arXiv:2603.01737v1 [eess.SP] 2 Mar 2026

任意噪声分布下的弱信号检测

J. Zschetzsche,[1,2,*] M. Weimar,[3,*] O. Lang,[1,2] S. Schuster,[4] A. Haberl,[4] S. Schertler,[1,2] B. Lehner,[5] J. Reisinger,[4] M. Huemer,[1] 以及 S. Rotter[3]

[1] *Institute of Signal Processing, Johannes Kepler University Linz, Austria*

[2] *Christian Doppler Laboratory for Steel Industry Signal Processing and Machine Learning, Austria*

[3] *Institute for Theoretical Physics, Vienna University of Technology, Austria*

[4] *voestalpine Stahl GmbH, Linz, Austria* [5] *Silicon Austria Labs GmbH, Linz, Austria*

在复杂的非高斯噪声中检测微弱信号，是科学与工程领域的一项根本性挑战，应用范围涵盖雷达系统、通信、工业监测和引力波探测。Rao检测器是这一领域的关键概念；随着测量值数量增加，它能够实现渐近最优性能，但需要精确了解数据的统计特性，而这往往依赖于简化的噪声模型。我们提出了一种混合框架，将轻量级神经网络与Rao检测框架相结合，以解决这一局限。该神经网络在纯噪声数据上训练，学习最优的多变量非线性，将含噪数据转换为更易检测信号的形式。新引入的LRao检测器随后充分提取信号信息，即使在严苛的噪声条件下也能实现渐近最优性能。通过在模拟磁传感器数据和真实磁传感器数据上进行验证，我们的方法显著优于传统方法。通过结合数据驱动技术与基于模型的信号处理，该方法为各种应用中的信号检测提供了一种稳健且可解释的解决方案。

I. 引言

当我们试图在嘈杂的环境中（例如拥挤的房间里）听清一段对话时，大脑会完成一项令人惊叹的信号检测任务：从复杂的声学背景中分离出微弱而有结构的信号。这个直观的例子展现了一项更广泛挑战的本质，这项挑战遍及科学与工程领域：检测嵌入在复杂且往往呈非高斯分布的噪声环境中的弱信号。从雷达 [1–3] 到纳米机械传感器 [4]和引力波检测 [5], , 从噪声数据中辨识微弱信号的能力是许多领域取得进展的基础。然而，与能够根据声学条件变化调整内部滤波器的人类听觉系统不同，大多数传统检测器依赖于对底层噪声统计特性的简化或理想化假设。

克服这一局限是现代检测理论的核心。常见策略是设计实验和数据采集流程，使信噪比（SNR）最大化，从而提升后续处理中的检测性能。然而，在许多情况下，决定信噪比的物理参数并不能自由选择。当信号被复杂的非高斯噪声掩盖时，有效的后处理也会变得极具挑战性。例如，在雷达应用 [6、海军测量8], [9, 10]以及复杂光散射 [11]中，声波或电磁波会受到不可预测的散射事件干扰。其他情形还包括振动分析 [12], 、采用量化数据的无线传感器网络 [13–15], 以及引力波测量 [16–19]。非高斯噪声通常表现为传感器数据中的信号含有异常值。传统做法是去除这些异常值，这样虽然能简化统计特性，但同时也会丢失有关信号的信息。

所谓的 Rao 得分检验或 Rao检测器 [20], 提供了一种精确且广泛应用的表达式，用于实现弱信号检测的渐近最优数据处理。Rao检测器可解释为[21, 22]对测量数据施加一种通常为非线性的变换，使变换后的数据对所关注信号具有最大敏感度。这种非线性就是所谓的得分函数或 Fisher 得分，它是对数似然关于信号参数的梯度，因此表明了对数似然对参数变化的敏感度。

Rao检测器在独立同分布（IID）噪声或高斯噪声过程下已得到广泛确立，但当概率分布未知时，其实际适用性便会受限。若只能获得有限的数据集，按照频率学派统计的理念，可以采用参数方法或非参数方法估计相应的概率密度函数（PDF），然后直接将其代入Rao检测器的解析表达式。参数方法的问题在于，通常要求分布模型非常准确。另一方面，非参数方法（例如对数据拟合直方图）无需模型，但会受到维数灾难的制约 [23]。鉴于这些问题，通常会假设测量值是独立同分布的，从而得到一维概率密度函数和一个标量非线

* 这些作者对本项工作的贡献相同。代码和数据见: https://github.com/jonaslindenberger/LRao-detector

性，用于表征非高斯性 [21, 22, 24]。为证明独立性假设合理，可以通过白化变换对数据进行预处理，而白化变换通常限于线性变换。然而，这种方法忽略了非线性依赖，而非线性依赖可能包含大量有关目标信号的信息。一个简单的例子是：信号受到相关噪声污染，同时还存在一些不呈现这种相关性的离群值，例如由突发干扰或传感器暂时性故障造成的离群值 [25–27]。

在本工作中，我们通过引入一种方法克服了所有这些困难：该方法以数据驱动方式，在复杂噪声条件下根据数据执行检测与估计，如图 1 所示。我们的核心思想是将测量数据输入神经网络，通过最小化合适的代价函数，提升置于网络之后的检测器的性能，直至其收敛到渐近最优性能。该方法还具有可解释性优势，因为代价函数以直观且具有理论依据的方式与检测器的性能直接相关。

我们在弱信号检测领域的一个问题上展示了新方法的优势。该问题是在近似高斯环境噪声和非高斯尖峰地磁噪声存在的情况下，从磁传感器数据中检测弱周期信号。我们设计了一个小型卷积神经网络，仅使用噪声数据（不含任何信号）进行训练，并表明我们方法的检测性能大幅超越标准检测器，收敛到渐近最优性能。我们的神经网络学会通过将重尾噪声分布变换为短尾分布来有效处理离群值，如图 1 所示。噪声特性从数据中学习，而信号知识则以基于模型的方式纳入。

II. 局部参数估计与测试理论

我们首先回顾经典的 Rao 检测统计量。噪声测量数据通常建模为一个随机变量 $\mathbf{x} \in \mathbb{R}^N$，其服从由概率密度函数 PDF $\mathbf{x} \sim p(\mathbf{x};\boldsymbol{\theta})$ 确定的概率分布，其中 $\boldsymbol{\theta} \in \mathbb{R}^l$ 是由 l 个未知参数组成的向量。相应的对数似然函数定义为 $L(\mathbf{x}\ \boldsymbol{\theta}) = \log\ p(\mathbf{x};\boldsymbol{\theta})$。为便于阅读，以下公式中省略其自变量。用于衡量数据中关于感兴趣参数 $\boldsymbol{\theta}$ 的信息量的量，称为费歇尔信息（FI） [28] ，定义为

$$\mathbf{F} = -\mathbb{E}\left[\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^{\top} L\right] \ , \tag{1}$$

其中， $\mathbb{E}[\cdot]$ 表示期望， $\nabla_{\boldsymbol{\theta}} = [\frac{\partial}{\partial[\boldsymbol{\theta}]_1} \cdots \frac{\partial}{\partial[\boldsymbol{\theta}]_l}]$ 是纳布拉算子。作为似然函数的期望局部曲率，费歇尔信息不仅在参数估计中发挥着关键作用，在局部参数测试 [21, 22, 29, 30]中也同样如此。随着我们讨论 Rao 检测器，这一点将变得十分明显，因为该检测器将费歇尔信息纳入其检测统计量。在双侧参数检验中，Rao 检测器在所谓的原假设 $\mathcal{H}_0$ 与备择假设 $\mathcal{H}_1$ 之间作出判决，该检验定义为

$$\mathcal{H}_0 :\ \boldsymbol{\theta} = \boldsymbol{\theta}_0 \tag{2}$$
$$\mathcal{H}_1 :\ \boldsymbol{\theta} \neq \boldsymbol{\theta}_0 \ . \tag{3}$$

假设在 $\mathcal{H}_1$, $\boldsymbol{\theta}_1$ 下的参数向量与 $\mathcal{H}_0$ 下的参考参数向量 $\boldsymbol{\theta}_0$ “接近”。众所周知，Rao 检测统计量为 [20]

$$T_{\mathrm{Rao}}(\mathbf{x}) = (\nabla_{\boldsymbol{\theta}} L_0)\mathbf{F}_0^{-1}(\nabla_{\boldsymbol{\theta}} L_0)^{\top}, \tag{4}$$

其中，下标 0 表示这些表达式是在 $\boldsymbol{\theta}_0$ 处计算的（求导之后）。为了对待检验的假设作出判定，将 T_{Rao} 与一个阈值进行比较，以权衡真正率（TPR）和假正率（FPR）。在 $\mathcal{H}_0$ 下， T_{Rao} 渐近服从中心卡方 χ_l^2 分布。因此，无论 $\mathcal{H}_1$ 下参数向量的未知取值如何，假正率都可以保持在固定值。在 $\mathcal{H}_1$ 下， T_{Rao} 渐近服从非中心卡方 $\chi_l'^2(\lambda_{\mathrm{Rao}})$ 分布，其非中心参数为 $\lambda_{\mathrm{Rao}} = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\top}\mathbf{F}_0(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$。随着 λ_{Rao} 增大， $\chi_l'^2(\lambda_{\mathrm{Rao}})$ 分布向较大值偏移，因此渐近真正率会随着费舍尔信息增大。尽管 Rao 检测器在某些条件下渐近最优 [22], ，但其实际适用性受到严重限制，因为数据的底层概率密度函数通常未知。在这种情况下，通常会对概率密度函数作出适当假设，使 Rao 检测器能够作为一种稳健且实用的检测器发挥作用 [21]。

我们的方法是首先大胆假设数据服从多元正态分布 $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_0)$，其中 $\boldsymbol{\mu}$ 是依赖于参数的均值向量，而 $\boldsymbol{\Sigma}_0$ 是在参考参数 $\boldsymbol{\theta}_0$处计算的协方差。该假设与“等效悲观系统”的概念相关联，因为与该“悲观概率密度函数”相关的费舍尔信息在一大类数据分布 [31–33]中是最低的。由于悲观概率密度函数假设通常并不成立，基于它设计检测器将产生次优结果。然而，与上述传统方法不同，我们现在通过将数据输入神经网络来对数据进行变换。关键在于，我们仍然可以使用变换后的数据的悲观概率密度函数来构造一个（次优的）检测器。然而，其性能通常与原始数据的性能不同。于是我们能够建立一个优化问题：我们选择神经网络的参数，使得对于变换后的数据，悲观概率密度函数假设能够得到统计意义上的最优结果。接下来，我们介绍该性能指标，我们

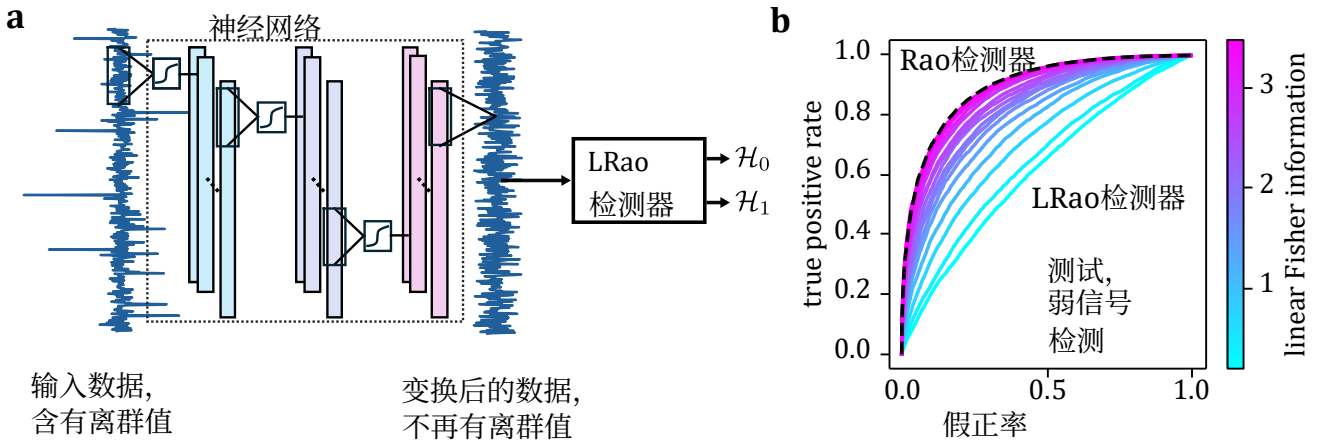


图 1：**方法及其性能示意图。a**，通过训练轻量级神经网络，可以对复杂的非高斯数据进行最优处理。通过最小化本研究提出的代价函数，神经网络学会以最适合后续处理的方式表示数据，从而不再包含任何离群值。结合加性信号模型时，神经网络仅使用纯噪声数据进行训练，而这类数据通常十分充足。**b**，随着代价函数不断减小，所提出的 LRao 检测器应用于变换后的数据时，其性能逐渐提升，直至达到渐近最优 Rao 检测器的水平。与我们的方法不同，后者需要精确了解数据概率密度函数。代价函数并非通过启发式方法惩罚离群值，而是自然地作为线性 Fisher 信息的负值产生。这一量从形式上衡量了变换后数据中易于提取的信息，无需进行进一步的复杂处理。

将其用作调整神经网络权重的优化目标。事实证明，该目标存在一个自然的选择，它由所谓的线性费舍尔信息（LFI） [34–41]决定。

LFI与在悲观概率密度函数假设下计算的FI完全一致，并为与原始概率密度函数相关的FI提供了下界 [31]。在 $\boldsymbol{\theta}_0$ 处求值时，LFI表示为

$$\mathbf{J}_0 = (\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0). \tag{5}$$

其逆矩阵为局部无偏线性估计量的协方差矩阵提供了可达到的下界 [37]。局部最佳线性无偏估计量（LBLUE）的协方差矩阵等于逆LFI，因此达到了该下界，其表达式为

$$\hat{\boldsymbol{\theta}}_{\mathrm{LBLUE}} = \boldsymbol{\theta}_0 + \mathbf{J}_0^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\mathbf{x} - \boldsymbol{\mu}_0). \tag{6}$$

