arXiv:2601.11461v1 [stat.CO] 16 Jan 2026

平滑 SCAD：用于小波去噪的升余弦型 SCAD 阈值规则

Radhika Kulkarni[1], Aluisio Pinheiro[2], Brani Vidakovic[1], 和 Abdourrahmane M. Atto[3]

[1]德克萨斯农工大学统计系，美国德克萨斯州大学城，邮编 77843[2]坎皮纳斯州立大学统计系，巴西坎皮纳斯，邮编 13083-859[3]萨瓦-勃朗峰大学阿讷西-尚贝里理工学院，法国阿讷西，邮编 74944

2026年1月19日

摘要

我们通过将 SCAD 阈值规则中的分段线性过渡替换为升余弦，提出了一种用于小波去噪的平滑变体。由此得到的收缩函数为奇函数，在 $\mathbb{R}$ 上连续，且除主阈值外处处连续可微，同时保留了 SCAD 的标志性特性：小系数具有稀疏性，大系数则近乎无偏。这种光滑性使该规则属于连续阈值化类，因此 Stein 无偏风险估计适用于该规则。由此，无偏风险计算、稳定的数据驱动阈值选择以及 Kudryavtsev 和 Shestakov 的渐近理论均可应用。我们还构造了相应的非凸先验，其后验众数与该估计量一致，从而提供了清晰的贝叶斯解释。我们给出明确的 SURE 风险表达式，讨论最优阈值的预言机尺度，并介绍全局自适应版本和依赖层级的自适应版本。因此，平滑 SCAD 规则为 SCAD 提供了一种易于处理的改进，在单一小波收缩程序中兼具低偏差、精确稀疏性和解析上的便利。

关键词： 平滑 SCAD 阈值化；小波去噪；Stein 无偏风险估计（SURE）；非凸收缩；贝叶斯后验众数。

1 引言

小波阈值处理是非参数噪声去除和信号恢复的重要工具。经过正交小波变换后，经验系数 $d_{j,\ k}$ 服从高斯序列模型

$$d_{j,k} \sim N(\theta_{j,k}, \sigma^2), \tag{1}$$

而任务是估计未知系数 $\theta_{j,\ k}$ 同时保持估计量稳定，并使其对真正较大的系数基本无偏。自 Donoho 和 Johnstone 的研究以来，人们已研究了范围广泛的小波阈值规则，从简单的硬阈值和软阈值规则，到受惩罚似然和贝叶斯思想启发的非凸规则；例如参见 Antoniadis 和 Fan（2001年）；Antoniadis 和 Sapatinas（2007年）；Antoniadis 等（2007年）。

Fan 和 Li（2001年）提出的 SCAD 惩罚是一项影响深远的贡献，最初用于采用非凹惩罚似然进行变量选择。SCAD 旨在兼具三种性质：通过精确的零值区域实现稀疏性、估计量的连续性，以及对较大系数近乎无偏。这些特点使其对应的阈值非常适合小波收缩，正如 Antoniadis 和 Fan（2001年）所强调的那样。SCAD 阈值在主阈值附近的行为类似于软阈值处理，而对较大系数则表现为恒等映射，从而在偏差和方差之间实现实用的折中。这一构造启发了其他相关的非凸惩罚方法，包括极小极大凹惩罚（Zhang，2010年）、单步稀疏估计量（Zou 和 Li，2008年），以及基于坐标下降的方法，例如 SparseNet 及其变体（Mazumder 等，2011年；Breheny 和 Huang，2011年；Huang 等，2013年）。在高维和超高维数据的小波特征筛选中，SCAD 和其他阈值处理方法同样是值得关注的问题（Fonseca 等，2024年）。

在小波与逆问题领域，连续且非凸的阈值规则已在多种情形下得到研究。例如，线性逆问题中的小波正则化和迭代阈值算法（Daubechies 等人，2004年）、稳健小波收缩和贝叶斯改进方法（Sardy 等人，2001年；Sardy 和 Bruce，2003年），以及用于去噪和回归的非凸惩罚估计量（Zhou 和 Shen，2008年；Li 等人，2011年）。此外，也有人提出平滑截断和光滑的 ℓ_0–型惩罚，以获得具有良好稀疏性特征的连续收缩规则（Xue，2009年；Gasso 等人，2011年）。这些研究成果凸显了收缩函数的连续性和光滑性对于数值稳定性和统计性能的益处。

从这一角度看，经典 SCAD 规则适用于各种应用，但在分析上存在一个小小的局限。由此得到的收缩函数是连续的，但在其内部折点处并非连续可微。这不会造成数值上的困难；然而，这使该规则无法纳入 Kudryavtsev 和 Shestakov（2016年；2020年；2024年）提出的连续 SURE 框架，我们将其统称为 KS 框架。他们的理论通过 Stein 恒等式，为一大类具有如下形式的连续阈值规则提供无偏风险估计

$$\rho_h(d,\lambda) = \begin{cases} d - h(d,\lambda), & |d| > \lambda, \\ 0, & |d| \le \lambda, \end{cases}$$

其中，辅助函数 $h(d,\ \lambda)$ 满足三个结构性要求：(i) 奇对称性$h(-d,\ \lambda) = -h(d,\ \lambda)$；(ii) 有界性 $0 \le h(d,\ \lambda) \le \lambda$对 $d \ge 0$成立；以及 (iii) 阈值匹配$h(\lambda,\ \lambda) = \lambda$。此外，KS 施加了形式为 $|\partial h/\partial d| \le C\lambda^{\mu}$、$|\partial h/\partial \lambda| \le C\lambda^{\mu}$ 和 $|\partial^2 h/(\partial d\,\partial \lambda)| \le C\lambda^{\mu}$ 的一致导数界，从而确保 SURE 所选阈值的渐近理论成立。由于 SCAD 收缩剖面存在折点，从形式上看，它恰好超出了其渐近分析所要求的假设范围；而其他框架，例如 Sigmoid 收缩（Atto 等人，2008年），已经实现了稳健的 SURE 风险估计，并利用该风险推导出最优的 Sigmoid 参数（Atto 等人，2009年）。

本说明旨在将 SCAD 的核心思想确切地置于连续 SURE 可容许类之中。我们通过将λ与 $a\lambda$ 之间的线性转变替换为升余弦，提出一种平滑变体。这样得到的收缩量 $h(d,\ \lambda)$ 具有奇对称性且一致有界，并且在整个过渡区域 $\lambda < |d| < a\lambda$ 内连续且连续可微（按照 KS 框架的要求，在连接点 $|d| = \lambda$ 和 $|d| = a\lambda$处也连续）。

所得估计量在系数较小和较大的区间表现得类似 SCAD，但在二者之间是平滑的。由于余弦过渡在 $\lambda < |d| <a\lambda$上具有有界导数，且在该区域之外 h 为常数/零，该规则满足结构条件（奇对称性、有界性以及在 $|d| = \lambda$处的连续性）和 Kudryavtsev 和 Shestakov（2024年）所要求的导数界。具体而言，平滑 SCAD 规则属于他们的连续阈值函数容许类；对于该类，已有风险界、SURE 风险的渐近正态性以及强相合性结果。

小波域长期以来一直是连续阈值和半软阈值方案的自然应用场景。稳健自适应方法（Sardy 等人，2001年；Sardy 和 Bruce，2003年）、非凸惩罚（Zhou 和 Shen，2008年；Li 等人，2011年）以及经验贝叶斯构造（Figueiredo 和 Nowak，2001年）都利用了小波系数的稀疏性或近似稀疏性。

平滑 SCAD 遵循同样的理念。它保留了 SCAD 的理想特性，即稀疏性、连续性以及大系数下近乎无偏的性质，同时消除了会增加解析处理难度的斜率突变。

明确的贡献。 尽管 SCAD 已被反复倡用于小波收缩，但此前尚无研究证明某种连续可微的 SCAD 型规则满足连续 SURE 框架的全部可容许性条件。本文填补了这一空白。本文的主要贡献如下：

