb{E}[Z_v^2]$（对于 $\varphi := \mathsf{h}\pi_{0,1}(\cdot)\, v\, \mathsf{i}_2$，，见 (A.4) 和 (A.5)）。这意味着 $Z_v = \delta_0$，从而对于超平面 $H := \{u \in \mathbb{R}^d \mid u^\intercal v = 0\}$，有 $\zeta_{0,1}(H) = 1$。□

D.2. **超平面、逆矩阵与缩并张量。** 以下辅助观察有助于分析定理证明所依据的内容 5.8。

Lem引理 D.2. 设 $\vartheta \in \mathbb{R}^{d\times d}$ 可逆，且$(e_i)^{i\in[d]}$ 是 $\mathbb{R}^d$ 的标准基。那么 *set*

$$\big\{u \in \mathbb{R}^d \;\big|\; \exists\, i,j \in [d],\, i \neq j \,:\, \langle \vartheta^\intercal u, e_i\rangle = \kappa_{ij}\langle \vartheta^\intercal u, e_j\rangle\big\} \tag{D.41}$$

对于任意 $\kappa \equiv (\kappa_{ij}) \in \mathbb{R}^{d\times d}$，其勒贝格测度为零。

证明。 对于任意固定的 $\kappa \equiv (\kappa_{ij}) \in \mathbb{R}^{d\times d}$，考虑以下集合

$$H_{ij} := \big\{u = (u_\nu) \in \mathbb{R}^d \;\big|\; u_i = \kappa_{ij} \cdot u_j\big\}$$

对于满足 $i \neq j$ 的 i、$j \in [d]$。由于对于非零线性映射 $\varphi_{ij}\colon u \mapsto u_i - \kappa_{ij} \cdot u_j$，$H_{ij} = \ker(\varphi_{ij})$，每个集合 H_{ij} 的余维均为一（根据秩-零度定理），因此是 $\mathbb{R}^d$ 中的勒贝格零测集。因此，$H := \bigcup_{i,j\in[d],\, i\neq j} H_{ij}$ 的并集也是 $\mathbb{R}^d$ 中的勒贝格零测集，其微分同胚像 $(\vartheta^\intercal)^{-1}(H)$ 也如此。后者等于 (D.41)，从而证明了该结论。□

以下两个引理适用于$\|\cdot\|$ 任意次乘性矩阵范数。

引理 D.3. 设 $\Lambda \in \mathrm{GL}_d(\mathbb{R})$ 且 $\Delta \in \mathbb{R}^{d\times d}$ 满足 $\epsilon := \|\Lambda^{-1}\Delta\| < 1$。则 $\Lambda + \Delta$ 可逆，且 $\|(\Lambda+\Delta)^{-1} - \Lambda^{-1}\| \leq \|\Lambda^{-1}\| \frac{\epsilon}{1-\epsilon}$。

证明。 简记 $\rho := -\Lambda^{-1}\Delta$，注意到 $\Lambda + \Delta$ 的可逆性源于 $\Lambda + \Delta = \Lambda(\mathrm{I} - \rho)$ 以及诺伊曼级数的收敛（例如 [46, Ex. 5.6.P26]）。事实上

$$(\Lambda+\Delta)^{-1} = (\mathrm{I}-\rho)^{-1}\Lambda^{-1} = \big(\mathrm{I} + \textstyle\sum_{n=1}^\infty \rho^n\big)\Lambda^{-1}$$

因此，根据 $\epsilon < 1$，并且对于 $\eta := \|\Lambda^{-1}\|$，我们有 $\|(\Lambda+\Delta)^{-1} - \Lambda^{-1}\| \leq \eta \sum_{n=1}^\infty \|\rho\|^n \leq \frac{\eta\epsilon}{1-\epsilon}$。□

引理 D.4. 令 $C \equiv (C_0, \ldots, C_d) \in V$，并记 $\Lambda_\nu := \mathrm{ddiag}[C_\nu]$ 以及 $C := (0, C_1, \ldots, C_d)$和 $\Lambda := (0, \Lambda_1, \ldots, \Lambda_d)$。若 $\Lambda_0 \in \mathrm{GL}_d(\mathbb{R})$ 且 $\delta_1 := \big\|\Lambda_0^{-1} \cdot (C_0 - \Lambda_0)\big\| < 1$，则 C_0 可逆，且该收缩

$$C^c_\odot := c_0 C_0^{-1} \cdot \textstyle\sum_{\nu=1}^d c_\nu C_\nu,$$

其中 $c \equiv (c_0, c) \in \mathbb{R}^{1+d}$ 任意，其形式为

$$C^c_\odot = c_0\Lambda_0^{-1}\Lambda_c + \Delta_c \quad \textit{with} \quad \|\Delta_c\| \leq \beta_c(\delta_1, \delta_2)$$

对于对角矩阵 $\Lambda_c := \sum_{\nu=1}^d c_\nu \Lambda_\nu$、数值 $\delta_2 := \|C - \Lambda\|_\mathcal{V}$ 以及界函数 $\beta_c(a,b) := |c_0|\eta\Big[\|\Lambda_c\|\gamma^c_{a,b} + 1 + \gamma^c_{a,b}\Big]\gamma^c_{a,b}$ 其中 $\eta := \|\Lambda_0^{-1}\|$ 且 $\gamma^c_{a,b} := \text{最大值}\,\big\{\frac{a}{1-a}, b|\bar{c}|\big\}$。

证明。任取 $c \equiv (c_0, c) \in \mathbb{R}^{1+d}$。根据定义，$C_\odot^c = c_0 \tilde{D}_0 \cdot D_1$ 对于因子 $\tilde{D}_0 := C_0^{-1}$ 和 $D_1 := \sum_{\nu=1}^d c_\nu C_\nu$，其中 $D_1 = \Lambda_c + \Delta_1$ 由构造可知 $\mathbf{k}\Delta_1\mathbf{k} \le |c|\delta_2$。由于 $\delta_1 < 1$，引理 D.3 表明 $\tilde{D}_0 = \Lambda_0^{-1} + \tilde{\Delta}_0$，其中 $\|\tilde{\Delta}_0\| \le \eta\tilde{\delta}_1$ 对于 $\tilde{\delta}_1 := \frac{\delta_1}{1-\delta_1}$。由此可得，“非对角项” $\Delta_c :=$ $C_\odot^c - c_0\Lambda_0^{-1}\Lambda_c = c_0\big(\tilde{\Delta}_0\Lambda_c + \Lambda_0^{-1}\Delta_1 + \tilde{\Delta}_0\Delta_1\big)$ 可由以下式子界定

$$\begin{aligned} |c_0|^{-1}\|\Delta_c\| &\le \|\tilde{\Delta}_0\|\|\Lambda_c\| + \eta\|\Delta_1\| + \|\tilde{\Delta}_0\|\|\Delta_1\| \\ &\le \eta\tilde{\delta}_1\|\Lambda_c\| + \eta\tilde{\delta}_2 + \eta\tilde{\delta}_1\tilde{\delta}_2 \le \eta\Big[\|\Lambda_c\| + 1 + \bar{\delta}\Big]\cdot\bar{\delta} \end{aligned}$$

对于 $\delta_2 := |c|\delta_2$ 和 $\delta := $，最大化{$\delta_1$, δ_2}，如所需。□

D.3. **特征值间隙**。给定一个向量 $w \equiv (w_1, \ldots, w_d) \in \mathbb{C}^d$，我们将 w 的间隙 称为数

$$\gamma[w] := \min_{i,j\in[d]\,:\,i\neq j} |w_i - w_j| \qquad (d \ge 2)$$

并将 $\ell(w) := \|w - (w_{i_0}, \ldots, w_{i_0})\|_2$，，其中 $i_0 = \min\big[\arg\min_{i\in[d]} |w_i|\big]$，称为 w 的展布长度。

引理 D.5。在 $\mathbb{R}^d$ 中所有展布长度固定为 $\ell > 0$ 的向量中，间隙在任意满足 $|w^*_{i+1} - w^*_i| = |w^*_{j+1} - w^*_j|$ *for* $i, j \in [d-1]$ 的展布长度为 ℓ 的向量 $w^* \equiv (w^*_i) \in \mathbb{R}^d$ 处取得最大值。特别地：对于 $k_d := \sqrt{\frac{1}{6}(d-1)d(2d-1)}$，

$$\max_{w\in\mathbb{R}^d\,:\,\ell(w)=\ell} \gamma[w] = \ell/k_d. \tag{D.42}$$

证明。令 $w = (w_1, \ldots, w_d) \in \mathbb{R}^d$ 为任意一个固定的展布长度为 $\ell(w) =: \ell > 0$的向量。由于间隙 γ在置换和常数偏移下保持不变（即，对于任意 $c \in \mathbb{R}$和 $\tau \in S_d$，$\gamma[w] = \gamma[\tau(w) + (c, \ldots, c)]$），不失一般性，我们可以假设 $w_1 = 0$且对所有 $\nu \in [d-1]$都有 $w_\nu \le w_{\nu+1}$。现在假设对每个 $i, j \in [d-1]$，都有 $|w_{i+1} - w_i| = |w_{j+1} - w_j| =: \delta$。那么 $w_2 = |w_2 - w_1| = \delta$，同样地，对每个 $\nu \in [d]$，都有 $w_\nu = \sum_{j=1}^{\nu-1} (w_{j+1} - w_j) = (\nu - 1)\delta$，这意味着

$$\ell^2 = \ell(w)^2 = \|w\|_2^2 = \delta^2\textstyle\sum_{j=0}^{d-1} j^2 = \delta^2 k_d^2$$

因此 $\delta = \frac{\ell}{k_d}$。此处，由于根据上述假设，对于任意满足 $i \neq j$ 的$i, j \in [d]$，都有 $|w_i - w_j| \ge |w_{j+1} - w_j| = \delta$，我们得到 $\gamma[w] = \delta$。现在假设存在一个展布长度为 $\ell(\tilde{w}) = \ell$ 的 $\tilde{w} \equiv (\tilde{w}_i) \in \mathbb{R}^d$（其中$0 = \tilde{w}_1 \le \tilde{w}_2 \le \ldots \le \tilde{w}_d$如上所述可不失一般性），使得 $\gamma[\tilde{w}] > \delta \equiv \ell/k_d$，由此可推出

$$\ell^2 = \ell(\tilde{w})^2 = \sum_{i=1}^d \tilde{w}_i^2 = \sum_{i=1}^d \left[\sum_{j=1}^{i-1} \big|\tilde{w}_{j+1} - \tilde{w}_j\big|\right]^2 > \sum_{i=1}^d \big[(i-1)\delta\big]^2 = \delta^2 k_d^2 = \ell^2.$$

既然这就构成了矛盾，（D.42）便如所述成立。□

在接下来的两个小节中，所有特征值均按各自的重数计数，而 $\kappa(B) := \|B\|\|B^{-1}\|$ 表示矩阵 B 的条件数（相对于给定范数$\|\cdot\|$）。

D.4. 非退化特征子空间上的扰动界。作为鲁棒性分析中的一个重要组成部分，我们引用以下关于受扰动矩阵（不一定是对称矩阵）的单特征值和特征向量的界。

引理 D.6. 设 $C, E \in \mathbb{C}^{d\times d}$和 $\tilde{C} := C + E$。假设 C 的特征值 $\lambda_1, \ldots, \lambda_d$ 两两不同，令 $v_1, \ldots, v_d$ 为其对应的单位范数特征向量（即 $kv_ik_2 = 1$

对所有 $i \in [d]$），并令 (u_1, λ_1) …，(u_d, λ_d) 为 C 对应的“相对范数化”左特征对（即 $u_i \cdot C = \lambda_i u_i^\intercal$，其中 $u_i v_i^\intercal = 1, i \in [d]$）。[40] 若

$$\begin{aligned} \|E\|_{\mathrm{F}} \;&<\; \tfrac{1}{2} \min_{i\in[d]} \frac{s_i}{\kappa_i} \qquad \textit{with} \\ s_i := \sigma_{d-1}\big([\mathrm{I}_d - v_i v_i^\intercal]\,(\lambda_i \mathrm{I}_d - C)\big) \quad &\textit{and} \quad \kappa_i := \|u_i\|_2 + \sqrt{\|u_i\|_2^2 - 1}, \end{aligned} \tag{D.43}$$

其中 $\sigma_{d-1}(\cdot)$ 表示其自变量的第 2nd 小奇异值，则有

$$\begin{aligned} &\forall\, i \in [d] \;:\; \exists \mathrm{EV} \;\; \tilde{v}_{\sigma(i)} \;\; \textit{of} \;\; \tilde{C} \quad \textit{such that} \\ &u_i^\intercal \tilde{v}_{\sigma(i)} = 1 \quad \textit{and} \quad \|\tilde{v}_{\sigma(i)} - v_i\|_2 \;\le\; \frac{2\kappa_i}{s_i}\|E\|_{\mathrm{F}} \end{aligned} \tag{D.44}$$

~ ~对于 $\sigma :\; [d] \to [d]$ 某个赋值，其中每个对应的特征对 $(\lambda_j, \tilde{v}_j)$ 都满足

$$|\tilde{\lambda}_{\sigma(i)} - \lambda_i| \;\le\; |u_i^\intercal E v_i + \ell_E^i| \quad \textit{with} \quad |\ell_E^i| \;\le\; \frac{2\|u_i\|_2 \kappa_i}{s_i}\|E\|_{\mathrm{F}}^2. \tag{D.45}$$

证明。这可直接由——实际上基本上是引用——[54, 定理 4.1 (ii)]。 □

D.5. **若干矩阵函数的连续性与界。**

引理 D.7. 映射 $\lambda : \mathrm{Sym}_d \to \mathbb{R}^d,\ \mathfrak{a} \mapsto (\lambda_1(\mathfrak{a}), \cdots\ \lambda_d(\mathfrak{a}))$ 将一个对称$d \times d$矩阵映射为其特征值（按降序排列，重数重复计算）组成的向量；而“逆平方根”函数 $R : \mathrm{Sym}_d^+ \to \mathrm{Sym}_d^+$，$\mathfrak{a} \mapsto \mathfrak{a}^{-1/2}$二者均连续。此外，对于任意酉不变且次乘性的矩阵范数 $k \cdot k$，

$$\Big\|\mathcal{R}(\mathfrak{a}_1)^{-1} - \mathcal{R}(\mathfrak{a}_2)^{-1}\Big\| \le \Big[\lambda_d^{1/2}(\mathfrak{a}_1) + \lambda_d^{1/2}\big(\mathfrak{a}_2\big)\Big]^{-1} \cdot \|\mathfrak{a}_1 - \mathfrak{a}_2\| \tag{D.46}$$

对于每个 $\mathfrak{a}_1, \mathfrak{a}_2 \in$ 对称$_d^+$。特别地，对于任意固定的 $A \in$ 广义线性群$_d$和 $S(\mathfrak{a}) := \frac{1}{2}(\mathfrak{a} + \mathfrak{a})$，两者都

$$\textit{the maps} \quad \mathbb{R}^{d\times d} \ni \mathfrak{a} \longmapsto \lambda\big(A \cdot \mathcal{S}(\mathfrak{a}) \cdot A^\intercal\big) \in \mathbb{R}^d \qquad \textit{and} \tag{D.47}$$

$$\mathrm{Pos}_d \ni \mathfrak{a} \longmapsto \mathcal{R}\big(A \cdot \mathcal{S}(\mathfrak{a}) \cdot A^\intercal\big)^{-1} \in \mathbb{R}^{d\times d} \quad \textit{are continuous.} \tag{D.48}$$