在某些条件下，其渐近分布为 $\mathcal{N}(\boldsymbol{\theta},\ \mathbf{J}_0^{-1})$，如补充信息（SI）A4中所证明。据我们所知，我们首次给出了LBLUE的多变量形式。对于标量参数这一特殊情况，LBLUE最早由 [35]，推导出，也称为局部最优线性估计量，或等价地称为局部最优线性解码器 [36, 37]。

我们现在将 LFI 框架中的局部无偏估计概念扩展到局部参数检验或检测。通过在悲观概率密度函数假设下评估 Rao 检测统计量 (4)，我们提出一种

“悲观 Rao 检测器”，我们称之为线性 Rao（LRao）检测器。如 SI A 5 所示，LRao检测统计量由下式给出

$$T_{\mathrm{LRao}}(\mathbf{x}) = (\mathbf{x} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0)\mathbf{J}_0^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\mathbf{x} - \boldsymbol{\mu}_0). \tag{7}$$

尽管它关于数据是二次的，我们仍将“线性”纳入其名称，以与LFI和LBLUE保持一致，并强调悲观得分函数（ LRao检测器的核心）实际上是线性的，详见SI A3。与Rao检测器类似，在温和的假设下，LRao检测统计量在 $\mathcal{H}_0$ 下渐近服从中心 χ_l^2 分布；在 $\mathcal{H}_1$ 下则服从非中心参数为 $\lambda_{\mathrm{LRao}} = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^\top \mathbf{J}_0 (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$ 的 $\chi_l'^2(\lambda_{\mathrm{LRao}})$ 分布。因此，其真正率会随着LFI的增大而提高；当LFI等于FI时，其真正率与Rao检测器相同。

通过用 LBLUE 和 LFI 表示检测统计量，可以从另一个角度理解 LRao 检测器：

$$T_{\mathrm{LRao}} = (\hat{\boldsymbol{\theta}}_{\mathrm{LBLUE}} - \boldsymbol{\theta}_0)^\top \mathbf{J}_0 (\hat{\boldsymbol{\theta}}_{\mathrm{LBLUE}} - \boldsymbol{\theta}_0). \tag{8}$$

因此，LRao检测器可解释为：先用LBLUE估计参数向量，再用其协方差（即逆LFI）对偏离参考参数的平方偏差进行归一化。由此得到归一化平方距离度量，用于检测参数向量的变化。注意

平方运算之所以必要，是因为式(3)中假设检验的双侧性质。对于单侧参数检验，其中备择假设 $\mathcal{H}_1$ 定义为 $\theta > \theta_0$ （且该参数通常为标量信号强度度量 [21, 24, 42]），则无需平方运算。此时，（归一化的）LBLUE本身即可用作局部最优线性检测器，我们称之为线性局部最大功率（LLMP）检测器，详见SI A 7。如上所述，尽管标量 LBLUE 最早在 [35], 中导出，但据我们所知，LLMP 检测器在检测理论背景下系首次在此提出。

我们现在讨论一种变换，它能使 LBLUE（因此也使 LLMP 检测器）和 LRao 检测器达到渐近最优。非线性变换（例如神经网络执行的变换）可以改变 LFI [43], ，而且可以优化网络，使其输出端的 LFI 最大化。然而，由于 LFI 是 FI 的下界，只要网络能够学习到最优变换，最大化 LFI 实际上就能得到 FI。通过将（5）中的均值和协方差矩阵替换为样本估计值，并使用中心差分商近似导数，可以估计 LFI [44]。对于标量参数，可以直接使用神经网络输出端的估计 LFI 进行优化。然而，对于向量参数，由于需要标量代价函数，因此不能直接使用估计的 LFI。我们发现，估计 LFI 的迹特别有效；如补充信息 A 6 所示，当数据经过得分函数（或得分函数的可逆仿射变换）变换时，该迹达到最大值。如果将这种变换作为预处理步骤应用于数据，那么局部估计和检测就分别转化为可由 LBLUE 和 LRao 检测器渐近最优地解决的问题。因此，通过最大化神经网络输出端估计的 LFI 的迹 $\widehat{\mathbf{J}}_0$ ，或等价地最小化代价函数

$$C = -\mathrm{tr}(\widehat{\mathbf{J}}_0), \tag{9}$$

我们有效地从数据中学习得分函数，从而绕过概率密度估计的需要。

现在，我们给出一个模拟示例，说明使用得分函数变换数据，如何提升利用 LBLUE 和 LRao检测器进行局部参数估计与检测的性能。我们的目标是估计并检测一个未知的加性移位参数，该参数作用于一组来自高斯混合分布的独立同分布噪声样本向量；该混合分布中的高斯分量均值相同，但方差不同。本例不使用神经网络，而是采用现成的最优非线性变换，即得分函数。图2a展示了原始数据分布（红色）以及经得分函数变换后的数据分布（蓝色）。原始分布的尾部被压缩，而混合分布中的两个高斯分量从各自占主导的区域出发，近似线性地映射到同一区间。

在图 2b 中，展示了原始数据和得分变换数据的 LBLUE 均值及第 95 百分位数。应用得分函数后，LBLUE 在局部达到克拉美-罗下界（CRLB）（即费舍尔信息（FI）的倒数），因为 LBLUE 能在局部提取有关移位参数的全部信息，无需再进行其他非线性变换。在图 2c 中，展示了 LRao检测统计量的均值及第 95 百分位数。结合得分函数后，LRao检测器等同于 Rao检测器，并且是渐近最优的。

在此，我们强调，对于弱信号检测任务而言，局部性并非粗略近似，而恰恰是基于以下原因可以预期的结果：首先，当信噪比较小时，即信号较弱时，检测信号最为困难。若信噪比较高，检测则相对简单，因为人们可以满足于次优解。其次，许多具有实际意义的得分函数即使在信号并不弱的情况下，也能对大规模数据记录产生可靠结果。如图 2c 所示，对于较大的偏移，原始数据产生的统计量大于变换后的数据；然而，变换后的数据仍位于阈值之上。用于信号检测任务的实用得分函数通常类似于某种“限幅函数” [21], ，这使得当数据向量大小较大时，Rao 检测统计量会超过合理设定的阈值。值得注意的是，较大的数据向量大小对于可靠的检测性能至关重要，尤其是在低信噪比环境中。

我们强调，我们的LFI优化框架兼容用于局部估计的LBLUE、用于局部双侧检测的LRao检测器，以及用于局部单侧检测的LLMP（归一化局部最佳线性无偏估计量）。在这三种情况下，提高LFI都会改善性能，且每种方法最终都会趋近于各自渐近最优的极限。本文余下部分将聚焦于LRao检测器，用于检测加性平稳噪声中的弱信号。

III. 弱信号检测

现在，我们将所提出的理论应用于弱信号检测，展示如何通过数据驱动策略，直接从测量值中学习最优的多变量非线性。我们采用神经网络对这种非线性进行参数化，并通过最大化估计的LFI的迹来选择其权重。文中还讨论了可加性假设（使得能够使用纯噪声数据进行训练）和平稳性假设（可提高训练过程的效率和稳健性）。详细推导见补充信息 B。

局部参数测试最重要的实际应用之一，是在非高斯噪声中进行弱信号检测 [21, 24, 42, 45]。对于上述双侧参数测试，局部参数向量 $\boldsymbol{\theta}$通常表示待检测信号分量的振幅， $\boldsymbol{\theta}_0 = \mathbf{0}$ 则表示局部工作

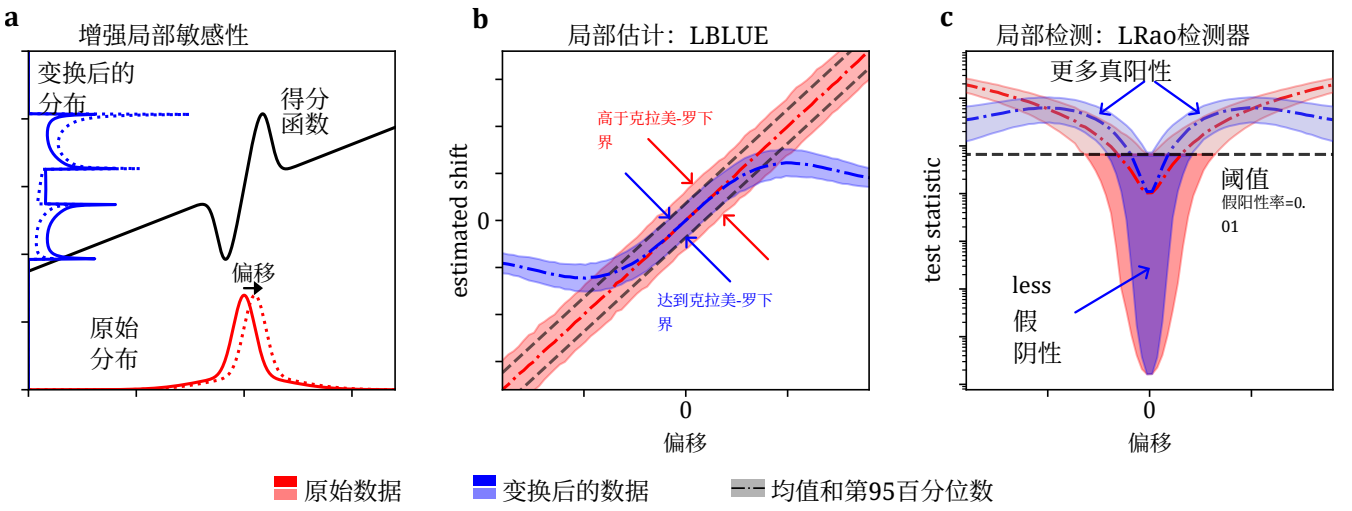


图2：**从高斯混合分布中抽取的噪声样本**。**a**，在我们估计并检测独立同分布噪声样本向量中偏移的任务中，原始分布（高斯混合分布，红色）通过得分函数（黑色）变换为新的分布（蓝色），从而最大化 LBLUE 所能提取的 LFI。**b**，原始数据与变换后的数据上 LBLUE 估计的均值和第 95 百分位数。值得注意的是，应用于变换后的数据的 LBLUE 在局部达到克拉美-罗下界，但从全局来看会产生有偏估计。应用于原始数据的 LBLUE 在全局上是无偏的，但其方差高于克拉美-罗下界。**c**，局部 Rao 检测统计量与阈值。此处，应用于得分变换数据的 LRao检测器（等价于原始数据上的 Rao检测器）与应用于原始数据的 LRao检测器相比，展现出更高的局部检测概率。这一点可通过比较蓝色曲线与红色曲线分别位于表示阈值的水平虚线之上和之下的彩色填充区域看出。

点。最简单且应用最广泛的模型是加性噪声中的线性信号模型，其定义如下：

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w}, \tag{10}$$

其中，$\mathbf{H} \in \mathbb{R}^{N\times l}$ 是已知的观测矩阵，其列张成信号空间；$\mathbf{w} \in \mathbb{R}^{N}$ 是服从未知分布的噪声向量。如补充信息 B 1 中所详述，加性模型 (10) 的优势在于，均值的雅可比矩阵（对数据进行神经网络变换后）仅需利用纯噪声（$\mathcal{H}_0$）数据即可计算，而这类数据通常十分充足。所提出的方法也适用于观测矩阵$\mathbf{H}$ 在训练期间（部分）未知的情况。例如，当干扰参数仅在备择假设 $\mathcal{H}_1$ 下存在时，就会出现这种情况；未知周期长度 [46, 47] 的周期信号检测便是一例。在此类情况下，可通过应用链式法则，将优化问题改写为针对模型 $\mathbf{x} = \boldsymbol{\theta}^* + \mathbf{w}$的优化问题，其中均值的雅可比矩阵右乘观测矩阵，详见补充信息 B1。

当噪声来自平稳过程时，通过最大化 LFI 构造最优检测器会变得简单得多；在这种过程中，统计特性随时间保持不变。时间序列数据通常满足这一假设，在实际应用中也经常如此，因此可以使用强大的分析工具[48]。在这种情况下，数据的协方差矩阵是一个对称托普利兹矩阵，其逆矩阵可近似为 [21, 49]

$$\boldsymbol{\Sigma}^{-1} \approx \sum_{i=0}^{N-1} \frac{\mathbf{v}_i \mathbf{v}_i^H}{P_{xx}(f_i)}\,, \tag{11}$$

其中 $\mathbf{v}_i^H = \frac{1}{\sqrt{N}}\left[1 \quad e^{-\mathrm{j}2\pi f_i} \quad e^{-\mathrm{j}4\pi f_i} \quad \dots \quad e^{-\mathrm{j}2\pi(N-1)f_i}\right]$ 表示共轭转置，$\mathbf{v}_i$ 是第 i 个离散傅里叶变换基向量，其归一化频率为 $f_i = i/N$，对应于 $i = 0, 1, \dots, N-1$，而 $P_{xx}(f_i)$ 是数据的功率谱密度。使用式 (11) 可以通过两种方式简化 LFI (5) 的计算。首先，它可以直接计算协方差矩阵的逆矩阵，无需对样本协方差矩阵进行数值求逆。其次，它大幅减少了需要估计的参数数量。补充信息 B 2 对这种方法进行了详细描述。将式 (11) 与式 (9) 和式 (5) 结合后，通过一组可学习参数 W 参数化的神经网络 $\Psi_W(\mathbf{x})$ 可通过最小化下式进行优化

$$C(\Psi_W(\mathbf{x})) = -\sum_{j=1}^{l}\sum_{i=0}^{N_\Psi-1} \frac{|\mathbf{v}_i^H \frac{\partial \boldsymbol{\mu}_\Psi}{\partial \theta_j}|^2}{P_{\Psi\Psi}(f_i)} \tag{12}$$

其中，$\boldsymbol{\mu}_\Psi$ 是输出数据 $\Psi(\mathbf{x})$的均值（为使符号简洁，省略下标 W），$P_{\Psi\Psi}(f_i)$是 $\Psi(\mathbf{x})$的功率谱密度，N_Ψ 是 $\Psi(\mathbf{x})$的维数（我们将其设为与输入维数 N相等），$|\cdot|$是复数的模。

