arXiv:2303.01353v3 [stat.ML] 9 Nov 2023

在范数正则化中惩罚偏置项可促成稀疏性

Etienne BoursierINRIA CELESTE、LMO、法国奥尔赛 etienne.boursier@inria.fr

Nicolas FlammarionTML 实验室、洛桑联邦理工学院、瑞士 nicolas.flammarion@epfl.ch

摘要

训练神经网络时，控制参数范数往往能带来良好的泛化效果。除了直观理解之外，参数范数正则化与所得估计器之间的关系在理论上仍未得到充分阐明。对于具有一层隐藏 ReLU 层且输入为一维数据的网络，本文表明，表示一个函数所需的参数范数由其二阶导数的全变差乘以一个 $\sqrt{1+x^2}$ 加权因子给出。值得注意的是，当偏置项的范数不进行正则化时，这个加权因子便会消失。研究表明，这个额外加权因子的存在至关重要，因为它能确保最小范数插值器的唯一性和稀疏性（以折点数量衡量）。相反，不对偏置项的范数进行惩罚会导致非稀疏解。因此，在正则化中显式或隐式地惩罚偏置项会带来稀疏估计量。

1 引言

尽管现代神经网络在参数数量方面并未受到特别严格的限制，但应用于真实世界数据时，仍展现出显著的泛化能力 [Belkin 等人，2019年，Zhang 等人， 2021]。耐人寻味的是，理论和实证研究均表明，决定网络泛化性质的关键因素并非参数的绝对数量，而是这些参数的范数 [Bartlett，1996，Neyshabur 等人， 2014]。通常通过结合显式正则化技术（例如权重衰减 [Krogh 和 Hertz， 1991], ）以及训练算法所产生的某种隐式正则化 [Soudry 等人，2018，Lyu 和 Li，2019，Ji 和 Telgarsky，2019，Chizat 和 Bach， 2020]，来控制这一范数。

具有大量参数的神经网络可以逼近紧集上的任意连续函数 [Barron， 1993]。因此，如果不控制范数，估计函数的空间就涵盖了所有连续函数。在参数空间中，这意味着要考虑具有无限宽度且权重无界的神经网络 [Neyshabur 等人， 2014]。然而，当对权重进行控制时，参数空间（即网络的参数 θ ）与函数空间（即网络输出所产生的估计函数 f_θ ）之间的确切对应关系便不再明晰。厘清这种对应关系，对于理解过参数化神经网络的泛化性质至关重要。由此产生两个基本问题。

问题1。 神经网络的参数范数对应于函数空间中的什么量？

Qu问题2。 在以最小参数范数拟合训练数据时，学习到的是什么函数？

我们在带有跳跃连接的单隐藏 ReLU 层网络的背景下研究这些问题。先前的研究 [Kurková 和 Sanguineti，2001，Bach， 2017] 考察了表征代价较小的函数的泛化保证，其中表征代价指的是

第37届神经信息处理系统大会（NeurIPS 2023）。

对函数进行参数化所需的范数。然而，使用经典分析工具来解释这种表征代价，并确定相应的函数空间，仍然颇具挑战。为解决这一问题，问题1旨在确定这种表征代价能否转化为更易于解释的泛函（伪）范数。请注意，问题1研究的是在整个定义域上拟合一个函数所需的参数范数。相比之下，训练神经网络执行回归任务时，我们只拟合训练数据给出的有限个点。问题2旨在探究：无论是显式还是隐式地对参数范数进行正则化，在最小化某种经验损失时，学习得到的函数具有哪些性质。

与我们的工作相关的是，Savarese 等人 [2019], Ongie 等人 [2019] 分别针对单隐层 ReLU 神经网络中的单变量函数和多变量函数研究了问题1。关于这项研究的全面综述，建议参阅 Parhi 和 Nowak 的综述 [2023]。另一方面，Parhi 和 Nowak [2021], Debarre 等人 [2022], Stewart 等人[2022] 专门研究了单变量情形下的问题2。此外，Sanford 等人 [2022]考察了一个特定的多维情形。然而，现有研究在考虑 ℓ_2 正则化项时，都忽略了神经网络的偏置参数。省略偏置后，对这些问题的分析和求解会变得更加简单。

与此形成鲜明对比的是，我们的工作针对单变量函数研究问题1和问题2，同时也纳入了对偏置参数的正则化。这看似只是一个小细节——人们通常认为，无论是否对偏置范数[1] 施加惩罚 [（例如，参见 Ng， 2011]），得到的估计器都相似。尽管如此，我们的研究表明，对偏置项施加惩罚能够促成稀疏性，并确保估计函数的唯一性；若不纳入偏置正则化，则无法实现这一点。这两种显式正则化在实际中的相似性，可以归因于隐式正则化的存在，因为隐式正则化也将偏置项考虑在内。一阶优化方法所执行的更新并不区分偏置参数和权重参数，这表明二者受到相同的隐式正则化。因此，尽管在实际应用中，两种正则化方法可能会产生相似的估计量，但我们认为，采用偏置项正则化得到的理论估计量能够体现观察到的隐式正则化效应。因此，研究在处理问题1和问题2时对偏置项施加惩罚的影响至关重要，因为在这种情况下得到的答案与不施加偏置惩罚时的答案存在显著差异。

还值得一提的是，Shevchenko 等人 [2022], Safran 等人 [2022]证明，对于具有 ReLU 激活的网络，梯度流能够学习稀疏估计量。不过，这些稀疏性保证要弱得多（激活方向的数量更多），因为他们还考虑了优化方面的问题（与我们以及上述研究脉络中的工作直接考察优化问题的极小化解不同）。

贡献。 在第2节介绍问题设定后，我们在第 3 节通过采用与 Savarese 等人 [2019]类似的分析方法来回答问题1。核心结果定理1表明，当允许使用一个自由的跳跃连接时，函数的表征代价由其二阶导数的加权全变差给出，其中包含一个 $\sqrt{1+x^2}$ 项。值得注意的是，对偏置项施加惩罚会在全变差中引入一个 $\sqrt{1+x^2}$ 乘法权重，这与不施加偏置惩罚形成对比。

这种加权方式从根本上影响了问题2的答案。具体而言，它破坏了函数表征代价的平移不变性，使 Debarre 等人 [2022] 提出的分析方法[2]变得不适用。为解决这一问题，我们在第4节和第5节中深入研究优化问题的解的计算及其性质：

$$\inf_f \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} \quad \text{subject to } \forall i \in [n],\ f(x_i) = y_i.$$

在第4节中，我们将这个问题重新表述为连续动态规划，从而能够更简单地分析最小化问题。利用这一动态规划的重新表述，第 5 节证明了解的唯一性。此外，在某些数据假设下，我们证明，按折点数量衡量，极小化解属于最稀疏的插值器之列。

[1]尽管 Goodfellow 等人 [2016, 第 7] 章声称，对偏置进行惩罚可能会导致欠拟合，但我们的工作并不关注优化方面，并假设会发生插值。[2]虽然平移不变性对于分析目的很有用，但在实践中未必理想。值得注意的是，Nacson 等人 [2022] 表明，处理未中心化数据可能有利于学习相关特征，这一结论正是基于缺乏平移不变性。

值得注意的是，稀疏脉冲反卷积领域也研究过类似结果[Candès 和 Fernandez-Granda，2014年，Fernandez-Granda，2016，Poon 等人， 2019], 而我们的问题可视为基追踪 [Chen 等人， 2001] 在无限维参数空间中的泛化。然而，稀疏脉冲反卷积的经典技术并不适用于解决问题2，因为在我们的设定中，最稀疏插值器的集合是无限的。
最后，第6节通过玩具示例说明了偏置项正则化在神经网络训练过程中实现更稀疏估计器的重要性。为简洁起见，正文仅给出证明概要，完整证明则见附录。

2 无限宽网络

本节介绍所考虑的设定，将一维函数表示为无限宽网络。此处省略了一些精确的数学论证，因为这一构造可直接沿用 Savarese 等人 [2019],Ongie 等人 [2019]的思路。本文考虑如下由 ReLU 激活函数的单隐藏层神经网络参数化的一维函数 $f_\theta : \mathbb{R} \to \mathbb{R}$

$$f_\theta(x) = \sum_{j=1}^{m} a_j \sigma(w_j x + b_j),$$

其中 $\sigma(z) = \max(0, z)$ 是 ReLU 激活函数，而 $\theta = (a_j, w_j, b_j)_{j\in[m]} \in \mathbb{R}^{3m}$ 是定义该神经网络的参数。向量 $\mathbf{a} = (a_j)_{j\in[m]}$ 表示最后一层的权重，而 $\mathbf{w}$ 和 $\mathbf{b}$ 分别表示隐藏层的权重和偏置。对于任意宽度 m和参数 θ，重要的量是参数的欧几里得范数平方：
$\|\theta\|_2^2 = \sum_{j=1}^{m} a_j^2 + w_j^2 + b_j^2$。

我们回顾一下，与 Savarese 等人 [2019], Ongie 等人 [2019], 不同，此处考虑的范数包含偏置项。现在我们定义函数 f 的表征代价： $\mathbb{R} \to \mathbb{R}$ 为

$$R(f) = \inf_{\substack{m\in\mathbb{N} \\ \theta\in\mathbb{R}^{3m}}} \frac{1}{2}\|\theta\|_2^2 \quad \text{such that} \quad f_\theta = f.$$

根据参数化的齐次性，一种典型的重缩放技巧 [（例如参见 Neyshabur 等人，2014年，定理 1]）可以将其改写为

$$R(f) = \inf_{m,\theta\in\mathbb{R}^{3m}} \|\mathbf{a}\|_1 \quad \text{such that} \quad f_\theta = f \text{ and } w_j^2 + b_j^2 = 1 \text{ for any } j \in [m].$$

注意，只有当函数 f 恰好可以表示为有限宽神经网络时， $R(f)$ 才是有限的。我们的目标是将此定义扩展到更大的函数空间，也就是说，扩展到任何可由有限宽网络任意良好地逼近、同时参数范数保持（一致）有界的函数。尽管使用有限宽网络逼近函数，但当近似误差趋于 0 时，网络宽度必然趋于无穷。与 Ongie 等人 [2019], 类似，我们定义

$$\overline{R}(f) = \lim_{\varepsilon\to 0^+} \left(\inf_{m,\theta\in\mathbb{R}^{3m}} \frac{1}{2}\|\theta\|_2^2 \quad \text{such that} \quad |f_\theta(x) - f(x)| \le \varepsilon \text{ for any } x \in [-1/\varepsilon, 1/\varepsilon] \right).$$

注意，逼近必须限制在紧集 $[-1/\varepsilon, 1/\varepsilon]$ 上，以避免出现有问题的退化情形。使 $R(f)$ 有限的函数空间远大于使 R 有限的函数空间，并且包含每个紧支撑利普希茨函数；当后者有限时，该空间与 R 重合。

再次利用重缩放论证，我们可以假设隐藏层参数 (w_j, b_j) 位于 $\mathbb{S}_1$ 中，并转而考虑输出层权重的 ℓ_1 范数。网络的参数因而可以视为单位球面 $\mathbb{S}_1$ 上的离散符号测度。当宽度趋于无穷时，极限便有了恰当的定义，并对应于一个可能连续的符号测度。具体而言，定义 $\mathcal{M}(\mathbb{S}_1)$ 为 $\mathbb{S}_1$ 上具有有限全变差 $\|\mu\|_{\mathrm{TV}}$ 的符号测度 μ 空间。遵循 Bengio 等人 [2005], 和 Bach [2017], 的典型构造，无限宽网络由测度 $\mu \in \mathcal{M}(\mathbb{S}_1)$ 参数化为[3]

$$f_\mu : x \mapsto \textstyle\int_{\mathbb{S}_1} \sigma(wx + b)\mathrm{d}\mu(w, b). \tag{1}$$

与 Ongie 等人 [2019] 类似， $\overline{R}(f)$ 满足等式

$$\overline{R}(f) = \inf_{\mu\in\mathcal{M}(\mathbb{S}_1)} \|\mu\|_{\mathrm{TV}} \text{ such that } f = f_\mu.$$

[3]为简化记法，我们将 f_θ 和 f_μ都这样表示；根据上下文，可以清楚判断下标表示的是向量还是测度。

右侧的项定义了 $\mathcal{F}_1$ 范数 [Kurková 和 Sanguineti， 2001], 即 $R(f) = \|f\|_{\mathcal{F}_1}$。人们认为， $\mathcal{F}_1$ 范数对于神经网络在经验上的成功具有重要意义。具体而言，Kurková 和 Sanguineti [2001], Bach [2017], 推导出了小 $\mathcal{F}_1$ 范数估计器的泛化性质；与此同时，许多理论结果表明，使用梯度下降训练单隐藏层神经网络，往往会对估计器的 $\mathcal{F}_1$ 范数产生隐式正则化 [Lyu 和 Li，2019，Ji 和 Telgarsky，2019，Chizat 和 Bach，2020，Boursier 等人， 2022]。然而，这种对 $\mathcal{F}_1$ 范数的隐式正则化并非普遍存在；一些研究表明，在特定示例中，隐式正则化的可能是另一种量 [Razin 和 Cohen，2020，Vardi 和 Shamir，2021，Chistikov 等人， 2023]。尽管如此，$\mathcal{F}_1$ 范数似乎与隐式偏置密切相关，其重要性正是本文的主要动机。以往研究也考察了神经网络表示函数的代价 [Savarese 等人，2019，Ongie 等人， 2019],但它们并未在参数范数中惩罚偏置项，研究的是一种与 $\mathcal{F}_1$ 范数略有不同的函数范数。这一细微差别导致了惩罚与不惩罚偏置项时所得估计量的稀疏程度不同，详见第5节和第6节；此处，估计器的稀疏性指网络表示该函数所需的最小宽度（或者等价地，公式（1）中 μ 的支撑的基数）。与参数的稀疏性相比，这种稀疏性概念更有意义 θ 因为不同的 θ （稀疏程度不同）可以表示完全相同的估计函数。