此外，对于每个 $\mathfrak{a} \equiv (\tilde{a}_{ij}) \in \mathbb{R}^{d\times d}$，满足 最小值$_{i\in[d]} |\tilde{a}_{ii}| > 0$ 且 $C_\mathfrak{a} := A\mathcal{S}(\mathfrak{a})\intercal A \in$ 对称$_d^+$ 且$N_\mathfrak{a} := \mathrm{ddiag}(|\tilde{a}_{11}|, \cdots\ |\tilde{a}_{dd}|)^{1/2}$，并且 $A_{\mathcal{R}(C_\mathfrak{a})}$ 和 $\bar{A}_{\mathcal{R}(C_\mathfrak{a})}$ 的定义与上文类似，我们有

$$\big\|A_{\mathcal{R}(C_\mathfrak{a})}^{-1}\big\|_{\mathrm{F}} \;\le\; \|A^{-1}\|_{\mathrm{F}} \max_{i\in[d]} \lambda_i^{1/2}(C_\mathfrak{a}) \;=:\; \beta_1(\mathfrak{a}; A), \tag{D.49}$$

$$\kappa_2\big(A_{\mathcal{R}(C_\mathfrak{a})}\big) \;\le\; \frac{2\det(\mathrm{ddiag}[\lambda(C_\mathfrak{a})]^{1/2})\|A\|_{\mathrm{F}}^d}{\det(A)\big(\sqrt{d}\min_{i\in[d]} \lambda_i^{1/2}(C_\mathfrak{a})\big)^d} \;=:\; \beta_2(\mathfrak{a}; A). \tag{D.50}$$

证明。这些都是矩阵分析中的经典事实和不等式的直接推论。

确实如此：映射 λ 的连续性（它是良定义的，因为其（闭）定义域 Sym_d 中的元素均可对角化且具有实特征值），不过是一个众所周知事实的 [特例化] 重述：矩阵的谱连续依赖于其系数，参见例如 [46, 附录 D]。不等式 (D.46)（它蕴含连续性；请注意，矩阵求逆是（局部 Lipschitz）连续的）正是安藤–赫门不等式 [44, 定理 6.2]。至于其余断言：注意到，(D.49) 由 A_R 的定义、表示式 (73) 以及标准不等式立即得到：$\|\mathfrak{a}\mathfrak{b}\|_{\mathrm{F}} \le \|\mathfrak{a}\|_{\mathrm{F}}\|\mathfrak{b}\|_2$ 对每个$\mathfrak{a}, \mathfrak{b} \in \mathbb{R}^{d\times d}$均成立。后者结合 [39] 中的界即可得到 (D.50)。□

[40] 在引理 D.6的上下文中，$v := \bar{v}$表示向量 $v \in \mathbb{C}^d$的复共轭。53

D.6. **极值函数的连续性。**

引理 D.8。 设 $k\cdot k$ 是 $\mathbb{R}^n$ 上的任意范数，且 $\varphi:\ \mathbb{R}^n\to\mathbb{R}$ 是连续函数。再设 $\bar{B}_r:=\{x\in\mathbb{R}^n\mid kxk\le r\}$ 是半径为 $r\ge 0$ 的以零为中心的闭 $k\cdot k$ 球。则函数 $\phi,\psi:\ [0,\infty)\to\mathbb{R}$ 分别由 $\phi(r):=$最小值$_{y\in\bar{B}_r}\varphi(y)$ 和 $\psi(r):=$最大值$_{y\in\bar{B}_r}\varphi(y)$给出，二者均连续。

证明。由于对每个 $r\ge 0$，$\psi(r)=-\min_{y\in\bar{B}_r}(-\varphi(y))$，只需证明 ϕ 的连续性。

由于开射线是 $\mathbb{R}$ 上欧几里得（即序）拓扑的一个子基，若对每个 $a\in\mathbb{R}$，集合 $\ell_a:=\{r\mid\phi(r)<a\}$ 和 $\rho_a:=\{r\mid\phi(r)>a\}$ 都在 $[0,\infty$ 中是开集，则 ϕ 的连续性随即成立。为证明这一点，先考虑 ℓ_a，任取一个 $s\in\ell_a$。于是 $m_s:=\min_{y\in\bar{B}_s}\varphi(y)<a$，且由于 φ 连续，对于某个 $x_s\in B_s$，有 $\varphi(x_s)=m_s$。记 $B_u(z)$ 为半径为 u、以 z 为中心的开 $\mathrm{k}\cdot\mathrm{k}$ 球，连续性还意味着存在 $\delta_s>0$，使得对每个 $y\in B_{\delta_s}(x_s)$，$\varphi(y)<a$，因而对于所有 $s'\in(\mathrm{k}x_s\mathrm{k}-\delta_s/2,s]\cap[0,\infty)$，都有 $\phi(s')<a$。显然，对于所有 $s'\ge s$，$\phi(s')<a$，因此我们找到了 $\phi|_{J_s}<a$，使得开集（在 $[0,\infty$ 中）$J_s:=(\mathrm{k}x_s\mathrm{k}-\delta_s/2,\infty)\cap[0,\infty)\ni s$。因此 $J_s\subset\ell_a$，这说明 ℓ_a 是开集。

ρ_a 的开性可用类似方法证明：任取 $s\in\rho_a$。那么 $\phi(s)>a$，因此显然对每个 $s'\le s$，也有$\phi(s')>a$。此外：$\min_{y\in\partial B_s}\varphi(y)\ge\phi(s)>a$，由此根据$\varphi$ 的连续性可知，对每个 $z\in\partial B_s$，存在 $\delta_z>0$ 使得：对每个 $y\in B_{\delta_z}(z)$，都有 $\varphi(y)>a$。因此，对于开超集 $G_s:=\bigcup_{z\in\partial B_s}B_{\delta_z}(z)\cup B_s$ of $\bar{B}_s$，有$\varphi\big|_{G_s}>a$。现在，由于 $\bar{B}_s$是紧集，且G_s^{c}（即 G_s的补集）是闭集，距离 $\hat{}\,\delta_s:=d(\bar{B}_s,G_s^{\mathrm{c}})=\inf_{u\in\bar{B}_s,v\in G_s^{\mathrm{c}}}\mathrm{k}u-v\mathrm{k}$为正，这意味着对于任意 $0\le\delta<\hat{}\,\delta_s$，都有 $\bar{B}_{s+\delta}\subset G_s$。因此，对于开（在$[0,\infty$中）超集 $J_s':=(-\infty,s+\hat{}\,\delta_s)\cap[0,\infty)$ of s，有 $\phi\big|_{J_s'}>a$，这表明 ρ_a 是开集，证毕。□

有了以上结果，我们终于可以证明第 5 节的主要结论。

E. 定理 5.8 和5.12 的证明

E.1. **定理 5.8 的证明。对于** $A\equiv(a_1|\cdots|a_d)\in\mathrm{GL}_d$ 以及 $\zeta\equiv(\zeta^1,\ldots,\zeta^d)\in\mathfrak{D}$，假设 se

$$\chi\ =\ A\cdot\zeta\quad\text{ and }\quad\varepsilon\equiv\delta_{\perp\!\!\!\perp}(\zeta)<\varepsilon_0 \tag{E.51}$$

与 (87) 中的 ε_0 相同，符合要求。将 ζ 按 N_ζ^{-1} 缩放，并注意到对于 $\tilde{\zeta}:=N_\zeta^{-1}\cdot\zeta$，有 $N_{\tilde{\zeta}}=\mathrm{I}$ 且 $\delta_{\perp\!\!\!\perp}(\tilde{\zeta})=\delta_{\perp\!\!\!\perp}(\zeta)$，因此以 (ζ,A) 和 $(\tilde{\zeta},A:=AN_\zeta)$ 替换后，(E.51) 仍然成立。于是对每个 $\nu\in[d]$，由引理 5.4 和 (76) 可得，

$$\begin{aligned}\mathfrak{s}_0\ &:=\ \mathfrak{r}_0(A^{-1})=\mathfrak{r}_0\big(\tilde{A}^{-1}\big)=N_{\tilde{\zeta}}^{-1}\big[\tilde{\zeta}\big]_0N_{\tilde{\zeta}}^{-1}=\big[\tilde{\zeta}\big]_0\qquad\text{and}\\ \mathfrak{s}_\nu\ &:=\ \mathfrak{r}_\nu(A^{-1})=\mathfrak{r}_\nu\big(\tilde{A}^{-1}\big)=\big[\tilde{\zeta}\big]_\nu/\sqrt{\langle\tilde{\zeta}\rangle_{\nu\nu}}\ =\ \big[\tilde{\zeta}\big]_\nu.\end{aligned} \tag{E.52}$$

此外，对于来自 (72) 的白化矩阵 $R\equiv R_{(\chi)}$，以及预处理观测量$\bar{\chi}:=R\cdot\chi$，我们发现对于每个 $\nu\in[d]_0$ 以及任意 $\theta\in\Xi_1$，

$$N_{\theta\cdot\bar{\chi}}\ =\ \mathrm{I}\quad\text{ and hence }\quad\mathfrak{r}_\nu(\theta R)=\big[\theta R\cdot\chi\big]_\nu. \tag{E.53}$$

事实上：直接计算迭代积分（28）可得，对于每个 $\mu\in\mathfrak{D}$，矩阵 $C_\mu\equiv\frac{1}{2}([\mu]_0+[\mu]_0)$与经典矩阵 $\big(\frac{1}{2}\mathbb{E}[\mu_{0,1}^i\mu_{0,1}^j]\big)$相同。因此，由于 $C_{\bar{\chi}}=\mathrm{I}$是根据 $R_{(\chi)}$的选择，任意 $\theta\equiv(\theta_1|\cdots|\theta_d)^{\mathsf{T}}\in\Xi_1$均满足

$$\begin{aligned}\langle\theta\cdot\bar{\chi}\rangle_{ii}\ &=\ \tfrac{1}{2}\mathbb{E}\Big[\big(\textstyle\sum_{j=1}^d\theta_{ij}\bar{\chi}_{0,1}^j\big)^2\Big]\ =\ \tfrac{1}{2}\displaystyle\sum_{j=1}^d\theta_{ij}^2\mathbb{E}\big[(\bar{\chi}_{0,1}^j)^2\big]+\sum_{j<k}\theta_{ij}\theta_{ik}\mathbb{E}[\bar{\chi}_{0,1}^j\bar{\chi}_{0,1}^k]\\ &=\ \sum_{j=1}^d\theta_{ij}^2\ =\ \|\theta_i\|_2^2\ =\ 1\qquad\text{for each } i\in[d],\end{aligned}$$

由此 $N_{\theta\cdot\bar{\chi}} = \mathrm{I}$，进而 $\mathfrak{x}_\nu(\theta R)_\nu = [\theta\cdot\bar{\chi}]_\nu$ 对所有 $\nu \in [d]_0$均成立，正如所述。由（E.51）可得

$$\mathfrak{s}_\nu = \Lambda_\nu^\zeta + \Delta_\nu^\zeta \quad \text{with} \quad \Lambda_\nu^\zeta \text{ diagonal and } \left\|\Delta_\nu^\zeta\right\| \le \varepsilon \tag{E.54}$$

fo分别 $\nu \in [d]_0$。事实上，根据定义（44）以及引理 4.8 或注 4.5，我们有

$$\varepsilon_0^2 \;>\; \varepsilon^2 = \delta_\perp(\zeta)^2 = \delta_\perp(\tilde{\zeta})^2 \;=\; \left\|\mathfrak{s}_0 - \Lambda_0^\zeta\right\|^2 + \sum_{\nu=1}^d \left\|\mathfrak{s}_\nu - \Lambda_\nu^\zeta\right\|^2 \tag{E.55}$$

对于对角矩阵

$$\Lambda_0^\zeta \;:=\; \mathrm{ddiag}\big([\tilde{\zeta}]_0\big) \;=\; N_{\tilde{\zeta}}^2 \;=\; \mathrm{I} \quad \text{and} \quad \Lambda_\nu^\zeta \;:=\; \big(\langle\tilde{\zeta}\rangle_{\nu\nu\nu}\cdot\delta_{ij\nu}\big)_{ij}\,; \tag{E.56}$$

令 $\Delta_\nu^\zeta := \mathfrak{s}_\nu - \Lambda_\nu^\zeta$于是得到（E.54），如所述。现在由（E.51），关系式 $\mathfrak{x}_\nu\left(A^{-1}\right) = \mathfrak{s}_\nu$ 由（E.52）及对比函数 78的定义 ϕ_χ 及其尺度不变性（79），可得

$$\min_{\theta\,\in\,\Xi_1}\phi_\chi(\theta) \;\le\; \phi_\chi(\bar{B}_R) = \phi_\chi(B_R) \;=\; \sum_{\nu=0}^d \|\mathfrak{s}_\nu\|_\times^2 \;\le\; \sum_{\nu=0}^d \left\|\Delta_\nu^\zeta\right\|^2 \;\le\; \varepsilon^2\,, \tag{E.57}$$

其中 $B_R := (RA)^{-1} =: (b_1|\cdots|b_d)$，且$\intercal$ $\bar{B}_R := \mathsf{ddiag}\big(|b_1|,\ \ldots,\ |b_d|\big)^{-1}\cdot B_R \in \Xi_1$将 B_R 的各行缩放至单位长度。现在令 $\theta_\star \in \Xi_1$ 为任意满足以下条件的对象

$$\phi_\chi(\theta_\star) \;\le\; \varepsilon^2\,, \quad \text{and set} \quad \tilde{\theta}_\star := \theta_\star R \tag{E.58}$$

（注意 $\theta_\star$ 由（E.57）以及 Ξ_1 的紧致性可知其存在）。根据构造，参见（E.53），我们有

$$\mathfrak{x}_\nu(\tilde{\theta}_\star) = \big[\tilde{\theta}_\star\cdot\chi\big]_\nu \quad \text{for all } \nu \in [d]_0, \tag{E.59}$$

并由（E.58）以及 ϕ_χ 的定义可得，对每个 $\nu \in [d]_0$，都有

$$\|\Delta_\nu^\star\| \;\le\; \varepsilon \quad \text{with} \quad \Delta_\nu^\star \;:=\; \mathfrak{x}_\nu(\tilde{\theta}_\star) - \Lambda_\nu^\star \quad \text{and} \quad \Lambda_\nu^\star := \mathrm{ddiag}(\mathfrak{x}_\nu(\tilde{\theta}_\star)).$$

因此，由（E.59）以及特征签名坐标的仿射等变性（C.40），可得

$$\begin{aligned}
\tilde{\theta}_\star\cdot[\chi]_0\cdot\tilde{\theta}_\star^\intercal \;&=\; \Lambda_0^\star + \Delta_0^\star \quad \text{and} \\
\tilde{\theta}_\star\cdot\Big[\textstyle\sum_{j=1}^d \tilde{\theta}_{\nu j}[\chi]_j\Big]\cdot\tilde{\theta}_\star^\intercal \;&=\; \Lambda_\nu^\star + \Delta_\nu^\star \quad \text{for all } \nu \in [d], \\
\text{as well as} \quad \left\|(\Lambda_0^\star)^{-1}\Delta_0^\star\right\| \;&=\; \|\Delta_0^\star\| \quad (\text{since } \Lambda_0^\star = N_{\tilde{\theta}_\star\cdot\chi}^2 = \mathrm{I}).
\end{aligned} \tag{E.60}$$