在优化过程中，通过将 $\boldsymbol{\mu}_\Psi$ 近似为批次平均值、将 $\partial\boldsymbol{\mu}_\Psi/\partial\theta_j$ 近似为中心差分商，并将 $P_{\Psi\Psi}(f_i)$近似为批次平均周期图 [50, 51]，来计算式（12）。直观而言，代价函数是在离散傅里叶变换域中计算的 LFI 的迹。其形式简单，是因为协方差矩阵 $\boldsymbol{\Sigma}_{\Psi\Psi}$ 的特征向量渐近地与离散傅里叶变换向量对齐，而均值的雅可比矩阵与协方差矩阵的乘积（见式（5））会得到雅可比矩阵各列的离散傅里叶变换的平方。在可加性和平稳性的假设下，我们只需少量数据，就能高效地优化轻量级神经网络，如下文的真实世界示例所示。

现在我们来说明神经网络的架构。我们的模型利用平稳数据的时间平移对称性，采用卷积层作为滤波器处理数据，只考虑较小时间步范围内的数据点，如图 1 的示意图所示，更多细节见补充信息 B7。此外，我们假设噪声分布关于零对称，因此得分函数为奇函数。网络设计通过使用双曲正切函数作为激活函数，并省略卷积层中的偏置参数，将这一特性嵌入其中。为确保尺度不变性，输入数据使用信号尺度的稳健估计进行归一化，从而通过不变输入 [52] 强制实现不变性。

尽管该架构仅包含少量可训练参数，我们在补充信息 B7 中表明，与更大的模型相比，它的性能并未降低。结合计算成本低廉的代价函数（12），这一神经网络模型能够快速训练，因此适用于实时应用和在线自适应策略。

A. 仿真：柯西噪声

首先，我们在模拟噪声数据上验证所提出的方法。我们从一个分布中采样数据，该分布的费舍尔信息（FI）以及因此代价函数的最优值都是已知的，从而为客观比较提供了依据。具体而言，我们选择柯西噪声，以证明我们的方法即使对于这种高阶矩未定义的重尾分布也有效。为了生成噪声序列，我们对柯西噪声进行独立同分布向量采样，随后应用线性滤波器引入样本间的依赖关系，从而能够对加性噪声模型 (10) 的费舍尔信息进行闭式计算，详见补充信息 B 4。与下文讨论的真实世界示例类似，我们要设计一个用于检测已知周期的弱周期信号的检测器。这些信号被建模为傅里叶级数，并且可以表示为信号模型 (10)，详见 SI B 3。

在图3a中，展示了待检测信号以及受噪声污染的信号。由于柯西分布具有重尾，纵坐标的刻度在区间 $[-1, 1]$ 内为线性，而在该区间之外则为对数，直至interval。图3b表明，在训练过程中，神经网络会增大其输出处估计的LFI的迹，直至其接近费舍尔信息的迹。当LFI等于费舍尔信息时，LRao检测器与Rao检测器一致，因而渐近最优。图3c展示了序列长度为$N = 128$ 时，LRao检测器在不同信噪比下的接收机工作特性（ROC）曲线以及渐近最优ROC曲线。LRao检测器尚未达到其渐近性能，这一点也可从图3d中看出。与渐近分布相比，$\mathcal{H}_1$ 下的直方图略微向左偏移，表明在本例中，弱信号假设在该信噪比水平下已略显不准确。相比之下，对于下文给出的真实世界示例，可以看到LRao检测器在 $N = 128$ 时已能以良好近似达到其渐近性能。图3e展示了序列长度增至 $N = 1024$时的ROC曲线。此时LRao检测器几乎达到渐近最优性能，这也可由图3f所示的分布推断得出。需要注意的是，对于不同的序列长度，使用的是同一个轻量级卷积神经网络（未经重新训练），因为噪声样本的平稳性使得该网络所执行的最优非线性滤波操作与序列长度无关。

B. 真实世界示例：磁传感器数据

现在，我们使用真实世界测量数据展示所提出方法的有效性，如图4a所示。这些由磁传感器采集的数据包含近似高斯环境噪声以及尖峰地磁噪声。数据采集过程的详细说明见 [47, 第 545 页及后续页]„ 因为这些数据取自本书的补充材料。我们将数据序列分割为较小的非重叠序列，以构建数据集，随后通过 8-1-1 嵌套交叉验证程序将其用于训练、验证和测试。任务是基于现有的纯噪声测量数据，设计一个针对已知周期长度的弱周期信号的检测器。与第 III A 小节的仿真示例相同，周期信号被建模为傅里叶级数，并且可以表示为模型 (10)，详见 SI B 3。

当我们转向真实世界数据时，会将我们的方法与标准参考检测器进行比较。这些参考检测器遵循一系列处理步骤：采用稳健的线性白化变换来减轻噪声数据中的相依性，并使独立性假设合理化；采用标量函数处理异常值；并在独立同分布高斯噪声假设下计算广义似然比检验（GLRT）统计量（见 SI B 6）。值得注意的是，对于独立同分布非高斯噪声，在独立同分布

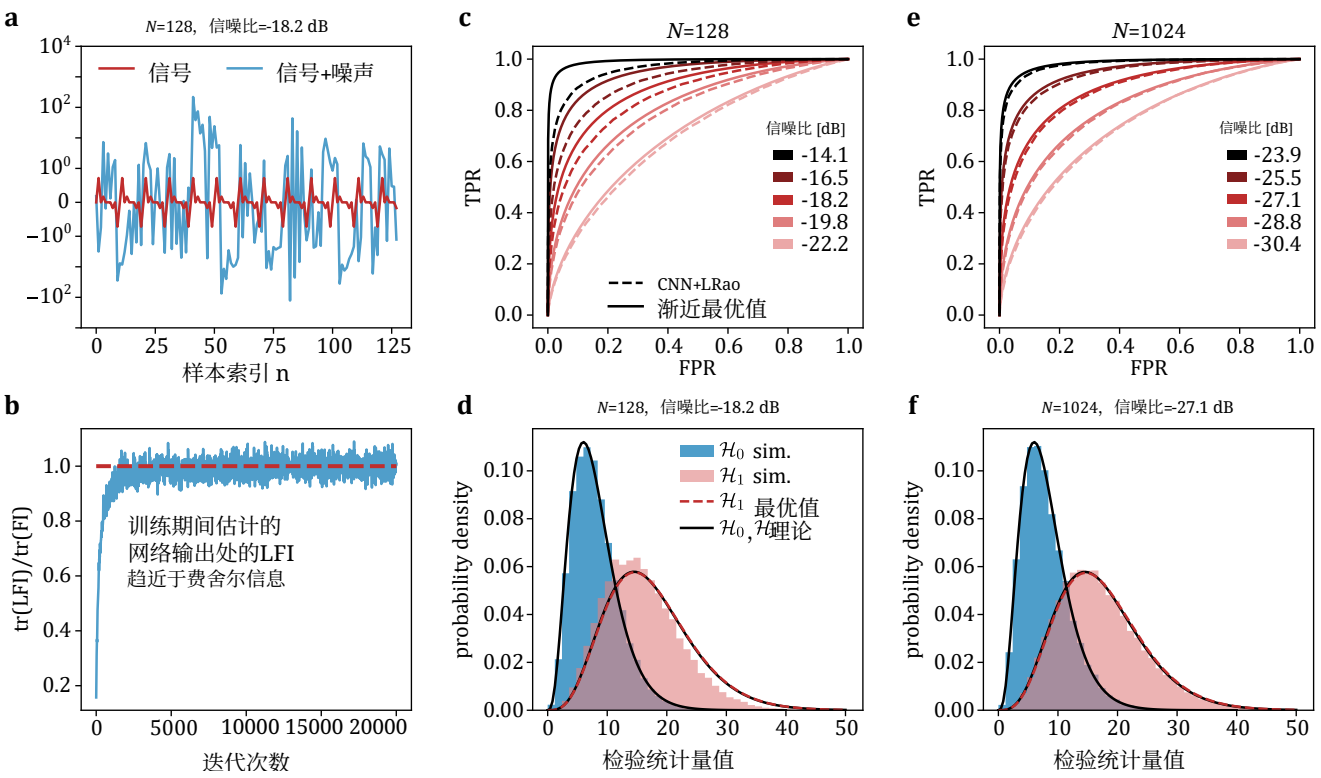


图 3：**模拟的柯西噪声**。**a**，待检测的周期信号以及受加性重尾柯西噪声污染的信号。为提高可视性，在区间 $[-1, 1]$ 内，纵坐标采用线性刻度；在该区间之外则采用对数刻度。**b**，在神经网络输出端估计的 LFI 迹，并以 FI 的迹归一化。**c**，序列长度为 $N = 128$时，不同信噪比下模拟得到的 ROC 曲线和渐近最优 ROC 曲线。LRao检测器尚未达到其渐近性能。**d**，LRao检测器的模拟分布和解析分布。在 $\mathcal{H}_1$条件下，与解析渐近分布相比，LRao检测统计量略微向左偏移。**e**，$N = 1024$的 ROC 曲线。LRao检测器逐渐接近其渐近性能，这一点也可从**f**中看出。

高斯噪声假设下计算经得分变换的数据的广义似然比检验统计量，会得到 Rao 检测器；该检测器是渐近最优的 [21]。因此，参考方法可以视为采用人工设计得分函数的 Rao 检测器。具体而言，我们使用符号函数（它等于拉普拉斯分布的得分函数；该分布用于取自的教科书 [47] ）、（启发式）限幅函数（将大于三倍标准差的数值截断），以及恒等函数（它等于标准高斯分布的得分函数）。

尽管数据的某些片段可以用高斯分布合理近似（见补充材料 B 5），但我们预计，与稳健方法相比，高斯假设（即不存在非线性）会导致次优结果，因为即使单个异常值也可能触发假阳性。与神经网络的训练、验证和测试相一致，参考检测器通过交叉验证程序进行评估，采用 9-1 的训练测试划分。由于不优化超参数，因此无需验证集。白化矩阵由训练数据得出，随后在测试数据上评估性能。

在图4b中，比较了所提出检测器（CNN+LRao，蓝色）的ROC曲线、其渐近性能以及基准检测器。渐近性能使用由训练数据加验证数据估计的LFI计算，并对所有交叉验证运行取平均。我们的CNN+LRao检测器以良好的近似达到其渐近性能，因此在CNN输出处的LFI与原始数据的费舍尔信息相匹配的程度上是最优的。我们所提出的检测器大幅优于所有基准检测器，包括 [47], 中使用的“符号检测器”，尽管可用的训练数据有限。高斯GLRT检测器（无非线性）的较差性能凸显了处理重尾噪声的必要性。我们的方法本质上对这些重尾进行惩罚，并直接从纯噪声测量数据中学习最优多变量非线性，而不是依赖启发式限幅器设计或假设特定的噪声概率分布。

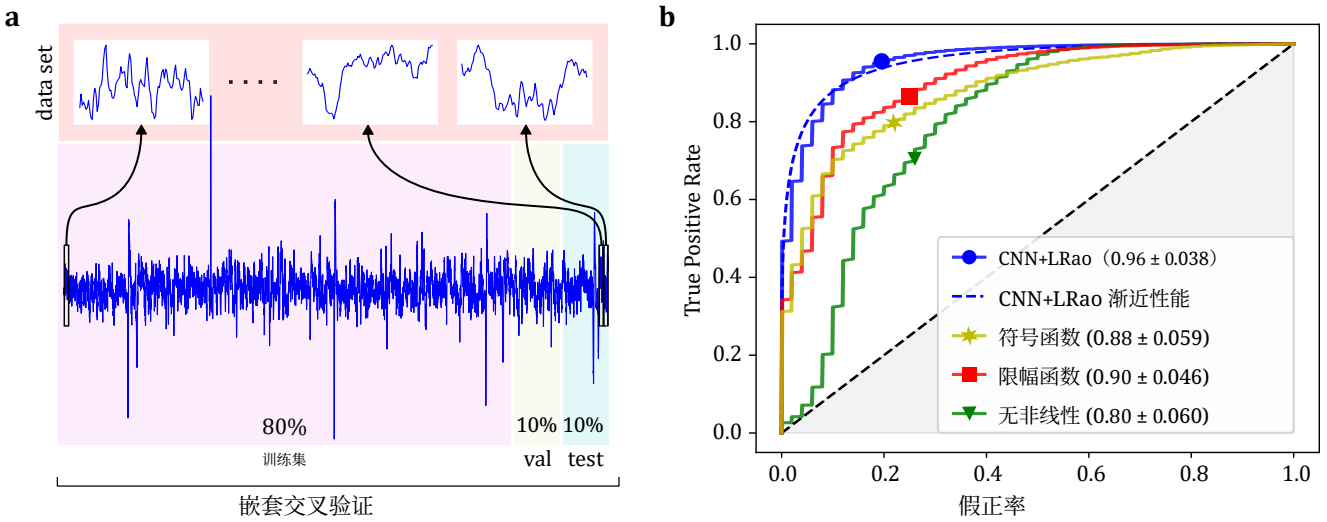


图4：**来自磁传感器的实验数据**。**a**，由磁传感器采集的尖峰噪声和相关噪声数据，可从 [47]的补充材料中获取。长度为10000的数据序列被拆分为长度为128的较小序列，形成包含78个序列的数据集。采用嵌套交叉验证程序评估检测器的性能，该检测器旨在检测在后处理中添加的弱周期信号。**b**，输入信噪比为 -25.5 dB时的ROC曲线和曲线下面积（括号内给出第 95 百分位数）。所提出的LRao检测器与轻量级卷积神经网络（CNN+LRao）相结合，明显优于参考方法。CNN+LRao检测器以较好的近似达到其渐近性能，因此在网络输出处的LFI与原始数据的FI相匹配的程度上，它是最优的。尽管训练数据量有限，我们仍强调该方法的有效性。

IV. 总结与展望

我们的方法以最大化数据集的线性 Fisher 信息为核心，能够检测淹没在强的、相关的非高斯噪声中的信号。通过求解该优化问题，我们实现了对含噪数据的变换，有效处理非高斯性，从而使数据中的全部信息易于由LBLUE和LRao检测器提取。该变换使用专为处理平稳时间序列数据而设计的轻量级神经网络架构进行参数化。一旦网络训练良好，我们对变换后的数据应用LRao检验，这等价于对原始数据执行Rao检验。我们证明，对于淹没在模拟柯西噪声中的信号，我们的方法收敛到最优解；当应用于磁传感器测量的实验数据时，即使训练数据有限，也显著优于传统方法。与其他方法不同，我们的方法是渐近最优的，并且不需要明确知道底层概率分布，因为最优多变量非线性直接从纯噪声数据中学习得到。