2.1 无惩罚的跳跃连接

我们的目标现在是刻画一维函数和最小范数插值器的 $\mathcal{F}_1$ 范数；在使用范数正则化训练神经网络时，可以近似得到这些插值器。尽管是在一维情形下，分析和结果仍然较为复杂。在神经网络表示中允许存在一个不受惩罚的仿射项，可以更简洁地刻画范数并描述最小范数插值器。因此，在本文余下部分，我们将有限宽网络和无限宽网络参数化如下：

$$f_{\theta,a_0,b_0} : x \mapsto a_0 x + b_0 + f_\theta(x), \quad \text{and} \quad f_{\mu,a_0,b_0} : x \mapsto a_0 x + b_0 + f_\mu(x),$$

其中 $(a_0,\ b_0) \in \mathbb{R}$。仿射部分 $a_0 x + b_0$ 实际上对应于神经网络架构 [He 等人， 2016]中的一个自由跳跃连接，并允许我们在函数 f 的表征代价中忽略仿射部分。我们现在将其定义为

$$\overline{R}_1(f) = \lim_{\varepsilon \to 0^+} \left(\inf_{\substack{m,\theta \in \mathbb{R}^{3m} \\ (a_0,b_0) \in \mathbb{R}^2}} \frac{1}{2}\|\theta\|_2^2 \quad \text{such that} \quad |f_{\theta,a_0,b_0}(x) - f(x)| \le \varepsilon \text{ for any } x \in [-1/\varepsilon, 1/\varepsilon] \right).$$

表征代价 $R_1(f)$ 与 $R(f)$ 类似，但允许网络架构中包含一个自由仿射项。与 $R(f)$ 类似，沿着 Savarese 等人 [2019], Ongie 等人 [2019], 的思路可以证明， $R_1(f)$ 满足

$$\overline{R}_1(f) = \inf_{\substack{\mu \in \mathcal{M}(\mathbb{S}_1) \\ a_0,b_0 \in \mathbb{R}}} \|\mu\|_{\mathrm{TV}} \text{ such that } f = f_{\mu,a_0,b_0}.$$

本文其余部分将更深入地研究代价 $R_1(f)$。第 3 节中的定理1可以直接扩展到代价 $R(f)$, 即不包含不惩罚跳跃连接的情况。为完整起见，其相应版本见附录 C 中的定理4。

许多工作也考虑自由跳跃连接，因为这可以简化分析 [例如 Savarese 等人，2019年、Ongie 等人，2019年、Debarre 等人，2022年、Sanford 等人， 2022]。由于一个跳跃连接可以由两个 ReLU 神经元表示（$z = \sigma(z) - \sigma(-z)$），人们通常认为，考虑自由跳跃连接不会改变所得结果的性质。第6节和附录 B 中的实证证据进一步支持了这一观点：无论是否使用自由跳跃连接，我们的发现都成立。

3 表征代价

下面的定理1刻画了任意单变量函数的表征代价 $R_1(f)$。**定理1**。 对于任意利普希茨函数 $f : \mathbb{R} \to \mathbb{R}$,

$$\overline{R}_1(f) = \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} = \int_{\mathbb{R}} \sqrt{1+x^2}\, \mathrm{d}|f''|(x).$$

对于任意非 *Lipschitz* 函数，$\overline{R}_1(f) = \infty$。

在定理1中，f'' 是 f的分布意义下的二阶导数，对于利普希茨函数而言定义良好。若不惩罚偏置项，表征代价由 f'' [的全变差给出（Savarese 等人， 2019]）。定理1指出，对偏置项施加惩罚会为 f''增加权重$\sqrt{1+x^2}$。如第 5 节所示，这种加权在训练神经网络时会倾向于得到更稀疏的估计器。此外，在忽略偏置项时，无限宽神经网络能够表示且参数范数有限的函数空间，对应于二阶导数具有有界全变差的函数。将这些偏置项纳入表征代价后，二阶导数还需满足轻尾条件。若没有自由的 仿射项，附录 C 中的定理4刻画了 $R(f)$，其中会出现一个额外项，用于计入 f的仿射部分。

我们注意到，E 和 Wojtowytsch 的注4.2 [2021]以及 Li 等人的定理1 [2020b] 与定理1和定理4密切相关。然而，这些结果仅证明范数$R(f)$与另一种范数等价，后者用于量化 $\sqrt{1+x^2}f''$的全变差（仿射项除外）。另一方面，我们的结果给出了这两个范数之间的精确等式，这在分析最小范数插值器时尤为有用。

例1. 若函数 f 由有限宽网络 $f(x) = \sum_{i=1}^n a_i\sigma(w_i x + b_i)$ 给出，其中a_i、 $w_i \neq 0$ 且 $\frac{b_i}{w_i}$ 两两不同；定理*1*得出 $R_1(f) = \sum_{i=1}^n |a_i|\sqrt{w_i^2 + b_i^2}$。这恰好

对应于向量（$c_i a_i$， $\frac{w_i}{c_i}$， $\frac{b_i}{c_i}$）$_{i=1,\ \ldots,\ n}$ 的平方 ℓ_2 范数的一半，其中 $c_i = \sqrt{\frac{\sqrt{w_i+b_i^2}}{|a_i|}}$。
因此，该向量是函数 f 的一种最小表示。

根据定理1，训练带有 ℓ_2 正则化的单隐藏层 ReLU 无限宽神经网络时所考虑的最小化问题，等价于以下最小化问题

$$\inf_f \sum_{i=1}^n (f(x_i) - y_i)^2 + \lambda \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}}. \tag{2}$$

哪些类型的函数能够使该问题达到极小值？正则化项倾向于哪些解？这些基本问题将在接下来的章节中进行研究。我们将证明，这种正则化倾向于选择能够由小型（有限宽）神经网络表示的函数。相反，如果权重衰减项不惩罚神经网络的偏置，那么这种稀疏性就不会特别受到偏好，正如第6节所强调的那样。

4 计算最小范数插值器

To s研究通过带有隐式或显式权重的数据 t
dec正则化训练所得到的解的性质，我们考虑最小范数插值器问题

$$\inf_{\theta, a_0, b_0} \frac{1}{2}\|\theta\|_2^2 \quad \text{such that } \forall i \in [n],\ f_{\theta,a_0,b_0}(x_i) = y_i, \tag{3}$$

其中（x_i， y_i）$_{i\in[n]} \in \mathbb{R}^{2n}$ 是一个训练集。不失一般性，以下假设观测值 x_i 已排序，即 $x_1 < x_2 < \ldots < x_n$。根据定理1，在允许使用无限宽网络时，该问题等价于

$$\inf_f \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} \quad \text{such that } \forall i \in [n],\ f(x_i) = y_i. \tag{4}$$

当宽度大于某个小于 $n-1$ 的阈值时，下面的引理 1 实际上就使这些问题等价。此时，当正则化参数 λ 无限趋小时，方程（4）对应于公式（2）。

引理 1。 公式 （4）中的问题存在一个极小化解。此外，对于 $i_0 :=$ 最小值$\{i \in [n] | x_i \geq 0\}$，任意极小化解都具有以下形式

$$f(x) = ax + b + \sum_{i=1}^{n-1} a_i (x - \tau_i)_+$$

其中 $\tau_i \in$（x_i， x_{i+1}] 对于任意 $i \in \{1, \ldots, i_0 - 2\}$， $\tau_{i_0-1} \in$（x_{i_0-1}， x_{i_0}）以及 $\tau_i \in [x_i, x_{i+1})$ 对于任意$i \in \{i_0, \ldots, n-1\}$。

引理 1 已经对方程（4）的任意极小化解的稀疏性给出了初步保证。该极小化解至多包含 $n-1$ 个折点。相比之下，当偏置项不受正则化时，存在具有无限多个

折点的最小范数插值器 [Debarre 等人， 2022]。第 5 节给出了更强的稀疏恢复结果。引理 1 可视为 Wang 等人定理1 的一个特例 [2021]。在多变量情形下，若没有自由跳跃连接，该定理指出，最小范数插值器在权重的每个激活锥中至多有一个折点（即神经元），[4] 且总折点数不超过 $n+1$。我们的证明思路是，同一个激活锥中的多个折点可以合并为该锥中的一个折点。所得函数仍然能够插值，但表征代价更低。

引理 1 表明，每个区间（x_i， x_{i+1}）（可能在一端闭合）只需考虑 2 个参数。实际上，这类区间上的自由度只有 1：由于插值约束，选定 a_i 便会确定 τ_i（反之亦然）。下文的引理 2 利用这一思路，将最小化问题 (4) 重写为一个状态变量为一维的动态规划 $s_i \in \mathbb{R}$，适用于任意 $i \in [n]$。

引理 2. 若 $x_1 < 0$ 且 $x_n \geq 0$，则对于 $i_0 = \min\{i \in [n] | x_i \geq 0\}$，以下优化问题等价：

$$\min_{\substack{f \\ \forall i \in [n], f(x_i) = y_i}} \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} = \min_{(s_{i_0-1}, s_{i_0}) \in \Lambda} g_{i_0}(s_{i_0}, s_{i_0-1}) + c_{i_0-1}(s_{i_0-1}) + c_{i_0}(s_{i_0}) \tag{5}$$

其中，集合 Λ 以及函数 g_i 和 c_i 定义见下文公式 (6) 至 (8)。

下面介绍定义函数 c_i的动态规划。根据引理 2，该动态规划刻画了最小范数插值器。首先，对任意 $i \in [n-1]$，定义斜率 $\delta_i := \frac{y_{i+1}-y_i}{x_{i+1}-x_i}$；函数

$$g_{i+1}(s_{i+1}, s_i) := \sqrt{\left(x_{i+1}(s_{i+1}-\delta_i) - x_i(s_i-\delta_i)\right)^2 + \left(s_{i+1}-s_i\right)^2} \text{ for any } (s_{i+1}, s_i) \in \mathbb{R}^2; \tag{6}$$

$$\text{and the intervals} \quad S_i(s) := \begin{cases} (-\infty, \delta_i] \text{ if } s > \delta_i \\ \{\delta_i\} \text{ if } s = \delta_i \\ [\delta_i, +\infty) \text{ if } s < \delta_i \end{cases} \quad \text{for any } s \in \mathbb{R}.$$

集合 Λ 是以下三个乘积空间的并：

$$\Lambda := (-\infty, \delta_{i_0-1}) \times (\delta_{i_0-1}, +\infty) \cup \{(\delta_{i_0-1}, \delta_{i_0-1})\} \cup (\delta_{i_0-1}, +\infty) \times (-\infty, \delta_{i_0-1}). \tag{7}$$

最后，我们将函数 c_i：$\mathbb{R} \to \mathbb{R}_+$ 递归定义如下：$c_1 = c_n \equiv 0$，以及

$$\begin{aligned} c_{i+1} : s_{i+1} &\mapsto \min_{s_i \in S_i(s_{i+1})} g_{i+1}(s_{i+1}, s_i) + c_i(s_i) \quad \text{for any } i \in \{1, \ldots, i_0 - 2\} \\ c_i : s_i &\mapsto \min_{s_{i+1} \in S_i(s_i)} g_{i+1}(s_{i+1}, s_i) + c_{i+1}(s_{i+1}) \quad \text{for any } i \in \{i_0, \ldots, n-1\}. \end{aligned} \tag{8}$$

公式 (8) 定义了一个具有连续状态空间的动态规划。直观地说，对于 $i \geq i_0$，变量 s_i 表示点 x_i 处的左导数。项 $g_{i+1}(s_{i+1}, s_i)$ 表示在已知点 (x_i, y_i) 处的左斜率为 s_i 的情况下，以斜率 s_{i+1} 到达点 (x_{i+1}, y_{i+1}) 所需的最小代价（按神经元范数计）。类似地，区间 $S_i(s_i)$ 给出了在已知 x_i 中的斜率为 s_i 时，x_{i+1} 处的可达斜率 [5]。最后，$c_i(s_i)$ 表示在点 (x_i, y_i) 处的左导数为 s_i 时，拟合所有点 $(x_{i+1}, y_{i+1}) \ldots, (x_n, y_n)$ 的最小代价。其递归定义为：将以斜率 s_{i+1} 到达下一个点 (x_{i+1}, y_{i+1}) 的代价 $g_{i+1}(s_{i+1}, s_i)$，与拟合 x_{i+1} 之后所有点的代价 c_{i+1} 相加，并使总和最小。下方的图 1 展示了这一递归定义。对 $i < i_0$ 也有对称的定义。

推导公式（6）的思路是，首先利用引理 1 得到极小函数 f^*的有限表示。接下来，通过在点 (x_{i+1}, y_{i+1}) 和 (x_i, y_i) 之间设置一个折点，求得连接这两点的最小代价。由 s_i 和 s_{i+1} 给出的限制条件，进而可确定唯一可能的折点。将其神经元范数最小化，便得到公式（6）

备注 1. 方程(5) 实际上考虑了两个动态规划的连接点：第一个对应于具有负 x 值的点，第二个对应于具有正值的点。围绕 $x = 0$ 进行这种划分并非引理 *2* 所必需，但能让第 *5* 节中的分析更加简洁。引理 *1* 和引理 *2* 对任意选择的 i_0也成立。特别是对于 $i_0 = 1$，方程(5) 将不再考虑两个动态规划的连接点，而只考虑一个动态规划。