接下来，我们利用由（E.51）成立的核心坐标恒等式（C.40），通过（E.52）将关系式（E.60）与源统计量联系起来：在（E.60）中使用（C.40）代入矩阵 $[\chi]_\nu$，得到

$$\vartheta\cdot\diamondsuit_\nu^\zeta\cdot\vartheta^\intercal \;=\; \diamondsuit_\nu^\star \quad (\nu \in [d]_0) \quad \text{with} \quad \vartheta := \tilde{\theta}_\star A \tag{E.61}$$

其中 $\diamondsuit_\nu^\zeta$、$\diamondsuit_\nu^\star$ 定义为基于核心坐标的统计量

$$\diamondsuit_0^\zeta := [\zeta]_0 \quad \text{and} \quad \diamondsuit_\nu^\zeta := \textstyle\sum_{\gamma=1}^d \vartheta_{\nu\gamma}[\zeta]_\gamma \quad (\nu \in [d]), \quad \text{and} \quad \diamondsuit_\nu^\star := \Lambda_\nu^\star + \Delta_\nu^\star \quad (\nu \in [d]_0).$$

对于任意向量 $c \equiv (c_0,\ c) \in \mathbb{R}_\times \times \mathbb{R}^d$，记 $c \equiv (\tilde{c}_1,\ \ldots,\ \tilde{c}_d)_\perp$ 则可将合同关系（E.61）合并为以下“精简”关系

$$\vartheta\cdot c_0\diamondsuit_0^\zeta\cdot\vartheta^\intercal \;=\; c_0\diamondsuit_0^\star \quad \text{and} \quad \vartheta\cdot\diamondsuit_{\underline{c}}^\zeta\cdot\vartheta^\intercal \;=\; \diamondsuit_{\underline{c}}^\star \tag{E.62}$$

其中涉及 c 加权线性组合

$$\diamondsuit_{\underline{c}}^\zeta := \textstyle\sum_{\nu=1}^d \tilde{c}_\nu\diamondsuit_\nu^\zeta \quad \text{and} \quad \diamondsuit_{\underline{c}}^\star := \sum_{\nu=1}^d \tilde{c}_\nu\diamondsuit_\nu^\star.$$

因此（将（E.62）中的第一个恒等式左乘以逆，再与（E.62）中的第二个恒等式结合），我们得到c-加权共轭关系

$$\vartheta^{-\intercal}\cdot\hat{\diamondsuit}_c^\zeta\cdot\vartheta^\intercal \;=\; \hat{\diamondsuit}_c^\star\,, \qquad \forall\, c \in \mathbb{R}_\times \times \mathbb{R}^d, \tag{E.63}$$

对于复合统计量 $\hat{\diamondsuit}_c^\zeta := \tilde{c}_0\big[\diamondsuit_0^\zeta\big]^{-1}\cdot\diamondsuit_c^\zeta$ 和 $\hat{\diamondsuit}_c^\star := \tilde{c}_0\big[\diamondsuit_0^\star\big]^{-1}\cdot\diamondsuit_c^\star$，其中 $\tilde{c}_0 := c_0^{-1}$。根据引理 D.4 （参见（E.54）、（E.60）以及 $\varepsilon_0 < 1$），这些复合统计量具有如下形式

$$\hat{\diamondsuit}_c^\zeta \;=\; \hat{\Lambda}_c^\zeta + \hat{\Delta}_c^\zeta \quad \text{and} \quad \hat{\diamondsuit}_c^\star \;=\; \hat{\Lambda}_c^\star + \hat{\Delta}_c^\star$$

对于对角矩阵，记作 $\eta \in \{\zeta, \star\}$，

$$\hat{\Lambda}_c^\eta := \tilde{c}_0\big[\Lambda_0^\eta\big]^{-1}\sum_{\nu=1}^d \tilde{c}_\nu \bar{\Lambda}_\nu^\eta \qquad \text{with} \qquad \bar{\Lambda}_\nu^\star := \Lambda_\nu^\star \quad \text{and} \quad \bar{\Lambda}_\nu^\zeta := \sum_{\gamma=1}^d \vartheta_{\nu\gamma}\Lambda_\gamma^\zeta, \tag{E.64}$$

$\hat{}\,\hat{}\,\hat{}$ 和非对角项 $\Delta_c^\eta \equiv \diamondsuit_c^\eta - \Lambda_c^\eta$，其有界于

$$\begin{aligned} \big\|\hat{\Delta}_c^\zeta\big\| \;&\le\; |\tilde{c}_0|\left[\big\|\Lambda_c^\zeta\big\|_2 + 1 + \gamma^c_{\epsilon_0^\zeta,\epsilon_1^\zeta}\right]\cdot\gamma^c_{\epsilon_0^\zeta,\epsilon_1^\zeta} \quad \text{and} \\ \big\|\hat{\Delta}_c^\star\big\| \;&\le\; |\tilde{c}_0|\left[\big\|\Lambda_c^\star\big\|_2 + 1 + \gamma^c_{\epsilon_0^\star,\epsilon_1^\star}\right]\cdot\gamma^c_{\epsilon_0^\star,\epsilon_1^\star} \end{aligned} \tag{E.65}$$

对于 $\Lambda_c^\zeta := \sum_{\nu=1}^d c_\nu\bar{\Lambda}_\nu^\zeta$ $\big(= c_0\hat{}\,\Lambda_c^\zeta\big)$ 和 $\Lambda_c^\star := c_0\hat{}\,\Lambda_c^\star$ 以及 $\underline{\gamma}_{a,b} := \max\{\frac{a}{1-a}, |c|\cdot b\}$，以及标量

$$\epsilon_0^\eta := \big\|\Delta_0^\eta\big\| \quad \text{and} \quad \epsilon_1^\eta := \big\|(0, \Delta_1^\eta, \ldots, \Delta_d^\eta)\big\|_{\mathcal{V}}.$$

注意，由 (E.55) 和 (E.58) 可知

$$(\epsilon_0^\eta)^2 + (\epsilon_1^\eta)^2 \;\le\; \varepsilon^2 \quad \text{for both} \quad \eta = \zeta \quad \text{and} \quad \eta = \star. \tag{E.66}$$

现在我们区分两种情况。

首先，假设 $\varepsilon = 0$，即 $\delta_\perp(\zeta) = 0$，这意味着源的核心坐标 $([\zeta]_\nu)^{\nu\in[d]_0}$ 与乘积信号 $\zeta \equiv (\zeta^1, \ldots, \zeta^d)$ 的核心坐标 $\zeta^1\otimes\cdots\otimes\zeta^d$ 相同。在这种情况下（参见 (E.64),(E.65),(E.66)），矩阵 $\hat{}\,\diamondsuit_c^\star$ 和 $\hat{}\,\diamondsuit_c^\zeta$ 均为对角矩阵，后者可写为

$$\hat{\diamondsuit}_c^\zeta = \hat{\Lambda}_c^\zeta = \tilde{c}_0\sum_{\gamma,\nu=1}^d \vartheta_{\nu\gamma}\tilde{c}_\nu\Lambda_\gamma^\zeta = \operatorname{ddiag}\big(\lambda_1^c, \lambda_2^c, \cdots, \lambda_d^c\big) \tag{E.67}$$

对于 c 相关的特征值 $\lambda_1^c, \ldots, \lambda_d^c$，由下式给出

$$\lambda_i^c = \alpha_i(c)\cdot\big(\langle\zeta\rangle_{iii}/\langle\zeta\rangle_{ii}^{3/2}\big) \quad \text{for} \quad \alpha_i(c) := \tilde{c}_0\vartheta_i^\intercal\cdot\underline{c} \tag{E.68}$$

其中 $\vartheta =: (\vartheta_1|\cdots|\vartheta_d)$。现在，由于 ϑ 可逆，引理 D.2 表明集合

$$\mathcal{N} := \big\{c\in\mathbb{R}^{1+d} \;\big|\; \exists\, i,j\in[d],\, i\neq j \,:\, \lambda_i^c = \lambda_j^c\big\} \tag{E.69}$$

has 具有零勒贝格测度，这可由将上述引理应用于系数矩阵得出

$$(\kappa_{ij}) := \left(\frac{\langle\zeta\rangle_{jjj}\langle\zeta\rangle_{ii}^{3/2}}{\langle\zeta\rangle_{jj}^{3/2}\langle\zeta\rangle_{iii}}\cdot(1-\delta_{ii_0})\right) \tag{E.70}$$

其中 $i_0\in[d]$ 是 $\zeta_{i_0i_0i_0} = 0$ 成立时唯一的指标。（根据关于 ζ 的假设，若指标 i_0 存在，则它是唯一的；否则，也就是说，如果 $\langle\zeta\rangle_{iii}$ 均不为零，则令 $(\kappa_{ij}) := \left(\frac{\langle\zeta\rangle_{jjj}\langle\zeta\rangle_{ii}^{3/2}\langle\zeta\rangle_{jj}^{3/2}\langle\zeta\rangle_{iii}}{\zeta}\right)$。）

因此，(E.69) 的补集必定非空，我们可以找到一个 $c_\star\in\mathbb{R}_\times\times\mathbb{R}^d$，使得 $\lambda_1^{c_\star}, \ldots, \lambda_d^{c_\star}$ 在 $\hat{}\,\diamondsuit_c^\zeta$ 的 (E.67) 中的特征值——根据矩阵共轭 (E.63)，它们与 $\hat{}\,\diamondsuit_{c_\star}^\star$ 的特征值相同——两两不同，因而均为单重的。由于 $\hat{}\,\diamondsuit_c^\zeta$ 和 $\hat{}\,\diamondsuit_{c_\star}^\star$ 都是对角矩阵，$_\intercal$ 因此我们可以得出 $\vartheta\in\mathrm{M}_d$，于是

$$\vartheta\in\mathrm{M}_d \quad \text{and hence} \quad \tilde{\theta}_\star\in\mathrm{M}_d\cdot A^{-1}$$

正如定理所述。

接下来我们考虑 $\varepsilon > 0$ 这一情形，即源信号的分量 ζ^i $\zeta = (\zeta^1, \ldots, \zeta^d)$ 允许存在统计相关的情形。此时，恒等式 (E.63) 可写为

$$\tilde{\vartheta}^{-1}\cdot\Big[\hat{\Lambda}_c^\zeta + E_c\Big]\cdot\tilde{\vartheta} \;=\; \hat{\Lambda}_c^\star \quad \text{for} \quad \tilde{\vartheta} := \vartheta^\intercal, \tag{E.71}$$

其中 $\hat{}\,\Lambda_c^\zeta$ 和 $\hat{}\,\Lambda_c^\star$ 如 (E.64) 所示，而对于非对角矩阵

$$E_c \;:=\; \hat{\Delta}_c^\zeta - \tilde{\vartheta}\hat{\Delta}_c^\star\tilde{\vartheta}^{-1}. \tag{E.72}$$

利用（E.65），我们注意到，“扰动” E_c 可以由以下内容控制：

$$\|E_c\| \;\le\; \big\|\hat{\Delta}_c^\zeta\big\| + \kappa_2(\tilde{\vartheta})\big\|\hat{\Delta}_c^\star\big\| \;\le\; |\tilde{c}_0|\,\big[\kappa_c\cdot\hat{\varphi}_{\underline{c}}(\varepsilon) + (1+\kappa_0)\hat{\varphi}_{\underline{c}}(\varepsilon)^2\big] \tag{E.73}$$

对于界函数

$$\hat{\varphi}_{\underline{c}}(a) \;:=\; \max\big\{\tfrac{a}{1-a}, |\underline{c}|\cdot a\big\} \tag{E.74}$$

以及依赖于 c 的常数

$$\kappa_c := \big\|\tilde{\vartheta}\big\|_\varsigma|\underline{c}| + 1 + \kappa_2(\tilde{\vartheta})\big[\xi d|\underline{c}| + 1\big] \le \|A_R\|_\varsigma\sqrt{d}|\underline{c}| + 1 + \kappa_0\kappa_2(A_R)\big[\xi d|\underline{c}| + 1\big]$$

对于 $\xi := \big\|(0,\ [\bar{\chi}]_1,\ \ldots\ldots,\ [\bar{\chi}]_d)\big\|_\nu$ 以及 $\varsigma := \sqrt{\sum_{i=1}^d\langle\zeta\rangle_{iii}^2/\langle\zeta\rangle_{ii}^3}$ 以及 $A_R := RA$。事实上：令 ϑ_i 表示 i^{th} 的 ϑ 列，对（E.64）应用柯西–施瓦茨不等式，并结合（E.56），可得

$$\big\|\Lambda_c^\zeta\big\| \;\le\; |\underline{c}|\big\|\big(0,\big(\bar{\Lambda}_\nu^\zeta\big)_{\nu=1}^d\big)\big\|_\nu \;=\; |\underline{c}|\sqrt{\textstyle\sum_{i=1}^d\|\vartheta_i\|_2^2\langle\tilde{\zeta}\rangle_{iii}^2} \;\le\; |\underline{c}|\big\|\tilde{\vartheta}\big\|_\varsigma,$$

而包含关系 $\theta_\star \equiv (\theta_{ij}^\star) \in \Xi_1$ (单位行) 意味着 $\sum_{\alpha,\,\beta,\,\gamma=1}^d\big(\theta_{i\alpha}^\star\theta_{i\beta}^\star\theta_{\nu\gamma}^\star\big)^2 = 1$ 因此

$$\begin{aligned}\big\|\Lambda_c^\star\big\| \;&\le\; |\underline{c}|\big\|\big(0,\big(\mathrm{ddiag}[\mathfrak{r}_\nu(\theta_\star R)]\big)_{\nu=1}^d\big)\big\|_\nu \;=\; |\underline{c}|\sqrt{\textstyle\sum_{\nu,i=1}^d\big[\sum_{\alpha,\beta,\gamma=1}^d\theta_{i\alpha}^\star\theta_{i\beta}^\star\theta_{\nu\gamma}^\star\langle\bar{\chi}\rangle_{\alpha\beta\gamma}\big]^2}\\ &\le\; |\underline{c}|\sqrt{d^2\textstyle\sum_{\alpha,\beta,\gamma=1}^d\langle\bar{\chi}\rangle_{\alpha\beta\gamma}^2} = d|\underline{c}|\xi.\end{aligned}$$

因此，（E.73）中的第二个不等式由（E.65）和 $\max_{\eta\in\{\zeta,\star\}}\gamma_{\epsilon_0^\eta,\epsilon_1^\eta}^c \le \hat{}\,\varphi_c(\varepsilon)$ 得出，后者由（E.66）和单调性成立。矩阵 $\hat{}\,\Lambda_c^\zeta$ 的扰动（E.72）（参见（E.67））得到控制后，我们现在可以应用引理 D.6，找到收缩参数 $\mathcal{c} = (c_0, c)$；根据（E.71），这意味着 ϑ 与 M_d 中的某个元素接近。为此，回顾（E.67）和（E.68）（以及 $N_{\tilde{\zeta}} = \mathrm{I}$），$\hat{}\,\Lambda_c^\zeta$ 的对角元素 $\lambda_c^\vartheta := (\lambda_1^c, \ldots, \lambda_d^c)$ 为

$$\underline{\lambda}_c^\vartheta \;=\; \tfrac{1}{c_0}H_\vartheta\cdot\underline{c} \quad\text{for}\quad H_\vartheta := \mathrm{ddiag}\big(\langle\tilde{\zeta}\rangle_{111},\ldots,\langle\tilde{\zeta}\rangle_{ddd}\big)\cdot\tilde{\vartheta},$$