我们的方法为关于局部非线性检测统计量的大量研究提供了有前景的推进 [21, 22, 24, 42], , 为基于模型的方法提供了一种强大且计算高效的替代方案，而后者通常需要对数据进行详尽的分析和统计建模。不同于传统的深度学习策略，该方法不必充当数据驱动黑箱（尽管也可以如此），而是支持混合使用，因为模型知识在工程系统中通常既可获得又准确，并且能够轻松集成。尽管本文聚焦于加性噪声中的弱信号，但该方法也可以很容易地调整，以适应其他信号模型。

我们的方法有一项显著特性，使其有别于其他技术：它非常适合扩展为在线自适应策略。由于神经网络规模较小，无需大量数据，因而能够对新的测量值保持较高的敏感度。这对于许多检测任务至关重要，例如物理实验或工业过程中，必须通过持续监测将实验装置的漂移和外部扰动纳入考量。

致谢

我们感谢 Lukas M. Rachbauer 富有启发性的讨论和建设性建议，这些讨论和建议帮助促成了本研究。我们衷心感谢 BMWET、国家研究、技术与发展基金会以及 Christian Doppler 研究协会提供的资金支持。本项工作获得了 Silicon Austria Labs（SAL）的支持；该机构由奥地利共和国、施蒂里亚州商业促进机构

机构（SFG）、克恩顿州、上奥地利研究（UAR）以及
奥地利电气与电子工业协会（FEEI）共同所有。

补充信息 A：理论

下文对第 II 节中给出的方程提供了更详细的推导。此外，我们还提供了有关局部估计与检测这一主题的一些背景信息。

1. 局部似然展开

在本工作中，我们关注局部参数估计和局部假设检验，例如弱信号检测。所谓局部，是指假定参数向量 $\boldsymbol{\theta}$ 接近参考参数向量 $\boldsymbol{\theta}_0$。在这种情况下，我们可以利用对数似然函数 $L = \log\, p(\mathbf{x};\boldsymbol{\theta})$在参考参数 $\boldsymbol{\theta}_0$附近的二阶泰勒级数展开

$$L \approx L_0 + \nabla_{\boldsymbol{\theta}} L_0(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\top \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}^\top L_0(\boldsymbol{\theta}-\boldsymbol{\theta}_0), \tag{A1}$$

其中，下标 0 表示在应用梯度和海森矩阵之后，这些表达式在 $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ 处求值。在非常一般的情形下，[53], 量 $\nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}^\top L_0$ 可由其期望（即负的费舍尔信息）代替，从而得到

$$L \approx L_0 + \nabla_{\boldsymbol{\theta}} L_0(\boldsymbol{\theta}-\boldsymbol{\theta}_0) - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\top \mathbf{F}_0(\boldsymbol{\theta}-\boldsymbol{\theta}_0). \tag{A2}$$

对数似然的这一局部表达式为理解费舍尔信息及其在局部参数估计与检测中的作用提供了有价值的洞见。

我们首先讨论局部最大似然估计量（LMLE），它与 LBLUE 十分相似，并自然引出 Rao 检测统计量。一般而言，$\boldsymbol{\theta}$ 的最大似然估计满足 $\nabla_{\boldsymbol{\theta}} L|_{\boldsymbol{\theta}_{\mathrm{MLE}}} = \mathbf{0}^\top$。对 (A2) 关于 $\boldsymbol{\theta}$ 求导，可得到得分函数的近似表达式

$$\nabla_{\boldsymbol{\theta}} L \approx \nabla_{\boldsymbol{\theta}} L_0 - (\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\top \mathbf{F}_0, \tag{A3}$$

它是关于 $\boldsymbol{\theta}$的线性函数。令其为零并求解 $\boldsymbol{\theta}$，即可得到 LMLE

$$\hat{\boldsymbol{\theta}}_{\mathrm{LMLE}} = \boldsymbol{\theta}_0 + \mathbf{F}_0^{-1}(\nabla_{\boldsymbol{\theta}} L_0)^\top. \tag{A4}$$

在 [54] 中，为实现最优数据压缩，这一估计量以准最大似然估计量之名进行了讨论。这是因为得分函数，或等价地，(A4) 中的 LMLE，是参数向量 $\boldsymbol{\theta}$ 的局部充分统计量，因为 $\nabla_{\boldsymbol{\theta}} L_0$ 是唯一依赖于数据的项。得分函数在保持数据中的费舍尔信息不变的同时，将数据维数从 $N \to l$降至更低。当迭代应用时，(A4) 通常称为评分法，或费希尔评分法 [29, 第 187] 页。如果最大似然估计趋近其大样本分布 $\mathcal{N}(\boldsymbol{\theta},\ \mathbf{F}^{-1})$，且初始参数向量接近真实参数向量，则评分法会收敛至最大似然估计（MLE）。

2. Rao检验的推导

Rao检测器最初由 [20], 提出，详细推导可参见例如 [21]。这里，我们基于前文介绍的局部似然展开，给出一种略有不同的推导。这有助于读者进一步理解相关内容，尤其是下文讨论的 LRao检测器。

为推导 Rao 检验统计量，我们将局部对数似然展开与似然比检验（也称为奈曼–皮尔逊检验）结合起来。对数似然比检验统计量为

$$\log T_{\mathrm{LRT}} = L(\boldsymbol{\theta}_1;\mathbf{x}) - L_0. \tag{A5}$$

由于 $\boldsymbol{\theta}_1$ 未知时，改用LMLE（A4）。将（A2）、（A4）和（A5）结合可得

$$\log T_{\mathrm{LRT}} \approx L(\hat{\boldsymbol{\theta}}_{\mathrm{LMLE}};\mathbf{x}) - L_0 \tag{A6}$$

$$\approx \nabla_{\boldsymbol{\theta}} L_0 \mathbf{F}_0^{-1}(\nabla_{\boldsymbol{\theta}} L_0)^\top - \frac{1}{2}\nabla_{\boldsymbol{\theta}} L_0 \mathbf{F}_0^{-1}\mathbf{F}_0\mathbf{F}_0^{-1}(\nabla_{\boldsymbol{\theta}} L_0)^\top \tag{A7}$$

$$= \frac{1}{2}\nabla_{\boldsymbol{\theta}} L_0 \mathbf{F}_0^{-1}(\nabla_{\boldsymbol{\theta}} L_0)^\top. \tag{A8}$$

因子 $1/2$ 可以并入判决阈值 [21] ，从而使Rao 检验统计量由下式给出

$$T_{\mathrm{Rao}} = \nabla_{\boldsymbol{\theta}} L_0 \mathbf{F}_0^{-1} (\nabla_{\boldsymbol{\theta}} L_0)^{\top}, \tag{A9}$$

并且可以证明其渐近分布 [21, 第 239] 页在 $\mathcal{H}_0$下为 $T_{\mathrm{Rao}} \sim \chi_l^2$，在 $\mathcal{H}_1$下则为带有$\lambda_{\mathrm{Rao}} = (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)^{\top} \mathbf{F}_0 (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)$的 $T_{\mathrm{Rao}} \sim \chi_l'^2(\lambda_{\mathrm{Rao}})$。因此，假阳性率（$P_{\mathrm{FA}}$）和真正率（$P_{\mathrm{D}}$）的渐近表达式为

$$\mathrm{FPR} = Q_{\chi_l^2}(\gamma) \tag{A10}$$

$$\mathrm{TPR} = Q_{\chi_l'^2(\lambda_{\mathrm{Rao}})}(\gamma), \tag{A11}$$

其中 Q 表示下标所示分布的右尾概率。阈值 γ 可加以选取，以满足给定的假阳性率约束。由于 $\mathcal{H}_0$ 下的概率密度函数以及由此得到的假阳性率均与未知参数无关，该检测器相对于 $\boldsymbol{\theta}$[55]具有理想的恒虚警率特性。

可以很容易地验证，Rao检验可以等价地写为

$$T_{\mathrm{Rao}} = (\hat{\boldsymbol{\theta}}_{\mathrm{LMLE}} - \boldsymbol{\theta}_0)^{\top} \mathbf{F}_0 (\hat{\boldsymbol{\theta}}_{\mathrm{LMLE}} - \boldsymbol{\theta}_0). \tag{A12}$$

直观地说，Rao检测器首先用LMLE估计未知参数向量（这通常比最大似然估计 [21]容易计算得多），并将其相对于参考参数的平方偏差与其期望方差（即费舍尔信息的逆）进行比较。如果平方偏差远大于估计量的期望方差，检测统计量将超过合理设定的阈值，我们便判定为 $\mathcal{H}_1$。反之，如果平方偏差与期望方差相近，我们便断定相应的参数向量确实是参考参数，并判定为 $\mathcal{H}_0$。

3. LFI 和悲观概率密度函数

概率密度函数 $p(\mathbf{x};\boldsymbol{\theta})$ 通常是未知的，因此必须对其形式作出假设。在 [31], 中，所谓的悲观概率密度函数是指其费舍尔信息为原始概率密度函数的费舍尔信息提供下界的概率密度函数。我们将悲观概率密度函数的费舍尔信息称为线性 Fisher 信息（LFI），并记为 $\mathbf{J}$。从数学上看，该界可写为

$$\mathbf{F} \geq \mathbf{J}, \tag{A13}$$

其中，不等式表示 $\mathbf{F}$ –$\mathbf{J}$ 是一个半正定矩阵。在悲观概率密度函数假设下，数据服从均值为 $\boldsymbol{\mu}(\boldsymbol{\theta})$ 的高斯分布，该均值在局部包含关于参数向量 $\boldsymbol{\theta}$ 的全部信息；协方差矩阵为 $\boldsymbol{\Sigma}$，我们假设它在局部与 $\boldsymbol{\theta}$无关。LFI 为

$$\mathbf{J} = (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}), \tag{A14}$$

其逆矩阵为局部无偏线性估计量 $\hat{\boldsymbol{\theta}}_{\mathrm{lin}}$ （关于数据向量 $\mathbf{x}$ 的线性形式）协方差矩阵提供了可达到的下界，该估计量适用于某个参考参数向量 $\boldsymbol{\theta}_0$ [37] 附近，即

$$\mathrm{Cov}[\hat{\boldsymbol{\theta}}_{\mathrm{lin}}] \geq \mathbf{J}_0^{-1}. \tag{A15}$$

与可逆变换不会改变费舍尔信息这一事实类似，可逆仿射变换不会改变 LFI [56], , 如下所示。

令 $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$，其中 $\mathbf{A}$ 是一个可逆矩阵。则 $\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{A}\boldsymbol{\mu} + \mathbf{b}$，且 $\boldsymbol{\Sigma}_{\mathbf{y}} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}${v23>。$\mathbf{y}$的 LFI 为

$$\mathbf{J}_{\mathbf{y}} = (\nabla_{\boldsymbol{\theta}} (\mathbf{A}\boldsymbol{\mu} + \mathbf{b}))^{\top} (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\top})^{-1} (\nabla_{\boldsymbol{\theta}} (\mathbf{A}\boldsymbol{\mu} + \mathbf{b})) \tag{A16}$$

$$= (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu})^{\top} \mathbf{A}^{\top} \mathbf{A}^{-\top} \boldsymbol{\Sigma}^{-1} \mathbf{A}^{-1} \mathbf{A} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}) \tag{A17}$$

$$= (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu})^{\top} \boldsymbol{\Sigma}^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}) \tag{A18}$$

$$= \mathbf{J}. \tag{A19}$$

因此，LFI不仅为费舍尔信息提供了下界，其逆还界定了局部无偏线性估计量的方差；并且与在可逆变换下保持不变的费舍尔信息类似，LFI在可逆仿射变换下保持不变。接下来，我们利用局部泰勒级数

展开来推导LBLUE和LRao检测器，其方式类似于LMLE和Rao检测器的推导。

我们将悲观对数似然函数定义为 $\tilde{L}(\boldsymbol{\theta}) = \ln p(\mathbf{x}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma})$。我们假设$\boldsymbol{\Sigma}$ 与 $\boldsymbol{\theta}$相互独立。由于我们仅在某个参考参数向量 $\boldsymbol{\theta}_0$ 附近局部使用悲观概率密度函数描述，并在 $\boldsymbol{\theta}_0$处计算 $\boldsymbol{\Sigma}$ ，记为 $\boldsymbol{\Sigma}_0$，因此这并非限制。按照上述方法， L 在 $\boldsymbol{\theta}_0$ 附近的二阶泰勒˜级数展开为

$$\tilde{L} \approx \tilde{L}_0 + \nabla_{\boldsymbol{\theta}} \tilde{L}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^\top \tilde{L}_0(\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{A20}$$

在 $\boldsymbol{\theta}_0$ 处计算得到的悲观得分函数为

$$\nabla_{\boldsymbol{\theta}} \tilde{L}_0 = (\mathbf{x} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0, \tag{A21}$$

它可以视为局部线性充分统计量，与得分函数类似；得分函数是局部充分统计量[54]。悲观得分函数从 $N \to l$进行降维投影，同时保持 LFI 不变，这可由 LBLUE 的统计特性推得，如下所示。悲观海森矩阵为

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^\top \tilde{L}_0 &= -(\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0) + (\mathbf{x} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^\top \boldsymbol{\mu}_0 & \text{(A22)} \\ &= -\mathbf{J}_0 + (\mathbf{x} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} \nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^\top \boldsymbol{\mu}_0. & \text{(A23)} \end{aligned}$$

同上，将 (A20) 中的海森矩阵替换为其期望，即 $\mathbb{E}[\nabla_{\boldsymbol{\theta}}^2 \tilde{L}_0] = -\mathbf{J}_0$, ，得到

$$\tilde{L} \approx \tilde{L}_0 + (\mathbf{x} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{J}_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0). \tag{A24}$$