[4]数学定义见附录中的公式（21）。[5]根据引理 3，这里在区间 $[x_i, x_{i+1}]$ 内使用了一个折点。

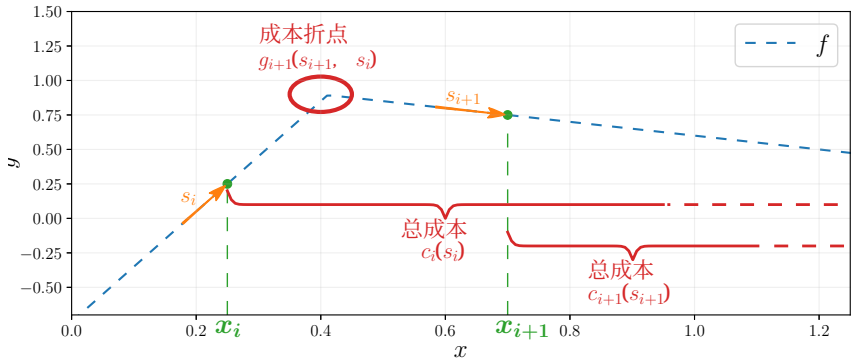


图 1： $i \geq i_0$的动态规划递归定义。

备注 2. 假设 $x_1 < 0$ 和 $x_n \geq 0$ 并非根本性的，只是为了正确定义备注 *1* 中提到的连接点。如果所有 x 值都为正（根据对称性，也可以都为负），则方程(5) 中右侧项的分析会更简单，因为无需考虑连接点。特别是，第 *5* 节中的所有结果在不作此假设的情况下仍然成立。这些结果在附录 *E* 中针对最困难的情形 $x_1 < 0$ 和 $x_n \geq 0$ 得到了证明，其他情形则可直接由此推得。

引理 2 将插值函数中表征代价的极小化表述为一个更简单的动态规划问题，该问题针对每个 x_i 处的斜率序列进行。这一等价关系是本文的关键技术结果，第 5 节据此推导出方程（4）的极小化者的诸多性质。

5 最小范数插值器的性质

借助引理 2 给出的动态规划表述，本节推导出表征代价最小的插值函数的关键性质。具体而言，本节证明方程（4）始终存在唯一的最小值。此外，在满足训练集的某些条件时，该极小化函数在所有插值器中具有最少的折点。

定理 2. 以下优化问题存在唯一极小化解：

$$\inf_f \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} \quad \textit{such that } \forall i \in [n],\ f(x_i) = y_i.$$

定理 2 的证明利用了插值函数与斜率序列 $(s_i)_{i\in[n]} \in \mathcal{S}$ 之间的对应关系，其中集合 $\mathcal{S}$ 由附录 D.2 中的方程（23）定义。具体而言，我们证明以下问题存在唯一极小化解：

$$\min_{\mathbf{s}\in\mathcal{S}} \sum_{i=1}^{n-1} g_{i+1}(s_{i+1}, s_i). \tag{9}$$

我们在下文中指出，$^* \in \mathcal{S}$ 是方程（9）中问题的唯一极小化解。从这个斜率序列 *可以恢复方程（4）的唯一极小化函数。此外，* 按照如下方式极小化由函数 c_i 给出的动态规划：

$$c_{i+1}(s^*_{i+1}) = g_{i+1}(s^*_{i+1}, s^*_i) + c_i(s^*_i) \quad \text{for any } i \in [i_0 - 2]$$
$$c_i(s^*_i) = g_{i+1}(s^*_{i+1}, s^*_i) + c_{i+1}(s^*_{i+1}) \quad \text{for any } i \in \{i_0, \dots, n-1\}.$$

利用附录 E 中引理 7 给出的函数 c_i 的简单性质，除了最小范数插值器的唯一性之外，还可以推导出关于 * 的性质。下面的引理 3 给出了这个极小化解的一条直观性质，这有助于证明本节的主要结果。

引理 3。 对于任意 $i \in [n]$, $s^*_i \in [\min(\delta_{i-1}, \delta_i), \max(\delta_{i-1}, \delta_i)]$, 其中 $\delta_0 := \delta_1$ 并且 $\delta_n := \delta_{n-1}$ 为约定。

引理 3 的几何解释是，x_i 中的最优（左或右）斜率介于连接 $(x_{i-1},\ y_{i-1})$ 与 $(x_i,\ y_i)$ 的直线，以及连接 $(x_i,\ y_i)$ 与 $(x_{i+1},\ y_{i+1})$ 的直线之间。

5.1 恢复最稀疏插值器

现在，我们旨在根据折点的数量，刻画方程（4）的极小化解何时属于最稀疏插值器集合。在描述拟合引理 4 中数据所需的最少折点数量之前

我们将 $[x_1, x_n)$ 划分为形如 $[x_{n_k}, x_{n_{k+1}})$ 的区间，其中 e

$$n_0 = 1 \text{ and for any } k \geq 0 \text{ such that } n_k < n,$$
$$n_{k+1} = \min\{j \in \{n_k + 1, \ldots, n-1\} \mid \operatorname{sign}(\delta_j - \delta_{j-1}) \neq \operatorname{sign}(\delta_{j-1} - \delta_{j-2})\} \cup \{n\}, \quad (10)$$

且按约定，$\operatorname{sign}(0) := 0$。如果我们将 f_{lin} 记作规范分段线性插值器，那么它在每个区间 $[x_{n_k-1}, x_{n_{k+1}}]$ 上均为凸、凹或仿射函数。因此，这种划分将 f_{lin}的空间分为凸、凹和仿射部分，如玩具示例中的图 2 所示。根据引理 4，这种划分对于描述最稀疏插值器至关重要。

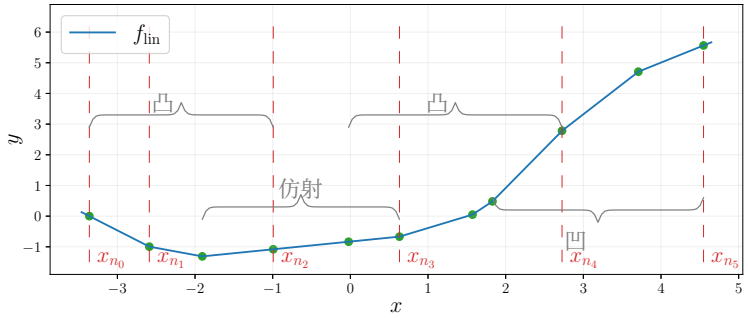


图 2：玩具示例中由 $(n_k)_k$ 给出的划分。

引理 4。 若我们将测度 f'' 的支集的基数记为 $\|f''\|_0$，

$$\min_{\substack{f \\ \forall i, f(x_i) = y_i}} \|f''\|_0 = \sum_{k \geq 1} \left\lceil \frac{n_{k+1} - n_k}{2} \right\rceil \mathbb{1}_{\delta_{n_k - 1} \neq \delta_{n_k}}.$$

引理 4 的证明思路是：对于任意区间 $[x_{k-1}, x_{k+1})$，若 f_{lin} 在该区间上是非仿射的凸函数（分别为凹函数），则任何函数都至少需要一个正折点（分别为负折点），才能拟合该区间内的三个数据点。随后，只需计算这类互不相交区间的数量，并证明某个特定插值函数恰好达到这一数量，即可得到结论。

插值数据所需的最少折点数由引理 4 给出。在给出本节的主要结果之前，我们先对数据 $(x_k,\ y_k)_{k \in [n]}$ 引入以下假设。

假设 1。 对于方程 (10) 中定义的序列$(n_k)_k$：

$$n_{k+1} - n_k \leq 3 \textit{ or } \delta_{n_k} = \delta_{n_k - 1} \quad \textit{for any } k \geq 0.$$

假设 1 的确切含义是，不存在连续 6 个（或更多）数据点 x_k，……，x_{k+5}，使得 f_{lin} 在 $[x_k, x_{k+5}]$ 上为凸函数（且不存在 3 个共线点）或凹函数。该假设在很大程度上取决于真实模型函数的结构（如果存在这样的函数）。例如，若真实函数是分段线性函数，则该假设成立；若真实函数是二次函数，则该假设未必成立。下面的定理 3 表明，在假设 1 下，按折点数量衡量，最小代价插值器属于最稀疏插值器。

定理 3。 若假设 *1* 成立，则

$$\operatorname*{argmin}_{\substack{f \\ \forall i, f(x_i) = y_i}} \|\sqrt{1 + x^2} f''\|_{\text{TV}} \in \operatorname*{argmin}_{\substack{f \\ \forall i, f(x_i) = y_i}} \|f''\|_0. \quad (11)$$

定理 3 给出了相应条件：在这些条件下，表示代价最小的插值函数 f $R_1(f)$在所有插值函数中也具有最少的折点，即 ReLU 隐藏神经元。这说明范数正则化，尤其是将偏置范数加入权重衰减中，会促使估计器只包含少量神经元。使用范数正则化训练神经网络时，最终估计器实际上可能有许多非零神经元，但它们都沿着少数几个关键方向排列。因此，得到的估计器实际上等价于一个窄网络，也就是说，对于每个输入，它们的输出都相同 $x \in \mathbb{R}$。

回想一下，当偏置项未进行正则化时，这种稀疏性并不成立。更准确地说，在这种情况下，某些最稀疏插值器具有最小表征代价，但也存在折点数量任意多（甚至无穷多）的最小代价插值器 [Debarre 等人，2022]。因此，在最小化表征代价而不惩罚偏置项时，并没有特别理由认为所得估计器是稀疏的

。第6节通过实验证明了所恢复估计量在稀疏性上的这一差异，具体取决于范数正则化中是否对偏置参数进行惩罚。

这种稀疏性带来的泛化优势尚不明确。事实上，文献中的泛化界往往依赖参数范数，而非网络宽度（即稀疏度）。在隐式正则化这一特定背景下，理解最稀疏插值器与最小范数插值器之间的关系十分重要。具体而言，Boursier 等人 [2022] 推测，回归问题中的隐式偏置倾向于最小范数插值器；而 Chistikov 等人 [2023]最近证明，隐式偏置有时反而会导向最稀疏插值器。我们的结果表明，最小范数插值器与最稀疏插值器往往会重合，这或许能解释此前为何认为模型会收敛到最小范数插值器。然而，定理 3 和 Chistikov 等人[2023]的结果则表明，至少在某些情况下，隐式偏置更倾向于最稀疏插值器，从而产生不同的估计器[6]。

注释3。 定理 *3* 指出，若假设 *1* 成立，则会发生由方程 (11)给出的稀疏恢复。当 $n_{k+1} - n_k \geq 4$，也就是说， f_{lin} 存在至少包含6 个点的凸区域时，附录*A*给出了一个反例，表明方程 (11)不成立。然而，在比假设 *1* 更弱的数据假设下，方程 (11)仍可能成立。具体而言，当凸区域恰好包含6 个点时，附录*A*给出了稀疏恢复的充要条件。但若允许凸区域至少包含7 个点，要推导出稀疏恢复仍会发生的条件则困难得多。

注释4。 附录*A*中 प्रस्तुत的反例揭示了一个出人意料的结果：最小表示成本插值器未必属于最稀疏的插值器。这一发现与这样一种观点密切相关：通常可能无法将梯度下降的隐式正则化刻画为最小化参数范数 *[Vardi* 和 *Shamir，2021*年，*Chistikov* 等人， *2023]*。具体而言，*Vardi* 和 *Shamir [2021], Chistikov* 等人 *[2023]*依赖于这样的示例：最小范数插值器并不是最稀疏的插值器；相反，隐式正则化更倾向于后者。我们认为，这一内在局限性是其他设置（例如矩阵分解）中观察到不同隐式正则化效应的根本原因之一 *[Gunasekar* 等人，*2017*，*Razin* 和 *Cohen*，*2020*年，*Li* 等人，*2020*年*]*。

5.2 分类的应用

在二元分类设置中，最大间隔分类器被定义为以下问题的极小化解

$$\min_f \overline{R}(f) \quad \text{such that } \forall i \in [n], y_i f(x_i) \geq 1, \tag{12}$$

已知是我们关注的估计量。事实上，对交叉熵损失$l(\hat{y}, y) = \log(1 + e^{-\hat{y}y})$进行梯度下降，其方向会收敛到这类估计量 [Lyu 和 Li，2019年，Chizat 和 Bach， 2020]。定理 3 可用于刻画最大间隔分类器，从而得到推论 1。**推论 1**。

$$\operatorname*{argmin}_{\forall i \in [n], y_i f(x_i) \geq 1} \overline{R}_1(f) \subset \operatorname*{argmin}_{\forall i \in [n], y_i f(x_i) \geq 1} \|f''\|_0.$$

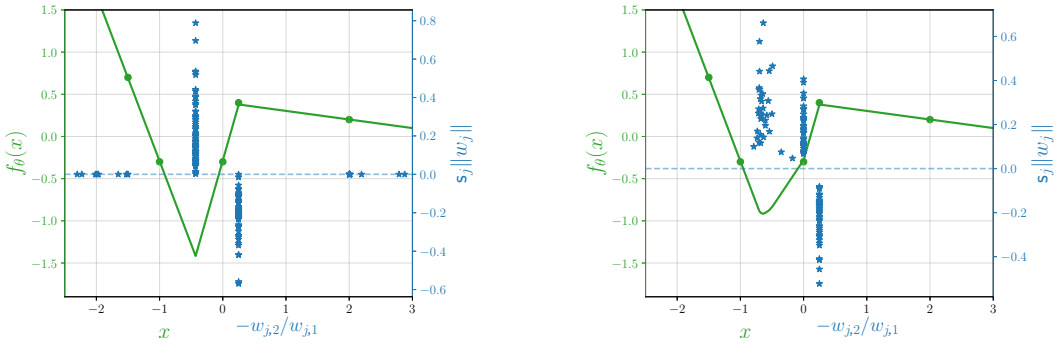
定理 3 表明，在允许使用自由跳跃连接时，最大间隔分类器属于最稀疏的间隔分类器。我们认为，左侧的最小化问题存在唯一解。然而，无法直接从定理 3 推导出唯一性，而是需要借助经过调整的动态规划表述进行另一番深入分析。由于唯一性这一性质并非特别重要，我们在此更倾向于关注定理 3 的一个直接推论。我们强调，对于分类任务，除了数据是单变量的这一条件外，无需作任何数据假设。