(i) 构造了一个升余弦平滑 SCAD 生成函数 $h(d, \lambda)$，并证明它满足 KS 框架的结构条件，同时具有一致导数界，从而使一种带有升余弦过渡的 SCAD 型规则纳入连续 SURE 类。

(ii) 推导出专用于平滑 SCAD 的显式 SURE 公式，从而能够针对任意固定阈值精确计算无偏风险。

(iii) 给出一种贝叶斯解释，该解释基于一个光滑非凸惩罚项，其后验众数能够精确复现该估计量。

(iv) 建立了阈值选择方法：全局 SURE 选择、在 $\sqrt{\log\ N}$ 尺度下的预言机渐近性质、简单的解析近似，以及依赖层级的自适应版本，使平滑 SCAD 与 SureShrink 等基准规则处于同一 SURE 方法框架之下。

在贝叶斯方面，我们识别出一种明确的平滑 SCAD 先验，其尾部实际上是平坦的，并且具有闭式惩罚项；其后验众数与升余弦阈值重合，并揭示了非零众数产生的确切位置 $|d| = \lambda$ 。

本文结构安排如下。第2节介绍小波模型和 SURE 框架。第3节定义平滑 SCAD，并给出升余弦构造。第4节验证 SURE 可容许性所需的结构假设和导数界。第5节推导专门的 SURE 表达式，并总结由此得到的渐近行为。第6节阐释贝叶斯解释和诱导先验。第7节讨论阈值和参数选择，包括预言机渐近性、由 SURE 驱动的调参、显式解析公式以及层级相关扩展。第8节概述数值示例。第9节总结全文。

2 模型与 SURE 框架

我们直接在小波域中开展分析。对长度为 N 的含噪信号进行正交离散小波变换后，经验小波系数满足

$$d_{j,k} = \theta_{j,k} + \epsilon_{j,k}, \quad \epsilon_{j,k} \sim N(0, \sigma^2), \tag{2}$$

其中 $\epsilon_{j,\,k}$ 相互独立。对 $(j,\ k)$ 对进行索引，分别表示分辨率层级 $j = J - L,\ \ldots\ldots,$ $J-1$ 和位置 $k = 0,\ \ldots\ldots,\ 2^j - 1$，而 $\theta_{j,\,k}$ 表示未知的真实系数。这里，$N = 2^J$ 和 L 表示小波变换的深度。在 L 细节多分辨率层级上，使用函数 $\rho(d,\ \lambda)$ 逐系数进行阈值化处理，其全局阈值为 $\lambda > 0$。根据 KS 框架，通过满足三个结构性质的辅助函数 $h(d,\ \lambda)$ 定义 ρ 更为方便。首先，h 是奇函数，

$$h(-d, \lambda) \;=\; -h(d, \lambda), \tag{3}$$

其次，对于 $d \geq 0$，它是有界的，

$$0 \leq h(d, \lambda) \leq \lambda, \tag{4}$$

第三，它满足阈值匹配条件，

$$h(\lambda, \lambda) \;=\; \lambda. \tag{5}$$

给定 h，收缩通过以下方式实现

$$\rho_h(d, \lambda) \;=\; \begin{cases} 0, & |d| \leq \lambda, \\ d - h(d, \lambda), & |d| > \lambda, \end{cases} \tag{6}$$

该公式包含软阈值处理、双曲正切和 S 形函数规则、混合规则，以及非凸 SCAD 型方法。

在分辨率深度 L下进行小波阈值处理的均方误差（风险）为

$$R_J(\lambda) \;=\; \sum_{j=J-L}^{J-1} \sum_{k=0}^{2^j-1} \mathbb{E}\Big(\rho_h(d_{j,k}, \lambda) - \theta_{j,k}\Big)^2. \tag{7}$$

由于在我们的构造中，h 关于 d 在 $|d| > \lambda$ 处可微，Stein恒等式给出了式 (7) 的无偏估计量。定义

$$g(d, \lambda) \;=\; d - \rho_h(d, \lambda), \tag{8}$$

它满足

$$g(d,\lambda) \;=\; \begin{cases} d, & |d| \le \lambda, \\ h(d,\lambda), & |d| > \lambda. \end{cases} \tag{9}$$

SURE估计量为

$$\widehat{R}_J(\lambda) \;=\; \sum_{j=J-L}^{J-1} \sum_{k=0}^{2^j-1} F(d_{j,k},\lambda), \tag{10}$$

其中

$$F(d,\lambda) \;=\; \begin{cases} d^2-\sigma^2, & |d| \le \lambda, \\ h(d,\lambda)^2+\sigma^2-2\sigma^2\dfrac{\partial h}{\partial d}(d,\lambda), & |d| > \lambda. \end{cases} \tag{11}$$

SURE 选择的阈值定义为

$$\lambda_S \;=\; \arg\min_{\lambda\in[0,\lambda_U]} \widehat{R}_J(\lambda), \tag{12}$$

其中 λ_U表示阶为 $\sigma\sqrt{2\log N}$ 的通用上界。

KS 框架为 $\widehat{\lambda_S}$和 $R_J(\lambda_S)$ 提供风险界、渐近正态性和相合性，只要 h 满足 (3)–(5)，且其导数满足温和的正则性条件。第3节中，我们将介绍一种平滑 SCAD 阈值化规则，并在第4节中验证它属于这一容许类。

3 通过升余弦实现平滑 SCAD

我们构造了一种具有 SCAD 定性行为的收缩规则，但以平滑转变取代了其分段线性部分。我们引入较低阈值 $\lambda > 0$和上乘法因子 $a > 1$，后者控制过渡区的宽度。在全文中，$a > 1$ 均视为固定值。

平滑 SCAD 通过辅助函数 $h(d,\ \lambda)$ 定义，该函数通过（6）生成阈值规则 ρ_h。我们首先为 $d \ge 0$ 指定 h，然后利用奇对称性将其扩展到 $\mathbb{R}$。对于 $d \ge 0$，我们设

$$h(d,\lambda) \;=\; \begin{cases} \lambda, & 0 < d \le \lambda, \\ \lambda\cos^2\left(\dfrac{\pi}{2}\,\dfrac{d-\lambda}{(a-1)\lambda}\right), & \lambda < d < a\lambda, \\ 0, & d \ge a\lambda, \end{cases} \tag{13}$$

其中 $h(0,\ \lambda) = 0$。用以下内容补充 h 的定义：

$$h(-d,\lambda) \;=\; -h(d,\lambda), \qquad d \geq 0, \tag{14}$$

保证 h 在 $\mathbb{R}$上为奇函数。选择分支 $h(d,\ \lambda) = \lambda$ for $0 < d \leq \lambda$ 以满足$h(\lambda,\ \lambda) = \lambda$ 并简化诱导惩罚项积分；这不会影响阈值规则本身，因为 $\rho_h(d,\ \lambda) = 0$ for $|d| \leq \lambda$。特别地，h 在 $d = 0$ 处的任何不连续性对估计量都无关紧要，因为 $\rho_h(d,\ \lambda) = 0$ for $|d| \leq \lambda$，而 KS 正则性分析只涉及 h for$|d| > \lambda$ 时的行为。

为解释（13），定义归一化过渡参数

$$s(d,\lambda) \;=\; \frac{d-\lambda}{(a-1)\lambda}, \qquad \lambda < d < a\lambda,$$

使得 $\lambda < d < a\lambda$ 对应于 $0 < s < 1$。则

$$h(d,\lambda) \;=\; \lambda\cos^2\!\Big(\frac{\pi}{2}s(d,\lambda)\Big), \qquad \lambda < d < a\lambda,$$

该函数从 $d = \lambda$ 处的 λ 平滑下降到 $d = a\lambda$处的 0。

给定 h，平滑 SCAD 阈值函数为

$$\rho_{\mathrm{ssc}}(d,\lambda) \;=\; \begin{cases} 0, & |d| \leq \lambda, \\ d - h(d,\lambda), & \lambda < |d| < a\lambda, \\ d, & |d| \geq a\lambda. \end{cases} \tag{15}$$