其中，在必要时对 $\tilde{\zeta}^1,\ldots,\tilde{\zeta}^d$ 重新排列后，我们可以假设来自（i_0E.70）的“消失”指标为 $i_0 = 1$。现在，对于给定的 $\alpha > 0$（将在下文具体指定），定义

$$\underline{\lambda}_\alpha \;:=\; \big(0, \alpha/k_d, 2\alpha/k_d, \ldots, (d-1)\alpha/k_d\big) \equiv \big(\lambda_i^\alpha\big)_{i=1}^d \tag{E.75}$$

对于引理 k_d 中定义的常数 D.5，并进一步设 $c_\alpha := (1, c_\alpha)$，其中满足

$$\underline{c}_\alpha \;:=\; \tilde{\vartheta}^{-1}\mathrm{ddiag}\big(0, \langle\tilde{\zeta}\rangle_{222}^{-1}, \ldots, \langle\tilde{\zeta}\rangle_{ddd}^{-1}\big)\cdot\underline{\lambda}_\alpha. \tag{E.76}$$

显然 $|\lambda_{\alpha+}| = \alpha$ 且 $\lambda_\alpha \equiv \lambda_{c_\alpha,-}^\vartheta$，根据引理 D.5，矩阵 $\hat{}\,\Lambda_{c_\alpha}^\zeta$ 具有如下性质：其任意两个特征值（即 λ_α 的各项）之间的最小距离，在集合 $\{\hat{}\,\Lambda_c^\zeta \mid c\in\mathbb{R}_\times\times\mathbb{R}^d:\ |\mathrm{diag}(\hat{}\,\Lambda_c^\zeta)| = \alpha\}$ 中的所有矩阵里达到最大。为简化下面的估计，我们选取 α，使得式（$\hat{}\,\varphi_{c_\alpha}(\varepsilon)$ E.74）中的 由其第二个自变量主导，即满足 $\hat{}\,\varphi_{c_\alpha}(\varepsilon) = |c_\alpha|\varepsilon$。显然，对于以下情形该条件成立：

$$\alpha > 0 \quad\text{such that}\quad |\underline{c}_\alpha| \;=\; \frac{1}{1-\varepsilon}. \tag{E.77}$$

注意，选择相应的 α 当然是可行的，因为参见（E.76），

$$|\underline{c}_\alpha| = \alpha\cdot\kappa_\zeta^\vartheta \quad\text{for}\quad \kappa_\zeta^\vartheta := \big|\tilde{\vartheta}^{-1}[D_\zeta\underline{\lambda}_1]\big| > 0, \tag{E.78}$$

其中 $D_\zeta := \mathrm{ddiag}\big[\big(0, \langle\tilde\zeta\rangle_{222}^{-1}, \ldots\ldots, \langle\tilde\zeta\rangle_{ddd}^{-1}\big)\big]$ 并且 κ_ζ^ϑ的正定性可由 $\tilde\vartheta^{-1}$ 的可逆性推出（因此 $\ker(\tilde\vartheta^{-1}) = \{0\} \not\ni D_\zeta \cdot \lambda_1$）。根据（E.73）和（E.77），我们可以将$E_\alpha := E_{c_\alpha}$ 界定为

$$\begin{aligned}\|E_\alpha\| &\le \kappa_{c_\alpha}|\underline{c}_\alpha|\cdot\varepsilon + (1+\kappa_0)|\underline{c}_\alpha|^2\cdot\varepsilon^2\\ &= \frac{\varepsilon}{(1-\varepsilon)^2}\big[1+\big\|\tilde\vartheta\big\|\varsigma + \kappa_2(\tilde\vartheta)\big[\xi d + (1-\varepsilon)\big] + \kappa_0\varepsilon\big]\\ &< \big[1+\kappa_0+\big\|\tilde\vartheta\big\|\varsigma + (1+\xi d)\kappa_2(\tilde\vartheta)\big]\cdot\frac{\varepsilon}{(1-\varepsilon)^2}. \end{aligned} \tag{E.79}$$

回顾 $A_R \equiv RA$ 和 $\|\tilde\vartheta\| = \|\vartheta\| \le \sqrt{d}\|A_R^{-}\|$（即 $\vartheta = \theta_\star A_R$，其中 $\theta_\star \in \Xi_1$），我们发现

$$K_\vartheta := 1+\kappa_0+\big\|\tilde\vartheta\big\|\varsigma + (1+\xi d)\kappa_2(\tilde\vartheta) \tag{E.80}$$

$$\le 1+\sqrt{d}\|A_R\|\varsigma + \big(1+(1+\xi d)\kappa_2(A_R)\big)\kappa_0 =: \bar K_0 \tag{E.81}$$

˜因为根据对 $\theta_\star$ 假设的条件数界，还有 $\kappa_2(\vartheta) \le \kappa_0\kappa_2(A_R)$。接下来我们结合引理 D.6。使用该引理中的术语，注意到根据上述对c_α的选取 (E.76)，对角矩阵 $C := \hat\Lambda_{c_\alpha}^\zeta$ 的特征值 $\lambda_i := \lambda_i^{c_\alpha}$, $i \in [d]$ 两两不同，其对应的单位范数特征向量为 $(v_i)^{i\in[d]} := (e_i)^{i\in[d]}$，即 $\mathbb{R}^d$ 的标准基。因此，我们有（采用引理 D.6 的符号约定）$(u_i)^{i\in[d]} = (e_i)^{i\in[d]}$，从而 $\kappa_i = 1$ 和

$$s_i = \sigma_{d-1}\big((\delta_{\mu\nu}-\delta_{i\mu\nu})_{\mu\nu}\cdot(C-\lambda_i\mathrm{I})\big) = \min_{j\in[d]\setminus\{i\}}\big|\lambda_j-\lambda_i\big| = \alpha/k_d \tag{E.82}$$

对每个 $i \in [d]$ (参见 (E.75))。我们注意到 E_α ($\equiv E$) 因此满足引理D.6的假设 (D.43)。事实上，首先注意到，比 (D.43) 更强的不等式

$$\|E_\alpha\| < \gamma^{-1}\min_{i\in[d]}\frac{s_i}{\kappa_i} = \frac{\alpha}{\gamma k_d}, \quad \gamma := 1+\sqrt{5}, \tag{E.83}$$

根据 (E.78)、(E.79) 和 (E.80)，可由（尚待证明的）不等式推出

$$\frac{\varepsilon}{1-\varepsilon} < \big(\gamma k_d\kappa_\zeta^\vartheta K_\vartheta\big)^{-1} =: q_\vartheta, \tag{E.84}$$

它 依次等价于 $\varepsilon < q_\vartheta/(1+q_\vartheta)$。但是根据定义，参见 (E.78) 和 (E.81)，

$$\kappa_\zeta^\vartheta\cdot K_\vartheta \le \kappa_\zeta^\vartheta(1+\kappa_0) + \kappa_2(\tilde\vartheta)\varsigma\varsigma_1 + (1+\xi d)\kappa_\zeta^\vartheta\kappa_2(\tilde\vartheta)$$

对于依赖于源的常数 $\varsigma_1 := \big\|D_\zeta\cdot\lambda_1\big\| = k_d^{-1}\sqrt{\sum_{i=1}^d (i-1)^2\,\langle\tilde\zeta\rangle_{iii}^2}$ 并且

$$\kappa(\tilde\vartheta) \le \kappa_0\kappa_2(A_R) \quad \text{as well as} \quad \kappa_\zeta^\vartheta \le \big\|\tilde\vartheta^{-1}\big\|\varsigma_1 \le \frac{\kappa_0\big\|A_R^{-1}\big\|}{\sqrt{d}}\varsigma_1$$

（因为 $\|\tilde\vartheta^{-1}\| = \|\vartheta^{-1}\| = \|A_R^{-1}\|\|\theta_\star^{-1}\|$ 且 $\|\theta_\star^{-1}\| \le \kappa_0/\|\theta_\star\| = \kappa_0/\sqrt{d}$)$^-$，我们得到

$$\kappa_\zeta^\vartheta K_\vartheta \le r_0 := \kappa_0\varsigma_1\left[\frac{\big\|A_R^{-1}\big\|}{\sqrt{d}}\big(1+\kappa_0+(1+\xi d)\kappa_0\kappa_2(A_R)\big) + \kappa_2(A_R)\varsigma\right] \tag{E.85}$$

因此 $q_\vartheta \ge \big(\gamma k_d r_0\big)^{-1} =: q_0$ 从而

$$\frac{q_\vartheta}{1+q_\vartheta} \ge \frac{q_0}{1+q_0}. \tag{E.86}$$

但由于 $\varepsilon < \varepsilon_0 = q_0/(1+q_0)$ 根据假设，我们得出（E.86）成立，因此（E.84）以及（E.83）也成立，尤其是（D.43）成立，正如所需。这保证引理 D.6适用于上述矩阵 $C \equiv \hat\Lambda_{c_\alpha}^\zeta$ 及其扰动 $\tilde C := C + E_\alpha$（参见（E.71））。将˜ ˜ ˜记为 λ_1、……、λ_d，它们是 C的特征值，按其在对角矩阵上从左到右出现的顺序排列

ˆ在（ $\Lambda^\star_{c_\alpha}$ ）的对角矩阵上出现 E.71。由于进一步由（E.82）和（E.83）可知，

$$\|E_\alpha\| + \frac{2\|u_i\|_2\kappa_i}{s_i}\|E_\alpha\|^2 \;<\; \frac{(2+\gamma)\alpha}{\gamma^2 k_d} = \frac{\alpha}{2k_d},$$

引理 D.6 (参见 (D.45)) 由此可知，存在一个指派 $\sigma : [d] \to [d]$ 使得

$$\big|\tilde\lambda_{\sigma(i)} - \lambda_i\big| \;<\; \frac{\alpha}{2k_d} \qquad \text{for each } i \in [d]. \tag{E.87}$$

˜ ˜ ˜由此可知， σ 实际上是一个置换，并且 $\lambda_1, \dots, \lambda_d$ 的特征值 C（均为实数）两两不同。事实上，反设存在 $i, j \in [d]$ ，满足 $i \neq j$ 使得˜ ˜$\lambda_{\sigma(i)} = \lambda_{\sigma(j)}$ 由此，根据 (E.87) 可得

$$|\lambda_i - \lambda_j| \;\le\; \big|\lambda_i - \tilde\lambda_{\sigma(i)}\big| + \big|\tilde\lambda_{\sigma(j)} - \lambda_j\big| \;<\; \alpha/k_d,$$

这与 $\min_{i,j\in[d],\, i\neq j} |\lambda_i - \lambda_j| = \alpha/k_d$ 矛盾（回忆 (E.75)）。因此，对每个 $i \in [d]$，与 $\tilde\lambda_i$ 对应的 V_i 的特征子空间 $\tilde C$ 的实维数为一，即 $V_i = \langle\tilde\vartheta_i\rangle_{\mathbb{R}}$ ，对应 $\tilde\vartheta_i$ 矩阵 i^{th} 的第 $\tilde\vartheta$列。因此，根据引理 D.6 (D.44)，可得，[41] 对于每个 $i \in [d]$ 以及上述的 $\sigma \in S_d$ ，

$$\big\|\beta_i^{-1}\cdot\tilde\vartheta_{\sigma(i)} - e_i\big\|_2 \;\le\; \frac{2k_d}{\alpha}\|E_\alpha\| \quad \text{for } \beta_i := \tilde\vartheta_{\sigma(i)i}. \tag{E.88}$$

利用（E.78）和（E.79），（E.88）中的右侧可估计为

$$\frac{2k_d}{\alpha}\|E_\alpha\| \;\le\; r_1\frac{\varepsilon}{1-\varepsilon} \quad \text{for} \quad r_1 := 2k_d\kappa_\zeta^\vartheta K_\vartheta, \tag{E.89}$$

when尤其是 $\frac{2k_d}{\alpha}\|E_\alpha\| < 2/\gamma < 1$ 根据（E.84）可得。至于对 β_i的估计，首先注意到 at

$$\beta_i \;=\; \max_{\nu\in[d]}\big|\tilde\vartheta_{\nu i}\big| \;\equiv\; \|\tilde\vartheta_i\|_\infty \qquad \text{for all } i \in [d],$$

否则 $\max_{\nu\neq\sigma(i)}\big|\beta_i^{-1}\cdot\tilde\vartheta_{\nu i}\big| \ge 1$ ，从而 $\big\|\beta_i^{-1}\cdot\tilde\vartheta_{\sigma(i)} - e_i\big\|_2 \ge 1$ ，这与$\frac{2k_d}{\alpha}\|E_\alpha\| < 1$矛盾。因此，对每个 $i \in [d]$,

$$\beta_i \;\ge\; d^{-1/2}\|\tilde\vartheta_i\|_2 \;\ge\; \frac{\|A_R\|_2}{\kappa_2(A_R)\sqrt{d}}$$

其中，第一个不等式源于最大值范数与欧几里得范数等价，˜最后一个不等式由 $\|\vartheta_i\|_2 = \|A_R\cdot\theta_i\|_2 \ge \sigma_d(A_R) = \|A_R\|_2/\kappa_2(A_R)$，其中我们利用了 θ_i 表示 $\theta_\star$的第 i^{th} 行，其欧几里得范数为 1。此外，

$$\beta_i \;\le\; \|\tilde\vartheta_i\|_2 = \|A_R^\intercal\cdot\theta_i\|_2 \;\le\; \|A_R\|_2 \tag{E.90}$$

对每个 $i \in [d]$。因此， $\Lambda := \mathsf{ddiag}[\beta_1^{-1}, \dots, \beta_d^{-1}]$ 的条件数有界于

$$\kappa_2(\Lambda) \;\le\; \sqrt{d}\kappa_2(A_R). \tag{E.91}$$

结合（E.88）和（E.89），我们得到

$$\big\|\tilde\vartheta\cdot\tilde M - \mathrm{I}\big\| \;\le\; r_2\frac{\varepsilon}{1-\varepsilon} \quad \text{for} \quad r_2 := d\cdot r_1, \tag{E.92}$$

˜其中 $M := P_\sigma\cdot\Lambda \in \mathbb{M}_d$ 如上所述，并且 $P_\sigma = (\delta_{\sigma(i)j})$ 是与 σ 相关联的置换矩阵˜ ˜ ˜。ᵀ因此，对于 $M := M^- \in \mathbb{M}_d$ 和 $E := (\vartheta M - \mathrm{I})$，我们得到

$$\vartheta \;=\; M(\mathrm{I}+E) \qquad \text{with} \qquad \|E\| \le \tfrac{r_2\varepsilon}{1-\varepsilon} \tag{E.93}$$

[41] 由于对于 $(\tilde\lambda_j, \tilde w_j)$ 的每个特征对 $\tilde C$，必有 $\tilde w_j = q\cdot\tilde\vartheta_j$, $q \in \mathbb{R}_\times$，又因为 $\tilde\lambda_j$ 是单重的，因此在（$\tilde v_j \equiv (\tilde v_j^1, \dots, \tilde v_j^d)$ 中对应于 $\tilde\lambda_j$ 、并通过 D.44 ）利用 $(\tilde v_j^i =)\; u_i\,\tilde v_j = 1$ 选出的特征向量为 $\tilde v_j = \tilde\vartheta_{ij}^{-1}\cdot\tilde\vartheta_j$。