6 实验

本节通过图 3 对比了训练单隐藏层 ReLU 神经网络时，在正则化中计入或不计入偏置项所得到的估计器。代码可在 github.com/eboursier/penalising_biases 获取。

[6]请注意，这一细微差别仅适用于回归任务。相反，众所周知，隐式正则化会在分类任务中偏向最小范数插值器 [Chizat 和 Bach， 2020], 而在单变量数据的情况下，这些插值器始终与最稀疏插值器重合（见第 5.2 节中的推论 1）。

在此实验中，我们通过最小化经验损失来训练神经网络，并使用参数的 ℓ_2 范数进行正则化（计入或不计入偏置项），正则化系数为$\lambda = 10^{-3}$。每个神经网络都有 $m = 200$个隐藏神经元，所有参数均独立同分布地初始化为均值为零、方差为 $1/\sqrt{m}$ 的高斯随机变量（采用较大的初始化尺度时也观察到类似结果）。[7] 此处没有自由跳跃连接，这体现了它无害的一面：上述理论所预期的结果，即使没有自由跳跃连接，也同样会出现。附录 B 给出了带有自由跳跃连接的实验，观察结果类似。



(a) 在 ℓ_2 正则化中对偏置项进行惩罚。　　(b) 在 ℓ_2 正则化中忽略偏置项。

图 3：使用 ℓ_2 正则化训练单隐藏层网络时的最终估计器。绿色圆点表示数据，绿色直线表示估计函数。每颗蓝色星号代表网络中的一个隐藏神经元（w_j，b_j）：其 x轴上的值由 $-b_j/w_j$给出，这与其对应 ReLU 的折点位置一致；其 y轴上的值由输出权重 a_j给出。

正如理论研究所预测的，在 ℓ_2 正则化中对偏置项进行惩罚，可以确保最终估计器具有稀疏性。图 3a 中的估计器确实只有 2 个折点（拟合数据所需的最少数量），而图 3b 中神经元的方向则分散开来。更具体地说，图 3b 中的估计器在 $x = -0.5$附近几乎是光滑的，而图 3a 中的稀疏估计器在此处显然不可微。此外，图 3b 中的估计器在 $x = 0$处还包含一个明显的额外折点。因此，图 3 展示了在正则化中计入偏置项能够得到更稀疏的估计器。

7 结论

本文研究了单变量情形下，参数范数对于单隐层 ReLU 神经网络的重要性。具体而言，在允许自由跳跃连接时，表示一个函数所需的参数范数为 $\left\|\sqrt{1+x^2}f''\right\|_{\mathrm{TV}}$。与权重衰减不同，权重衰减在范数中忽略了偏置参数，而表征代价中则额外出现了一个 $\sqrt{1+x^2}$ 加权项。这个加权项至关重要，因为它意味着最小范数插值器具有唯一性。此外，它还促使该插值器在折点数量上呈现稀疏性。因此，最小化参数范数（包括偏置），无论是通过显式正则化还是在训练神经网络时通过隐式正则化实现，都会得到稀疏插值器。我们认为，这种稀疏性是实践中观察到神经网络具有良好泛化性质的原因之一。

尽管这些结果有助于理解最小范数插值器，但将其推广到更一般、更困难的情形仍有待研究。即使在多变量情形下，表征代价 [也可以像 Ongie 等人，2019年所做的那样加以描述，但在这种情形下，若不进行偏置惩罚], ，要刻画最小范数插值器似乎极具挑战。此外，在没有自由跳跃连接的情况下刻画最小范数插值器，也是未来研究面临的一项重大挑战。

[7] 对于小初始化，两种方法都会得到稀疏估计量，因为在这种情况下，偏置项的隐式正则化作用显著。我们的目标仅在于阐明两个问题（有偏置惩罚与无偏置惩罚）的极小化解之间的差异，而不考虑任何优化问题。

致谢与经费披露

作者感谢 Gal Vardi 提出通过直接运用定理 3 来证明推论 1。作者还感谢 Claire Boyer、Julien Fageot 和 LoucasPillaud-Vivien，感谢他们进行了非常有益的讨论并提供了反馈。

参考文献

Francis Bach. 用凸神经网络打破维数灾难。《机器学习研究杂志》，18(1):629–681，2017。

Andrew R Barron. S型函数叠加的通用逼近界。*IEEE* 信息论汇刊，39(3):930–945，1993。

Peter Bartlett. 为实现有效泛化，权重大小比网络规模更重要。神经信息处理系统进展，9，1996。

Mikhail Belkin、Daniel Hsu、Siyuan Ma 和 Soumik Mandal. 调和现代机器学习实践与经典的偏差–方差权衡。《美国国家科学院院刊》，116(32):15849–15854，2019年。

Yoshua Bengio、Nicolas Roux、Pascal Vincent、Olivier Delalleau 和 Patrice Marcotte。凸神经网络。神经信息处理系统进展，18，2005年。

Etienne Boursier、Loucas Pillaud-Vivien 和 Nicolas Flammarion。平方损失和正交输入下浅层ReLU网络的梯度流动力学。*arXiv* 预印本 *arXiv:2206.00939*，2022年。

Emmanuel J Candès 和 Carlos Fernandez-Granda。迈向超分辨率的数学理论。《纯粹与应用数学通讯》，67(6)：906–956，2014年。

Scott Shaobing Chen、David L Donoho 和 Michael A Saunders。通过基追踪进行原子分解。《*SIAM*评论》，43(1)：129–159，2001。

Dmitry Chistikov、Matthias Englert 和 Ranko Lazic。通过浅层ReLU网络学习神经元：相关输入下的动力学与隐式偏置。*arXiv* 预印本 *arXiv:2306.06479*，2023年。

Lenaic Chizat 和 Francis Bach。使用逻辑损失训练的宽型双层神经网络中梯度下降的隐式偏置。载于学习理论会议，第 1305–1338 页。PMLR，2020。

Thomas Debarre、Quentin Denoyelle、Michael Unser 和 Julien Fageot。一维数据的最稀疏分段线性回归。计算与应用数学杂志，406：114044，2022。

Weinan E 和 Stephan Wojtowytsch。Barron函数的表示公式与逐点性质，2021。

Carlos Fernandez-Granda。通过凸规划实现点源超分辨率。信息与推断：*IMA*杂志，5(3)：251–303，2016。

Ian Goodfellow、Yoshua Bengio 和 Aaron Courville。深度学习。麻省理工学院出版社，2016。http: //www.deeplearningbook.org。

Suriya Gunasekar、Blake E Woodworth、Srinadh Bhojanapalli、Behnam Neyshabur 和 Nati Srebro。矩阵分解中的隐式正则化。神经信息处理系统进展，30，2017。

何恺明、张祥雨、任少卿和孙剑。用于图像识别的深度残差学习。载于*IEEE*计算机视觉与模式识别会议论文集，第770–778页，2016年。

Ziwei Ji 和 Matus Telgarsky。不可分数据上梯度下降的隐式偏置。载于学习理论会议，第 1772–1798 页。PMLR，2019年。

Anders Krogh 和 John Hertz。简单的权重衰减可以改善泛化。神经信息处理系统进展，第 4 卷，1991 年。Vera Kurková 和 Marcello Sanguineti。变基逼近和神经网络逼近速率的界。*IEEE* 信息论汇刊，47(6):2659–2665，2001。Zhiyuan Li、Yuping Luo 和 Kaifeng Lyu。探索如何消除矩阵分解中梯度下降的隐式偏置：贪心低秩学习。*arXiv* 预印本 *arXiv:2012.09839*，2020年。Zhong Li、Chao Ma 和 Lei Wu。使用基于路径的范数度量一般激活函数神经网络的复杂度。*arXiv* 预印本 *arXiv:2009.06132*，2020b。Kaifeng Lyu 和 Jian Li。梯度下降最大化齐次神经网络的间隔。载于国际学习表征会议，2019年。Mor Shpigel Nacson、Kavya Ravichandran、Nathan Srebro 和 Daniel Soudry。线性对角神经网络中步长的隐式偏置。载于国际机器学习会议，第 16270–16295 页。PMLR，2022年。Behnam Neyshabur、Ryota Tomioka 和 Nathan Srebro。探寻真正的归纳偏置：论隐式正则化在深度学习中的作用。*arXiv* 预印本 *arXiv:1412.6614*，2014年。Andrew Ng。稀疏自编码器。*CS294A* 讲义，72(2011):1–19，2011 年。

Greg Ongie、Rebecca Willett、Daniel Soudry 和 Nathan Srebro。 有界范数无限宽度 ReLU 网络的函数空间视角：多变量情形。载于 国际学习表征会议（*ICLR 2020*），2019年。

Rahul Parhi 和 Robert D Nowak。神经网络与脊样条的巴拿赫空间表示定理。机器学习研究杂志，22(43):1–40，2021。

Rahul Parhi 和 Robert D Nowak。深度学习遇上稀疏正则化：信号处理视角。*arXiv* 预印本 *arXiv:2301.09554*，2023年。

Clarice Poon、Nicolas Keriven 和 Gabriel Peyré。无网格稀疏正则化的支撑集定位与费 Fisher 度量。载于第*22*届国际人工智能与统计会议，第1341–1350页。PMLR，2019年。Noam Razin 和 Nadav Cohen。深度学习中的隐式正则化可能无法用范数解释。神经信息处理系统进展，33：21174–21187，2020年。Itay Safran、Gal Vardi 和 Jason D Lee。关于浅层单变量 ReLU 网络中线性区域的有效数量：收敛保证与隐式偏置。神经信息处理系统进展，35：32667–32679，2022年。Clayton Sanford、Navid Ardeshir 和 Daniel Hsu。 $\mathcal{R}$-范数归纳偏置的内在维数与泛化性质。*arXiv* 预印本 *arXiv:2206.05317*，2022年。Pedro Savarese、Itay Evron、Daniel Soudry 和 Nathan Srebro。无限宽度有界范数网络在函数空间中是什么样子？载于学习理论会议，第2667–2690页。PMLR，2019年。Alexander Shevchenko、Vyacheslav Kungurtsev 和 Marco Mondelli。宽 ReLU 网络分段线性解的平均场分析。《机器学习研究杂志》，23（1）：5660–5714，2022年。Daniel Soudry、Elad Hoffer、Mor Shpigel Nacson、Suriya Gunasekar 和 Nathan Srebro。梯度下降在可分数据上的隐式偏置。《机器学习研究杂志》，19（1）：2822–2878，2018年。Lawrence Stewart、Francis Bach、Quentin Berthet 和 Jean-Philippe Vert。将回归视为分类：任务表述对神经网络特征的影响。*arXiv* 预印本 *arXiv:2211.05641*，2022年。Gal Vardi 和 Ohad Shamir。平方损失下 ReLU 网络中的隐式正则化。载于学习理论会议，第4224–4258页。PMLR，2021。

Yifei Wang、Jonathan Lacotte 和 Mert Pilanci。正则化双层 ReLU 网络中隐藏的凸优化景观：最优解的精确刻画。载于国际学习表征会议，2021。

Chiyuan Zhang、Samy Bengio、Moritz Hardt、BenjaminRecht 和 Oriol Vinyals。《理解深度学习（仍然）需要重新思考泛化》。《美国计算机学会通讯》，64(3):107–115，2021。

讨论假设 1

定理 3 需要假设 1，该假设认为 f_{lin} 中不存在包含至少 6 个数据点的凸（或凹）区域。实际上，当存在一个恰好包含 6 个数据点的凸（或凹）区域，即 $n_{k+1} = n_k + 4$ 时，定理 3 对该区域成立，当且仅当对于 $i = n_k + 1$:

$$\frac{\langle u_i, w_{i-1}\rangle\langle u_{i+1}, w_{i+1}\rangle}{\|w_{i-1}\|\,\|w_{i+1}\|} - \langle u_i, u_{i+1}\rangle \le \sqrt{\|u_i\|^2 - \frac{\langle u_i, w_{i-1}\rangle^2}{\|w_{i-1}\|^2}}\sqrt{\|u_{i+1}\|^2 - \frac{\langle u_{i+1}, w_{i+1}\rangle^2}{\|w_{i+1}\|^2}} \tag{13}$$

$$\text{where} \quad u_i = (x_i, 1); \quad w_{i-1} = \frac{\delta_i - \delta_{i-1}}{\delta_i - \delta_{i-2}}(x_i, 1) + \frac{\delta_{i-1} - \delta_{i-2}}{\delta_i - \delta_{i-2}}(x_{i-1}, 1);$$

$$\text{and} \quad u_{i+1} = (x_{i+1}, 1); \quad w_{i+1} = \frac{\delta_{i+2} - \delta_{i+1}}{\delta_{i+2} - \delta_i}(x_{i+2}, 1) + \frac{\delta_{i+1} - \delta_i}{\delta_{i+2} - \delta_i}(x_{i+1}, 1).$$