4. LBLUE 的推导

在悲观概率密度函数假设下， $\boldsymbol{\theta}$ 的 LBLUE 就是 LMLE，因此可以根据$\nabla_{\boldsymbol{\theta}} \tilde{L}|_{\boldsymbol{\theta}_{\text{LBLUE}}} = \mathbf{0}^\top$求得，其中使用了 $\nabla_{\boldsymbol{\theta}} \tilde{L}$ 的局部展开。对 (A24) 关于 $\boldsymbol{\theta}$ 求导，得到

$$\nabla_{\boldsymbol{\theta}} \tilde{L} \approx (\mathbf{x} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0 - (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{J}_0. \tag{A25}$$

令其等于零并求解 $\boldsymbol{\theta}$ ，得到

$$\hat{\boldsymbol{\theta}}_{\text{LBLUE}} = \boldsymbol{\theta}_0 + \mathbf{J}_0^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0). \tag{A26}$$

现在我们将验证这一假设：LBLUE 的统计量是渐近高斯的，其均值对应参数 $\boldsymbol{\theta}$，协方差矩阵等于逆 LFI $\mathbf{J}_0^{-1}$，即 ˆ$\boldsymbol{\theta}_{\text{LBLUE}} \overset{a}{\sim} \mathcal{N}(\boldsymbol{\theta}, \mathbf{J}_0^{-1})$，其中 a 表示当样本量 $N \to \infty$趋于无穷大时，该关系渐近成立。

我们首先推导 LBLUE 的均值和协方差。利用 $\boldsymbol{\mu} \approx \boldsymbol{\mu}_0 + \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0)$，可得

$$\begin{aligned} \mathbb{E}[\hat{\boldsymbol{\theta}}_{\text{LBLUE}}] &\approx \boldsymbol{\theta}_0 + \mathbf{J}_0^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu}_0 + \nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0) - \boldsymbol{\mu}_0) \\ &= \boldsymbol{\theta}_0 + \mathbf{J}_0^{-1} \mathbf{J}_0 (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &= \boldsymbol{\theta}, \end{aligned} \tag{A27}$$

and

$$\begin{aligned} \text{Cov}[\hat{\boldsymbol{\theta}}_{\text{LBLUE}}] &= \mathbf{J}_0^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} \text{Cov}[\mathbf{x} - \boldsymbol{\mu}_0] \boldsymbol{\Sigma}_0^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0) \mathbf{J}_0^{-1} \\ &= \mathbf{J}_0^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\nabla_{\boldsymbol{\theta}} \boldsymbol{\mu}_0) \mathbf{J}_0^{-1} \\ &= \mathbf{J}_0^{-1} \mathbf{J}_0 \mathbf{J}_0^{-1} \\ &= \mathbf{J}_0^{-1}. \end{aligned} \tag{A28}$$

ˆ 接下来的任务是验证 $\boldsymbol{\theta}_{\text{LBLUE}}$ 的分布确实是渐近高斯的。由于随机变量$\mathbf{x}$ 的各个分量并不相互独立，因此无法直接从要求独立性的经典中心极限定理推出渐近正态性。不过，在处理时间序列数据时，随机变量的取值通常只会在时间上彼此足够接近时才相互依赖。这里，我们假设 m-相依性 [57], ，这意味着随机变量$\mathbf{x}$ 的分量 x_i 和 x_j 在 $|i - j| \leq m$ 时相依，并且

在 $|i-j|>m$时相互独立。 [57] 给出了适用于 m-相依随机变量的某个版本的中心极限定理的证明，该定理可直接推出 $\boldsymbol{\theta}_{\mathrm{LBLUE}}$的渐近正态性。

假设 m-相依性并非获得 LBLUE 正态性的唯一途径。一般而言，随机变量各分量之间的相依性有大量不同的条件，因此也存在多种适用于这些弱相依随机变量的中心极限定理变体。无论如何，应用我们的信号检测方法时，都应分析各分量之间的相依性。仅凭数据很难推断相依性的确切性质，但通过（稳健地）估计协方差矩阵，并比较其对角线元素与远离对角线的元素，我们可以对随机向量各项之间的相依关系形成较好的直观认识。对于平稳数据，可以稳健地估计自相关序列，以直观分析相依性。更先进但计算量更大的方法是估计各分量之间的互信息，直接评估其相依性 [58]。

关于相依性，我们补充一点：我们预计 LBLUE 的统计量可以用高斯分布很好地近似建模。因此，作为谨慎测量相依性的实用替代方法，可以估计 LBLUE 的概率密度函数，以确认其服从高斯分布。

5. LRao检验的推导

下面，我们推导所提出的 LRao检验。正如正文所述，只需在悲观概率密度函数假设下计算 Rao检验，即可得到 LRao 检验。这里，我们采用稍深入的方法，从该悲观概率密度函数假设下的对数似然比检验统计量入手

$$\log \widetilde{T}_{\mathrm{LRT}} = \tilde{L}(\boldsymbol{\theta}_1;\mathbf{x}) - \tilde{L}_0. \tag{A29}$$

与上文类似，由于 $\boldsymbol{\theta}_1$ 未知，我们代入 LBLUE（A26）。结合（A24）、（A26）和（A29）可得

$$\log \widetilde{T}_{\mathrm{LRT}} \approx \tilde{L}(\hat{\boldsymbol{\theta}}_{\mathrm{LBLUE}}) - \tilde{L}(\boldsymbol{\theta}_0) \tag{A30}$$

$$= \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0)\mathbf{J}_0^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}-\boldsymbol{\mu}_0). \tag{A31}$$

A如上，t 该因子可以转移到阈值中，因此 LRao统计量可表示为

$$T_{\mathrm{LRao}} = \nabla_{\boldsymbol{\theta}}\tilde{L}_0 \mathbf{J}_0^{-1}(\nabla_{\boldsymbol{\theta}}\tilde{L}_0)^\top \tag{A32}$$

$$= (\mathbf{x}-\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0)\mathbf{J}_0^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1}(\mathbf{x}-\boldsymbol{\mu}_0). \tag{A33}$$

其渐近分布可推导如下。容易验证，LRao检验统计量可改写为

$$T_{\mathrm{LRao}} = (\hat{\boldsymbol{\theta}}_{\mathrm{LBLUE}} - \boldsymbol{\theta}_0)^\top \mathbf{J}_0 (\hat{\boldsymbol{\theta}}_{\mathrm{LBLUE}} - \boldsymbol{\theta}_0). \tag{A34}$$

假设 LBLUE 达到其渐近分布 $\mathcal{N}\left(\boldsymbol{\theta}_0, \mathbf{J}_0^{-1}\right)$。在 $\mathcal{H}_0$ 下，$T_{\mathrm{LRao}} \sim \chi_l^2$，以及 $\mathcal{H}_1$ 下，$T_{\mathrm{LRao}} \sim \chi_l'^2(\lambda_{\mathrm{LRao}})$，其中 $\lambda_{\mathrm{LRao}} = (\boldsymbol{\theta}_1-\boldsymbol{\theta}_0)^\top \mathbf{J}_0 (\boldsymbol{\theta}_1-\boldsymbol{\theta}_0)$。因此，FPR 为 P_{FA} [21, 32,] 由此可直接得出 p. f.，在 下，以及 ( )，其中 ( )J( ) 在 下。因此，FPR ( ) 和 TPR (P_{D}) 为

$$\mathrm{FPR} = Q_{\chi_l^2}(\gamma) \tag{A35}$$

$$\mathrm{TPR} = Q_{\chi_l'^2(\lambda_{\mathrm{LRao}})}(\gamma), \tag{A36}$$

其中，Q 表示下标所示分布的右尾概率。与 Rao检验类似，可选择阈值 γ 以满足给定的 FPR 约束。由于 在 下的 PDF 与未知参数无关，因此 FPR 也与未知参数无关，该检测器关于 $\boldsymbol{\theta}$具有理想的恒虚警率（CFAR）特性。

6. 得分函数最大化 LFI 的迹的证明

近期文献 [43] 指出，得分变换数据可以最大化 LFI；本文进一步证明，对于向量参数，它可以最大化 LFI 的迹，并给出形式化证明。由得分函数$\mathbf{x}' = (\nabla_{\boldsymbol{\theta}} L_0)^\top$ （其不依赖于 $\boldsymbol{\theta}$）变换后的数据的 LFI 定义为

$$\mathbf{J}_0' = (\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0')^\top \boldsymbol{\Sigma}_0'^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0'). \tag{A37}$$

从均值的梯度开始，我们得到

$$\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_0' = (\nabla_{\boldsymbol{\theta}}\mathbb{E}[(\nabla_{\boldsymbol{\theta}}L_0)^\top])|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{A38}$$

$$= \int (\nabla_{\boldsymbol{\theta}}L_0)^\top \nabla_{\boldsymbol{\theta}} p(\mathbf{x};\boldsymbol{\theta})\mathrm{d}\mathbf{x}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{A39}$$

$$= \int (\nabla_{\boldsymbol{\theta}}L_0)^\top \frac{\nabla_{\boldsymbol{\theta}} p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta})} p(\mathbf{x};\boldsymbol{\theta})\mathrm{d}\mathbf{x}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{A40}$$

$$= \int (\nabla_{\boldsymbol{\theta}}L_0)^\top (\nabla_{\boldsymbol{\theta}}L) p(\mathbf{x};\boldsymbol{\theta})\mathrm{d}\mathbf{x}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{A41}$$

$$= \int (\nabla_{\boldsymbol{\theta}}L_0)^\top (\nabla_{\boldsymbol{\theta}}L_0) p(\mathbf{x};\boldsymbol{\theta}_0)\mathrm{d}\mathbf{x} \tag{A42}$$

$$= \mathbf{F}_0. \tag{A43}$$

继续处理协方差矩阵，我们得到

$$\boldsymbol{\Sigma}_0' = \mathbb{E}[(\nabla_{\boldsymbol{\theta}}L_0 - \mathbb{E}[\nabla_{\boldsymbol{\theta}}L_0])^\top (\nabla_{\boldsymbol{\theta}}L_0 - \mathbb{E}[\nabla_{\boldsymbol{\theta}}L_0])]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{A44}$$

$$= \mathbb{E}[(\nabla_{\boldsymbol{\theta}}L_0 - \mathbb{E}[\nabla_{\boldsymbol{\theta}}L_0]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0})^\top (\nabla_{\boldsymbol{\theta}}L_0 - \mathbb{E}[\nabla_{\boldsymbol{\theta}}L_0]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0})]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{A45}$$

$$= \mathbb{E}[(\nabla_{\boldsymbol{\theta}}L_0)^\top (\nabla_{\boldsymbol{\theta}}L_0)]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{A46}$$

$$= \mathbf{F}_0. \tag{A47}$$

其中我们使用了正则性条件 $\mathbb{E}[(\nabla_{\boldsymbol{\theta}}L_0)^\top]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{0}$ [29]。将 (A43) 和 (A47) 代入 (A37) 可得

$$\mathbf{J}_0' = \mathbf{F}_0\mathbf{F}_0^{-1}\mathbf{F}_0 = \mathbf{F}_0. \tag{A48}$$

根据 (A13)，这证明了 LFI 及其迹由得分函数最大化。此外，得分函数的每一个仿射变换都使 LFI 最大化。这是因为可逆仿射变换不会改变 LFI，正如补充信息 A3 中所示。

最后，我们指出，与最大化 LFI 的迹类似，近年来在生成式人工智能领域非常成功的得分匹配，也会在其最优值处产生得分函数 [59, 60]。由于生成模型学习数据的分布，得分匹配通常侧重于所谓的 Stein 得分（定义为 $\nabla_{\mathrm{x}}L$ ），而非 Fisher 得分 $\nabla_{\boldsymbol{\theta}}L$。

7. 单侧局部参数检验

我们在 LFI 框架下结合 LRao检验的检测方法，也可适用于单侧参数检验。尽管所得统计量（尺度不同，且未经 LFI 优化）最早在 [35], 中导出，但据我们所知，本文首次在检测理论的背景下给出该统计量。

我们考虑通常的标量单侧检验，其假设定义如下

$$\mathcal{H}_0: \ \theta = \theta_0 \tag{A49}$$

$$\mathcal{H}_1: \ \theta > \theta_0 \ . \tag{A50}$$

局部最优势（LMP）检验统计量在局部是最优的 [21, 第217], 页，由下式给出

$$T_{\mathrm{LMP}} = \frac{1}{\sqrt{F_0}} \frac{\partial L}{\partial \theta}\bigg|_{\theta=\theta_0}, \tag{A51}$$

其渐近分布在 $\mathcal{H}_0$ 下为 $\mathcal{N}(0,\ 1)$，在 $\mathcal{H}_1$ [21] 下为 $\mathcal{N}(\sqrt{F_0}(\theta_1 - \theta_0),\ 1)$。该检测器可解释为单侧 Rao 检验，其中省略了平方运算，因为 $\theta \geq \theta_0$。将 LMP 检验扩展到 LFI 框架，可得到线性局部最强功效（LLMP）检验统计量（也可以称为单侧 LRao 检验，但为与文献中的 LMP 检验 [21] 保持一致，我们称之为 LLMP 检验）

$$T_{\mathrm{LLMP}} = \frac{1}{\sqrt{J_0}} \frac{\partial \boldsymbol{\mu}^\top}{\partial \theta}\bigg|_{\theta=\theta_0} \boldsymbol{\Sigma}_0^{-1}(\mathbf{x} - \boldsymbol{\mu}_0), \tag{A52}$$

其渐近分布在 $\mathcal{H}_0$ 下为 $\mathcal{N}(0, 1)$，在 $\mathcal{H}_1$ 下为 $\mathcal{N}(\sqrt{J_0}(\theta_1 - \theta_0), 1)$，这可直接由渐近 LBLUE 统计量推出。

补充信息 B：弱信号检测

本补充材料的这一部分介绍了我们与第 III 节相关的方法细节。

1. 加性信号模型

在本研究的范围内，我们假设数据可以通过加性噪声模型准确建模

$$\mathbf{x} = \mathbf{f}(\boldsymbol{\theta}) + \mathbf{w} \ , \tag{B1}$$

其中$\mathbf{f}(\boldsymbol{\theta})$是一个已知的确定性函数，而$\mathbf{w} \sim p_{\mathbf{w}}(\mathbf{w})$是一个随机项。在此模型中，我们假设噪声与确定性贡献相互独立，但噪声的分布可以任意复杂。请注意，加性噪声这一假设并非严格要求，我们的许多结果仍适用于其他模型，或可相应地加以推广。正如下文所示，假设存在加性噪声对于弱信号检测任务尤其有用。