因此，ρ_{ssc} 具有以下特性：(i) 在主阈值 λ以下严格为零；(ii) 在区域 $\lambda < |d| < a\lambda$内进行平滑插值；(iii) 当 $|d| \geq a\lambda$时与恒等映射相同。

如果将 (13) 中的余弦过渡替换为在 $[\lambda, a\lambda]$ 上从 λ 到 0 的线性斜坡，所得的 ρ_h 便会复现 Fan 和 Li（2001年）的经典 SCAD 规则。因此，平滑 SCAD 可视为 SCAD 的可微类比，具有相同的小系数和大系数行为，但两者之间的转变是平滑的。

函数 ρ_{ssc} 继承了（13）–（14）中的奇对称性和连续性。特别地，ρ_{ssc} 在 $\mathbb{R}$ 上连续，在 $|d| \notin \{\lambda, a\lambda\}$ 上连续可微，并满足 $\rho_{\mathrm{ssc}}(d,\lambda) = d$当$|d| \geq a\lambda$时成立，因此在大系数区间内具有完全无偏性。下一步是验证 h 满足 SURE 理论在 KS 框架中所要求的假设（3）–（5）以及可微性和有界性条件。我们将在下一节总结这些性质。

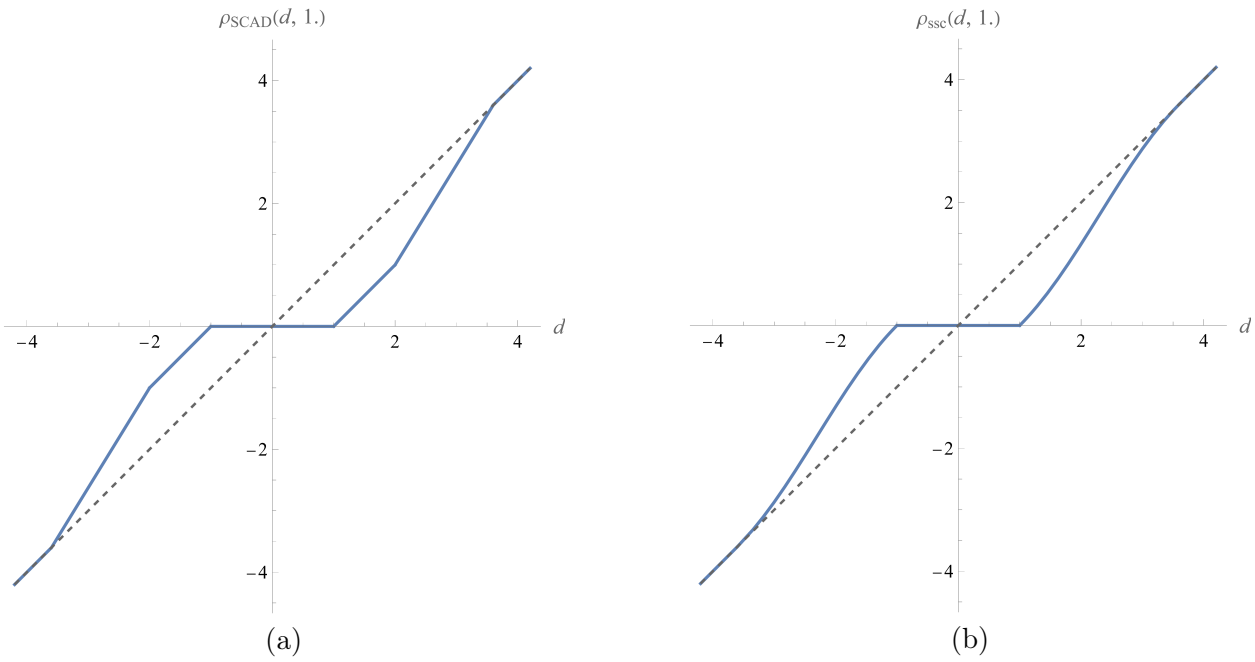


图1： $\lambda = 1$对应的经典 SCAD（a）和平滑 SCAD（b）阈值函数。

4 结构假设与正则性条件的验证

我们验证了（13）–（15）中定义的平滑 SCAD 构造满足连续 SURE 框架所要求的假设。所有符号和定义均沿用第 2 节和第 3 节中的介绍。

4.1 结构条件

可容许的 SURE 类由对收缩生成函数 h 的三项要求定义：奇对称性、在 $[0, \infty)$ 上有界，以及在主阈值 $d = \lambda$ 处连续（条件 (3)–(5)）；不要求 h 全局连续，因为对于 $|d| \leq \lambda$，$\rho_h(d, \lambda) = 0$，且导数条件仅施加于 $|d| > \lambda$。

奇对称性可由扩展规则 (14) 直接得出。对于 $d \geq 0$，(13) 中的余弦项满足 $0 \leq \cos^2(\cdot) \leq 1$，因此

$$0 \leq h(d, \lambda) \leq \lambda, \qquad d \geq 0,$$

因此有界性成立。

$d = \lambda$ 处的连续性可由以下方式验证：

$$\begin{aligned} \lim_{d \downarrow \lambda} h(d, \lambda) &= \lambda \cos^2(0) = \lambda, \\ \lim_{0 < d \uparrow \lambda} h(d, \lambda) &= \lambda, \end{aligned}$$

因此 $h(\lambda, \lambda) = \lambda$，且阈值处的连续性成立。由于 $h(a\lambda, \lambda) = 0$ 来自余弦分支，而 $h(d, \lambda) = 0$ 对于 $d \geq a\lambda$成立，因此 $d = a\lambda$ 处的连续性显而易见。

因此，平滑 SCAD 满足 SURE 理论的结构条件。

4.2 导数界

KS 框架对 h 在 d 和 λ中的导数施加有界性条件，即

$$\left|\frac{\partial h}{\partial d}(d, \lambda)\right| \leq C\lambda^{\mu}, \quad \left|\frac{\partial h}{\partial \lambda}(d, \lambda)\right| \leq C\lambda^{\mu}, \quad \left|\frac{\partial^2 h}{\partial d \partial \lambda}(d, \lambda)\right| \leq C\lambda^{\mu}, \tag{16}$$

其中常数 $C > 0$ 和 $\mu \geq 0$存在。只需计算 $d > 0$的这些导数；由对称性，利用式 (14) 可得到 $d < 0$的相应界。KS 框架不要求连接点 $d = \pm\lambda$, $\pm a\lambda$ 处的导数，因为它仅在开区域 $|d| > \lambda$上施加导数界。

在区域 $0 < |d| \leq \lambda$ 和 $|d| \geq a\lambda$上， h 为常数或零，因此

$$\frac{\partial h}{\partial d} = \frac{\partial h}{\partial \lambda} = \frac{\partial^2 h}{\partial d \partial \lambda} = 0.$$

在过渡区域 $\lambda < d < a\lambda$,

$$h(d, \lambda) = \lambda \cos^2(\alpha\, s(d, \lambda)), \qquad \alpha = \frac{\pi}{2}, \qquad s(d, \lambda) = \frac{d - \lambda}{(a-1)\lambda}.$$

对 d 求导，得到

$$\frac{\partial h}{\partial d}(d, \lambda) = -\frac{\alpha}{a-1} \sin\big(2\alpha\, s(d, \lambda)\big),$$

因此

$$\left|\frac{\partial h}{\partial d}(d, \lambda)\right| \leq \frac{\alpha}{a-1} = \frac{\pi}{2(a-1)},$$

其关于 d和 λ是一致的。因此，KS 条件所要求的导数界在指数 $\mu = 0$ 下成立。

对于 $\partial h/\partial\lambda$，对 $h(d,\ \lambda) = \lambda\cos^2(\alpha s(d,\ \lambda))$ 求导，并结合 $s(d,\ \lambda) = (d-\lambda)/((a-1)\lambda)$，可得到有限个如下形式的项

$$\cos(\alpha s), \quad \sin(\alpha s), \quad \frac{d-\lambda}{\lambda}, \quad \frac{d}{\lambda},$$