该结果的证明（此处省略）表明，问题

$$\min_{(s_i, s_{i+1}) \in [\delta_{i-1}, \delta_i] \times [\delta_i, \delta_{i+1}]} g_i(s_i, s^*_{i-1}) + g_{i+1}(s_{i+1,s_i}) + g_{i+2}(s^*_{i+2}, s_{i+1})$$

当且仅当公式（13）成立时，在 $(s_i, s_{i+1}) = (\delta_i, \delta_i)$ 处取得最小值；这对应于在该凸区域上插值数据的（唯一）最稀疏方式。为了证明最小值在该点取得，我们可以先注意到 $s^*_{i-1} = \delta_{i-2}$ 和 $s^*_{i+2} = \delta_{i+2}$。接下来，需要细致研究点 (δ_i, δ_i) 处（凸但不可微的）目标函数的方向导数。

下方的图4展示了一个包含6个数据点的情况，此时公式（13）的条件不成立。显然，在这种情况下，最小范数插值器与（唯一的）最稀疏插值器不同。

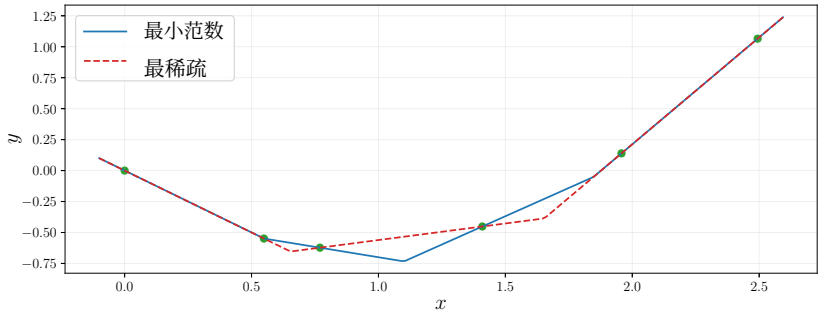


图4：最小范数插值器与最稀疏插值器之间存在差异的情况。

当考虑的点超过 6 个时，研究最小化问题会变得繁琐，也无法推导出简单的稀疏恢复条件。使用较大的凸区域（例如 35 个点）生成随机数据时，最小范数插值器很少是最稀疏插值器之一。此外，其折点数量似乎可以任意接近 34，即引理 3 给出的折点数量平凡上界；而最稀疏插值器只有 17 个折点。

B 补充实验

图5展示了玩具示例图 2 中方程（4）的极小化解。借助引理 2 给出的动态规划，可以计算出该极小化函数。尽管该动态规划的变量是连续的，但根据引理 3，我们可以将每个斜率 s_i 的约束空间近似为由 $[\delta_{i-1}, \delta_i]$ 构成的离散网格，从而高效地求解它。对于图5所使用的数据，假设 1 成立。显然，该极小化解非常稀疏，仅包含 4 个折点。图 2 给出的划分随后表明，根据引理 4，这确实是可能的最少折点数。另一方面，规范分段线性插值器 f_{lin} 并没有那么稀疏，在此处包含 7 个折点。

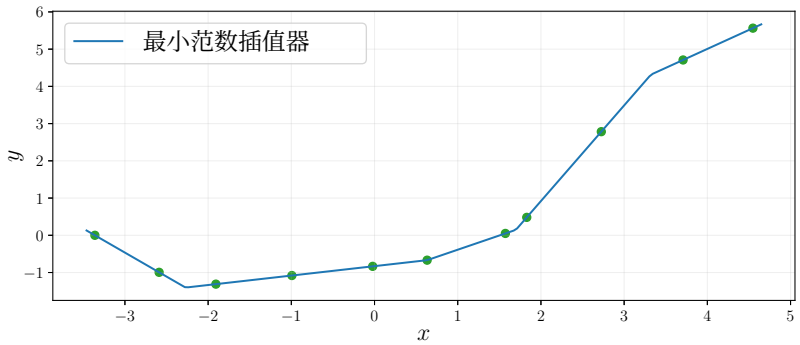
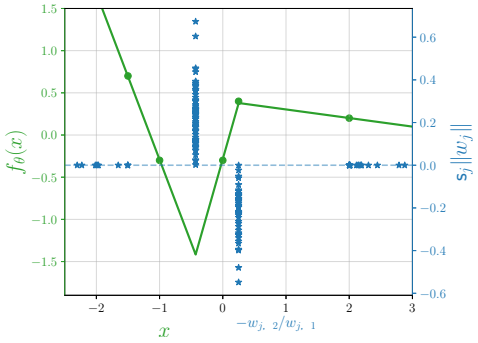
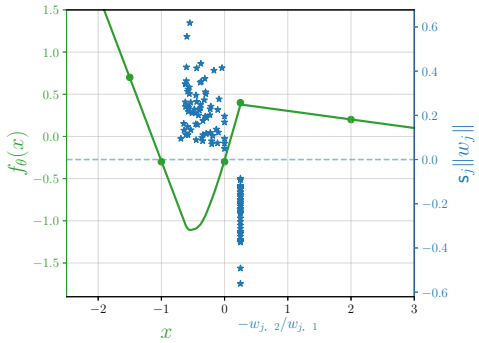


图5：玩具数据示例中方程（4）的极小化解。

图6考虑了与第6节中图 3 完全相同的设置。唯一的区别是，我们在此允许神经网络架构中存在一个自由跳跃连接，这正好对应于定理 3 所描述的设置。



（a）对 ℓ_2 正则化中的偏置项进行惩罚时的最终估计器。



（b）使用 ℓ_2 正则化训练单隐层 ReLU 神经网络**并采用自由跳跃连接**时的最终估计量。

图6：使用 ℓ_2 正则化训练单隐藏层ReLU神经网络时的最终估计器。绿色点对应数据，绿色直线为估计函数。每个蓝色星号代表网络的一个隐藏神经元（w_j，b_j）：其 x轴取值由 $-b_j/w_j$给出，这与其关联ReLU的折点位置一致；其 y轴取值由输出层权重 a_j给出。

可以观察到类似现象：仅在正则化中计入偏置项时得到的估计器只有2个折点，而在正则化中省略偏置后得到的估计器则平滑得多（因此在折点数量上也稀疏得多）。唯一的区别是，后一个估计器在此处于 $x = 0$处没有明显的折点，而是在区间 $[-0.5, 0]$上更加平滑。这可由该区间内存在更多分散的折点来解释。尽管存在这一细微差别，主要观察结果保持不变：计入偏置项时，估计器是一个最稀疏的估计器；而省略偏置时，它包含大量折点（甚至是平滑的）。

C 第3节证明

下面的定理4推广了定理1中表征代价 $R_1(f)$ 的刻画。

定理4。 对于任意利普希茨函数 $f:\ \mathbb{R} \to \mathbb{R}$,

$$\overline{R}_1(f) = \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} = \int_{\mathbb{R}} \sqrt{1+x^2}\, \mathrm{d}|f''|(x)$$

$$\textit{and} \quad \overline{R}(f) = \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} + D(x_f, \mathcal{C}_f),$$

其中

$$x_f = \left(f'(+\infty) + f'(-\infty), f(0) - \int_{\mathbb{R}} |x| \mathrm{d}f''(x) \right)$$
$$\mathcal{C}_f = \left\{ \int_{\mathbb{R}} \varphi(x) \mathrm{d}f''(x), - \int_{\mathbb{R}} x\varphi(x) \mathrm{d}f''(x) \mid \|\varphi\|_\infty \leq 1 \right\}$$
$$D(x_f, \mathcal{C}_f) = \inf_{x \in \mathcal{C}_f} \|x - x_f\|.$$

对于任意非 *Lipschitz* 函数，$\overline{R}_1(f) = \overline{R}(f) = +\infty$。

证明。我们这里只证明 $R(f)$ 上的等式。定理4的另一部分可直接由此证明得出。首先假设 $R(f)$ 有限。于是我们可以考虑某个 $\mu \in \overline{\mathcal{M}}(\mathbb{S}_1)$，使得对于任意 $x \in \mathbb{R}$, $f(x) = \int_{\mathbb{S}_1} \sigma(wx + b) \mathrm{d}\mu(w, b)$。注意，$f$必然是 $\|\mu\|_{\mathrm{TV}}$-利普希茨的，这就证明了定理4的第二部分。不失一般性，我们可以用 $T^{-1}(\theta) = (\cos\theta, \sin\theta)$ 将 $\mathbb{S}_1$ 参数化在 $\theta \in [-\frac{\pi}{2}, \frac{3\pi}{2})$ 上。若记 $\nu = T_{\#}\mu$ 为 μ经 T映射后的推前测度，则有

$$f(x) = \int_{-\frac{\pi}{2}}^{\frac{3\pi}{2}} \sigma(x\cos\theta + \sin\theta) \mathrm{d}\nu(\theta). \tag{14}$$

由于 μ的全变差有界，因而 ν 也有界，我们可以将积分号下的表达式写为：

$$f'(x) = \int_{-\frac{\pi}{2}}^{\frac{3\pi}{2}} \cos\theta \mathbb{1}_{x\cos\theta + \sin\theta \geq 0} \mathrm{d}\nu(\theta).$$

现在注意到 $x\cos\theta + \sin\theta \geq 0$ 当且仅当 $\theta \in [-\arctan(x), \pi - \arctan(x)]$，因为 $\theta \in [-\frac{\pi}{2}, \frac{3\pi}{2})$，即

$$f'(x) = \int_{-\arctan x}^{\pi - \arctan x} \cos\theta \mathrm{d}\nu(\theta).$$

在 x 上推导该表达式时，我们最终得到分布 f''

$$\begin{aligned} \mathrm{d}f''(x) &= -\frac{\cos(\pi - \arctan(x)) \, \mathrm{d}\nu(\pi - \arctan(x)) - \cos(-\arctan(x)) \, \mathrm{d}\nu(-\arctan(x))}{1 + x^2} \\ &= \frac{\cos(\arctan(x)) (\mathrm{d}\nu(\pi - \arctan(x)) + \mathrm{d}\nu(-\arctan(x)))}{1 + x^2}. \end{aligned}$$

对于连续分布 ν，这个等式显然成立。将其推广到任意分布ν 需要一些额外工作，但可以按照分布导数的典型定义推导得到。

定义 $\nu_+(\theta) = \nu(\theta) + \nu(\pi + \theta)$，对任意 $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2})$ 均成立，并注意到 $\cos(\arctan x) = \frac{1}{\sqrt{1+x^2}}$

$$\forall x \in \mathbb{R}, \sqrt{1 + x^2} \mathrm{d}f''(x) = \frac{\mathrm{d}\nu_+(-\arctan(x))}{1 + x^2}.$$

或者等价地，对任意 $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$

$$\frac{\mathrm{d}f''(-\tan\theta)}{\cos\theta} = \cos^2(\theta) \, \mathrm{d}\nu_+(\theta). \tag{15}$$

与 Savarese 等人 [2019] 的证明类似，f'' 固定了 ν_+，唯一的自由度在$\nu_\perp(\theta) := \nu(\theta) - \nu(\pi + \theta)$。现在，证明确定了哪个有效的 $\nu_\perp$ 使 $\|\mu\|_{\mathrm{TV}} = \|\nu\|_{\mathrm{TV}}$ 最小。方程 (14) 对 $\nu_\perp$给出了以下条件

$$\begin{aligned} f(x) &= \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sigma(-x\cos\theta - \sin\theta) \mathrm{d}(\nu_+ + \nu_\perp)(\theta) + \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sigma(x\cos\theta + \sin\theta) \mathrm{d}(\nu_+ - \nu_\perp)(\theta) \\ &= \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} |x\cos\theta + \sin\theta| \mathrm{d}\nu_+(\theta) + \frac{x}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos\theta \mathrm{d}\nu_\perp(\theta) + \frac{1}{2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin\theta \mathrm{d}\nu_\perp(\theta). \end{aligned}$$

虽然 ν_+ 由 f''给出，但对于 f的仿射部分，$\nu_\perp$ 成立。上述等式直接导出关于 $\nu_\perp$的以下条件

$$\begin{cases} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos\theta \, \mathrm{d}\nu_\perp(\theta) = f'(+\infty) + f'(\infty) \\ \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin\theta \, \mathrm{d}\nu_\perp(\theta) = f(0) - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} |\sin\theta| \mathrm{d}\nu_+(\theta) \end{cases}. \tag{16}$$

现在注意到 $2\|\mu\|_{\mathrm{TV}} = \|\nu_+ + \nu_\perp\|_{\mathrm{TV}} + \|\nu_+ - \nu_\perp\|_{\mathrm{TV}}$，因此 $\bar{R}(f)$由下式给出

$$2\bar{R}(f) = \min_{\nu_\perp} \|\nu_+ + \nu_\perp\|_{\mathrm{TV}} + \|\nu_+ - \nu_\perp\|_{\mathrm{TV}} \quad \text{such that } \nu_\perp \text{ verifies Equation (16).}$$

由下文的引理 5 可知[8]

$$\bar{R}(f) = \|\nu_+\|_{\mathrm{TV}} + D(x_f, \mathcal{C}_f),$$

其中，x_f 和 $\mathcal{C}_f$ 由定理4定义。当 $R(f)$有限时，对方程（15）作简单的变量替换可得

$$\bar{R}(f) = \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} + D(x_f, \mathcal{C}_f).$$