检测弱信号意味着，我们希望区分数据来自参数为$\boldsymbol{\theta}_0$还是参数为 $\boldsymbol{\theta}_0 + \delta\boldsymbol{\theta}$的分布，其中 $\delta\boldsymbol{\theta}$ 各分量的幅值都很小。这意味着我们可以将函数线性化

$$\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{f}(\boldsymbol{\theta}_0) + \nabla_{\boldsymbol{\theta}}\mathbf{f}(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \ , \tag{B2}$$

其中，$\nabla_{\boldsymbol{\theta}}\mathbf{f}(\boldsymbol{\theta}_0)$ 是在 $\boldsymbol{\theta}_0$ 处求得的$\mathbf{f}$ 的雅可比矩阵。一般性不失，我们可以从数据中减去偏移$\mathbf{f}(\boldsymbol{\theta}_0) - \nabla_{\boldsymbol{\theta}_0}\mathbf{f}(\boldsymbol{\theta})\boldsymbol{\theta}_0$，并用 $\boldsymbol{\theta}$ 替换 $\boldsymbol{\theta} - \boldsymbol{\theta}_0$，从而得到式 (10)；或者 $\mathbf{f}(\boldsymbol{\theta}) \approx \mathbf{H}\boldsymbol{\theta}$，其中 $\boldsymbol{\theta}_0 = \mathbf{0}$。

此外，加性噪声使我们能够在后处理中计算关于参数的导数，而无需在实验中实际对参数进行微小偏移，这带来了显著的实践优势。为了评估 LFI、LBLUE 和 LRao检验，我们关注神经网络 $\Psi_W(\mathbf{x})$ 输出处的雅可比矩阵$\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_{\Psi,0}$。具体而言，雅可比矩阵的第 j 列可以用中心差分商近似为

$$\frac{\partial \boldsymbol{\mu}_{\Psi,0}}{\partial \theta_j} \approx \sum_i \frac{\Psi_W(\mathbf{w}_i + \mathbf{h}_j\Delta\theta) - \Psi_W(\mathbf{w}_i - \mathbf{h}_j\Delta\theta)}{2\Delta\theta} \tag{B3}$$

其中，$\mathbf{w}_i$ 是第 i 个训练数据（仅含噪声）序列，$\mathbf{h}j$ 是$\mathbf{H}$ 的第 j 列，而 $\Delta\theta$ 是步长参数，补充信息 B7 中对此进行了详尽讨论。

加性模型 (10) 的另一个优点是，它可以方便地推广到$\mathbf{H}$ 事先未知，或包含其他未知参数（即在 $\mathcal{H}_0$ 下不存在的干扰参数）的情况。我们可以在采用不同数据模型$\mathbf{x} = \boldsymbol{\theta}^* + \mathbf{w}$的假设下优化神经网络，之后再与$\mathbf{H}$ 相乘。应用链式法则即可得到这一结果，具体如下。神经网络输出处均值的雅可比矩阵为

$$\begin{aligned}
\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_{\Psi,0} &= \int \Psi_W(\mathbf{x})\nabla_{\boldsymbol{\theta}} p_{\mathbf{w}}(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})|_{\boldsymbol{\theta}=\mathbf{0}}\mathrm{d}\mathbf{x} \\
&= \int \Psi_W(\mathbf{x})\nabla_{\boldsymbol{\theta}^*} p_{\mathbf{w}}(\mathbf{x} - \boldsymbol{\theta}^*)|_{\boldsymbol{\theta}^*=\mathbf{0}}\mathrm{d}\mathbf{x}\mathbf{H} \\
&= \nabla_{\boldsymbol{\theta}^*}\boldsymbol{\mu}_{\Psi,0}\mathbf{H},
\end{aligned} \tag{B4}$$

而 LFI 为

$$\begin{aligned}
\mathbf{J}_{\Psi,0} &= (\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_{\Psi,0})^\top \boldsymbol{\Sigma}_{\Psi,0}^{-1}(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}_{\Psi,0}) \\
&= \mathbf{H}^\top(\nabla_{\boldsymbol{\theta}^*}\boldsymbol{\mu}_{\Psi,0})^\top \boldsymbol{\Sigma}_{\Psi,0}^{-1}(\nabla_{\boldsymbol{\theta}^*}\boldsymbol{\mu}_{\Psi,0})\mathbf{H} \\
&= \mathbf{H}^\top \mathbf{J}^*_{\Psi,0}\mathbf{H},
\end{aligned} \tag{B5}$$

其中，$\mathbf{J}^*_{\Psi,0} = (\nabla_{\boldsymbol{\theta}^*}\boldsymbol{\mu}_{\Psi,0})^\top \boldsymbol{\Sigma}_{\Psi,0}^{-1}(\nabla_{\boldsymbol{\theta}^*}\boldsymbol{\mu}_{\Psi,0})$ 是与数据模型 $\mathbf{x} = \boldsymbol{\theta}^* + \mathbf{w}$相关的神经网络输出端的 LFI。如上所述，表达式 $\nabla_{\boldsymbol{\theta}^*}\boldsymbol{\mu}_{\Psi,0}$ 可以通过后处理计算得出，而无需实际改变任何参数值。因此，如果 $\mathbf{H}$ 先验未知，或包含其他未知参数，我们可以仅使用噪声数据，通过最大化 $\mathbf{J}^*_\Psi$ 的迹来优化网络，随后将其应用于任意观测矩阵 $\mathbf{H}$。对于标量参数情形，与 $\mathbf{J}^*_{\Psi,0}$ 相关的量是噪声分布 [21] 的所谓固有精度。它衡量相对于加性偏移参数的 FI 量，并在加性独立同分布噪声下的弱信号检测中发挥核心作用。

2. 平稳数据

平稳性是信号处理许多领域中的一项核心假设。在实践中，它通常近似满足，因此可以使用极具实用价值的工具 [48]。其中包括频域方法，其优势在于往往可以通过计算效率较高的快速傅里叶变换（FFT）算法 [61] 实现。接下来，我们将说明如何在平稳性假设下推导出相对简单的 LFI 及其迹的表达式。

从平稳过程 x_n [62] 中抽取的随机向量$\mathbf{x}$ 的协方差矩阵 $\mathbf{\Sigma}$ 具有托普利兹结构[49], ，因为两个样本之间的相关性仅取决于它们的间距。假设对于 $|n-l|>m,\quad x_n$ 和 x_l 不相关，其中 $N \gg m$ （N是 $\mathbf{\Sigma}$的维数），则 $\mathbf{\Sigma}$ 的特征值可近似为

$$\lambda_i \approx P_{xx}(f_i), \tag{B6}$$

其中

$$P_{xx}(f_i) = \sum_{k=-\infty}^{\infty} r_k \mathrm{e}^{-\mathrm{j}2\pi f_i k} \tag{B7}$$

是按归一化离散傅里叶变换频率 $f_i = i/N$采样的功率谱密度，其中 $i = 0, 1,..., N-1,\quad r_k = \mathbb{E}[x_n x_{n-k}]$是自协方差序列（假设随机过程为零均值），其特征向量

$$\mathbf{v}_i = \frac{1}{\sqrt{N}}\left[1 \;\; \mathrm{e}^{\mathrm{j}2\pi f_i} \;\; \mathrm{e}^{\mathrm{j}4\pi f_i} \;\; \ldots \;\; \mathrm{e}^{\mathrm{j}2\pi(N-1)f_i}\right]^\top \tag{B8}$$

与离散傅里叶变换向量 [63]对齐。关于这一特征值分解的直观解释，可参见 [21, 第 33 页及后续内容。]因此，协方差矩阵可近似为

$$\mathbf{\Sigma} \approx \sum_{n=0}^{N-1} P_{xx}(f_i)\mathbf{v}_i\mathbf{v}_i^H, \tag{B9}$$

其逆矩阵可近似为

$$\mathbf{\Sigma}^{-1} \approx \sum_{n=0}^{N-1} \frac{\mathbf{v}_i\mathbf{v}_i^H}{P_{xx}(f_i)}. \tag{B10}$$

因此，利用数据估计协方差矩阵归结为功率谱密度估计，因为特征向量是已知的。这种近似的一大优势是，功率谱密度仅有 $N/2$个未知参数（实数据向量的功率谱密度具有对称性），而样本协方差矩阵则有 $N(N-1)/2$个未知参数。关于功率谱密度估计的文献浩如烟海 [64–66]。我们采用一种最具代表性的非参数技术，即所谓的舒斯特周期图 [50]

$$\hat{P}_{xx}(f_i) = \left|\sum_{n=0}^{N-1} x_n \mathrm{e}^{-\mathrm{j}2\pi f_i n}\right|^2. \tag{B11}$$

为降低估计的方差，我们对多个数据向量的周期图取平均，其思路与 Bartlett 估计量类似 [51]。

利用 (B10)，LFI 可表示为

$$\mathbf{J} = (\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu})^\top \mathbf{\Sigma}^{-1} (\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}) \tag{B12}$$

$$\approx \sum_{i=0}^{N-1} \frac{(\nabla_{\boldsymbol{\theta}}\boldsymbol{\mu})^\top \mathbf{v}_i\mathbf{v}_i^H \nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}}{P_{xx}(f_i)}, \tag{B13}$$

其中 $\mathbf{v}_i^H \nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}$ 是雅可比矩阵沿时间维度的离散傅里叶变换（DFT）的第 i 个分量。计算该近似的迹可得

$$\mathrm{tr}(\mathbf{J}) \approx \sum_{j=1}^{l}\sum_{i=0}^{N} \frac{|\mathbf{v}_i^H \frac{\partial \boldsymbol{\mu}}{\partial \theta_j}|^2}{P_{xx}(f_i)}. \tag{B14}$$

两者 P_{xx}（f_i）和$\mathbf{v}_i^H \nabla_{\boldsymbol{\theta}}\boldsymbol{\mu}$均可使用 FFT 算法估计。

3. 周期信号模型

对于磁传感器数据中弱周期信号检测的实际示例，我们使用加性信号模型（10）。通常使用傅里叶级数 [66] 对周期信号进行建模。对于有限的 N，通常使用多谐波信号模型 [47, 67]，其形式为

$$s_n = \sum_{k=1}^{K} A_k \cos(2\pi k f_0 n + \phi_k), \tag{B15}$$

对于 $n = 0, 1, \ldots, N-1$，其中 $N = 128$ 为信号长度，$A_k > 0$ 和 $-\pi \le \phi_k < \pi$ 分别为振幅和相位，$f_0 = 0.1$ 为归一化基频，$K = 4$ 为模型阶数。通过如下表示，信号模型可写成式 (10) 的形式

$$s_n = \sum_{k=1}^{K} \alpha_k \cos(2\pi k f_0 n) + \beta_k \sin(2\pi k f_0 n), \tag{B16}$$

其中 $\alpha_k = A_k \cos(\phi_k)$ 和 $\beta_k = -A_k \sin(\phi_k)$，或者采用矩阵向量表示法

$$\mathbf{s} = \mathbf{H}\boldsymbol{\theta}, \tag{B17}$$

with

$$\mathbf{H} = \begin{bmatrix} \cos(2\pi f_0 0) & \ldots & \cos(2\pi f_0 (N-1)) \\ \vdots & & \vdots \\ \cos(2\pi K f_0 0) & \ldots & \cos(2\pi K f_0 (N-1)) \\ \sin(2\pi f_0 0) & \ldots & \sin(2\pi f_0 (N-1)) \\ \vdots & & \vdots \\ \sin(2\pi K f_0 0) & \ldots & \sin(2\pi K f_0 (N-1)) \end{bmatrix}^{\top} \tag{B18}$$

and

$$\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 & \ldots & \alpha_K & \beta_1 & \ldots & \beta_K \end{bmatrix}^{\top}. \tag{B19}$$

4. 柯西噪声

仿真研究所用的柯西噪声序列通过生成柯西噪声的独立同分布向量$\mathbf{z} \in \mathbb{R}^N$得到，其中 $z_i \sim \mathcal{C}(0, 1)$ 对于 $0 \le i < N$。随后，我们应用线性滤波器，使间隔较近的样本之间产生依赖。具体而言，我们使用一个随机生成的三阶稳定自回归滤波器。滤波噪声序列表示为 $\mathbf{w} = \mathbf{A}^{-1}\mathbf{z}$，其中 $\mathbf{A}$ 是一个下三角滤波矩阵。为推导加性信号模型 (10) 的费舍尔信息，我们在等式两边左乘 $\mathbf{A}$，得到 $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{H}\boldsymbol{\theta} + \mathbf{z}$；由于 $\mathbf{A}$ 可逆，这不会改变费舍尔信息。关于 $\boldsymbol{\theta}$ 的费舍尔信息为

$$\mathbf{F} = \mathbf{H}^{\top}\mathbf{A}^{\top}\mathbf{F}^{*}\mathbf{A}\mathbf{H}, \tag{B20}$$

其中，$\mathbf{F}^{*} = \frac{1}{2}\mathbf{I}$ 是与模型 $\mathbf{x}^{*} = \boldsymbol{\theta}^{*} + \mathbf{z}$相对应的费舍尔信息。

5. 磁传感器数据

图 4 展示了取自补充材料 [47] 中的磁传感器数据。该数据包含一个长度为 10 000 个样本的纯噪声时间序列。作为预处理步骤，我们采用稳健偏移和归一化变换

$$\frac{\mathbf{x} - \mathrm{median}(\mathbf{x})}{1.483 \cdot \mathrm{median}(|\mathbf{x} - \mathrm{median}(\mathbf{x})|)}, \tag{B21}$$