并且 $r_2 \leq 2dk_d r_0$，对于 r_0 如（E.85）所示；这证明了（88）。为了验证（88）中的最后一个不等式，~我们首先注意到，由于 $\vartheta = \theta_\star A$ 根据定义，方程（E.93）可得

$$\tilde{\theta}_\star \;=\; M \cdot A^{-1} + ME \cdot A^{-1}. \tag{E.94}$$

特别是，由此我们得到了相对误差界

$$\frac{\|\tilde{\theta}_\star - MA^{-1}\|}{\|MA^{-1}\|} \;\leq\; \frac{\|MEM^{-1}\|_2 \|MA^{-1}\|}{\|MA^{-1}\|} \overset{(E.91)}{\leq} r_3 \frac{\varepsilon}{1-\varepsilon} \;=:\; \varepsilon' \tag{E.95}$$

其中 $r_3 := \sqrt{d}\kappa_2(A_R) r_2$ （回顾（E.91）和（E.92））。证明至此完成。 □

E.2. **相关工作**。 如第 5.2 节所述，通过对观测信号的适当选取的等变张量统计量进行联合对角化，从而对线性混合随机源进行盲反演的总体策略——如（82）中所采用的方法——已成为当前盲源分离理论中成熟的一部分，概述可参见 [17, 18, 91] 或 [28, 68, 71, 77, 90, 91] 。该策略在盲源分离语境中的重要性，使得这一联合对角化问题（JDP）——它是对流形 Θ （以及诸如（81）所示的某些子集）上的对比函数（78）进行最小化的基础——在数值分析和优化领域引起了广泛关注并得到大量研究，参见所引文献及其中的参考文献。

然而，尽管这一问题显然十分重要且一直受到积极研究，关于 JDP 的扰动分析，以及一般而言针对张量分解的相关敏感性分析，却长期未得到充分探索，参见 [16, 第 1 节。4]在为数不多的现有扰动结果中，与我们的界（88）最为接近的是以下研究： [1], 其针对对称扰动的对称矩阵，得出了联合对角化矩阵的一阶扰动界； [86], 其采用与我们相同的行归一化约束，但同样仅考虑完全对称的情形，得出了对称扰动的联合对角化矩阵与精确联合对角化矩阵之间的界；最后还有 [16], 其在更一般的受扰联合块对角化背景下，得出了与我们的结果相关的上界。为求完整，我们还提及近期一项虽无直接关联的研究 [62], 该研究同样在对称约束下，考察了联合近似对角化矩阵的一些数值稳定性性质。

我们强调，尽管上述研究在性质上与我们的工作相似，但没有一项能使我们的界（88）或其基础扰动分析（第E.1节）变得多余：定理5.8的扰动结果是证明定理5.12的基石；在该定理中，我们通过预度量空间之间映射的连续性，确立了（84）的鲁棒性，而要做到这一点，就必须理解统计量（76）受到一般扰动（因而也包括非对称扰动）时，极小化子（82）会受到何种影响（仅有 [16] 适用）。基于近期针对一般扰动矩阵的特征子空间所得到的二阶界 [54]，第E.1节量身定制的扰动分析在方法论上独立于所有先前方法，并以清晰、简洁且基本自成一体的方式阐明了这一点。然而，如上所述，我们得到的界（88）并未经过优化，因此，将第E.1 节与所引用的某些方法结合起来，很可能会带来改进。

E.3. **定理 5.12的证明： 的连续性**。

证明。 根据注 4.12，并且由于 $(E,\ \mathrm{d})$ 是一个预度量空间，而 $(F,\ \mathfrak{d})$ 是一个度量空间，因此映射 $\mathfrak{I}$ 在 $\mathscr{I}$ 上连续，当且仅当（参见注 4.3），对每个 $\mathfrak{p}_\star \in \mathscr{I}$，都有：

$$\forall \varepsilon > 0 : \exists r > 0 : \forall \mathfrak{p} \in E : \ \text{if} \quad \mathrm{d}(\mathfrak{p}_\star, \mathfrak{p}) < r \quad \text{then} \quad \mathfrak{d}(\mathfrak{I}(\mathfrak{p}_\star), \mathfrak{I}(\mathfrak{p})) < \varepsilon. \tag{E.96}$$

为证明（E.96），任取 $\mathfrak{p}_\star \equiv (\mu_\star,\ A) \in \mathscr{I}$，并令 $\varepsilon > 0$ 为任意值。

进一步令 $0 < r < r_0$，对于 r_0 如引理 4.13中所述，并任取 $\mathfrak{p} \equiv (\mu,\ f)$ 属于 E 且满足$\mathrm{d}(\mathfrak{p}_\star,\ \mathfrak{p}) < r_\circ$。根据引理 4.13，可为该 $\mathfrak{p}$ 唯一指定一个信号 μ_ζ 属于 $\mathfrak{D}$，使得

$$\mu_\chi \equiv f_*\mu = A\mu_\zeta \quad \text{with} \quad \delta_{\perp\!\!\!\perp}(\mu_\zeta) \le K' r^{\alpha/2} \quad \text{and} \quad \big\|[\mu_\zeta]_{\mathfrak{c}} - [\mu_\star]_{\mathfrak{c}}\big\|_{\mathcal{V}} \le \hat{K} r^{\alpha} \tag{E.97}$$

对于（显式）常数 $K' \equiv K_r := K_r(\mu_\star,\ K_{\mathfrak{S}}) > 0$，其关于 r递增，并且$\hat{} K = \hat{} K(\mu_\star,\ K_{\mathfrak{S}}) > 0$。根据第5.1节，三元组 $(\mu_\chi,\ \mu_\zeta,\ A)$ 可与可识别性相关参数

$$\varepsilon(\mu_\zeta) := \frac{q(\mu_\zeta)}{1+q(\mu_\zeta)} \quad \text{with} \quad q(\mu_\zeta) := \big(\gamma k_d p(\mu_\zeta)\big)^{-1} \tag{E.98}$$

对于 $\gamma = 1+\sqrt{5}$、$k_d = \sqrt{\frac{d}{6}(d-1)(2d-1)}$ 以及函数 $p(\nu) = p(\nu;\ \mathfrak{p}_\star)$ $(\nu \in \mathfrak{D})$，其定义为

$$p(\nu) = \tilde{\kappa}_0 \varsigma_1(\nu)\left[\frac{\big\|A_{R_\nu}^{-1}\big\|}{\sqrt{d}}\big(1+\tilde{\kappa}_0 + (1+\xi(\nu)d)\tilde{\kappa}_0\kappa_2(A_{R_\nu})\big) + \kappa_2(A_{R_\nu})\varsigma(\nu)\right]$$

对于 $A_{R_\nu} := R_\nu A$ 以及白化矩阵 $R_\nu := \mathcal{R}(\mathcal{C}^{[\nu]_0})$ 如引理D.7中所示，令$\mathcal{C}_{\mathfrak{a}} := \mathcal{A}\mathcal{S}(\mathfrak{a})\ A$，对于上文固定的 A，以及辅助函数

$$\varsigma_1(\nu) := k_d^{-1}\sqrt{\sum_{i=1}^d \Big(\tfrac{(i-1)\langle\nu\rangle_{ii}^{3/2}}{\langle\nu\rangle_{iii}}\Big)^2} \quad \text{and} \quad \varsigma(\nu) := \sqrt{\textstyle\sum_{i=1}^d \langle\nu\rangle_{iii}^2/\langle\nu\rangle_{ii}^3} \tag{E.99}$$

并且 $\xi(\nu) := \sqrt{\sum_{j=1}^d \big\|[A_{R_\nu}\cdot\nu]_j\big\|^2}$ 以及 $\tilde{\kappa}_0 := b + \Delta_\kappa$，参见（94）。进一步回顾注5.11，存在半径 $r_* \equiv r_*(\mathfrak{p}_\star,\ \Delta_\kappa) > 0$，使得对于 $B_{(A,\nu)}$ 该注中定义的

$$\bar{B}_{(A,\nu)} \in \tilde{\Xi}_1 \quad \text{if} \quad \big\|[\nu]_0 - [\mu_\star]_0\big\| \le r_*. \tag{E.100}$$

利用（E.97）中关于 $[\mu_\zeta]_{\mathfrak{c}}$ 与 $[\mu_\star]_{\mathfrak{c}}$之间距离的界，我们看到系数（E.99）$|_{\nu=\mu_\zeta}$ 是良定义的——也就是说，对于所有 $\langle\mu_\zeta\rangle_{ii} > 0$ 和 $\langle\mu_\zeta\rangle_{iii} > 0$，当 $i \in [d]$ 时——如果

$$r < \big[(\rho_0 \wedge \varrho_0)/\hat{K}\big]^{1/\alpha} =: r_1, \quad \text{where} \quad \varrho_0 := \min_{i\ge 2}\langle\mu_\star\rangle_{iii};$$

注意，根据 $\mathscr{I}$ 的定义，$\varrho_0 = \varrho_0(\mu_\star) > 0$。由于分母在 $\{\nu \in \mathfrak{D} \mid \mathsf{k}[\nu]_{\mathfrak{c}} - [\mu_\star]_{\mathfrak{c}}\mathsf{k}_{\mathcal{V}} \le \mathfrak{q}\}$ 上与零有~界距离 =: 对于任意 $\mathfrak{q} < \rho_0 \wedge \varrho_0$，$B_{\mathfrak{q}}$，我们看到系数界 $\mathfrak{K}_1(\mathfrak{q}) := \sup^{\nu\in\tilde{B}_{\mathfrak{q}}}\varsigma_1(\nu)$ 和 $\mathfrak{K}_2(\mathfrak{q}) := \sup^{\nu\in\tilde{B}_{\mathfrak{q}}}\varsigma(\nu)$ 都是有限的。进一步，由于对称化矩阵 $\mathsf{C}_{\mathfrak{a}} = A\mathsf{S}(\mathfrak{a})A$ 的特征值 $(\lambda_i(\mathfrak{a};\ A)) :_{\mathsf{T}}= \lambda(A\mathsf{S}(\mathfrak{a})A\ )$ 关于 $\mathfrak{a} \in \mathbb{R}^{d\times d}$ 连续（引理 D.7），其最小值$\tilde{\sigma}(\mathfrak{a}) := \min_{i\in[d]}\lambda_i(\mathfrak{a};\ A)$ 也是连续的。因此，并且由于 $\sigma_\star \equiv \tilde{\sigma}([\mu_\star]_0) > 0$ [回忆 $\mathsf{C}^{[\mu_\star]_0}$ 与 $\mathsf{S}([\mu_\star]_0) \in$ 正定同余（正定$_d$-包含关系成立，因为 $C_{\mu_\star}$ 可逆），参见（74）]，，我们有$\tilde{\rho}_0 := \sup\{s \ge 0 \mid \min_{\{\mathfrak{a}\in\mathbb{R}^{d\times d}:\mathsf{k}\mathfrak{a}-[\mu_\star]_0\mathsf{k}\le s\}}\tilde{\sigma}(\mathfrak{a}) \ge \sigma_\star/2\} > 0$。由此，根据（73），可得

$$\sup_{\mathfrak{a}\in\mathfrak{B}_{\tilde{\rho}_0}}\big\|\mathcal{R}(\mathcal{C}_{\mathfrak{a}})\big\| \le \sqrt{2d/\sigma_\star} =: L_{\mathfrak{p}_\star} \quad \text{for} \quad \mathfrak{B}_{\tilde{\rho}_0} \equiv \mathfrak{B}_{\tilde{\rho}_0}([\mu_\star]_0) := \big\{\mathfrak{a}\in\mathbb{R}^{d\times d}\,\big|\,\|\mathfrak{a}-[\mu_\star]_0\| \le \tilde{\rho}_0\big\}$$

(注意到 $\{\mathcal{C}_{\mathfrak{a}} \mid \mathfrak{a} \in \mathfrak{B}_{\tilde{\rho}_0}\} \subset \mathrm{Sym}_d^+$ 根据 $\tilde{\rho}_0$ 的定义）。因此显然 $\mathfrak{K}_3 := $上确界$\{\xi(\nu) \mid \nu \in \mathfrak{D}:\ \|[\nu]_{\mathfrak{c}} - [\mu_\star]_{\mathfrak{c}}\|_{\mathcal{V}} \le \tilde{\rho}_0\}$ 是有限的，结合（62）可得

$$\sup_{(\tilde{\mu},\tilde{f})\in\mathbb{B}_r^\star}\big\|\mathcal{R}(\mathcal{C}_{[\tilde{\mu}_\zeta]_0})\big\| \le L_{\mathfrak{p}_\star} \ \text{ provided that } \ r < r_0 \wedge r_1 \wedge (\tilde{\rho}_0/\hat{K})^{1/\alpha} =: r_2'. \tag{E.101}$$

此处，泛函赋值 [如（E.97）和（E.101）中所用]

$$\mathbb{B}_r^\star \ni (\tilde{\mu},\tilde{f}) \longmapsto \tilde{\mu}_\zeta \in \tilde{B}_{\hat{K}r^\alpha} \tag{E.102}$$

正是引理4.13所断言的，参见（E.97）。从现在起假设

$$r < r_2 := r_2' \wedge (r_*/\hat{K})^{1/\alpha},$$

我们刚刚看到，对于每个 $\nu \in \tilde{B}_{\mathfrak{q}_r}$ 满足 $\mathfrak{q}_r := \hat{} Kr^{\alpha}$，都有（E.100）成立，且满足支配关系

$$|p(\nu)| \le \tilde{\kappa}_0 \mathfrak{K}_1(\mathfrak{q}_r)\left[\frac{\beta_1([\nu]_0)}{\sqrt{d}}\big(1+\tilde{\kappa}_0+(1+\mathfrak{K}_3 d)\tilde{\kappa}_0\beta_2([\nu]_0)\big)+\beta_2([\nu]_0)\mathfrak{K}_2(\mathfrak{q}_r)\right]$$

其中 $\beta_j \equiv \beta_j(\cdot; A)\,(j=1,2)$ 是引理 D.7 (D.49)–(D.50) 中的界函数。上述内容等价于（参见引理 4.6的证明）以下不等式

$$p(\cdot) \le \vartheta_r \circ [\,\cdot\,]_0 \quad \text{for} \quad \vartheta_r := \tilde{\kappa}_0 \mathfrak{K}_1(\mathfrak{q}_r)\left[\frac{\beta_1}{\sqrt{d}}\big(1+\tilde{\kappa}_0+(1+\mathfrak{K}_3 d)\tilde{\kappa}_0\beta_2\big)+\beta_2\mathfrak{K}_2(\mathfrak{q}_r)\right]$$

该不等式对 $\tilde{B}_{\mathfrak{q}_r}$ 中的每一点均成立，且其中 $[\cdot]_0\colon \mathfrak{D} \to \mathbb{R}^{d\times d}$ 是将信号映射到其基准层坐标的映射（参见 (29)）。注意，虽然 β_1 连续，但函数 β_2 是形如 $\beta_2 = \hat{}\beta/z$的分式，其中 $\hat{}\beta$ 是其定义（D.50）中的分子，z 是分母；$\hat{}\beta$ 和 z 在 $\mathbb{R}^{d\times d}$ 上均连续（引理 D.7）。还要注意，根据 r_2的定义（见 (E.101)），[，并回顾 (73)]，，分母 z 在$[\tilde{B}_{\mathfrak{q}_r}]_0 \equiv \{[\nu]_0 \mid \nu \in \tilde{B}_{\mathfrak{q}_r}\} =: \mathfrak{A}_r$ 上有正的下界，即 $\mathfrak{z}(r) := \inf\{z(\mathfrak{a}) \mid \mathfrak{a} \in \mathfrak{A}_r\} > 0$。因此 $\vartheta_r\big|_{\mathfrak{A}_r} \le \hat{}\vartheta_r\big|_{\mathfrak{A}_r}$，进而对于 $\hat{}\beta_r := \mathfrak{z}(r)^{-1}\hat{}\beta$，我们有以下 $\tilde{B}_{\mathfrak{q}_r}$-逐点支配

$$p(\cdot) \le \hat{}\vartheta_r \circ [\,\cdot\,]_0 \quad \text{对于 } \hat{}\vartheta_r := \tilde{\kappa}_0\mathfrak{K}_1(\mathfrak{q}_r)\left[\frac{\beta_1(r)}{\sqrt{d}}\big(1+\tilde{\kappa}_0+(1+\mathfrak{K}_3 d)\tilde{\kappa}_0\hat{}\beta_r\big)+\hat{}\beta_r\mathfrak{K}_2(\mathfrak{q}_r)\right]. \tag{E.103}$$

T　　ˆ上界函数 ϑ_r 在上述球 $\mathfrak{B}_{\tilde{\rho}_0} \supseteq \mathfrak{A}_r$ 上连续且严格为正　　.