乘以仅取决于 a 的常数。在过渡区域 $\lambda < d < a\lambda$ 上，有

$$1 < \frac{d}{\lambda} < a, \qquad 0 < \frac{d-\lambda}{(a-1)\lambda} < 1,$$

因此，所有此类项均一致有界，且与 d 和 λ 无关。于是

$$\left|\frac{\partial h}{\partial\lambda}(d,\lambda)\right| \le C(a).$$

混合导数 $\partial^2 h/(\partial d\,\partial\lambda)$ 涉及相同的有界因子，因此满足

$$\left|\frac{\partial^2 h}{\partial d\,\partial\lambda}(d,\lambda)\right| \le C(a).$$

因此，正则性条件 *(16)* 在指数为 $\mu = 0$ 时成立。

总之，平滑 SCAD 规则属于 KS 框架中的可容许连续阈值类。其 SURE 风险、SURE 选择的阈值以及渐近行为，均由针对该框架建立的一般理论所支配。

5 平滑 SCAD 的风险、SURE 与渐近行为

对于平滑 SCAD 规则，oracle 风险的定义与式 (7) 相同，其中 $\rho_h(d,\lambda) = \rho_{\mathrm{ssc}}(d,\lambda)$ 来自式 (15)。由于 $R_J(\lambda)$ 取决于未知系数 $\theta_{j,k}$，因此无法以闭式形式求得，但其无偏估计量可直接由式 (10)–(11) 得出。将式 (13)–(14) 中的 h 代入一般表达式，可得

$$\widehat{R}_J(\lambda) \;=\; \sum_{j=J-L}^{J-1}\sum_{k=0}^{2^j-1} F_{\mathrm{ssc}}(d_{j,k},\lambda), \tag{17}$$

其中

$$F_{\mathrm{ssc}}(d,\lambda) \;=\; \begin{cases} d^2-\sigma^2, & |d| \le \lambda, \\ h(d,\lambda)^2 + \sigma^2 - 2\sigma^2\dfrac{\partial h}{\partial d}(d,\lambda), & \lambda < |d| < a\lambda, \\ \sigma^2, & |d| \ge a\lambda, \end{cases} \tag{18}$$

中间区域上的 h 和 $\partial h/\partial d$ 均为显式表达式：

$$\begin{aligned} h(d,\lambda) &= \lambda\cos^2\!\left(\frac{\pi}{2}\,\frac{d-\lambda}{(a-1)\lambda}\right), \qquad \lambda < d < a\lambda, \\ \frac{\partial h}{\partial d}(d,\lambda) &= -\frac{\pi}{2(a-1)}\sin\!\left(\pi\frac{d-\lambda}{(a-1)\lambda}\right), \end{aligned}$$

对于负的 d，适用奇延拓（14）。因此，对于 $(d_{j,\,k})$ 的任何取值，均可无须近似地数值计算（17）。

SURE 选择的阈值为

$$\lambda_S \;=\; \arg\min_{\lambda\in[0,\lambda_U]} \widehat{R}_J(\lambda), \tag{19}$$

其中 λ_U 是通常的、阶为 $\sigma\sqrt{2\log N}$ 的通用上界。实际应用中，可在网格上计算 $\widehat{R}_J(\lambda)$，并选择使其最小的 λ。

由于平滑 SCAD 满足结构要求（3）–（5）以及导数界（16），其中 $\mu = 0$，因此 Kudryavtsev 和 Shestakov（2016年）；Shestakov（2020年）；Kudryavtsev 和 Shestakov（2024年）的渐近结果均适用。特别地：

定理1（Lipschitz正则性下的渐近行为）。 假设底层函数 f 的支集位于有限区间上，且具有指数为 $\gamma > 1/2$的统一利普希茨正则性。那么，其真实小波系数满足通常形式的稀疏性诱导衰减条件。对于固定 $a > 1$ 的平滑 SCAD 收缩，预言机最优风险 $R_J(\lambda_{\min})$ 满足与一般连续 SURE 类相同阶的上界；特别地，存在一个依赖于 γ 和 a 的常数 C，使得

$$R_J(\lambda_{\min}) \;\le\; C\,2^{J/(2\gamma+1)}\,J^{(2\gamma+2)/(2\gamma+1)}. \tag{20}$$

精确的常数和对数因子直接继承自 KS 分析，并且仅取决于 γ 和 a。

此外，SURE 选择的阈值 λ_S 以趋近于 1 的概率避免了次优的小值；当 $J \to \infty$ 时，$R_J(\lambda_S)$ 与 $R_J(\lambda_{\min})$ 同阶，并且中心化的 SURE 风险具有渐近正态极限和强相合性（Kudryavtsev 和 Shestakov，2024年）。

该定理概括了平滑 SCAD 从连续 SURE 可容许类中继承的性质。

6 平滑 SCAD 的贝叶斯解释

贝叶斯方法论为小波域推断提供了自然且连贯的框架。在这一方法中，贝叶斯规则充当收缩算子：先验设定决定系数衰减的程度，而损失函数决定规则的函数形式。在二次损失下，后验均值产生平滑收缩；其他损失结构则会导出阈值化类型的非线性收缩器，通常表现为后验众数或中位数。因此，贝叶斯收缩在统一范式下涵盖连续和选择性系数抑制，并由对信号和噪声的显式概率建模加以指导。

贝叶斯小波文献沿袭了这一思路。Vidakovic（1994年）引入了用于小波情境中自适应收缩的贝叶斯规则和贝叶斯因子。随后，Cutillo 等人（2008年）以及 Vidakovic 和 Ruggeri（2001年）提出了一种明确的规则，通过闭式后验众数展现出类似阈值的行为；Reményi 和 Vidakovic（2013年）对其扩展进行了综述。Vidakovic（1999年）为基于模型的小波分析提供了更为广泛的基础，将小波收缩置于多尺度推断框架之中。总体而言，这些工作确立了贝叶斯收缩在噪声降低、平滑和多尺度结构提取方面作为一种具有原则性的途径。

现在，我们将平滑 SCAD 纳入一个简单的贝叶斯框架。为清晰起见，我们先处理单个系数，然后说明如何将其扩展到完整的小波向量。

考虑正态均值模型

$$d \mid \theta \quad \sim \quad N(\theta, \sigma^2), \tag{21}$$

并回顾光滑 SCAD 估计量为

$$\delta_{\mathrm{ssc}}(d, \lambda) \quad = \quad \rho_{\mathrm{ssc}}(d, \lambda), \tag{22}$$

其中， ρ_{ssc} 的结构见式（15），并取决于式（13）–（14）中的 h 。目标是构造先验 $p(\theta)$，使得后验众数满足

$$\hat{\theta}(d, \lambda) = \rho_{\mathrm{ssc}}(d, \lambda).$$

6.1 由收缩轮廓诱导的先验

通过对收缩轮廓 h进行积分，可以建立平滑 SCAD 与贝叶斯最大后验估计之间的联系。将 $r = |\theta|$ 上的惩罚函数定义为

$$\Phi(r; \lambda) \quad = \quad \int_0^r h(u, \lambda)\, du, \qquad r \geq 0, \tag{23}$$

由于 h 有界且其支集为 $[0, a\lambda]$，该函数对所有 r 均为有限值。诱导先验为

$$p(\theta \mid \lambda, \sigma^2) \;\propto\; \exp\left\{-\frac{1}{\sigma^2}\,\Phi(|\theta|;\lambda)\right\}. \tag{24}$$

该先验关于零对称，且在零处为单峰。由于 $h(u,\lambda) = 0$ 对于 $u \geq a\lambda$，惩罚项$\Phi(r;\lambda)$在超过 $r = a\lambda$ 后变为常数，因此诱导先验在该区域实际上具有平坦尾部。因此，大系数实际上不受惩罚；按照标准的惩罚似然意义，应将该先验解释为不可积的 MAP 诱导先验，而非正规的概率分布。

结合 (21) 和 (24)，负对数后验（略去常数项）为

$$\mathcal{L}(\theta \mid d, \lambda) \;=\; \frac{(d-\theta)^2}{2\sigma^2} + \frac{1}{\sigma^2}\,\Phi(|\theta|;\lambda). \tag{25}$$