反过来，当 $\left\|\sqrt{1+x^2}f''\right\|_{\mathrm{TV}}$有限时，我们可以按照方程（15）定义 ν_+，并将 $\nu_\perp$ 定义为由 $-\frac{\pi}{2}$ 中的狄拉克测度和 0-构成的和，使其满足方程（16）。相应的 μ 具有有限全变差，因此$R(f)$有限。至此，定理4第一部分的证明结束：

$$\bar{R}(f) = \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}} + D(x_f, \mathcal{C}_f).$$

对于$R_1(f)$，分析更为简单，因为对 $\nu_\perp$没有约束，其最优取值为 $\nu_\perp = 0$。

□

引理 5。 最小化程序

$$\begin{aligned} &\min_{\nu_\perp} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\nu_+ + \nu_\perp| + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\nu_+ - \nu_\perp| \\ &\textit{such that } \left(\int_{-\pi/2}^{\pi/2} \cos\theta \, \mathrm{d}\nu_\perp(\theta), \int_{-\pi/2}^{\pi/2} \sin\theta \, \mathrm{d}\nu_\perp(\theta) \right) = (a, b) \end{aligned} \tag{17}$$

等价于

$$\begin{aligned} &2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\nu_+| + 2\min_{u\in\mathcal{C}} \|(a,b) - u\|, \\ &\textit{where } \mathcal{C} = \left\{ \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos(\theta)\varphi(\theta)\mathrm{d}\nu_+(\theta), \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin(\theta)\varphi(\theta)\mathrm{d}\nu_+(\theta) \right) \mid \|\varphi\|_\infty \le 1 \right\}. \end{aligned} \tag{18}$$

证明。 对于方程（17）的约束集中的任意 $\nu_\perp$，我们可以作分解 $\nu_\perp = \varphi\nu_+ + \mu_2$，其中 $\|\varphi\|_\infty \le 1$。于是逐点有

$$\begin{aligned} |\nu_+ + \nu_\perp| + |\nu_+ - \nu_\perp| &\le 2|\mu_2| + |(1+\varphi)\nu_+| + |(1-\varphi)\nu_+| \\ &= 2|\mu_2| + 2|\nu_+|. \end{aligned} \tag{19}$$

因此，若记 v 方程（17）给出的下确界：

$$v \le 2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\nu_+| + 2\min_{(\varphi,\mu_2)\in\Gamma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\mu_2|,$$

其中

$$\Gamma = \left\{ (\varphi, \mu_2) \;\middle|\; \|\varphi\|_\infty \le 1 \text{ and } \int_{-\pi/2}^{\pi/2} (\cos\theta, \sin\theta) \; (\varphi(\theta)\mathrm{d}\nu_+(\theta) + \mathrm{d}\mu_2(\theta)) = (a,b) \right\}.$$

此外，对于固定的 $\nu_\perp$，我们可以选择（φ，μ_2），使得：

$$\begin{cases} \varphi = \operatorname{sign}(\frac{\mathrm{d}\nu_\perp}{\mathrm{d}\nu_+}) \min\left(\left| \frac{\mathrm{d}\nu_\perp}{\mathrm{d}\nu_+} \right|, 1 \right), \\ \mu_2 = \nu_\perp - \varphi\nu_+, \end{cases}$$

其中 $\mathrm{d}\nu_\perp/\mathrm{d}\nu_+$ 是对拉东–尼科迪姆导数 $\mathrm{d}\nu_a\mathrm{d}\nu_+$ 的一种记号滥用，且勒贝格分解为 $\nu_\perp = \nu_a + \nu_s$，其中 $\nu_a \ll \nu_+$ 和 $\nu_s \perp \nu_+$。对于这一选择，方程（19）取等号，由此可直接推出

$$v = 2\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\nu_+| + 2\min_{(\varphi,\mu_2)\in\Gamma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\mu_2|.$$

[8]还需要进行变量替换，以观察到引理 5 中的最小值等于 $D(x_f, \mathcal{C}_f)$。

现在只需证明 $\min_{u\in\mathcal{C}} \|(a,b)-u\|^2 = \min_{(\varphi,\mu_2)\in\Gamma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\mu_2|$。下文固定 φ 使得 $\|\varphi\|_\infty \le 1$，并注意

$$\begin{cases} x = a - \int_{-\pi/2}^{\pi/2} \cos\theta\, \varphi(\theta)\mathrm{d}\nu_+(\theta), \\ y = b - \int_{-\pi/2}^{\pi/2} \sin\theta\, \varphi(\theta)\mathrm{d}\nu_+(\theta). \end{cases}$$

现在只需证明，对于任意固定的 φ:

$$\min_{\mu_2 \text{ s.t. } (\varphi,\mu_2)\in\Gamma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\mu_2| = \|(x,y)\|.$$

约束集实际上是 $\left\{\mu_2 \mid \int_{-\pi/2}^{\pi/2} (\cos\theta, \sin\theta)\, \mathrm{d}\mu_2(\theta) = (x,y)\right\}$。现在定义

$$\theta^* = \arcsin\left(\frac{\mathrm{sign}(x)y}{\sqrt{x^2+y^2}}\right) \quad \text{and} \quad \mu_2^* = \mathrm{sign}(x)\sqrt{x^2+y^2}\,\delta_{\theta^*},$$

其中，δ_{θ^*} 是位于 θ^*处的狄拉克分布。此定义仅当 $x \neq 0$时有效，否则我们取 $\mu_2^* = -y\delta_{-\frac{\pi}{2}}$。

注意，μ_2^*属于约束集，且 $\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\mu_2| = \|(x,y)\|$，即

$$\min_{\mu_2 \text{ s.t. } (\varphi,\mu_2)\in\Gamma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\mu_2| \le \|(x,y)\|.$$

现在考虑约束集中的任意 μ_2，并将 $\mu_2 = \mu_2^+ - \mu_2^-$分解为 $\mathcal{M}_+([-\frac{\pi}{2},\frac{\pi}{2}])^2$ 定义

$$(x_+,y_+) = \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos\theta\,\mathrm{d}\mu_2^+, \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin\theta\,\mathrm{d}\mu_2^+\right)$$

$$(x_-,y_-) = \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos\theta\,\mathrm{d}\mu_2^-, \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin\theta\,\mathrm{d}\mu_2^-\right)$$

根据柯西-施瓦茨不等式,

$$\int \cos^2(\theta)\,\mathrm{d}\mu_2^+(\theta) \int \mathrm{d}\mu_2^+(\theta) \ge x_+^2,$$

$$\int \sin^2(\theta)\,\mathrm{d}\mu_2^+(\theta) \int \mathrm{d}\mu_2^+(\theta) \ge y_+^2.$$

将这两个不等式相加可得

$$\int \mathrm{d}\mu_2^+ \ge \sqrt{x_+^2+y_+^2}.$$

同理，我们有

$$\int \mathrm{d}\mu_2^- \ge \sqrt{x_-^2+y_-^2}.$$

回忆一下，$\int \mathrm{d}|\mu_2| = \int \mathrm{d}\mu_2^+ + \int \mathrm{d}\mu_2^-$。根据三角不等式，可得:

$$\begin{aligned} \int \mathrm{d}|\mu_2| &\ge \|(x_+,y_+)\| + \|(x_-,y_-)\| \\ &\ge \|(x_+,y_+) - (x_-,y_-)\| = \|(x,y)\|. \end{aligned}$$

因此:

$$\min_{\mu_2 \text{ s.t. } (\varphi,\mu_2)\in\Gamma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\mu_2| \ge \|(x,y)\|.$$

最终，我们证明了

$$\min_{\mu_2 \text{ s.t. } (\varphi,\mu_2)\in\Gamma} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}|\mu_2| = \left\|(a,b) - \left(\int_{-\pi/2}^{\pi/2} \cos\theta\,\varphi(\theta)\mathrm{d}\nu_+(\theta), \int_{-\pi/2}^{\pi/2} \sin\theta\,\varphi(\theta)\mathrm{d}\nu_+(\theta)\right)\right\|.$$

对 φ取下确界后，即可得到引理 5。 □

第 D 节 第4节的证明

D.1 引理 1 的证明

我们首先需要证明最小值存在。根据 $R_1(f)$的定义和定理1，方程（4）等价于

$$\inf_{\mu,a,b} \|\mu\|_{\mathrm{TV}} \quad \text{such that for any } i \in [n],\ f_{\mu,a,b}(x_i) = y_i. \tag{20}$$

考虑一个序列（μ_j, a_j, b_j）$_j$，满足 $f_{\mu_j,a_j,b_j}(x_i) = y_i$对任意 i 和 j，且 $\|\mu_j\|_{\mathrm{TV}}$ 收敛于方程（20）的下确界。序列 $\|\mu_j\|_{\mathrm{TV}}$ 必然有界。这也意味着（a_j）和（b_j）有界[9]。由于 $\mathbb{S}_1$ 上有限符号测度构成的空间是巴拿赫空间，因此存在一个子序列弱收敛到某个（μ, a, b）。根据弱收敛，（μ, a, b）属于方程（20）的约束集，并且 $\|\mu\|_{\mathrm{TV}} = \lim_j \|\mu_j\|_{\mathrm{TV}}$。（$\mu, a, b$）因此是方程（20）的极小化解。由此，我们证明了方程（4）存在一个最小值，该最小值在 $f_{\mu,a,b}$ 处取得。

为证明之用，将激活锥 C_i 定义为

$$\begin{aligned}
C_0 &= \left\{\theta \in \mathbb{R}^2 \mid \forall i = 1, \ldots, n, \langle \theta, (x_i, 1)\rangle \geq 0\right\}, \\
C_i &= \left\{\theta \in \mathbb{R}^2 \mid \langle \theta, (x_{i+1}, 1)\rangle \geq 0 > \langle \theta, (x_i, 1)\rangle\right\} \quad \text{for any } i = 1, \ldots, i_0 - 2, \\
C_{i_0-1} &= \left\{\theta \in \mathbb{R}^2 \mid \langle \theta, (x_{i_0}, 1)\rangle > 0 > \langle \theta, (x_{i_0-1}, 1)\rangle\right\}, \\
C_i &= \left\{\theta \in \mathbb{R}^2 \mid \langle \theta, (x_{i+1}, 1)\rangle > 0 \geq \langle \theta, (x_i, 1)\rangle\right\} \quad \text{for any } i = i_0, \ldots, n - 1, \\
C_n &= \left\{\theta \in \mathbb{R}^2 \setminus \{0\} \mid \forall i = 1, \ldots, n, \langle \theta, (x_i, 1)\rangle \leq 0\rangle\right\}.
\end{aligned} \tag{21}$$

由于 x_i 是有序的，注意到 $(C_0,\ C_1,\ -C_1,\ \ldots,\ C_{n-1},\ -C_{n-1},\ C_n)$ 构成了 $\mathbb{R}^2$的一个划分。为证明引理 1，只需证明方程（20）的任意极小化解 $(\mu_a,\ b)$ 对应的函数 $f_{\mu,\ a,\ b}$具有如下形式

$$f_{\mu,a,b}(x) = \tilde{a}x + \tilde{b} + \sum_{i=1}^{n-1} \tilde{a}_i \sigma(\langle \theta_i, (x, 1)\rangle) \quad \text{where } \theta_i \in C_i.$$

设 f 是方程（4）的一个极小化解。设 $\mu,\ a,\ b$ 是方程（20）的一个极小化解，满足$f_{\mu,\ a,\ b} = f$。~将 $\tilde{\mu},\ \tilde{a},\ b$ 定义为

$$\mathrm{d}\tilde{\mu}(\theta) = \begin{cases} \mathrm{d}\mu(\theta) + \mathrm{d}\mu(-\theta) & \text{for } \theta \in C_i \text{ for any } i = 1, \ldots, n-1 \\ 0 & \text{for } \theta \in -C_i \text{ for any } i = 1, \ldots, n-1 \\ \mathrm{d}\mu(\theta) & \text{otherwise,} \end{cases}$$

$$\tilde{a} = a - \sum_{j=1}^{n-1} \int_{-C_j} \theta_1 \mathrm{d}\mu(\theta),$$

$$\tilde{b} = b - \sum_{j=1}^{n-1} \int_{-C_j} \theta_2 \mathrm{d}\mu(\theta).$$

根据恒等式 $\sigma(u) - u = \sigma(-u)$，$f_{\mu,\ a,\ b} = f_{\tilde{\mu},\ \tilde{a},\ \tilde{b}}$。此外，$\|\tilde{\mu}\|_{\mathrm{TV}} \leq \|\mu\|_{\mathrm{TV}}$，因此不失一般性，我们可以假设 μ 的支撑包含于 $\bigcup_{i=0}^{n} C_i$ 中。在这种情况下，对任意 $i =, \mathbf{1} \ldots,\ n$

$$\begin{aligned}
f(x_i) &= ax_i + b + \sum_{j=0}^{i-1} \int_{C_j} \langle \theta, (x_i, 1)\rangle \mathrm{d}\mu(\theta) \\
&= ax_i + b + \sum_{j=0}^{i-1} \langle \int_{C_j} \theta \mathrm{d}\mu(\theta), (x_i, 1)\rangle.
\end{aligned} \tag{22}$$

[9]为说明这一点，我们可以先考察差值 $f_{\mu_j,\ a_j,\ b_j}(x_1) - f_{\mu_j,\ a_j,\ b_j}(x_2)$，以证明 $(a_j)_j$ 有界。进而，考察 $f_{\mu_j,\ a_j,\ b_j}(x_1)$ 时，可得出 $(b_j)_j$ 有界。[10]这里我们对三元组 $(\mu,\ a,\ b)$ 进行了变换，但相应的函数 f保持不变。

首先，约化

$$\tilde{a} = a + \int_{C_0} \theta_1 \mathrm{d}\mu(\theta)$$
$$\tilde{b} = b + \int_{C_0} \theta_2 \mathrm{d}\mu(\theta)$$
$$\tilde{\mu} = \mu_{|\bigcup_{i=1}^{n-1} C_i},$$