特别地 [回想 $\mathfrak{q}_r \le \tilde{\rho}_0$]，对于函数 $\psi_r(\mathfrak{a}, \mathfrak{a}_1, \dots, \mathfrak{a}_d) := \hat{}\vartheta_r(\mathfrak{a})$，我们有

$$\psi_r \in C\big(\bar{B}^{\mathcal{V}}_{\mathfrak{q}_r}; \mathbb{R}_{>0}\big) \quad \text{on} \quad \bar{B}^{\mathcal{V}}_{\mathfrak{q}_r} \equiv \bar{B}^{\mathcal{V}}_{\mathfrak{q}_r}([\mu_\star]_{\mathfrak{c}}) := \{\tilde{\mathfrak{a}} \in \mathcal{V} \mid \|\tilde{\mathfrak{a}} - [\mu_\star]_{\mathfrak{c}}\|_{\mathcal{V}} \le r\},$$

a并由上述 $p(\cdot)$-支配（E.103），回顾（E.102）（另见（E.97））， 我们得到

$$p(\tilde{\mu}_\zeta) \le \psi_r([\tilde{\mu}_\zeta]_{\mathfrak{c}}) \quad \text{for each } (\tilde{\mu}, \tilde{f}) \in \mathbb{B}^\star_r. \tag{E.104}$$

令 $\psi := \psi_{\tilde{r}_2}$对于 $\tilde{r}_2 := 0.95\cdot r_2$，并引入函数 $\tau\colon [0, \tilde{r}_2] \to (0,\infty)$，其定义如下

$$\tau(r) := \min\Big\{Q\big((\gamma k_d \psi(\tilde{\mathfrak{a}}))^{-1}\big) \mid \tilde{\mathfrak{a}} \in \bar{B}^{\mathcal{V}}_{\mathfrak{q}_r}([\mu_\star]_{\mathfrak{c}})\Big\} \quad \text{for} \quad Q(u) := \frac{u}{1+u},$$

˜我们发现，对任意 $\mu_\zeta \in B_{\mathfrak{q}_r}$ 如（E.104） [回顾（E.102）] 由于（E.98）和（E.104），因此

$$\varepsilon(\mu_\zeta) = Q\big(g(\mu_\zeta)\big) \ge Q\big((\gamma k_d \psi([\mu_\zeta]_{\mathfrak{c}}))^{-1}\big) \ge \tau(r), \quad \text{for each } r \in [0, \tilde{r}_2]. \tag{E.105}$$

现在，由于 τ是连续的（引理 D.8），且 $\tau(0) = Q\big((\gamma k_d \psi([\mu_\star]_{\mathfrak{c}}))^{-1}\big) > 0$，因此存在一个正半径，使得在该范围内，$\tau$大于连续上界函数 υ：　$[0, r_0] \ni r \mapsto K' r^{\alpha/2}$的 $\delta_{\perp\!\!\!\perp}(\mu_\zeta)$ (参见（E.97）)。再次参见（E.97），可知必有

$$\begin{aligned} r_3 &:= \inf\big\{0 \le r \le \tilde{r}_2 \mid \tau(r) - \sup_{(\mu,f)\in\mathbb{B}^\star_r} \delta_{\perp\!\!\!\perp}(\mu_\zeta) \le 0\big\} \\ &\ge \inf\big\{0 \le r \le \tilde{r}_2 \mid \tau(r) - \upsilon(r) \le 0\big\} > 0. \end{aligned} \tag{E.106}$$

现在令 $0 \le r \le r_3$，并设 $\tilde{K}_1 := 2dk_d \max\{\psi(\tilde{\mathfrak{a}}) \mid \tilde{\mathfrak{a}} \in \bar{B}^{\mathcal{V}}_{\mathfrak{q}_{r_3}}\}$，再固定 $K' \equiv K_{r_3}(= K'|_{r=r_3})$。那么，对于任意满足（E.102）所关联的信号为 $\mu_\zeta \in \bar{B}^{\mathcal{V}}_{\mathfrak{q}_{r_3}}$ 的 $(\mu, f) \in \mathbb{B}_r(\mu_\star, A)$，由（E.106）和（E.105）可得 $\delta_{\perp\!\!\!\perp}(\mu_\zeta) \le \tau(r) \le \varepsilon(\mu_\zeta) < 1$ [，因此，根据定理 5.8 (88) [（由（E.100）可知该定理完全适用）]，我们得到

$$\begin{aligned} &\forall\, \theta_{\mathfrak{p}} \in \Phi(f_*\mu)\colon \text{存在} \quad M = M(\theta_{\mathfrak{p}}) \in \mathbf{M}_d \text{且} \quad E = E(\theta_{\mathfrak{p}}) \in \mathbb{R}^{d\times d} \text{使得} \\ &\qquad \theta_{\mathfrak{p}} = M(\mathrm{I}+E)A^{-1} \quad \text{with} \quad \|E\| \le \epsilon_r \quad \text{for} \quad \epsilon_r := \tilde{K}_1 \tfrac{\upsilon(r)}{1-\upsilon(r)} \end{aligned} \tag{E.107}$$

(另见E.93，可参阅定理 5.8 的证明）。显然，误差界 ϵ_r 关于 r 严格递增（回忆一下，$\upsilon(r) < 1$），且$\lim_{r\to 0+} \epsilon_r = 0$。

接下来，记常数 $K_2 := \kappa_2(A)L_{\mathfrak{p}_\star}$ 和 $\tilde{\gamma}_\star := \min_{i\in[d]} \left|e_i \cdot A^{-1}\mathcal{C}^{1/2}_{[\mu_\star]_0}\right|$，并考虑由下式给出的辅助函数 $\eta\colon [0,r_3]\to\mathbb{R}_+$

$$\eta(r) := \max\left\{\tilde{\gamma}_\star^{-1}\left\|A^{-1}\big(\mathcal{R}(\mathcal{C}_{\mathfrak{a}})^{-1}-\mathcal{R}_\star^{-1}\big)\right\| \;\middle|\; (\mathfrak{a},\mathfrak{a}_1,\cdots,\mathfrak{a}_d)\in \bar{B}^{\mathcal{V}}_{\mathfrak{q}_r}\right\} \tag{E.108}$$

对于 $\mathrm{R}_\star := \mathrm{R}(\mathrm{C}_{[\mu_\star]_0})$。根据 (D.48) 和引理 D.8，函数 η 连续，且 $\eta_i(0)=0$。进一步令 $\Lambda_{\mathfrak{a}} := \mathrm{ddiag}_{i=1,\dots,d}\big(|e_i\cdot A^{-1}\mathrm{R}(\mathrm{C}_{\mathfrak{a}})^{-1}|\big)^{\intercal}$，并通过 (E.102)（参见 (E.97)）设 $\Lambda_{(\mu,f)} := \Lambda_{[\mu_\zeta]_0}$。

For（任意） $\varepsilon>0$ 在本证明开始时固定的，参见 (E.96)，选择较大的 st

$$0<r_4\le r_3 \quad \text{such that} \quad \max_{r\in[0,r_4)}\big(\sqrt{2}\eta(r)+(\|A^{-1}\|_2\hat{\sigma}_r+1)\epsilon_r\big)<\varepsilon/K_2, \tag{E.109}$$

其中 $\hat{}\sigma^\star\colon [0,r_3]\ni r\mapsto \hat{}\sigma^\star_r :=$ 最大值$_{\mathfrak{a}\in\mathfrak{B}_{\mathfrak{q}_r}}\sqrt{\text{最大值}}(\mathcal{C}_{\mathfrak{a}})\in\mathbb{R}_+$（它是 [单调递增且] 根据 (D.47) 和引理 D.8) 连续。若能证明E.96，则证明即告结束（回顾 (

$$\mathfrak{d}\big(\mathfrak{I}(\mathfrak{p}_\star),\mathfrak{I}(\mathfrak{p})\big)\le\varepsilon, \quad \text{for each } \mathfrak{p}\in\mathbb{B}_{r_4}(\mathfrak{p}_\star). \tag{E.110}$$

˜ ˜确实如此：记 $\Xi_{\mathcal{R}} := \Xi_1\cdot\mathcal{R}$，由定理 5.8（根据 $\mathscr{I}$ 的定义，该定理适用）可知

$$\mathfrak{I}(\mathfrak{p}_\star)=\mathrm{M}_d\cdot A^{-1}\cap\tilde{\Xi}_{\mathcal{R}_\star}=\big\{P\Lambda^{-1}_{\mathfrak{p}_\star}A^{-1}\,\big|\,P\in\mathrm{P}^{\pm}_d\big\}=:\mathcal{A}_\star,$$

第二个等式可由上述定义轻易验证（回顾 (81)˜ 以及单位行条件 (80)）。为证明 (E.110) 成立，任取 $\mathfrak{p}\equiv(\tilde{\mu},f)\in\mathbb{B}_{r_4}(\mathfrak{p}_\star)$，先固定任意 $\mathfrak{a}_\star\equiv P_{\mathfrak{a}_\star}\Lambda^{-1}_{\mathfrak{p}_\star}A^{-1}\in\mathrm{A}_\star$，并为方便起见记 $\mathrm{R}^{(\tilde{\mu},\tilde{f})} := \mathrm{R}(\mathrm{C}_{[\tilde{\mu}_\zeta]_0})$（参见 (E.102)）˜。由于 $\mathfrak{I}(\mathfrak{p})$ 非空 (82)，存在某个 $\theta_{\mathfrak{p}}\in\Xi_{\mathcal{R}_{\mathfrak{p}}}$，使得 $\theta_{\mathfrak{p}}=M_{\mathfrak{p}}(\mathrm{I}+E_{\mathfrak{p}})A^{-1}$ 对 $M_{\mathfrak{p}}\equiv P_{\mathfrak{p}}\tilde{\Lambda}_{(\mathfrak{p})}\in\mathrm{M}_d$ 成立（其中 $P_{\mathfrak{p}}\in\mathrm{P}^{\pm}_d$ 且 $\tilde{\Lambda}_{(\mathfrak{p})}\equiv\mathrm{ddiag}(m^1_{\mathfrak{p}},\dots,m^d_{\mathfrak{p}})>0$），并且 $\mathrm{k}E_{\mathfrak{p}}\mathrm{k}\le\epsilon_{r_4}$，参见 (E.107)。由 $\phi_{\tilde{f}_\star\tilde{\mu}}$ 和 $\tilde{\Xi}_{\mathcal{R}_{\mathfrak{p}}}$ 对 $\mathrm{P}^{\pm}_d$ 不变（参见 (79) 和 (94)），可知 $\theta^\star_{\mathfrak{p}} := (P_{\mathfrak{a}_\star}P^{-1}_{\mathfrak{p}})\theta_{\mathfrak{p}}\in\hat{}\mathfrak{I}(\mathfrak{p})$ 也成立。因此（并且由于 $\mathrm{P}^{\pm}_d\subset\mathrm{O}_d$ 和 $\mathrm{k}A_1A_2\mathrm{k}\le\mathrm{k}A_1\mathrm{k}\mathrm{k}A_2\mathrm{k}_2$, $\forall\,A_1,A_2\in\mathbb{R}^{d\times d}$）

$$d\big(\mathfrak{a}_\star,\mathfrak{I}(\mathfrak{p})\big)\le\big\|\mathfrak{a}_\star-\theta^\star_{\mathfrak{p}}\big\|\le\|A^{-1}\|_2\big\|\Lambda^{-1}_{\mathfrak{p}_\star}-\tilde{\Lambda}_{(\mathfrak{p})}(\mathrm{I}+E_{\mathfrak{p}})\big\|. \tag{E.111}$$

由于 $\theta^\star_{\mathfrak{p}}\in\tilde{\Xi}_{\mathcal{R}_{\mathfrak{p}}}$，因此 $\theta^\star_{\mathfrak{p}}$实数集$^{-1}_p = P_{\mathfrak{a}_\star}\tilde{\Lambda}_{(\mathfrak{p})}(\mathrm{I}+E_{\mathfrak{p}})B_{\mathfrak{p}}\in\tilde{\Xi}_1$ 对于 $B_{\mathfrak{p}} :=$ (实数集$_{\mathfrak{p}}A)^{-1}$，我们有

$$1=\left|e_i^{\intercal}\cdot\theta^\star_{\mathfrak{p}}\mathcal{R}^{-1}_{\mathfrak{p}}\right|=m^{\sigma(i)}_{\mathfrak{p}}\left|e^{\intercal}_{\sigma(i)}\cdot\big(B_{\mathfrak{p}}+E_{\mathfrak{p}}B_{\mathfrak{p}}\big)\right| \quad \text{for each } i\in[d] \tag{E.112}$$

by t单位行要求 (80)；此处，σ 是 $P_{\mathfrak{a}_\star}$ 所对应的置换。因此 y,

$$\begin{aligned}\big\|\Lambda^{-1}_{\mathfrak{p}_\star}-\tilde{\Lambda}_{(\mathfrak{p})}(\mathrm{I}+E_{\mathfrak{p}})\big\| &\le \sqrt{\sum_{i=1}^d\left|\frac{1}{|e_i^{\intercal}\cdot B_{\mathfrak{p}_\star}|}-m^i_{\mathfrak{p}}\right|^2}+\|A\|_2L_{\mathfrak{p}_\star}\epsilon_{r_4}\\ &\overset{(E.112)}{\le}\sqrt{2}\sqrt{\sum_{i=1}^d\left[\frac{m^i_{\mathfrak{p}}|e_i^{\intercal}(B_{\mathfrak{p}}-B_{\mathfrak{p}_\star})|}{|e_i^{\intercal}\cdot B_{\mathfrak{p}_\star}|}\right]^2+\big(m^i_{\mathfrak{p}}|e_i^{\intercal}E_{\mathfrak{p}}B_{\mathfrak{p}}|\big)^2}+\|A\|_2L_{\mathfrak{p}_\star}\epsilon_{r_4}\\ &\le\frac{\sqrt{2}\|A\|_2L_{\mathfrak{p}_\star}\|B_{\mathfrak{p}}-B_{\mathfrak{p}_\star}\|}{\min_{i\in[d]}|e_i^{\intercal}\cdot B_{\mathfrak{p}_\star}|}+\|A\|_2\big(\|A^{-1}\|_2\,\hat{\sigma}^\star_{r_4}+1\big)L_{\mathfrak{p}_\star}\epsilon_{r_4};\end{aligned}$$