其中引入因子 1.483，以确保正态分布 [25, 68]中标准差尺度估计的一致性。随后，我们将数据划分为长度为 $N = 128$ 的非重叠序列，以获得一

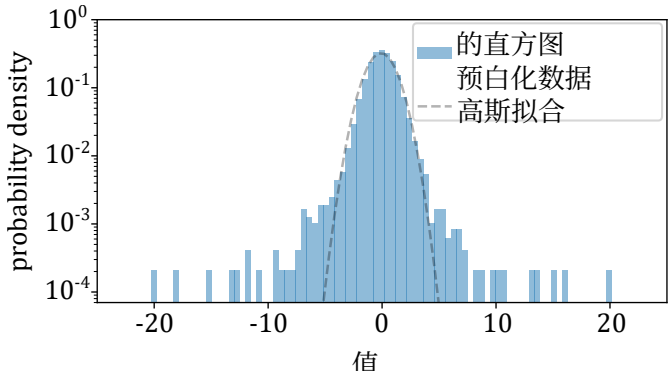


图5：预白化磁传感器数据的归一化直方图和高斯拟合。部分数据可以很好地用高斯分布近似。

共 78 个用于训练、验证和测试的数据向量。由于可用数据量有限，采用嵌套交叉验证程序：在训练数据上优化神经网络权重，在验证数据上优化超参数，以防止信息泄漏到模型 [69], ，并在测试数据上评估性能。外层交叉验证过程（测试集 ｜训练集 ＋ 验证集）采用 10 折，内层（训练集 |验证集）采用 9 折。完成一次内层交叉验证后，将训练集和验证集合并，用优化后的超参数训练网络，并在测试集上进行评估。训练和评估重复 100 次，每次使用随机初始化的权重（仅针对我们基于网络的方法，因为参考方法不需要初始化）以及随机选择的生成 $\mathcal{H}_1$ 数据，具体如下所述。

用于测试的 $\mathcal{H}_1$ 数据是通过将生成的周期信号（见补充信息B3）添加到测试集数据序列中得到的。为简便起见，所有振幅均取相同数值，具体取决于模拟所用的信噪比。相位从均匀分布 $\mathcal{U}(-\pi,\pi)$中随机抽取。这样，信号完全基于模型，而噪声来自真实世界测量。

我们现在分析图4中绘制的磁传感器数据，以说明选择参考方法和神经网络超参数的依据。首先从边缘数据分布入手：在图5中，将预白化数据的直方图与高斯拟合进行比较。预白化过程见SI B 6。在忽略离群值时，大部分数据可由高斯分布很好地近似。然而，对于在高斯噪声假设下推导的算法，即便单个异常值也可能导致严重的性能下降 [25]。因此，具有非线性限幅函数的参考方法显著优于不具有非线性的方法，如图4所示。

接下来考察平稳性与弱依赖假设。图6绘制了按(B21)预处理的测量数据经稳健估计（以中位数代替均值）得到的归一化自相关序列，以及将同一数据划分为4个连续且互不重叠的窗口后的相应序列。由于红色的4条自相关序列彼此相似，数据似乎是广义（二阶）平稳的。此外，自相关序列在大约20个样本后衰减至零，表明数据具有弱依赖（这是渐近分布有效性所要求的）。

6. 参考方法

对于两种参考方法，首先进行白化变换 [70] ，这需要已知协方差矩阵。如上所述，我们采用 10 折交叉验证程序（无需验证集，因为我们没有超参数）。协方差矩阵根据训练集估计，并在测试集上评估检测器的性能。为获得白化矩阵，我们将 (B10) 与功率谱密度的稳健估计结合使用。该稳健估计通过中位数而非均值对多个舒斯特周期图取平均，这是信号处理领域一种成熟的方法 [25, 71]。因此，参考方法的白化矩阵为

$$\sqrt{\mathbf{\Sigma}^{-1}} \approx \sum_{n=0}^{N-1} \frac{\mathbf{v}_i \mathbf{v}_i^H}{\sqrt{P_{xx}(f_i)}}, \tag{B22}$$

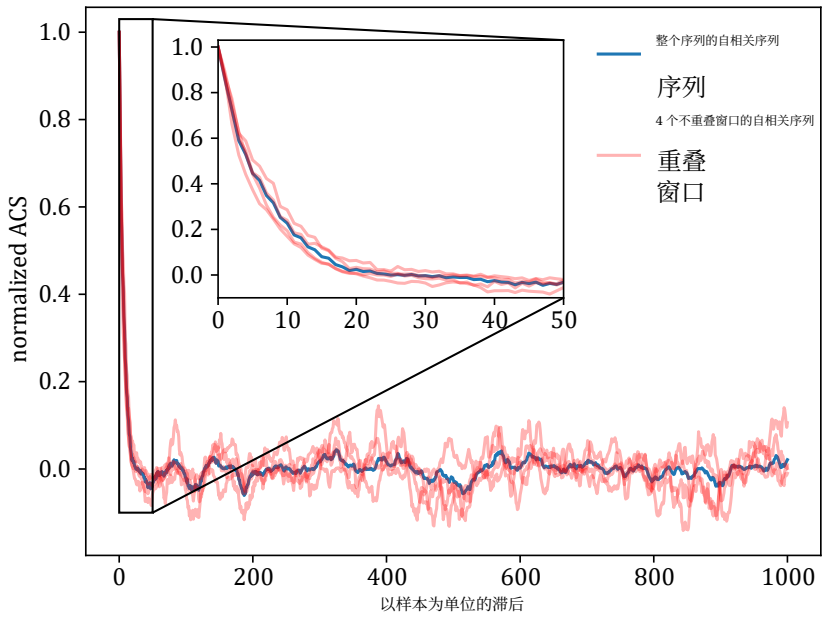


图6：磁传感器数据的稳健估计自相关序列（自相关序列）。蓝色曲线表示整个数据序列的自相关序列，红色曲线表示将原始序列划分为4个互不重叠的连续部分后得到的4个数据窗口的自相关序列。数据似乎具有二阶弱相关性且平稳。

其中 $P_{xx}(f_i)$ 被其稳健估计取代。由此，数据被变换为 $\mathbf{y} = g(\sqrt{\boldsymbol{\Sigma}^{-1}}\mathbf{x})$，其中 $g(\cdot)$ 是一个逐元素作用于其向量自变量的函数。对于“限幅函数”方法， $g(\cdot)$ 定义为

$$g(z) = \begin{cases} 3 & z > 3 \\ z & -3 < z < 3 \\ -3 & z < -3 \end{cases}, \tag{B23}$$

对于“符号函数”方法， $g(\cdot)$由下式给出

$$g(z) = \begin{cases} 1 & z > 0 \\ 0 & z = 0 \\ -1 & z < 0 \end{cases}, \tag{B24}$$

对于“无非线性”方法， $g(\cdot)$是恒等函数 $g(z) = z$。随后，计算具有观测矩阵$\mathbf{D} = \sqrt{\boldsymbol{\Sigma}^{-1}}\mathbf{H}$ 、处于白高斯噪声 [21, p. 274], 中的线性参数的广义似然比统计量

$$\mathbf{y}^\top \mathbf{D}(\mathbf{D}^\top \mathbf{D})^{-1}\mathbf{D}^\top \mathbf{y}, \tag{B25}$$

并将其与阈值进行比较。请注意，对于独立同分布非高斯噪声，将广义似然比统计量应用于得分变换数据，即可得到渐近最优的Rao检测器。

7. 神经网络训练与评估

以下将介绍神经网络训练与评估的详细内容，该过程使用深度学习库PyTorch [72]在Python中实现。

如补充信息 B5 所述，我们采用嵌套交叉验证程序。表 I 列出了各项超参数，其中一些通过在验证集上进行网格搜索来优化。学习

参数	值 / 方法
一维卷积层数	3
通道数	20
滤波器大小	3
优化器	AdamW
差商步长 $\Delta\theta$ 1e-2	
耐心早停	3个训练周期
批大小	数据集大小
学习率	1e-4
L2正则化权重	1e-5
训练周期数	早停

表 I：固定（上）和可变（下）超参数。训练周期数通过验证集（早停）进行优化。

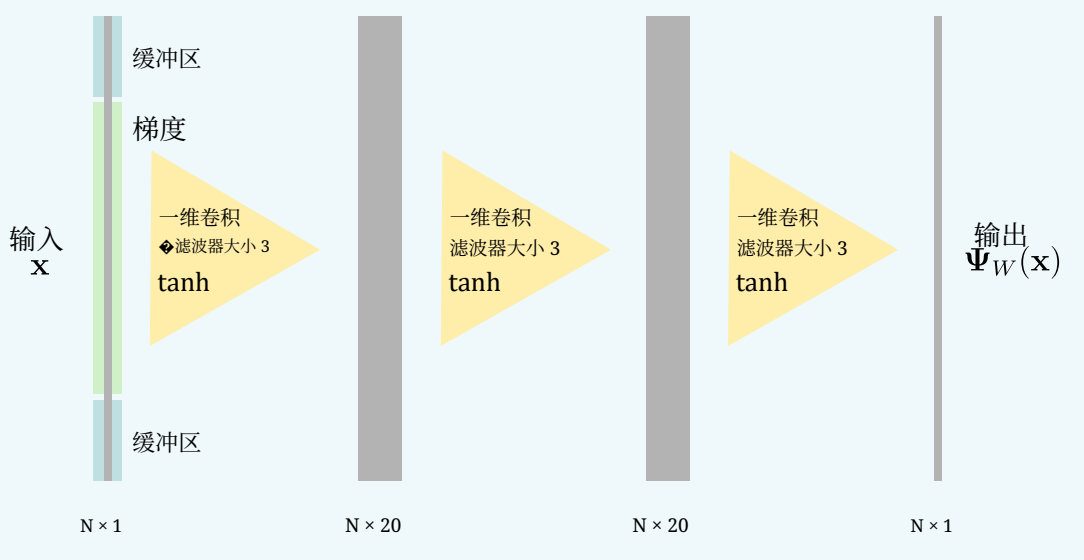


图7：卷积神经网络架构示意图。竖直的灰色条表示神经网络的各层，黄色三角形表示滤波操作和（tanh）激活函数的应用。训练期间，仅对绿色标示的缩短输入序列样本进行偏移，以减轻计算梯度时的边缘效应。

用于弱信号检测的局部最优非线性函数这一相当具体的任务，使我们能够对大多数超参数施加合理的约束。此外，由于可用数据有限，设置约束有助于在很大程度上避免过拟合。

图7展示了我们的小型卷积神经网络架构草图。受弱信号检测中典型非线性函数形状的启发 [21, 42], ，该架构类似于一种限幅函数，能够减小数据中离群值的影响，同时也能够捕捉输入数据中复杂的时间依赖关系。假设噪声分布对称，则该非线性函数为奇函数。我们通过使用 tanh 激活函数并省略一维卷积层中的偏置参数来确保这一点。总体网络规模设置得非常小，这是基于噪声分布依赖性较弱且平稳的假设；如补充材料 B 5 所述，这应该会产生复杂度有限的最优非线性函数。滤波器大小和层数均设为 3。如图8所示，小型模型能够使训练数据的 LFI 最大化，且其数值高于验证数据，表明出现了过拟合（若非早停则会如此）。因此，对于当前问题，无需使用更大的模型。

如果可用的训练数据量有限，设计一个小型、高效的网络至关重要。代价函数 (12) 具有渐近性质，因此，对于有限的序列长度 N，它只是 LFI 迹的（负）值的近似（即使在用估计的 LFI 替代 LFI 之前也是如此）。如果神经网络的架构过大，网络可能会学会无限降低代价函数，却并未真正提高 LFI（若没有早停的话）。这种潜在的不稳定性不仅由模型复杂度造成，也与代价函数中的噪声水平有关，以下将对此进行讨论。

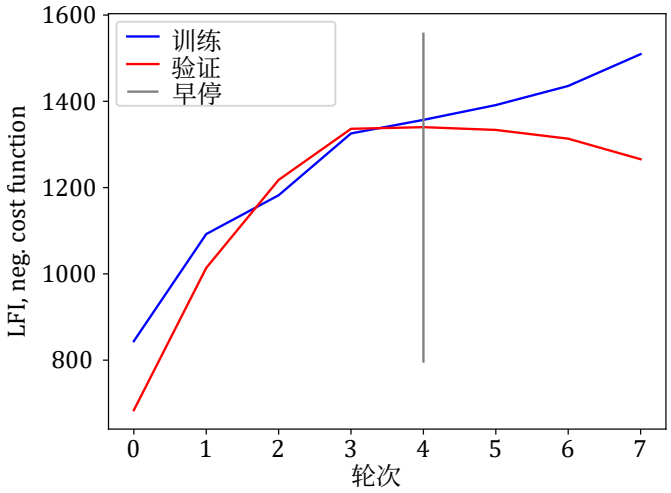


图8：早停的示例说明，用于限制过拟合对训练数据的影响，并确定网络训练的训练周期数。尽管LFI（及其迹）在理论上受费舍尔信息的约束，仍可能发生过拟合，因为我们仅在优化LFI渐近迹的一个估计。

可能最重要的单个超参数是步长 $\Delta\theta$ ，它用于通过中心差分商对均值的梯度进行数值计算。如果其值过低，LFI 迹的估计就会受到过多噪声干扰，训练过程也往往会不稳定。如果其值过高，通过中心差分商对梯度的近似就会过于粗糙，实质上我们优化的是粗LFI [37]。不过，为避免潜在的不稳定性，步长略微偏高可能优于偏低，即使这意味着在一定程度上偏离最优局部非线性。通过手动参数搜索，我们选取了 $\Delta\theta = 0.01$ 这一数值，以兼顾稳定性与中心差分商的有效性。请注意， $\Delta\theta$ 高度依赖于当前问题的具体情况。对于大量训练数据，较宽范围内的步长都可能取得良好结果。

对于学习率，我们选择能使训练过程可靠的值 10^{-4} 。这里所说的可靠，是指验证集上的负代价有一个清晰的最大值，如图8所示。

一般而言，随着训练数据集规模增大、序列长度 N 增加，过拟合和不稳定性不再那么突出。此时，由于我们不再那么依赖归纳偏置，因此可以考虑范围更广的架构。值得强调的是，即使只有 $\lfloor 10000/128 \rfloor = 78$个数据序列，我们的方法仍然有效；这些序列被划分为训练集、测试集和验证集。因此，具有在线网络自适应功能的实时应用似乎是一个很有前景的后续发展方向。这也是因为所需的反向传播步骤数相当少，见图8（由于批大小等于数据集大小，一个轮次等于一次反向传播步骤）。