对于 $\theta \neq 0$，求导可得

$$\frac{\partial \mathcal{L}}{\partial \theta}(\theta \mid d, \lambda) \;=\; -\frac{d-\theta}{\sigma^2} + \frac{1}{\sigma^2}\,\Phi'(|\theta|;\lambda)\,\operatorname{sign}(\theta). \tag{26}$$

由于 $\Phi'(r;\lambda) = h(r,\lambda)$，由驻点条件可得

$$d - \theta \;=\; h(|\theta|, \lambda)\,\operatorname{sign}(\theta). \tag{27}$$

这就是刻画非零区域中 ρ_{ssc} 的不动点方程：事实上，

$$d - \rho_{\mathrm{ssc}}(d,\lambda) = h(d,\lambda) \quad \text{for } \lambda < |d| < a\lambda,$$

^ 且对于 $|d| \geq a\lambda$ ，右侧为零，得到 $\theta = d$。

剩余情况 $|d| \leq \lambda$ 对应于后验在零附近的行为。在 $(0,\ \lambda]$上，有 $h(\theta,\ \lambda) = \lambda$，因此 $\mathcal{L}(\theta \mid d,\ \lambda)$ 在 $\theta = 0$ 处的右导数等于 $(-d+\lambda)/\sigma^2$。因此，$\theta = 0$ 是后验最大化点，当且仅当 $|d| \leq \lambda$，并且估计量在此区间上恰好为零。

这种构造将对 λ 的基于 SURE 的调参与光滑非凸先验下完全确定的 MAP 估计量联系起来，而不是将阈值视为纯粹的频率学调参参数。

6.2 惩罚项、先验与后验

式（23）中的惩罚项 Φ 可以显式写出。将式（13）中的 h 代入后，可得

$$\Phi(r;\lambda) \;=\; \begin{cases} \lambda r, & 0 \leq r \leq \lambda, \\ \lambda^2 + \dfrac{(a-1)\lambda^2}{\pi}\left[T + \dfrac{1}{2}\sin(2T)\right], & \lambda < r < a\lambda, \\ \dfrac{a+1}{2}\lambda^2, & r \geq a\lambda, \end{cases} \tag{28}$$

其中 $T = \dfrac{\pi \quad r-\lambda}{-2 \quad a \quad -1}$ ( )λ。将其代入式（24）即可得到平滑 SCAD 先验密度，结合似然与先验可得未归一化后验

$$\pi(\theta \mid d, \lambda, \sigma^2) \quad \propto \quad \exp\left\{ -\frac{1}{2\sigma^2}(d-\theta)^2 - \frac{1}{\sigma^2}\,\Phi(|\theta|;\lambda) \right\}. \tag{29}$$

ˆ 后验众数满足 (27)。当 $|d|$ 较小时，不存在非零解，且 $\theta = 0$；一旦$|d|$ 超过模型临界值 $|d|_{\text{crit}} = \lambda$，非零的最大化分支便会出现，并在尾部连续地趋向 d。

6.3 后验众数离开零点的阈值

为了分析　　这一转变，在 $\theta = 0$处进行局部分析。在 (0, λ], $h(\theta, \lambda) = \lambda$，因此由 (25) 得

$$\frac{\partial}{\partial\theta}\mathcal{L}(\theta \mid d, \lambda) \quad = \quad \frac{\theta - d}{\sigma^2} + \frac{\lambda}{\sigma^2},$$

零点处的右导数为

$$\left.\frac{\partial}{\partial\theta}\mathcal{L}(\theta \mid d, \lambda)\right|_{\theta\downarrow 0} \quad = \quad \frac{-d+\lambda}{\sigma^2}.$$

因此，当 $d < \lambda$时，后验在零点右侧立即递减；当 $d = \lambda$时，后验保持平坦；当 $d > \lambda$时，后验递增。直接比较 $\mathcal{L}(\theta \mid d, \lambda)$ 在 $\theta = 0$ 和 $\theta = \varepsilon > 0$ 处的取值可知，只要 $d > \lambda$，那么在一个较小的正 θ 处，后验值就严格更大；因此全局最大化点会在以下位置离开零点

$$|d|_{\text{crit}} \quad = \quad \lambda, \tag{30}$$

独立于 a 和 σ^2。对于 $d > \lambda$，非零模态满足 $\theta + h(\theta, \lambda) = d$；对于$d < -\lambda$，解则为其相反数。

6.4 推广到完整小波模型

在小波模型 (2) 中，可以对系数 $\{\theta_{j,k}\}$ 赋予 (24) 型的独立先验。由于高斯似然和平滑 SCAD 先验均可分解，负对数后验可拆分为单变量项，而后验众数则通过对每个坐标应用 ρ_{ssc} 得到。第 5 节中的 SURE 分析随后为这一贝叶斯估计量提供频率学派风险刻画；由于平滑 SCAD 满足第4节中验证的结构条件和正则性条件，因此 KS 框架的渐近保证也适用。

注（零点处的最大后验与次梯度解释）。由于平滑 SCAD 先验取决于 $|\theta|$，后验对数密度在 $\theta = 0$ 处不可微，因此驻点条件必须从次梯度的角度来解释。具体而言，当且仅当 $\mathcal{L}(\theta \mid d,\ \lambda)$ 在零点处的次梯度包含 0 时，$\theta = 0$ 才是后验众数；这恰好发生在 $|d| \le \lambda$ 时。对于 $|d| > \lambda$，零次梯度条件不成立，后验存在一个非零最大值。这解释了为何平滑 SCAD 估计量在 $[-\lambda, \lambda]$ 上恰好为零，并在 $|d|$ 越过临界阈值时连续地偏离零。

7 平滑 SCAD 的阈值选择

我们讨论当收缩规则为参数为 $a > 1$ 的平滑 SCAD 时，主阈值 λ 的选择，并探讨最优 λ 与信号长度 N、光滑性指数 γ 及形状参数 a 的关系。随后，我们将这一选择与 SURE 最小化联系起来。

回顾 $N = 2^J$，并且在小波模型 (2) 下，

$$d_{j,k} \;=\; \theta_{j,k} + \epsilon_{j,k}, \qquad \epsilon_{j,k} \sim N(0, \sigma^2), \tag{31}$$

其具有独立的 $\epsilon_{j,\ k}$。假设底层函数 f 的支撑位于有限区间上，且一致满足指数为 $\gamma > 0$ 的利普希茨正则性，因此真实系数满足如下形式的衰减界

$$|\theta_{j,k}| \;\le\; C_f\, 2^{-j(\gamma + 1/2)}, \tag{32}$$

这体现了小波域中的稀疏性。

对于任意选择的 $h(d,\ \lambda)$ 及由 (6) 定义的相应收缩规则 $\rho_h(d,\ \lambda)$，均方误差由 (7) 给出。在 $\lambda \in [0, \lambda_U]$ 上使 $R_J(\lambda)$ 最小的阈值 $\lambda_{\min}$，可视为依赖于未知系数的神谕选择。

Kudryavtsev 和 Shestakov（2024年）表明，对于一般的连续阈值规则类以及 Lipschitz 指数 $\gamma > 0$，oracle 风险具有如（20）所示的上界，其中 C 取决于 f、γ 以及具体规则，但不取决于 J。在其分析过程中，对偏差项和方差项进行平衡表明，忽略低阶修正后，oracle 阈值具有如下渐近尺度

$$\lambda_{\text{oracle}}(N, \gamma) \;\approx\; \sigma \sqrt{\frac{4\gamma}{2\gamma + 1} \log N}. \tag{33}$$

对于每个 $\gamma > 0$，因子 $4\gamma/(2\gamma + 1)$ 都严格小于 2，因此预言阈值小于通用阈值 $\lambda_U = \sigma\sqrt{2 \log N}$。规则形状的影响通过风险中的乘法常数体现，但不会影响主导的 $\sqrt{\log N}$ 尺度。