˜ ˜第一个不等式源于 $\mathrm{k}\Lambda_{(\mathfrak{p})}E_{\mathfrak{p}}\mathrm{k}\le\mathrm{k}\Lambda_{(\mathfrak{p})}\mathrm{k}_2\epsilon_r$，以及 $\mathrm{k}\Lambda_{(\mathfrak{p})}\mathrm{k}_2\le\mathrm{k}A\mathrm{k}_2\,\mathrm{k}\mathrm{R}_{\mathfrak{p}}\mathrm{k}$（回忆 (E.90)）；对于最后一个不等式，我们还用到了 $\mathrm{k}B_{\mathfrak{p}}\mathrm{k}_2\le\mathrm{k}A^{-1}\mathrm{k}_2\mathrm{k}\mathrm{R}^{-1}_{\mathfrak{p}}\mathrm{k}_2\le\mathrm{k}A^{-1}\mathrm{k}_2\hat{}\sigma^\star_r$（参见 (73)）。结合 (E.111) 和 (E.109)（同时回忆 (E.102)），以上结果意味着

$$d\big(\mathfrak{a}_\star,\mathfrak{I}(\mathfrak{p})\big)\le K_2\Big(\sqrt{2}\eta(r_4)+\big(\|A^{-1}\|_2\,\hat{\sigma}^\star_{r_4}+1\big)\epsilon_{r_4}\Big)\le\varepsilon$$

因此，由于 $\mathfrak{a}_\star\in\mathcal{A}_\star$ 是任意的，$\sup_{\mathfrak{a}\in\hat{\mathfrak{I}}(\mathfrak{p}_\star)}d(\mathfrak{a},\mathfrak{I}(\mathfrak{p}))\le\varepsilon$。对于任意 $\mathfrak{p}\in\mathbb{B}_{r_4}(\mathfrak{p}_\star)$，$\sup_{\mathfrak{b}\in\hat{\mathfrak{I}}(\mathfrak{p})}d(\mathfrak{I}(\mathfrak{p}_\star)\;\mathfrak{b})\le\varepsilon$ 也可类似地推出。事实上，取任意一个 $\mathfrak{b}\in\mathfrak{I}(\mathfrak{p})$，例如

$\mathfrak{b} \equiv m_{\mathfrak{b}}(\mathrm{I}+E_{\mathfrak{b}})A^{-1} \in \tilde{\Xi}_{\mathcal{R}_{\mathfrak{p}}}$ (参见 (E.107))，其中 $\mathbf{k}E_{\mathfrak{b}}\mathbf{k} \leq \epsilon_{r_4}$，且 $m_{\mathfrak{b}} \equiv P_\sigma \tilde{\Lambda}_{(\mathfrak{b})} \in \mathbf{M}_d$ 对于 $P_\sigma \in \mathbf{P}_d^{\pm}$ 和 $\tilde{\Lambda}_{(\mathfrak{b})} \equiv \mathbf{ddiag}[m_{\mathfrak{b}}^1, \dots, m_{\mathfrak{b}}^d] > 0$ 成立。如上所述，单位行再次意味着 (E.112)（但将 $(\theta_{\mathfrak{p}}^\star, E_p)$ 替换为 $(\mathfrak{b}, E_{\mathfrak{b}})$），因此对于 $\tilde{\mathfrak{a}} := P_\sigma \Lambda_{\mathfrak{p}_\star}^{-1} A^{-1} \in \mathfrak{I}(\mathfrak{p}_\star)$，我们可以如上所述得到

$$d(\mathfrak{I}(\mathfrak{p}_\star), \mathfrak{b}) \leq \|\tilde{\mathfrak{a}} - \mathfrak{b}\| \leq \|A^{-1}\|_2 \big\|\Lambda_{\mathfrak{p}_\star}^{-1} - \tilde{\Lambda}_{(\mathfrak{b})}(\mathrm{I}+E_{\mathfrak{b}})\big\| \leq \varepsilon.$$

因此上确界 $_{\mathfrak{b}\in\mathfrak{I}(\mathfrak{p})} d(\mathfrak{I}(\mathfrak{p}_\star), \mathfrak{b}) \leq \varepsilon$（由于 $\mathfrak{b} \in \hat{}\mathfrak{I}(\mathfrak{p})$ 是任意的），从而得到（E.110），如所需。 □

注 E.1。 (i) 为简化常数，我们可以通过下式利用（D.46）控制（E.108）：

$$\eta(r) \leq \kappa_2^2(A) L_{\mathfrak{p}_\star} \frac{\|A\|_2}{\check{\gamma}_\star \sigma_\star} \sup_{\mathfrak{a}\in\mathfrak{B}_{\mathfrak{q}_r}} \big\|\mathfrak{a} - [\mu_\star]_0\big\| \leq \check{K}_1 r^\alpha$$

对于与 $\mathfrak{p}_\star$ 相关的常数 $\check{} K_1 := \kappa_2^2(A) L_{\mathfrak{p}_\star} \mathbf{k}A\mathbf{k} \check{\gamma}_\star \sigma_\star^{\ 2} \hat{} K$；~~另令~~ $\check{} K_2 := \mathbf{k}A^{-1}\mathbf{k}_2 \hat{} \sigma_{r_3} + 1$。相应地，对于任意给定的 $\varepsilon > 0$，我们可以不采用（E.109），而改为选择 $\hat{} r = \hat{} r(\varepsilon; \mathfrak{p}_\star)$ 为

$$\hat{r} := \sup\big\{0 \leq r \leq r_3 \,\big|\, \max_{s\in[0,r)}(\sqrt{2}\check{K}_1 s^\alpha + \check{K}_2 \epsilon_s) \leq \varepsilon/(\kappa_2(A) L_{\mathfrak{p}_\star})\big\} \in (0, r_4]$$

以确保连续性关系（E.96）$|_{r=\hat{r}}$ 成立。

(ii) 如果对于任意固定的 $A \in \mathbf{GL}_d$，我们关注映射在 $\mathscr{I}$ 处的连续性

$$\mathfrak{I}_A : \big(\mathfrak{D}, \delta\big) \ni \mu \longmapsto \hat{\Phi}(A \cdot \mu) \in (F, \mathfrak{d}), \tag{E.113}$$

即，在 $E' = \mathfrak{D} \times \mathbf{GL}_d$ 的分段可识别性条件 $\mathscr{I}_A := \mathscr{I} \cap (\mathfrak{D} \times \{A\})$ 下，解映射 $\mathfrak{I}(\cdot\ A)$ 的（部分）连续性，那么定理 5.12 的上述证明表明，无需对 $\mathfrak{D}$ 施加任何额外的可积性假设，该连续性仍然成立（参见（39））。更准确地说，根据注 4.14，可以通过简化的量化方式表述（E.113）的连续性：在可识别性域 $E_A := \mathfrak{D} \times \{A\}$ 上进行操作，可以将上述 r_0 替换为 ρ_0/ρ_1，然后从经调整的前提开始，沿用定理 5.12 证明中的论证

$$\mu_\chi \equiv A_* \mu \quad \text{with} \quad \delta_\perp(\mu) \leq L_r \sqrt{r} \quad \text{and} \quad \big\|[\mu]_{\mathfrak{c}} - [\mu_\star]_{\mathfrak{c}}\big\|_{\mathcal{V}} \leq \hat{L} r, \tag{E.97'}$$

对于任意 $(\mu,\ A) \in E_A$ 满足 $\delta(\mu_\star,\ \mu) < r \leq \rho_0/\rho_1$ 的情况，由注 4.14 (66) 可知成立。根据 (E.97')，上述定理 5.12 的证明可原样进行，只需将其替换为

$$\text{the (global}^{42}\text{) replacements: } K' \leftrightarrow L_r,\ \hat{K} \leftrightarrow \hat{L},\ \alpha \leftrightarrow 1,$$

由此得到基于 (40) 中独立常数构造的连续模得到放宽。 ◆

F. Proof of Proposition 6.6 (Cases $i = 2, 3$)

证明。现在我们来证明 (109)，针对乘性噪声的情形 $i = 2,\ 3$。为此，我们首先考虑情形 $i = 2$，即三元组 $(\check{X}^{(2)},\ A,\ S^\eta)$ 其中 $\check{X}^{(2)}$ 如 (106) 所示，且 S^η 如 (105) 所示。

那么，由于勒贝格–斯蒂尔杰斯积分满足乘积法则 $\int_a^b \varphi\, \mathrm{d}(g_1 g_2) = \int_a^b \varphi g_1\, \mathrm{d}g_2 + \int_a^b \varphi g_2\, \mathrm{d}g_1$，对于任意 $\varphi,\ g_1,\ g_2 \in \mathcal{C}_1^1$ 以及每个 $0 \leq a \leq b \leq 1$，我们有

$$\begin{aligned}
\langle S^\eta\rangle_{ij} &= \langle \eta^i, \eta^j \,|\, S^i, S^j\rangle + \langle S^i, \eta^j \,|\, \eta^i, S^j\rangle + \langle \eta^i, S^j \,|\, S^i, \eta^j\rangle + \langle S^i, S^j \,|\, \eta^i, \eta^j\rangle \quad \text{and} \\
\langle S^\eta\rangle_{ijk} &= \langle \eta^i, \eta^j, \eta^k \,|\, S^i, S^j, S^k\rangle + \langle \eta^i, \eta^j, S^k \,|\, S^i, S^j, \eta^k\rangle + \langle S^i, \eta^j, \eta^k \,|\, \eta^i, S^j, S^k\rangle \\
&\quad + \langle S^i, \eta^j, S^k \,|\, \eta^i, S^j, \eta^k\rangle + \langle \eta^i, S^j, \eta^k \,|\, S^i, \eta^j, S^k\rangle + \langle \eta^i, S^j, S^k \,|\, S^i, \eta^j, \eta^k\rangle \\
&\quad + \langle S^i, S^j, \eta^k \,|\, \eta^i, \eta^j, S^k\rangle + \langle S^i, S^j, S^k \,|\, \eta^i, \eta^j, \eta^k\rangle
\end{aligned} \tag{F.114}$$

[42] 因此 $r_1 = r_1(\rho_0,\ \varrho_0,\ \hat{} K,\ \alpha)$ 被替换为 $r_1(\rho_0,\ \varrho_0,\ \hat{} L,\ 1)$ 等。

其中，对于 $Y,\ \tilde{Y},\ \hat{Y},\ \check{Y},\ Z,\ \tilde{Z} \in \{S,\ \eta\}$，我们记$\langle Y^i,\ \tilde{Y}^j \,|\, \hat{Y}^i,\ \check{Y}^j\rangle := \mathbb{E}\big[\int_0^1\int_0^t \boldsymbol{Y}_s^i\tilde{\boldsymbol{Y}}_t^j \,\mathrm{d}\,\hat{\boldsymbol{Y}}_s^i\mathrm{d}\,\check{\boldsymbol{Y}}_t^j\big]$ 以及

$$\langle Y^i, \tilde{Y}^j, Z^k \,|\, \hat{Y}^i, \check{Y}^j, \tilde{Z}^k\rangle := \mathbb{E}\left[\int_0^1\int_0^t\int_0^s \boldsymbol{Y}_r^i\tilde{\boldsymbol{Y}}_s^j\boldsymbol{Z}_t^k \,\mathrm{d}\hat{\boldsymbol{Y}}_r^i\mathrm{d}\check{\boldsymbol{Y}}_s^j\mathrm{d}\tilde{\boldsymbol{Z}}_t^k\right]. \tag{F.115}$$

以与前文类似的方式逐项计算统计量（F.114）的加数，我们得到

$\langle \eta^i,\ \eta^j \,|\, S^i,\ S^j\rangle = \int_{\Delta_2}\mathbb{E}[\boldsymbol{\eta}_s^i\boldsymbol{\eta}_t^j]\mathbb{E}[\dot{\boldsymbol{S}}_s^i\dot{\boldsymbol{S}}_t^j]\,\mathrm{d}^2\boldsymbol{t}$ 以及$\langle S^i,\ \eta^j \,|\, \eta^i,\ S^j\rangle = \int_{\Delta_2}\mathbb{E}[\dot{\boldsymbol{\eta}}_s^i\boldsymbol{\eta}_t^j]\mathbb{E}[\boldsymbol{S}_s^i\dot{\boldsymbol{S}}_t^j]$

$\mathrm{d}^2\boldsymbol{t}$等等，因此，结合 $\boldsymbol{S}$ 是一个具有 $\mathbb{E}[\boldsymbol{S}]=0$的均值平稳 IC源这一事实，可知每当 $i\neq j$时，$\mathbb{E}[\boldsymbol{S}_s^i\dot{\boldsymbol{S}}_t^j] = \mathbb{E}[\boldsymbol{S}_s^i]\mathbb{E}[\dot{\boldsymbol{S}}_t^j] = 0 = \mathbb{E}[\boldsymbol{S}_s^i\boldsymbol{S}_t^j]$，由此得到

$$\langle S^\eta\rangle_{ij} = \left[\int_{\Delta_2}\phi^i_{(S,\eta)}(\boldsymbol{t})\,\mathrm{d}^2\boldsymbol{t}\right]\cdot\delta_{ij} \qquad \text{for each}\quad ij\in[d]^\star_2, \tag{F.116}$$

其中 $\phi^i_{(S,\eta)}(s,\ t) := \mathbb{E}[\boldsymbol{\eta}_s^i\boldsymbol{\eta}_t^i]\mathbb{E}[\dot{\boldsymbol{S}}_s^i\dot{\boldsymbol{S}}_t^i]+\mathbb{E}[\dot{\boldsymbol{\eta}}_s^i\boldsymbol{\eta}_t^i]\mathbb{E}[\boldsymbol{S}_s^i\dot{\boldsymbol{S}}_t^i]+\mathbb{E}[\boldsymbol{\eta}_s^i\dot{\boldsymbol{\eta}}_t^i]\mathbb{E}[\dot{\boldsymbol{S}}_s^i\boldsymbol{S}_t^i]+\mathbb{E}[\dot{\boldsymbol{\eta}}_s^i\dot{\boldsymbol{\eta}}_t^i]\mathbb{E}[\boldsymbol{S}_s^i\boldsymbol{S}_t^i]$。三阶统计量（F.115）的处理类似，因此得到

$$\langle S^\eta\rangle_{ijk} = \left[\int_{\Delta_3}\psi^i_{(S,\eta)}(\boldsymbol{t})\,\mathrm{d}^3\boldsymbol{t}\right]\cdot\delta_{ijk} \qquad \text{for each}\quad ijk\in[d]^\star_3, \tag{F.117}$$