[1] Mark A. Richards。 *Fundamentals of Radar Signal Processing*，第1卷。McGraw-Hill，2005年。[2] Angelo Coluccia。 *Adaptive Radar Detection: Model-Based， Data-Driven and Hybrid Approaches*。Artech House，2022年。[3] David Ram´ırez、Ignacio Santamar´ıa 和 Louis Scharf。 *Coherence: In Signal Processing and Machine Learning*。Springer Nature，2023年。[4] Adrian Bachtold、Joel Moser 和 M. I. Dykman。纳米机械系统的介观物理。 *Reviews of Modern Physics* 94 4 045005, 2022 ，（）：。

[5] Denis V Martynov、E. D. Hall、B. P. Abbott、R. Abbott、T. D. Abbott、C. Adams、R. X. Adhikari、R. A. Anderson、S. B. Anderson、K. Arai 等。引力波天文学起步阶段先进激光干涉引力波天文台探测器的灵敏度。*Physical Review D*，93(11)：112004，2016年。[6] Simon Watts 和 Luke Rosenberg。雷达海杂波建模面临的挑战。 *IET Radar， Sonar & Navigation*，16(9)：1403–1414，2022年。[7] G. V. Trunk 和 S. F. George。非高斯海杂波中的目标检测。 *IEEE Transactions on Aerospace and Electronic Systems* 5 620 628, 2007 ，（）：–。

[8] J Barrie Billingsley、Alfonso Farina、Fulvio Gini、Maria Valerina Greco 和 Lucio Verrazzani。实测雷达地杂波数据的统计分析。 *IEEE Transactions on Aerospace and Electronic Systems*，35(2)：579–593，2002年。[9] Douglas A. Abraham。 *Underwater Acoustic Signal Processing: Modeling， Detection， and Estimation*。施普林格，2019年。[10] Douglas A. Abraham、James M. Gelb 和 Andrew W. Oldag。主动声呐统计中的背景与杂波混合分布。 *IEEE Journal of Oceanic Engineering*，36(2)：231–247，2011年。[11] Ilya Starshynov、Maximilian Weimar、Lukas M. Rachbauer、G¨unther Hackl、Daniele Faccio、Stefan Rotter 和 Dorian Bouchet。复杂介质中深度学习显微成像的克拉美-罗下界无模型估计。 *Nature Photonics*，第1–8页，2025年。[12] Shahin Jafarzadeh、David B. Jess、Marco Stangalini、Samuel D. T. Grant、Jonathan E. Higham、Martin E. Pessah、Peter H. Keys、Sergey Belov、Daniele Calchetti、Timothy J. Duckenfield 等。波分析工具。 *Nature Reviews Methods Primers*，5 1 21，2025( )：。[13] Abdolreza Mohammadi、Domenico Ciuonzo、Ali Khazaee 和 Pierluigi Salvo Rossi。用于分布式稀疏信号检测的广义局部最强检验。 *IEEE Transactions on Signal and Information Processing over Networks*，8 528：–542，2022年。[14] Domenico Ciuonzo 和 P. Salvo Rossi。无线传感器网络中广义局部最优检测器的量化器设计。 *IEEE Wireless Communications Letters*，7(2)：162–165，2017年。[15] Xueqian Wang、Gang Li 和 Pramod K. Varshney。传感器网络中利用量化测量检测稀疏随机信号。 *IEEE Transactions on Signal Processing*，67(8)：2210–2220，2019年。[16] Nikolaos Karnesis、Argyro Sasli、Riccardo Buscicchio 和 Nikolaos Stergioulas。利用重尾似然刻画非高斯随机信号。 *Physical Review D*，111(2)：022005，2025年。[17] Benjamin P. Abbott、Richard Abbott、Thomas D. Abbott、Matthew R. Abernathy、Fausto Acernese、Kendall Ackley、Carl Adams、Thomas Adams、Paolo Addesso、Rana X. Adhikari 等。双黑洞并合产生引力波的观测。 *Physical Review Letters*，116(6)：061102，2016年。[18] Benjamin P. Abbott、R. Abbott、T. D. Abbott、M. R. Abernathy、Fausto Acernese、K. Ackley、M. Adamo、C. Adams、T. Adams、Paolo Addesso 等。先进激光干涉引力波天文台中与引力波信号 GW150914 相关的瞬态噪声表征。 *Classical and Quantum Gravity*，33(13)：134001，2016年。[19] Hang Yu 和 Rana X. Adhikari。利用卷积神经网络清除引力波探测器中的非线性噪声。 *Frontiers in Artificial Intelligence*，5：811563，2022年。[20] C. Radhakrishna Rao。关于涉及多个参数的统计假设的大样本检验及其在估计问题中的应用。载于 *Mathematical Proceedings of the Cambridge Philosophical Society*，第44卷，第50–57页。剑桥大学出版社，1948年。[21] Steven M. Kay。 *Fundamentals of Statistical Signal Processing: Detection Theory*。Prentice Hall，1998年。[22] Erich Leo Lehmann 和 Joseph P. Romano。 *Testing Statistical Hypotheses*。施普林格，1986年。[23] Larry Wasserman。 *All of Nonparametric Statistics*。施普林格，2006年。[24] Saleem A. Kassam。 *Signal Detection in Non-Gaussian Noise*。施普林格，2012年。[25] Abdelhak M. Zoubir、Visa Koivunen、Yacine Chakhchoukh 和 Michael Muma。信号处理中的稳健估计：基本概念教程式讲解。 *IEEE Signal Processing Magazine*，29(4)：61–80，2012年。[26] Anthony J. Fox。时间序列中的离群值。 *Journal of the Royal Statistical Society Series B: Statistical Methodology*，34(3)：350–363，1972年。[27] Si-Yi Chen、You-Wu Wang 和 Yi-Qing Ni。采用基于湮灭滤波器的汉克尔结构稳健主成分分析方法，对动态响应的结构健康监测数据进行粗大离群值剔除与故障数据恢复。 *Structural Control and Health Monitoring*，29 12( )：e3144，2022年。[28] Ronald A. Fisher。论理论统计学的数学基础。 *Philosophical Transactions of the Royal Society of London. Series A Containing Papers of a Mathematical or Physical Character* 222 594 604 309 368, 1922 ，，（–）：–。[29] Steven M. Kay。 *Fundamentals of Statistical Signal Processing: Estimation Theory*。Prentice-Hall，1993 年。[30] L. Scharf 和 C. Demeure。 *Statistical Signal Processing: Detection， Estimation， and Time Series Analysis.* Addison-Wesley，1991 年。[31] Manuel Stein、Amine Mezghani 和 Josef A. Nossek。费 Fisher 信息度量的下界。 *IEEE Signal*

Processing Letters, 21(7):796–799, 2014. [32] Manuel S. Stein、Josef A. Nossek 和 Kurt Barb´e。费歇尔信息下界及其在硬件感知非线性信号处理中的应用。 *arXiv preprint arXiv:1512.03473*，2015。[33] Petre Stoica 和 Prabhu Babu。高斯数据假设导出最大的 Cram´er-Rao 界 [讲义]。*IEEE Signal Processing Magazine*，28(3):132–133，2011。[34] Peggy S´eri`es、Peter E. Latham 和 Alexandre Pouget。方向选择性的调谐曲线锐化：编码效率及相关性的影响。 *Nature Neuroscience*，7(10):1129–1135，2004。[35] Rub´en Moreno-Bote、Jeffrey Beck、Ingmar Kanitscheider、Xaq Pitkow、Peter Latham 和 Alexandre Pouget。信息限制性相关性。 *Nature Neuroscience*，17(10):1410–1417，2014。[36] Ingmar Kanitscheider、Ruben Coen-Cagli、Adam Kohn 和 Alexandre Pouget。准确测量相关神经元群体中的费歇尔信息。 *PLoS Computational Biology*，11(6):e1004218，2015。[37] Adam Kohn、Ruben Coen-Cagli、Ingmar Kanitscheider 和 Alexandre Pouget。相关性与神经元群体信息。 *Annual Review of Neuroscience*，39(1):237–256，2016。[38] Zi-Jian Cheng、Lingxiao Yang、Wen-Hao Zhang 和 Ru-Yuan Zhang。表征几何揭示响应相关性对神经生理学与功能性磁共振成像中群体编码的差异影响。 *Journal of Neuroscience* 43 24 4498 4512, 2023 ，( ): – 。[39]

Oren Weiss 和 Ruben Coen-Cagli。通过通信子空间测量神经元群体之间的刺激信息传递。 *bioRxiv*，第 2024–11 页，2024。[40] Joel Zylberberg、Jon Cafaro、Maxwell H. Turner、Eric Shea-Brown 和 Fred Rieke。方向选择性电路塑造噪声，以确保精确的群体编码。 *Neuron*，89(2):369–383，2016。[41] Yu Hu、Joel Zylberberg 和 Eric Shea-Brown。符号规则及其拓展：神经元群体编码中的边界效应、灵活性与噪声相关性。 *PLoS Computational Biology*，10(2):e1003469，2014。[42] Iickho Song、Elbio Yong Dagotto、Jinsoo Bae 和 S. Y. Kim。 *Advanced Theory of Signal Detection*。施普林格，2002。[43] Maximilian Weimar、Lukas M Rachbauer、Ilya Starshynov、Daniele Faccio、Linara Adilova、Dorian Bouchet 和 Stefan Rotter。人工神经网络中的费 Fisher 信息流。 *Physical Review X*，15(3):031072，2025。[44] Benjamin Friedlander。到达方向估计中 Cramer-Rao 下界的基于数据的估计。载于 *2022 56th Asilomar Conference on Signals Systems and Computers* 101 105 2022 ，，，第 – 页。

IEEE，。[45] Jerome J. Sheehy。非高斯噪声中信号的最优检测。 *The Journal of the Acoustical Society of America*，63 1 81( ): –90，1978。[46] Robert B Davies。仅在备择假设下存在干扰参数时的假设检验。 *Biometrika*，74 1 33 43，1987 ( ): – 。[47] Steven M. Kay。 *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*，第 3 卷。培生教育，2013。[48] Athanasios Papoulis 和 S. Unnikrishna Pillai。 *Probability、 Random Variables、 and Stochastic Processes*。McGraw-Hill，2002。[49] Robert M. Gray。托普利茨矩阵与循环矩阵：综述。 *Foundations and Trends in Communications and Information ®* ，( ): – 。[50] *Theory* 2 3 155 239, 2006
Arthur Schuster。隐含周期性的研究及其在气象现象中一个假定的 26 天周期上的应用。 *Terrestrial Magnetism*，3(1):13–41，1898。[51] Maurice S. Bartlett。对连续谱时间序列的周期图进行平滑。 *Nature*，161(4096):686–687，1948。[52] Angelo Coluccia、Alessio Fascista 和 Giuseppe Ricci。恒虚警率特征平面：用于分析和设计雷达检测器的新框架。 *IEEE Transactions on Signal Processing*，68:3903–3916，2020。[53] Yaakov Bar-Shalom。基于相关观测所得最大似然估计的渐近性质。 *Journal of the Royal Statistical Society Series B: Statistical Methodology*，33(1):72–77，1971。[54] Justin Alsing 和 Benjamin Wandelt。广义大规模最优数据压缩。 *Monthly Notices of the Royal Astronomical Society: Letters* 476 1 60 64, 2018 ，( ):L –L 。[55]

H. L. Van Trees、K. Bell 和 Z. Tiany。 *Detection、 Estimation、 and Modulation Theory、 Part I: Detection、 Estimation、 and Filtering Theory* 2013 。威利父子出版公司，。[56] Ram Zamir。通过数据处理论证证明费歇尔信息不等式。 *IEEE Transactions on Information Theory* 44 3 1246 1250, 1998 ，( ): – 。[57]

Steven Orey。关于 m 依赖随机变量的中心极限定理。 *Duke Mathematical Journal*，25(4):543–546，1958 年 12 月。[58] Alexander Kraskov、Harald St
¨
Physical Review E: ogbauer 和 *Statistical Nonlinear and Soft Matter Physics* 69 6 066138, 2004
Peter Grassberger。互信息估计。，，，( ): 。[59] Aapo Hyv
¨
Journal of Machine arinen 和 Peter *Learning Research* 6 4 2005
Dayan。通过得分匹配估计非归一化统计模型。，( )，。[60] Christopher M. Bishop 和 Hugh Bishop。 *Deep Learning: Foundations and Concepts*。Springer Nature，2023。[61] James W. Cooley 和 John W. Tukey。用于计算机计算复数傅里叶级数的算法。*of Computation* 19 90 297 301, 1965 *Mathematics* ，( ): – 。[62] Steven Kay。 *Intuitive Probability and Random Processes using MATLAB®*。Springer Science & Business Media，2006。[63] Robert Gray。关于托普利茨矩阵的渐近特征值分布。 *IEEE Transactions on Information Theory*，18 6 725 730，1972 ( ): – 。[64] Petre Stoica 和 Randolph L. Moses。 *Spectral Analysis of Signals*，第 452 卷。Citeseer，2005。[65] Steven M. Kay。 *Modern Spectral Estimation*。培生教育印度，1988。

[66] John G. Proakis。 *Digital Signal Processing: Principles, Algorithms, and Applications*。培生教育印度，2007。[67] Mads Græsbøll Christensen 和 Andreas Jakobsson。 *Multi-Pitch Estimation*，第 5 卷。Morgan & Claypool出版社，2009。[68] Abdelhak M. Zoubir、Visa Koivunen、Esa Ollila 和 Michael Muma。 *Robust Statistics for Signal Processing*。剑桥大学出版社，2018。[69] Gavin C. Cawley 和 Nicola L. C. Talbot。论模型选择中过拟合及其后续对性能评估中选择偏差的影响。 *The Journal of Machine Learning Research*，11:2079–2107，2010。[70] Agnan Kessy、Alex Lewin 和 Korbinian Strimmer。最优白化与去相关。 *The American Statistician*，72 4 309 314，2018 ( ): – 。[71] R. Douglas Martin 和 David J. Thomson。稳健抗干扰谱估计。 *Proceedings of the IEEE*，70(9):1097–1115，1982。[72]Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga、Alban Desmaison、Andreas Kopf、Edward Yang、Zachary DeVito、Martin Raison、Alykhan Tejani、Sasank Chilamkurthy、Benoit Steiner、Lu Fang、Junjie Bai 和 Soumith Chintala。PyTorch：一种命令式风格、高性能的深度学习库。 *arXiv preprint arXiv:1912.01703*，2019。