对于平滑 SCAD，参数 $a > 1$ 决定过渡区 $[\lambda, a\lambda]$的宽度，在该区域内，规则从强收缩过渡到无收缩，见公式 (13) 和 (15)。改变 a会修改风险表达式中的常数，但主要的偏差-方差平衡仍由以下形式的项决定

$$\lambda^2 2^{J/(2\gamma+1)} J^{1/(2\gamma+1)}$$

and

$$\lambda \exp\Big\{-\frac{\lambda^2}{2\sigma^2}\Big\} 2^J,$$

因此，平衡阈值具有形式（33）。对于固定的 a ，最优阈值对 N 和 γ 的主导依赖关系不变，而 a 只影响速率前的常数。

在 SURE 第 5 节的框架中，平滑 SCAD 的无偏风险估计量 is

$$\widehat{R}_J(\lambda) \;=\; \sum_{j=J-L}^{J-1} \sum_{k=0}^{2^j-1} F_{\mathrm{ssc}}(d_{j,k}, \lambda), \tag{34}$$

其中 F_{ssc} 由（18）给出。SURE 阈值为

$$\lambda_S \;=\; \arg\min_{\lambda \in [0, \lambda_U]} \widehat{R}_J(\lambda), \tag{35}$$

（12）的一个特例。Kudryavtsev 和 Shestakov 证明，对于其容许类中的任意阈值函数（包括平滑 SCAD），SURE 阈值 λ_S 以高概率不会过小，并且 $R_J(\lambda_S)$与 $R_J(\lambda_{\min})$渐近同阶。

这些结果展现出如下图景。对于固定的 $a > 1$ 和 $\gamma > 0$，风险最优阈值（从预言机视角来看）处于$\sigma\sqrt{\log N}$ 尺度上，即

$$\lambda_{\mathrm{oracle}}(N, \gamma, a) \;\asymp\; \sigma \sqrt{\frac{4\gamma}{2\gamma+1} \log N}, \tag{36}$$

而通用阈值对应于将 $4\gamma/(2\gamma+1)$ 替换为 2。SURE 阈值λ_S是在 $[0, \lambda_U]$ 中的一种数据驱动选择，其风险接近该区间上的最小风险；随着 N 增大， λ_S 的表现类似于对形式如（33）的预言机阈值进行随机扰动。

7.1 平滑 SCAD 的 λ 近似选择

对于平滑 SCAD 收缩，当信号包含 N 小波系数时，获得一个关于 λ 和$\widehat{\ }$的简单解析建议很有用。虽然可以通过网格轻松计算 $R_J(\lambda)$ 的精确极小值，但显式近似能提供直观认识，并可作为默认选择。

经典的通用阈值 λ_U 虽然较为保守，但可以保证去除纯噪声。稀疏预言机论证表明，对于许多近乎无偏的阈值规则，最优阈值可近似为

$$\lambda_{\text{oracle}} \;\approx\; \sigma\sqrt{2\log\!\left(\frac{N}{K}\right)}, \tag{37}$$

其中， K 是非零系数的数量。由于 K 未知，通用规则将 K 替换为 1，从而得到 λ_U。由于平滑 SCAD 相较于软阈值处理降低了尾部偏差，其风险最优阈值小于 λ_U。

一种方便的经验启发式规则为

$$\lambda_{\text{SCAD}} \;\approx\; \sigma\, c(a)\sqrt{2\log N}, \qquad c(a)=\sqrt{\frac{a-2}{a-1}}, \tag{38}$$

其中， $a>2$ 为 SCAD形状参数。因子 c（a）反映了当 $|d|$ 超过 λ后，估计量放松惩罚的速度有多快。当 a较大时，估计量趋近于软阈值处理，且 c（a）$\to 1$；当 a 接近 2 时，收缩趋近于硬阈值处理，且 c（a）变小。对于常用取值 $a=3.7$，式（38）给出

$$\lambda_{\text{SCAD}} \;\approx\; 0.8\,\sigma\sqrt{2\log N}, \tag{39}$$

与通用阈值相比，约降低百分之二十。对于采用升余弦过渡的平滑 SCAD，额外的光滑性主要影响基于导数的量（例如通过 h'影响 SURE表达式），而有效阈值的 $\sigma\sqrt{2\log}\ N$尺度关系与经典SCAD类规则相同。实践中，对于适中的 a范围，近似式（38）可提供合理的初始选择。

总之，平滑 SCAD 收缩建议采用如下形式的阈值

$$\lambda \;\approx\; \sigma\sqrt{2\log N}\,\sqrt{\frac{a-2}{a-1}}, \tag{40}$$

该阈值小于通用阈值，但由于尾部偏差较低，因此具有稳定性。这为选择 λ 提供了一条简单实用的规则，并且与预言机尺度（33）一致。在应用中，可以将（40）作为起始值，并结合 SURE 最小化。

7.2 形状参数 a 的选择

平滑 SCAD 规则还包含形状参数 $a>1$，用于控制收缩从强烈衰减转变为几乎不衰减的过程。阈值 λ 控制

哪些系数会被设为零，而 a 则控制收缩逐渐减弱时 (λ, $a\lambda$) 的宽度和光滑性。a 接近 1 时会产生类似硬阈值处理的急剧转变，而 a 较大时则会产生更长、更平滑的转变。

从理论角度来看，a 仅受温和的限制。SURE 理论要求收缩轮廓 $h(d, \lambda, a)$在 d和 λ 上都具有有界的一阶导数，并且 a 保持在一个远离 1 的紧区间内。这些要求已在第4节中验证；对于范围如 $[1.5, 4]$内的任意固定 a，均能满足这些要求。与原始 SCAD 建议 $a \approx 3.7$（Fan 和 Li，2001年）的思路相同，分析并不要求 a 依赖于 σ^2、N或 λ。

一种自然的策略是固定 a，并让 SURE 确定 λ。a 取 2 到 3 之间的值时，会产生平滑且稳定的收缩；模拟（此处未报告）表明，性能对具体取值并不十分敏感。固定 a 后，SURE 选择过程与式（35）中的相同。

如果希望 N 和 σ之间存在轻微的依赖关系，可以将过渡区的末端与通用阈值对齐。由于 $\lambda_U = \sigma\sqrt{2\log N}$，要求 $a\lambda \approx \lambda_U$可得

$$a \approx \frac{\sigma\sqrt{2\log N}}{\lambda}. \tag{41}$$

取由 SURE 确定的 λ，即可得到一种数据驱动的规则来选择 a，并满足诸如 $a \in [1.5, 4]$ 这样的界限。更复杂的替代方案是通过在一个小型网格 $a \in \{2, 2.5, 3, 3.5\}$ 上最小化 SURE，联合选择 a 和 λ；对于每个 a，求出 $\lambda_S(a)$，并选择 SURE 风险最小的参数对。

总而言之，a 的选择没有 λ 那么微妙。对 N 和 σ 的主要渐近依赖关系由 λ 决定，而 a 则用于调节转变的光滑性。将固定的适中 a 与通过 SURE 选择 λ 相结合，已经能够得到有效的方法。

7.3 具有层级相关阈值的自适应平滑 SCAD

到目前为止，参数 λ 和 a 都是全局的。在许多情况下，经验小波系数的变异性会随层级 j 发生显著变化。通常，采用层级相关阈值是有益的。与之前一样，我们考虑来自分解层级 $j = J-L$、.....、$J-1$ 的细节小波系数，其中 $J-1$ 对应细节系数的最精细尺度，$J-L$ 对应纳入收缩的最粗尺度。

为纳入 a 和 λ的异质性，我们引入平滑 SCAD 的层级依赖版本。对于第 j 层的每个小波系数 d，我们选择一个主阈值 λ_j

和一个过渡参数 $a_j > 1$，并定义

$$h_j(d, \lambda_j) \;=\; \begin{cases} \lambda_j \,\mathrm{sign}(d), & 0 < |d| \le \lambda_j, \\ \lambda_j \cos^2\!\left(\dfrac{\pi}{2}\,\dfrac{|d| - \lambda_j}{(a_j - 1)\lambda_j}\right) \mathrm{sign}(d), & \lambda_j < |d| < a_j\lambda_j, \\ 0, & |d| \ge a_j\lambda_j, \end{cases} \tag{42}$$