其中，此时被积函数为

$$\begin{aligned}\psi^i_{(S,\eta)}(r,s,t) :=\ & \mathbb{E}[\boldsymbol{\eta}_r^i\boldsymbol{\eta}_s^i\boldsymbol{\eta}_t^i]\mathbb{E}[\dot{\boldsymbol{S}}_r^i\dot{\boldsymbol{S}}_s^i\dot{\boldsymbol{S}}_t^i] + \mathbb{E}[\boldsymbol{\eta}_r^i\boldsymbol{\eta}_s^i\dot{\boldsymbol{\eta}}_t^i]\mathbb{E}[\dot{\boldsymbol{S}}_r^i\dot{\boldsymbol{S}}_s^i\boldsymbol{S}_t^i] + \mathbb{E}[\dot{\boldsymbol{\eta}}_s^i\boldsymbol{\eta}_s^i\boldsymbol{\eta}_t^i]\mathbb{E}[\boldsymbol{S}_r^i\dot{\boldsymbol{S}}_s^i\dot{\boldsymbol{S}}_t^i]\\ &+ \mathbb{E}[\dot{\boldsymbol{\eta}}_r^i\eta_s^i\dot{\boldsymbol{\eta}}_t^i]\mathbb{E}[\boldsymbol{S}_r^i\dot{\boldsymbol{S}}_s^i\boldsymbol{S}_t^i] + \mathbb{E}[\boldsymbol{\eta}_r^i\dot{\boldsymbol{\eta}}_s^i\boldsymbol{\eta}_t^i]\mathbb{E}[\dot{\boldsymbol{S}}_r^i\boldsymbol{S}_s^i\dot{\boldsymbol{S}}_t^i] + \mathbb{E}[\boldsymbol{\eta}_r^i\dot{\boldsymbol{\eta}}_s^i\dot{\boldsymbol{\eta}}_t^i]\mathbb{E}[\dot{\boldsymbol{S}}_r^i\boldsymbol{S}_s^i\boldsymbol{S}_t^i]\\ &+ \mathbb{E}[\dot{\boldsymbol{\eta}}_r^i\dot{\boldsymbol{\eta}}_s^i\boldsymbol{\eta}_t^i]\mathbb{E}[\boldsymbol{S}_r^i\boldsymbol{S}_s^i\dot{\boldsymbol{S}}_t^i] + \mathbb{E}[\dot{\boldsymbol{\eta}}_r^i\dot{\boldsymbol{\eta}}_s^i\dot{\boldsymbol{\eta}}_t^i]\mathbb{E}[\boldsymbol{S}_r^i\boldsymbol{S}_s^i\boldsymbol{S}_t^i].\end{aligned}$$

按上述显示顺序枚举构成 $\phi^i_{(S,\eta)}$ 的各项 $\phi^i_j \equiv \phi^i_{\ell,j}\cdot\phi^i_{r,j}$ ——以及构成 $\psi^i_{(S,\eta)}$ 的各项 $\psi^i_j \equiv \psi^i_{\ell,j}\cdot\psi^i_{r,j}$ ——（例如 $\phi^i_2 = \mathbb{E}[\dot{\boldsymbol{\eta}}_s^i\boldsymbol{\eta}_t^i]\mathbb{E}[\boldsymbol{S}_s^i\dot{\boldsymbol{S}}_t^i]$ 和 $(\phi^i_{\ell,2},\phi^i_{r,2}) = \big(\mathbb{E}[\dot{\boldsymbol{\eta}}_s^i\boldsymbol{\eta}_t^i],\mathbb{E}[\boldsymbol{S}_s^i\dot{\boldsymbol{S}}_t^i]\big)$ 等），由 (F.116) 和 Hölder不等式可得，对每个 $i\in[d]$,

$$\begin{aligned}\big|\langle S^\eta\rangle_{ii} - \langle S\rangle_{ii}\big| &= \left|\int_{\Delta_2}\big(\phi^i_{\ell,1}(\boldsymbol{t})-1\big)\phi^i_{r,1}(\boldsymbol{t}) + \textstyle\sum_{j=2}^4\phi^i_j(\boldsymbol{t})\,\mathrm{d}^2\boldsymbol{t}\right|\\ &\le \left[\int_{\Delta_2}\big(\phi^i_{\ell,1}(\boldsymbol{t})-1\big)^2\mathrm{d}^2\boldsymbol{t}\right]^{1/2}\cdot\tilde{c}_{i,1} + \sum_{j=2}^4\left[\int_{\Delta_2}\big(\phi^i_{\ell,j}(\boldsymbol{t})\big)^2\mathrm{d}^2\boldsymbol{t}\right]^{1/2}\cdot\tilde{c}_{i,j}\end{aligned} \tag{F.118}$$

对于依赖于源的常数 $\tilde{c}_{i,\ j} := \big(\int_{\Delta_2}\phi^i_{r,\ j}(\boldsymbol{t})^2\mathrm{d}^2\boldsymbol{t}\big)^{1/2}$。同样，对每个 $i\in[d]$,

$$\big|\langle S^\eta\rangle_{iii} - \langle S\rangle_{iii}\big| \le \left[\int_{\Delta_3}\big(\psi^i_{\ell,1}(\boldsymbol{t})-1\big)^2\mathrm{d}^3\boldsymbol{t}\right]^{1/2}\cdot\check{c}_{i,1} + \sum_{j=2}^8\left[\int_{\Delta_3}\big(\psi^i_{\ell,j}(\boldsymbol{t})\big)^2\mathrm{d}^3\boldsymbol{t}\right]^{1/2}\cdot\check{c}_{i,j} \tag{F.119}$$

对于依赖于源的常数 $\check{c}_{i,j} := \big(\int_{\Delta_3}\psi^i_{r,j}(\boldsymbol{t})^2\mathrm{d}^3\boldsymbol{t}\big)^{1/2}$。为方便起见，我们将式 (F.118) 中的表达式简记为 $\mathrm{h}\tilde{b}_i,\tilde{c}_i\mathrm{i} \equiv \sum_{j=1}^4\tilde{b}_{i,j}\cdot\tilde{c}_{i,j}$，其中 $\tilde{b}_i := (\tilde{b}_{i,1},\tilde{b}_{i,2},\tilde{b}_{i,3},\tilde{b}_{i,4})$，且 $\tilde{c}_i := (\tilde{c}_{i,1},\dots,\tilde{c}_{i,4})$，以及 $\tilde{b}_{i,j} := \big(\int_{\Delta_2}(\phi^i_{\ell,j}(\boldsymbol{t}) - \delta_{1j})^2\,\mathrm{d}^2\boldsymbol{t}\big)^{1/2}$；同样，将 (F.119) 中的右端简记为 $\mathrm{h}\check{b}_i,\check{c}_i\mathrm{i} \equiv \sum_{j=1}^8\check{b}_{i,j}\cdot\check{c}_{i,j}$，其中 $\check{b}_i := (\check{b}_{i,1},\dots,\check{b}_{i,8})$，且$\check{c}_i := (\check{c}_{i,1},\dots,\check{c}_{i,8})$，以及

$\check{b}_{i,j} := \big(\int_{\Delta_3} (\psi^i_{\ell,j}(\boldsymbol{t}) - \delta_{1j})^2 \,\mathrm{d}^3\boldsymbol{t}\big)^{1/2}$。由 (F.116)、(F.117)、(F.118)、(F.119) 以及 (44),

$$\begin{aligned}
\delta(S,S^\eta)^2 &= \big\|N_S^{-1}\big([S^\eta]_0 - [S]_0\big)N_S^{-1}\big\|^2 + \sum_{\nu=1}^d \langle S\rangle_{\nu\nu}^{-1}\big\|N_S^{-1}\big([S^\eta]_\nu - [S]_\nu\big)N_S^{-1}\big\|^2 \\
&= \sum_{i=1}^d \big(\langle S\rangle_{ii}^{-1}\big|\langle S^\eta\rangle_{ii} - \langle S\rangle_{ii}\big|\big)^2 + \sum_{\nu=1}^d \langle S\rangle_{\nu\nu}^{-1}\big(\langle S\rangle_{\nu\nu}^{-1}\big|\langle S^\eta\rangle_{\nu\nu\nu} - \langle S\rangle_{\nu\nu\nu}\big|\big)^2 \\
&\le \sum_{i=1}^d \gamma_i^{-2}\big(\langle \tilde{b}_i, \tilde{c}_i\rangle^2 + \gamma_i^{-1}\langle \check{b}_i, \check{c}_i\rangle^2\big) =: \beta_2(S,\eta), \qquad \text{for} \quad \gamma_i := \langle S\rangle_{ii}.
\end{aligned}$$

由于 $\beta(S,\eta) \le \beta_\varepsilon^2$ 根据关于 η 的假设，我们得到噪声情形下的（109） $i=2$，如所需。

情形 $i=3$ 几乎可以完全类似地处理：根据乘积法则，统计量$\langle X^\eta\rangle_{ij}$ 和 $\langle X^\eta\rangle_{ijk}$ 也具有（F.114）的形式，只是将 S 替换为 X。因此，依照上述相同的论证，我们可以得出，对于每个 $ijk \in [d]^\star_3$,

$$\begin{aligned}
\langle X^\eta\rangle_{ij} &= \langle \eta^i, \eta^j \,\big|\, X^i, X^j\rangle + \langle X^i, X^j \,\big|\, \eta^i, \eta^j\rangle \quad \text{and} \qquad \text{(F.120)}\\
\langle X^\eta\rangle_{ijk} &= \langle \eta^i, \eta^j, \eta^k \,\big|\, X^i, X^j, X^k\rangle + \langle X^i, \eta^j, X^k \,\big|\, \eta^i, X^j, \eta^k\rangle + \langle \eta^i, X^j, X^k \,\big|\, X^i, \eta^j, \eta^k\rangle \\
&\quad + \langle X^i, X^j, \eta^k \,\big|\, \eta^i, \eta^j, X^k\rangle + \langle X^i, X^j, X^k \,\big|\, \eta^i, \eta^j, \eta^k\rangle
\end{aligned}$$

where 我们利用了 η 是独立同分布且中心化的、与 X 独立，并且均值平稳。此外，利用 $\mathbb{E}[\eta] \equiv \mathrm{I}$,

$$\big\|[X^\eta]_0 - [X]_0\big\|^2 = \sum_{i=1}^d \big|\langle \eta^i, \eta^i \,|\, X^i, X^i\rangle - \langle X\rangle_{ii} + \langle X^i, X^i \,|\, \eta^i, \eta^i\rangle\big|^2 = \sum_{i=1}^d \bigg|\int_{\Delta_2} \xi_i(\boldsymbol{t})\,\mathrm{d}^2\boldsymbol{t}\bigg|^2,$$

$$\big\|[X^\eta]_\nu - [X]_\nu\big\|^2 = \tilde{\Delta}^2_{\nu\nu\nu} + \sum\nolimits_{i\in[d]\setminus\{\nu\}}\big(\tilde{\Delta}^2_{i\nu\nu} + \tilde{\Delta}^2_{\nu i\nu}\big) = \bigg|\int_{\Delta_3} \Xi_\nu(\boldsymbol{t})\,\mathrm{d}^3\boldsymbol{t}\bigg|^2 + \sum_{i\neq\nu}\big(\tilde{\Xi}^{(1)}_{i,\nu} + \tilde{\Xi}^{(2)}_{i,\nu}\big)$$

对于 $\nu = 1,\dots,d$和 $\tilde{\Delta}_{ijk} := \big|\langle X^\eta\rangle_{ijk} - \langle X\rangle_{ijk}\big|$ 以及由（F.120）导出的量

$$\begin{aligned}
\xi_i(s,t) &:= \big(\mathbb{E}[\boldsymbol{\eta}^i_s\boldsymbol{\eta}^i_t] - 1\big)\mathbb{E}[\dot{\boldsymbol{X}}^i_s\dot{\boldsymbol{X}}^i_t] + \mathbb{E}[\dot{\boldsymbol{\eta}}^i_s\dot{\boldsymbol{\eta}}^i_t]\mathbb{E}[\boldsymbol{X}^i_s\boldsymbol{X}^i_t] \quad \text{and} \\
\Xi_\nu(r,s,t) &:= \big(\mathbb{E}[\boldsymbol{\eta}^\nu_r\boldsymbol{\eta}^\nu_s\boldsymbol{\eta}^\nu_t] - 1\big)\mathbb{E}[\dot{\boldsymbol{X}}^\nu_r\dot{\boldsymbol{X}}^\nu_s\dot{\boldsymbol{X}}^\nu_t] + \mathbb{E}[\dot{\boldsymbol{\eta}}^\nu_r\dot{\boldsymbol{\eta}}^\nu_s\dot{\boldsymbol{\eta}}^\nu_t]\mathbb{E}[\boldsymbol{X}^\nu_r\boldsymbol{X}^\nu_s\boldsymbol{X}^\nu_t] + \mathcal{R}_\nu
\end{aligned}$$

对于 R_ν:

$= \mathbb{E}[\dot{\boldsymbol{\eta}}^\nu_r\boldsymbol{\eta}^\nu_s\dot{\boldsymbol{\eta}}^\nu_t]\mathbb{E}[\boldsymbol{X}^\nu_r\dot{\boldsymbol{X}}^\nu_s\boldsymbol{X}^\nu_t] + \mathbb{E}[\boldsymbol{\eta}^\nu_r\dot{\boldsymbol{\eta}}^\nu_s\dot{\boldsymbol{\eta}}^\nu_t]\mathbb{E}[\dot{\boldsymbol{X}}^\nu_r\boldsymbol{X}^\nu_s\boldsymbol{X}^\nu_t] + \mathbb{E}[\dot{\boldsymbol{\eta}}^\nu_r\dot{\boldsymbol{\eta}}^\nu_s\boldsymbol{\eta}^\nu_t]\mathbb{E}[\boldsymbol{X}^\nu_r\boldsymbol{X}^\nu_s\dot{\boldsymbol{X}}^\nu_t]$ 和$\tilde{\Xi}^{(1)}_{i\,\nu,} := \big|\int_{\Delta_3}(\mathbb{E}[\boldsymbol{\eta}^\nu_s\boldsymbol{\eta}^\nu_t] - 1)$ $\mathbb{E}[\dot{\boldsymbol{X}}^i_r\dot{\boldsymbol{X}}^\nu_s\dot{\boldsymbol{X}}^\nu_t] + \mathbb{E}[\dot{\boldsymbol{\eta}}^\nu_s\boldsymbol{\eta}^\nu_t]\mathbb{E}[\dot{\boldsymbol{X}}^i_r\boldsymbol{X}^\nu_s\boldsymbol{X}^\nu_t]\,\mathrm{d}^3\boldsymbol{t}\big|^2$，其中 $\tilde{\Xi}^{(2)}_{i\,\nu,}$ 也以类似方式定义。

简记 $\hat{}\,\xi_i := \big|\int_{\Delta_2}\xi_i(\boldsymbol{t})\,\mathrm{d}^2\boldsymbol{t}\big|^2$ 和 $\hat{}\,\Xi_\nu := \big|\int_{\Delta_3}\Xi_\nu(\boldsymbol{t})\,\mathrm{d}^3\boldsymbol{t}\big|^2$，由此得到该控制

$$\|[X^\eta]_{\mathfrak{c}} - [X]_{\mathfrak{c}}\|^2 = \sum_{i=1}^d\big(\hat{\xi}_i + \hat{\Xi}_i\big)^2 + \sum_{\nu=1}^d\sum_{i\neq\nu}\big(\tilde{\Xi}^{(1)}_{i,\nu} + \tilde{\Xi}^{(2)}_{i,\nu}\big) =: \beta_3(X,\eta). \qquad \text{(F.121)}$$

现在，利用 A^{-1}在$\mathcal{V}$上的诱导张量作用的连续性，参见命题 64 [的前半部分]，存在一个可显式计算的阈值 $\gamma_\varepsilon \equiv \gamma_\varepsilon(\beta_\varepsilon,\ A) > 0$ 足够小，使得

$$\beta_3(X,\eta) \le \gamma_\varepsilon \quad \text{implies} \quad \delta(S, A^{-1}X^\eta) \le \beta_\varepsilon.$$

这就证明了109）在 $i=3$的情形下成立，正如所需。 □