不会增加 μ 的全变差，并且仍然插值数据。因此， μ 的支撑包含于 $\bigcup_{i=1}^{n-1} C_i$。现在令 $\mu = \mu_+ - \mu_-$ 为 μ 的约当分解，并对任意 $i \in [n-1]$定义

$$\alpha_i = \int_{C_i} \theta \mathrm{d}\mu_+(\theta) \quad \text{and} \quad \beta_i = \int_{C_i} \theta \mathrm{d}\mu_-(\theta).$$

Not注意， α_i 和 β_i 都属于正凸锥 C_i。对于 $\theta_i := \alpha_i - \beta_i$，公式（22）改写为 s

$$f(x_i) = ax_i + b + \sum_{j=1}^{i-1} \langle \theta_i, (x_i, 1) \rangle.$$

若 $\theta_i \in \overline{C}_i \cup \overline{-C_i}$，则我们可以定义

$$\tilde{\mu} = \mu - \mu_{|C_i} + \|\theta_i\| \delta_{\frac{\theta_i}{\|\theta_i\|}}.$$

根据公式（22），函数 $f_{\tilde{\mu},\, a,\, b}$ 仍然插值数据，并且

$$\|\tilde{\mu}\|_{\mathrm{TV}} \leq \|\mu\|_{\mathrm{TV}} - \left\|\mu_{|C_i}\right\|_{\mathrm{TV}} + \|\theta_i\|.$$

由 $\|\mu\|_{\mathrm{TV}}$的极小化可知，这是一个等式。此外，由于 μ 是球面上的测度，

$$\left\|\mu_{|C_i}\right\|_{\mathrm{TV}} = \int_{C_i} \|\theta\| \mathrm{d}\mu_+(\theta) + \int_{C_i} \|\theta\| \mathrm{d}\mu_-(\theta)$$
$$\geq \left\| \int_{C_i} \theta \mathrm{d}\mu_+(\theta) \right\| + \left\| \int_{C_i} \theta \mathrm{d}\mu_-(\theta) \right\|$$
$$= \|\alpha_i\| + \|\beta_i\| \geq \|\theta_i\|.$$

由极小化可知，所有不等式均为等式。詹森等号情形表明，对于第一个不等式，$\mu_{+|C_i}$和$\mu_{-|C_i}$ 都是狄拉克测度；而第二个不等式则意味着 $\alpha_i = 0$或$\beta_i = 0$。总之，$\mu_{|C_i}$至多是一个狄拉克测度。

现在若 $\theta_i \notin \overline{C}_i \cup \overline{-C_i}$，首先假设 $\langle \theta_i, (x_{i+1}, 1) \rangle > 0$。这意味着 $\langle \theta_i, (x_i, 1) \rangle > 0$，因为 $\theta_i \notin C_i \cup -C_i$。由此可知， $\alpha_i \in C_i$ 或 $\beta_i \in C_i$，具体取决于 $i \geq i_0$ 还是 $i < i_0$。首先假设 $\beta_i \in C_i$ $(i \geq i_0)$，并定义

$$t = \sup \left\{ t' \in [0,1] \mid t'\alpha_i - \beta_i \in \overline{-C_i} \right\}.$$

By 连续性， $t\alpha_i - \beta_i \in \overline{-C_i}$。此外， $0 < t < 1$，因为 $\beta_i \in C_i$ 且 $\theta_i \notin \overline{-C_i}$ 现在我们定义 e

$$\tilde{\mu} = \mu - \mu_{|C_i} + (1-t)\|\alpha_i\| \delta_{\frac{\alpha_i}{\|\alpha_i\|}} + \|t\alpha_i - \beta_i\| \delta_{\frac{t\alpha_i - \beta_i}{\|t\alpha_i - \beta_i\|}}.$$

函数 $f_{\tilde{\mu},\, a,\, b}$ 仍然插值数据。与情形 $\theta_i \in C_i \cup -C_i$ 类似，对 $\|\mu\|_{\mathrm{TV}}$ 的极小化意味着 $\mu_{|C_i}$ 至多是一个狄拉克测度。

若改为 $\alpha_i \in C_i$，则通过定义可得类似的论证

$$t = \sup \left\{ t' \in [0,1] \mid \alpha_i - t'\beta_i \in \overline{C}_i \right\}.$$

若 $\langle \theta_i, (x_{i+1}, 1) \rangle < 0$，对称的论证同样成立。无论如何，$\mu_{|C_i}$ 至多是一个狄拉克测度。这对任意的 $i = 1, \ldots, n-1$ 均成立，最终可得引理 1。

D.2 引理 2 的证明

在证明引理 2 之前，我们先说明引理1给出的参数化与由点 x_i 处的斜率序列给出的参数化之间存在一一映射。我们定义

sets

$$\mathcal{S} = \Big\{ (s_1, \dots, s_n) \in \mathbb{R}^n \mid \forall i = 1, \dots, i_0 - 2, s_i \in S_i(s_{i+1}), \\ (s_{i_0-1}, s_{i_0}) \in \Lambda \quad \text{and} \quad \forall i = i_0, \dots, n-1, s_{i+1} \in S_i(s_i) \Big\} \tag{23}$$

and

$$\mathcal{I} = \Big\{ (a, b, (a_i, \tau_i)_{i=1,\dots,n-1}) \mid \forall j = 1, \dots, n,\ a x_j + b + \sum_{i=1}^{n-1} a_i (x_j - \tau_i)_+ = y_j, \quad \tau_{i_0-1} \in (x_{i_0-1}, x_{i_0}), \\ \tau_i = \frac{x_i + x_{i+1}}{2} \text{ if } a_i = 0, \quad \forall i \in \{1, \dots, i_0 - 2\}, \tau_i \in (x_i, x_{i+1}] \\ \text{and} \quad \forall i \in \{i_0, \dots, n-1\}, \tau_i \in [x_i, x_{i+1}) \Big\}.$$

定义 $\mathcal{I}$ 中的条件 $\tau_i = \frac{x_i + x_{i+1}}{2}$ if $a_i = 0$ 只是为了避免冗余，因为 τ_i 的任意取值都会得到相同的插值函数。下面的引理 6 给出了这两个集合之间的一一映射。

引理 6. 函数

$$\psi : \begin{array}{l} \mathcal{I} \to \mathcal{S} \\ (a_0, b_0, (a_i, \tau_i)_{i=1,\dots,n-1}) \mapsto (\sum_{j=0}^{i-1} a_j)_{i=1,\dots,n-1} \end{array}$$

是一个一一映射。其逆映射为

$$\psi^{-1} : \begin{array}{l} \mathcal{S} \to \mathcal{I} \\ (s_i)_{i \in [n]} \mapsto (a_0, b_0, (a_i, \tau_i)_{i \in [n-1]}) \end{array}$$

其中

$$a_0 = s_1; \qquad b_0 = y_1 - s_1 x_1; \qquad a_i = s_{i+1} - s_i \quad \textit{for any } i \in [n-1];$$
$$\tau_i = \begin{cases} \frac{s_{i+1} - \delta_i}{s_{i+1} - s_i} x_{i+1} + \frac{\delta_i - s_i}{s_{i+1} - s_i} x_i & \textit{if } s_{i+1} \neq s_i \\ \frac{x_i + x_{i+1}}{2} & \textit{otherwise} \end{cases}.$$

证明。 对于 $(a_0, b_0, (a_i, \tau_i)_{i=1,\dots,n-1}) \in \mathcal{I}$，令 f 为相关插值函数：

$$f(x) = a_0 x + b_0 + \sum_{i=1}^{n-1} a_i (x - \tau_i)_+$$

并令 $(s_i)_{i \in [n]} = \psi(a_0, b_0, (a_i, \tau_i)_i)$。根据 ψ 的定义，不难验证 s_i 对应于 f 在 $x_i \geq 0$ 处的左导数（分别为右导数），以及在 $x_i < 0$ 处的右导数（分别为左导数）。实际上，对于任意 $i \in [n-1]$，参数 $(a_0, b_0, (a_i, \tau_i)_i)$ 和 $(s_i)_i$ 满足以下两个不等式：

$$s_i + a_i = s_{i+1},$$
$$y_i + s_i (x_{i+1} - x_i) + a_i (x_{i+1} - \tau_i) = y_{i+1}.$$

第一个等式来自 f 在 x_i 处的（左或右）导数，第二个等式则源于 f 对数据的插值。这两个等式表明，在 $\mathcal{I}$ 中具有 ReLU 参数的插值函数（即 f）可以等价地由其在每个 x_i 处的（左或右）导数来描述。通过直接计算即可证明 ψ 和 ψ^{-1} 定义良好，且确实满足 $\psi \circ \psi^{-1} = I_{\mathcal{S}}$ 和 $\psi^{-1} \circ \psi = I_{\mathcal{I}}$。 □

利用从 $\mathcal{I}$ 到 $\mathcal{S}$ 的这一双射，我们现在可以证明引理 2。注意，在证明的其余部分中 $\alpha = \min_{\forall i \in [n]\ f,\ (x_i) = y_i} \int_{\mathbb{R}} \sqrt{1 + x^2}\, \mathrm{d}|f''(x)|$。根据引理 1，我们得到第一个等价关系：

$$\alpha = \min_{(a_0, b_0, (a_i, \tau_i)_{i=1,\dots,n-1}) \in \mathcal{I}} \sum_{i=1}^{n-1} |a_i| \sqrt{1 + \tau_i^2}.$$

对于任意 $(a_0\ b_0\ a_i\ \tau_i)_{i=1\ \dots\ n-1} \in \mathcal{I}$，$(s_i)_i = \psi(a_0, b_0, (a_i, \tau_i)_i) \in \mathcal{S}$，根据引理 6 可以定义 $(a_0, b_0, (a_i, \tau_i)_i) = \psi^{-1}((s_i)_i)$。于是有 $(a_0, b_0, (a_i, \tau_i)_i) \in \mathcal{I}$。此外，根据定义

对于 ψ^{-1}，我们可以很容易地验证

$$\begin{aligned}|a_i|\sqrt{1+\tau_i^2} &= \sqrt{a_i^2+(a_i\tau_i)^2}\\ &= \sqrt{(s_{i+1}-s_i)^2+((s_{i+1}-\delta_i)x_{i+1}+(s_i-\delta_i)x_i)^2}\\ &= g_{i+1}(s_{i+1},s_i).\end{aligned} \tag{24}$$

由于 ψ 是一一映射，对于任意函数 h，我们有等价关系 $\min_{u\in\psi^{-1}(\mathcal{S})}h(u) = \min_{s\in\mathcal{S}} h(\psi^{-1}(s))$。特别地，由公式（24）可得：

$$\min_{(a_0,b_0,(a_i,\tau_i)_{i=1,\dots,n-1})\in\mathcal{I}} \sum_{i=1}^{n-1}|a_i|\sqrt{1+\tau_i^2} = \min_{(s_i)_i\in\mathcal{S}}\sum_{i=1}^{n-1} g_{i+1}(s_{i+1},s_i). \tag{25}$$

据此，对任意 $i\geq i_0$，定义

$$d_i(s_i) = \min_{\substack{(\tilde{s})_j\in\mathcal{S}\\ \text{s.t. } \tilde{s}_i=s_i}} \sum_{j=i}^{n-1} g_{j+1}(\tilde{s}_{j+1},\tilde{s}_j);$$

对于任意 $i<i_0$

$$d_i(s_i) = \min_{\substack{(\tilde{s})_j\in\mathcal{S}\\ \text{s.t. } \tilde{s}_i=s_i}} \sum_{j=1}^{i-1} g_{j+1}(\tilde{s}_{j+1},\tilde{s}_j).$$

显然，由公式（25）和 $\mathcal{S}$ 的定义可得

$$\alpha = \min_{(s_{i_0-1},s_{i_0})\in\Lambda} g_{i_0}(s_{i_0},s_{i_0-1}) + d_{i_0-1}(s_{i_0-1}) + d_{i_0}(s_{i_0}). \tag{26}$$

接下来只需通过归纳法证明，对于任意 i，都有 $c_i = d_i$。当 $i=n$ 时显然成立。现在考虑 $i\in\{i_0,\dots,n-1\}$。根据 d_i 的定义可得

$$\begin{aligned}d_i(s_i) &= \min_{\substack{(\tilde{s})_j\in\mathcal{S}\\ \text{s.t. } \tilde{s}_i=s_i}} \sum_{j=i}^{n-1} g_{j+1}(\tilde{s}_{j+1},\tilde{s}_j)\\ &= \min_{s_{i+1}\in S_i(s_i)} \min_{\substack{(\tilde{s})_j\in\mathcal{S}\\ \text{s.t. } \tilde{s}_i=s_i\\ \tilde{s}_{i+1}=s_{i+1}}} g_{i+1}(s_{i+1},s_i) + \sum_{j=i+1}^{n-1} g_{j+1}(\tilde{s}_{j+1},\tilde{s}_j)\\ &= \min_{s_{i+1}\in S_i(s_i)} g_{i+1}(s_{i+1},s_i) + \min_{\substack{(\tilde{s})_j\in\mathcal{S}\\ \text{s.t. } \tilde{s}_i=s_i\\ \tilde{s}_{i+1}=s_{i+1}}} \sum_{j=i+1}^{n-1} g_{j+1}(\tilde{s}_{j+1},\tilde{s}_j).\end{aligned} \tag{27}$$