以及自适应平滑 SCAD 规则

$$\rho_{\mathrm{ssc},j}(d, \lambda_j) \;=\; \begin{cases} 0, & |d| \le \lambda_j, \\ d - h_j(d, \lambda_j), & \lambda_j < |d| < a_j\lambda_j, \\ d, & |d| \ge a_j\lambda_j. \end{cases} \tag{43}$$

施加于 $d_{j,k}$ 的收缩为 $\rho_{\mathrm{ssc},j}(d_{j,k}, \lambda_j)$。

这种构造在每个层级都保留了奇对称性、连续性和有界导数性质，因为每组三元组 (h_j, λ_j, a_j) 都满足与全局版本相同的结构假设。所有 SURE 计算均可扩展到各个层级。在层级 j，无偏风险估计为

$$\widehat{R}_j(\lambda_j) \;=\; \sum_{k=0}^{2^j-1} F_{\mathrm{ssc}}(d_{j,k}, \lambda_j), \tag{44}$$

其中 F_{ssc} 与式（18）中的相同，但将 h 及其导数替换为 h_j 及其导数。全局风险估计为

$$\widehat{R}_J(\lambda_{J-L}, \ldots, \lambda_{J-1}) \;=\; \sum_{j=J-L}^{J-1} \widehat{R}_j(\lambda_j). \tag{45}$$

对于固定的过渡参数 a，一种简单的自适应方案是让 SURE 在每个层级独立地选取 λ_j，即

$$\lambda_{j,S} \;=\; \arg\min_{\lambda_j \in [0, \lambda_U]} \widehat{R}_j(\lambda_j). \tag{46}$$

这种对 SureShrink 的扩展使每个层级都能根据自身的系数结构进行自适应。可容许性条件仍然成立，因为 KS 分析会分别控制每个层级，且 SURE 项之和仍保持良好的性质。

为了提高稳定性，尤其是在系数数量较少、SURE 曲面变化较大的较粗尺度层级上，对选定阈值设置一个下限是自然的做法。一种实用的选择是

$$\lambda_j^{\star} \;=\; \min\Big\{\lambda_U,\; \max\{\lambda_{j,S},\; c\,\sigma\}\Big\}, \tag{47}$$

其中 c 是一个固定常数。实践中， $c \in [2,3]$ 范围内的取值较为合理，既能充分避免保留过多噪声，又允许在 SURE 选出的阈值本身已较为保守时由其主导。如果不设置这样的下限，$\lambda_{j,\ s}$ 偶尔可能过小，导致系数较少的层级中残留噪声。因此，这个下限是对逐层 SURE 规则进行简单而有效稳定化的一种方式；还可以通过纳入关于底层函数光滑性的信息，构造更复杂的方法。

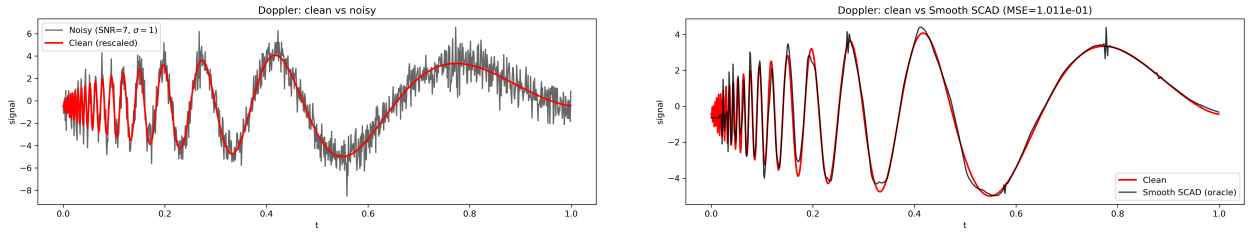
第二种方案是让 λ_j 和 a_j 都随 j变化。较大的 a_j 会拓宽过渡区域 $[\lambda_j, a_j\lambda_j]$ ，并在较粗尺度下产生更平缓的收缩，因为此时系数往往包含真实信号；较小的 a_j 则会使该规则更快地趋近于硬阈值处理，这可能适用于最细、由噪声主导的层级。可以为 a_j固定一组较小的候选值，例如 $a_j \in \{1.5, 2, 3\}$，并在一个小网格上通过 SURE 选择 a_j ；也可以设定确定性的逐层方案，例如在细尺度下取较小的 a_j ，在粗尺度下取较大的 a_j 。只要 $a_j > 1$且一致有界，结构性假设便仍然成立，SURE表达式也保持原有形式。

8 数值示例

为了展示平滑 SCAD 去噪器的实际表现，我们采用四个经典的 Donoho-Johnstone 测试函数开展了受控的蒙特卡洛研究：Doppler、Blocks、HeaviSine 和 Bumps。这些函数是小波去噪文献中广泛采用的基准信号（Donoho 和 Johnstone，1994年、1995年），各自体现了不同类型的非参数难度：Doppler 呈现空间变化的振荡；Blocks 包含尖锐的不连续点；HeaviSine 兼具平滑与突变特征；Bumps 则包含多个异质的局部尖峰。

对每个信号，均加入具有受控信噪比的独立高斯白噪声。按照统计学惯例，信噪比定义为信号方差与噪声方差之比；我们固定 Var(ϵ) $= 1$ ，并对每个无噪信号重新缩放，使得 Var(f)/Var(ϵ) $=$ SNR。实验考察了信噪比 $= 3$、5、7 和 10，以及样本量 $N = 512$、1024 和 $N = 2048$；为简洁起见，表 1 仅展示了具有代表性的结果。所有情况下均进行了 $M = 1000$ 次独立蒙特卡洛重复试验，并通过平均均方误差（AMSE）及其蒙特卡洛标准差汇总结果。

小波变换采用去噪文献中常用的正交归一紧支撑滤波器计算：Doppler 和 HeaviSine 使用 Symmlet 4，Blocks 使用 Haar



(a) (b)

图 2：(a) 含加性高斯噪声的长度为 $N = 1024$ 的多普勒测试信号，经重新缩放后达到信噪比 $= 7$ （方差比）；(b) 使用预言机选择的阈值 $\lambda^* = \arg\min_\lambda \mathrm{MSE}(\lambda) \approx 2.3485$ 进行的平滑 SCAD 重建，显示出有效的结构恢复和高频伪影抑制。预言机选择的 λ^* 小于通用阈值，本例中的通用阈值为 $\lambda_U = \sqrt{2\log(1024)} = 3.724$。

（以提供无偏的不连续性表示），Bumps 使用 Daubechies 3。去噪通过对各个系数分别应用四种收缩规则完成：通用硬阈值化、通用软阈值、经典 SCAD（Fan 和 Li，2001 年），并用预言选择 λ，以及其升余弦平滑对应形式（平滑 SCAD）。这里采用 SCAD 和平滑 SCAD 的预言调参，以隔离收缩形状的影响；通用硬阈值和软阈值则作为固定的插件基线。

图 2 展示了一个受高斯噪声污染、信噪比为 $= 7$ 的多普勒数据集，以及使用平滑 SCAD 收缩规则得到的去噪重建结果。

在所有配置下，平滑 SCAD 的表现始终优于 SCAD，而两者都显著优于通用阈值化，尤其是在高频或不连续的情形下。平均均方误差的改善（基于 $M = 1000$ 次运行）在 Blocks 和 Bumps 上最为显著，因为在这两类信号中，抑制噪声且不过度引入偏差至关重要。表 1 报告了 $N = 1024$、信噪比 $= 7$，以及 $\sigma = 1$ 的代表性结果。

仿真引擎使用 Python 实现，且完全可复现；Jupyter 笔记本可在 https://github.com/BraniV/SmoothSCAD 找到。

9 结论

本文提出的平滑 SCAD 规则通过用升余弦取代经典 SCAD 阈值的分段线性过渡，对其进行了改进。这一修改保留了 SCAD 的基本特性，同时生成的收缩剖面在 $\mathbb{R}$ 上连续，且除主阈值外处处连续可微。由此，该规则完全属于连续阈值化类，因此适用 Stein 无偏风险估计。

表 1： $M = 1000$ 次运行的平均均方误差和均方误差的标准差， $N = 1024$ 且信噪比为 $= 7.0$。

信号	Hard	Soft	SCAD	平滑 SCAD
Doppler	1.5513e-01	4.0264e-01	1.3505e-01	1.3188e-01
	(1.8200e-02)	(3.2646e-02)	(1.4472e-02)	(1.4609e-02)
Blocks	2.0078e-01	6.6805e-01	1.5286e-01	1.4484e-01
	(2.7987e-02)	(4.4013e-02)	(1.8624e-02)	(1.9502e-02)
HeaviSine	7.7260e-02	1.0477e-01	6.8234e-02	6.5714e-02
	(1.5162e-02)	(1.1461e-02)	(1.0011e-02)	(1.0515e-02)
Bumps	2.9867e-01	9.5999e-01	2.3165e-01	2.2799e-01
	(3.2696e-02)	(5.2262e-02)	(2.0762e-02)	(2.1343e-02)

由于平滑 SCAD 收缩剖面满足 SURE 框架所要求的结构与正则性条件，因此无需进一步调整即可无偏地计算风险及其极小值。针对连续阈值化规则得出的渐近结论——最小风险的预言机阶收敛、中心化 SURE 风险的渐近正态性，以及归一化下的强相合性——均可直接推广。这些性质对每个层级都成立，并可扩展至依赖层级的构造和自适应构造。

贝叶斯解释进一步阐明了该方法。升余弦过渡产生了一种光滑非凸惩罚项，其后验众数恰好重现了光滑 SCAD 估计量，将惩罚似然论证与小波收缩联系起来，同时也凸显了过渡宽度在调节稀疏性和尾部行为方面的作用。升余弦过渡是一种简便的光滑选择；其他光滑插值函数（例如基于样条的过渡）也可能满足相同的 SURE 可容许条件，同时改变高阶导数行为，值得进一步探索。

在实践中，光滑性能够带来计算优势。在 λ 上进行 SURE最小化具有数值稳定性，并且临界系数的收缩比基于分段线性 SCAD 的收缩更加规则。阈值尺度的简单解析表达式可提供一个合理的初始值，而 SURE 原理则能在不同尺度间实现数据自适应。

可以自然地沿着几个方向进一步拓展。首先，升余弦构造可用于获得其他非凸惩罚的平滑类比，包括 MCP 和平滑截断 ℓ_0 型规则，这些规则同样属于连续 SURE 类。其次，可将平滑 SCAD 嵌入线性逆问题的迭代阈值化和近端梯度方案中，此时收缩算子的可微性有助于进行收敛性分析。第三，可以将 Stein 恒等式替换为

指数族类比或 Tweedie型类比，从而有望为重尾或异方差数据推导出 SURE变体。

据我们所知，平滑 SCAD 是首个同时满足 KS 框架连续 SURE 条件，并具有显式非凸先验（其后验众数与估计量一致）的 SCAD 型小波收缩规则。由此得到的收缩规则透明、灵活且实用，能够自然融入基于小波的统计方法。

参考文献

Antoniadis, A., Bigot, J., and Sapatinas, T.（2007年）。非参数回归与反卷积中的小波估计量。统计学讲义，192:37–64。Antoniadis, A. and Fan, J.（2001年）。小波近似的正则化。美国统计学会会刊，96(455):939–967。Antoniadis, A. and Sapatinas, T.（2007年）。非参数回归与逆问题中的小波方法。统计学综述，1:16–55。Atto, A. M., Pastor, D., and Mercier, G.（2008年）。用于非参数估计的平滑S形小波收缩。载于*2008*年*IEEE*国际声学、语音与信号处理会议，第3265–3268页。

Atto, A. M.、Pastor, D. 和 Mercier, G.（2009）。基于S形收缩的平滑自适应。*EURASIP*图像与视频处理期刊，第532312页。

Breheny, P. 和 Huang, J.（2011）。用于非凸惩罚回归的坐标下降算法及其在生物特征选择中的应用。应用统计年鉴，5(1):232–253。

Cutillo, L.、Jung, Y. Y.、Ruggeri, F. 和 Vidakovic, B.（2008）。更大的后验众数小波阈值处理及其应用。统计规划与推断杂志，138(12):3758–3773。

Daubechies, I.、Defrise, M. 和 De Mol, C.（2004）。线性逆问题的迭代阈值算法。纯粹与应用数学通讯，57(11):1413–1457。

Donoho, D. L. 和 Johnstone, I. M.（1994）。通过小波收缩实现理想的空间自适应。*Biometrika*，81(3):425–455。Donoho, D. L. 和 Johnstone, I. M.（1995）。通过小波收缩适应未知光滑性。美国统计学会会刊，90(432):1200–1224。Fan, J. 和 Li, R.（2001）。通过非凹惩罚似然进行变量选择及其预言机性质。美国统计学会会刊，96(456):1348–1360。Figueiredo, M. A. T. 和 Nowak, R. D.（2001）。基于小波的图像估计：采用Jeffrey非信息先验的经验贝叶斯方法。*IEEE*图像处理汇刊，10(9):1322–1331。Fonseca, R.、Morettin, P. 和 Pinheiro, A.（2024）。小波特征筛选。计算与图形统计学杂志，33(4):1310–1319。Gasso, G.、Rakotomamonjy, A. 和 Cannesson, M.（2011）。使用平滑的 ℓ_0 范数恢复稀疏信号。*IEEE Transactions on Signal Processing*，59(7):5640–5650。Huang, J.、Breheny, P. 和 Ma, S.（2013）。高维模型中组选择的精选综述。统计科学，28(4):558–575。Kudryavtsev, A. 和 Shestakov, O.（2016）。计算小波系数时使平均错误概率最小化的阈值的渐近行为。数学学报，93(3):295–299。Kudryavtsev, A. 和 Shestakov, O.（2024）。在小波收缩中使用连续阈值函数时SURE估计的性质。*Mathematics*，12(3):3969。Li, G.、Peng, H.-C. 和 Zhang, H.（2011）。用于信号去噪的非凸惩罚小波估计。计算统计学与数据分析，55(4):2024–2038。Mazumder, R.、Friedman, J. 和 Hastie, T.（2011）。SparseNet：非凸惩罚回归的坐标下降。美国统计学会会刊，106(495):1125–1138。Reményi, N. 和 Vidakovic, B.（2013）。贝叶斯小波收缩策略——新进展。载于 Shen, X. 和 Zayed, A. I. 编，多尺度信号分析与建模，电气工程讲义，第317–346页。Springer-Verlag。

Sardy, S. 和 Bruce, A. G.（2003）。稳健小波收缩与迭代贝叶斯估计。*Statistics and Computing*，13(2):129–146。Sardy, S.、Tseng, P. 和 Bruce, A. G.（2001）。稳健小波去噪。*IEEE Transactionson Signal Processing*，49(6):1146–1152。Shestakov, O.（2020）。随机样本与相关噪声模型的小波阈值风险估计。*Mathematics*，8(3):377。Vidakovic, B.（1994）。基于贝叶斯规则和贝叶斯因子的非线性小波收缩。技术报告94-24，统计与决策科学研究所，杜克大学。Vidakovic, B.（1999）。小波统计建模。John Wiley & Sons，纽约。Vidakovic, B. 和 Ruggeri, F.（2001）。BAMS方法：理论与模拟。*Sankhy¯a,Series B*，63(2):234–249。Xue, L.（2009）。平滑截断绝对偏差惩罚。*Statistics and ProbabilityLetters*，79(15):1752–1758。Zhang, C.-H.（2010）。极小极大凹惩罚下近似无偏的变量选择。*Annals of Statistics*，38(2):894–942。Zhou, Q. 和 Shen, X.（2008）。非凸惩罚小波估计。*Journal of theAmerican Statistical Association*，103(482):672–681。Zou, H. 和 Li, R.（2008）。非凹惩罚似然模型中的一步稀疏估计。*Annals of Statistics*，36(4):1509–1533。