现在注意到，对于任意 $s_{i+1}\in S_i(s_i)$，集合 $\{(\tilde{s}_j)_{j\geq i+1} \mid (\tilde{s})_{j\in[n-1]}\in\mathcal{S} \text{ 满足 } \tilde{s}_i=s_i \text{ 且 } \tilde{s}_{i+1}=s_{i+1}\} = \{(\tilde{s}_j)_{j\geq i+1} \mid (\tilde{s})_{j\in[n-1]}\in\mathcal{S} \text{ 满足 } \tilde{s}_{i+1}=s_{i+1}\}$由于公式（27）中的最后一项仅取决于 $(\tilde{s}_j)_{j\geq i+1}$，这意味着

$$\begin{aligned}d_i(s_i) &= \min_{s_{i+1}\in S_i(s_i)} g_{i+1}(s_{i+1},s_i) + \min_{\substack{(\tilde{s})_j\in\mathcal{S}\\ \text{s.t. } \tilde{s}_{i+1}=s_{i+1}}} \sum_{j=i+1}^{n-1} g_{j+1}(\tilde{s}_{j+1},\tilde{s}_j)\\ &= \min_{s_{i+1}\in S_i(s_i)} g_{i+1}(s_{i+1},s_i) + d_{i+1}(s_{i+1}).\end{aligned}$$

根据归纳法，这自然可由 c_i 的定义得出：对于任意 $i\geq i_0$，均有 $c_i = d_i$。对任意 $i<i_0$，对称论证同样成立，最终得到对于任意 $i\in[n]$，均有 $c_i = d_i$。由公式（26）即可得到引理 2。

E 第 5 节的证明

本节的证明是在 $x_1 < 0$ 且 $x_n \geq 0$的情况下给出的。当所有 x 均为正数，即 $x_1 \geq 0$时，引理 2 的相应版本将给出 $i_0 = 1$的等价关系[11]

$$\min_{\substack{f \\ \forall i\in[n], f(x_i)=y_i}} \int_{\mathbb{R}} \sqrt{1+x^2}\mathrm{d}|f''(x)| = \min_{s_{i_0}\in\mathbb{R}} c_{i_0}(s_{i_0}).$$

附录 E 中的证明可以很容易地调整到这种情况（ $x_n < 0$ 时也类似）。本节末尾的附录 E.5 更具体地说明了如何将这些证明调整到这种情况。

E.1 定理 2 的证明

在证明定理 2 之前，下面的引理 7 给出了由公式 (8) 中定义的函数 c_i 所满足的重要性质。

Le引理 7. 对于每个 $i \in \{i_0, \dots, n-1\}$，函数 c_i 是凸函数，在 $\mathbb{R}$ 上是 $\sqrt{1+x_i^2}$-利普希茨的，且 d *min*在 $s_i = \delta_i$ 处取最小值。

此外，在这两个区间 $(-\infty, \delta_i]$ 和 $[\delta_i, +\infty)$ 上：

1. 要么对所考虑区间内所有 s_i，$c_i(s_i) = \sqrt{1+x_i^2}|s_i - \delta_i| + c_{i+1}(\delta_i)$；要么 c_i 在所考虑区间上严格凸；

2. 对所考虑区间内所有 s_i, s_i'，$|c_i(s_i) - c_i(s_i')| \geq \frac{1+x_i x_{i+1}}{\sqrt{1+x_{i+1}^2}}|s_i - s_i'|$。

同样，对于每个 $i \in \{1, \dots, i_0 - 2\}$，函数 c_{i+1} 是凸函数，在 $\mathbb{R}$ 上是 $\sqrt{1+x_{i+1}^2}$-利普希茨函数，并且在 $s_{i+1} = \delta_i$ 处取最小值。

此外，在这两个区间 $(-\infty, \delta_i]$ 和 $[\delta_i, +\infty)$ 上：

1. 要么对所考虑区间内所有 s_{i+1}，$c_{i+1}(s_{i+1}) = \sqrt{1+x_{i+1}^2}|s_{i+1} - \delta_i| + c_i(\delta_i)$；要么 c_{i+1} 在所考虑区间上严格凸；*2.* 对所考虑区间内所有 s_{i+1}、s_{i+1}'，$|c_{i+1}(s_{i+1}) - c_{i+1}(s_{i+1}')| \geq \frac{1+x_i x_{i+1}}{\sqrt{1+x_i^2}}|s_{i+1} - s_{i+1}'|$。

证明。对于任意 $i \in \{1, \dots, i_0 - 2\}$，我们通过（反向）归纳法证明该结论。由于 $c_n = 0$，直接计算可得[12]

$$c_{n-1}(s_{n-1}) = \sqrt{1+x_{n-1}^2}|s_{n-1} - \delta_{n-1}|,$$

由此得到 $i = n-1$ 所需的性质。

现在考虑满足以下条件的 $i \in \{i_0, \dots, n-2\}$：$c_{i+1}$ 满足引理 7 第一部分中的所有性质。我们先证明 c_i 的利普希茨性质。首先令 $s_i, s_i' < \delta_i$。根据归纳假设，函数$s_{i+1} \mapsto g_{i+1}(s_{i+1}, s_i) + c_{i+1}(s_{i+1})$ 在 $[\delta_i, +\infty)$ 上达到最小值。考虑满足以下条件的 $s_{i+1} \geq \delta_i$

$$c_i(s_i) = g_{i+1}(s_{i+1}, s_i) + c_{i+1}(s_{i+1}).$$

Als因此，通过极小化，$c_i(s_i') \leq g_{i+1}(s_{i+1}, s_i') + c_{i+1}(s_{i+1})$。对于向量 $u = (x_{i+1}, 1)$ 以及 d $v = (x_i, 1)$，此时成立：

$$\begin{aligned} c_i(s_i') - c_i(s_i) &\leq g_{i+1}(s_{i+1}, s_i') - g_{i+1}(s_{i+1}, s_i) \\ &= \|(s_{i+1} - \delta_i)u - (s_i' - \delta_i)v\| - \|(s_{i+1} - \delta_i)u - (s_i - \delta_i)v\| \\ &\leq \|(s_i - s_i')v\| = \sqrt{1+x_i^2}|s_i - s_i'|. \end{aligned}$$

第一个等式来自将 g_{i+1} 定义为范数，第二个不等式来自三角不等式。根据对称性，我们证明了 $|c_i(s_i') - c_i(s_i)| \leq \sqrt{1+x_i^2}|s_i - s_i'|$ 当 $s_i, s_i' < \delta_i$。

[11]请注意，在这种情况下 $c_1 \not\equiv 0$。相反，c_1 是通过公式 (8) 中给出的递推关系定义的。[12]这一计算利用了 x_{n-1} 和 x_n 均为正数这一事实，由此可知，c_{n-1} 定义中的最小 s_n 为 δ_{n-1}

注意，如果 $s_i = \delta_i$，那么 $s_{i+1} = \delta_i$，并且我们可以类似地证明 $c_i(s_i') - c_i(\delta_i) \le \sqrt{1+x_i^2}|\delta_i - s_i'|$。此外，

$$\begin{aligned} c_i(s_i') - c_i(\delta_i) &= \min_{s_{i+1}' \ge \delta_i} \|(s_{i+1}' - \delta_i)u - (s_i' - \delta_i)v\| + c_{i+1}(s_{i+1}') - c_{i+1}(\delta_i) \\ &\ge \min_{s_{i+1}' \ge \delta_i} \|(s_{i+1}' - \delta_i)u - (s_i' - \delta_i)v\| - \|(s_{i+1}' - \delta_i)u\| \\ &\ge 0. \end{aligned}$$

第一个不等式来自 c_{i+1} 的利普希茨性质。第二个不等式来自这样一个事实：$(s_{i+1}' - \delta_i)u$ 和 $(s_i' - \delta_i)v$ 负相关，因为 x_i 和 x_{i+1} 都为正。因此，c_i 在 $(-\infty, \delta_i]$ 上是 $\sqrt{1+x_i^2}$-利普希茨的。对称地，它在$[\delta_i, +\infty)$ 上也是 $\sqrt{1+x_i^2}$-利普希茨的，这最终说明它在 $\mathbb{R}$ 上是 $\sqrt{1+x_i^2}$-利普希茨的。此外，最后的计算还表明，c_i 在 $s_i = \delta_i$ 处取最小值。

现在我们证明 c_i 在 $(-\infty, \delta_i]$ 上满足第一点。由连续性可知，我们只需证明它在 $(-\infty, \delta_i)$ 上成立。令 $s_i \in (-\infty, \delta_i)$，根据定义可得

$$c_i(s_i) = \min_{s_{i+1} \ge \delta_i} g_{i+1}(s_{i+1}, s_i) + c_{i+1}(s_{i+1}).$$

若 $\delta_{i+1} \le \delta_i$，注意到两个函数 $g_{i+1}(\cdot, s_i)$ 和 c_{i+1} 在 $[\delta_i, +\infty)$[13] 上递增。因此，最小值在 $s_{i+1} = \delta_i$ 处取得，且

$$c_i(s_i) = \sqrt{1+x_i^2}|s_i - \delta_i| + c_{i+1}(\delta_i).$$

若 $\delta_{i+1} > \delta_i$，两个函数 $g_{i+1}(\cdot, s_i)$ 和 c_{i+1} 在 $[\delta_{i+1}, +\infty)$ 上递增。因此，我们可以将其改写为

$$c_i(s_i) = \min_{s_{i+1} \in [\delta_i, \delta_{i+1}]} g_{i+1}(s_{i+1}, s_i) + c_{i+1}(s_{i+1}). \tag{28}$$

首先假设在 $[\delta_i, \delta_{i+1}]$ 上，$c_{i+1}(s_{i+1}) = \sqrt{1+x_{i+1}^2}|s_{i+1} - \delta_{i+1}| + c_{i+2}(\delta_{i+1})$。根据三角不等式，我们实际上有

$$g_{i+1}(s_{i+1}, s_i) \ge g_{i+1}(\delta_{i+1}, s_i) - \sqrt{1+x_{i+1}^2}|\delta_{i+1} - s_{i+1}|.$$

这使得对于 $s_{i+1} \in [\delta_i, \delta_{i+1}]$，有

$$g_{i+1}(s_{i+1}, s_i) + c_{i+1}(s_{i+1}) \ge g_{i+1}(\delta_{i+1}, s_i) + c_{i+2}(\delta_{i+1}).$$

The 公式 (28) 中的最小值因此在 $s_{i+1} = \delta_{i+1}$处取得，最终对于任意 $s_i \le \delta$，得到 $_i$

$$c_i(s_i) = g_{i+1}(\delta_{i+1}, s_i) + c_{i+2}(\delta_{i+1}).$$

由于 $\delta_{i+1} > \delta_i$，容易验证 g_{i+1} δ_{i+1}，$\cdot$ 在 $(-\infty)$，$\delta_i)$ 上严格凸，c_i 也如此。

现在假设最后一种情况，即 c_{i+1} 在 $[\delta_i, \delta_{i+1}]$上严格凸。反证，假设在 $(-\infty, \delta_i]$ 上的第一个点不成立。注意，在下文中 $h(s_{i+1}, s_i) = g_{i+1}(s_{i+1}, s_i) + c_{i+1}(s_{i+1})$。对于 $s_i, s_i' < \delta_i$，由 h的连续性，取 $s_{i+1}, s_{i+1}' \in [\delta_i, \delta_{i+1}]$ 使得

$$c_i(s_i) = h(s_{i+1}, s_i) \quad \text{and} \quad c_i(s_i') = h(s_{i+1}', s_i').$$

对于任意 $t \in (0,1)$，由 h的凸性可得：

$$\begin{aligned} c_i(ts_i + (1-t)s_i') &\le h(t(s_{i+1}, s_i) + (1-t)(s_{i+1}, s_i)) \\ &\le th(s_{i+1}, s_i) + (1-t)h(s_{i+1}', s_i') \\ &= tc_i(s_i) + (1-t)c_i(s_i'). \end{aligned}$$

因此，c_i 在 $(-\infty, \delta_i]$上是凸的。此外，等号成立的情况对应于 g_{i+1} 和 c_{i+1}同时取等号的情况：

$$\begin{aligned} g_{i+1}(t(s_{i+1}, s_i) + (1-t)(s_{i+1}, s_i)) &= tg_{i+1}(s_{i+1}, s_i) + (1-t)g_{i+1}(s_{i+1}', s_i') \\ c_{i+1}(ts_{i+1} + (1-t)s_{i+1}') &= tc_{i+1}(s_{i+1}) + (1-t)c_{i+1}(s_{i+1}'). \end{aligned}$$

前者导致向量（$s_{i+1} - \delta_i$，$s_i - \delta_i$）和（$s_{i+1}' - \delta_i$，$s_i' - \delta_i$）共线；后者则由 c_{i+1}的严格凸性得出$s_{i+1} = s_{i+1}'$。因此有两种可能的情况

$$\begin{cases} \text{either } s_{i+1} = \delta_i = s_{i+1}' \\ \text{or } s_i = s_i'. \end{cases}$$

[13]这里我们再次利用了 x_i 和 x_{i+1} 为正 数。

于是，前一种情况意味着 $c_i(s_i) = \sqrt{1+x_i^2}|s_i - \delta_i| + c_{i+1}(\delta_i)$。由于 $c_i(\delta_i) = c_{i+1}(\delta_i)$，$c_i$在（$-\infty$，$\delta_i$]）上是 $\sqrt{1+x_i^2}$-利普希茨且为凸函数，因此对任意 $s \in (-\infty, \delta_i]$，都有 $c_i(s) = \sqrt{1+x_i^2}|s-\delta_i| + c_{i+1}(\delta_i)$。这与第一个条件在（$-\infty$，$\delta_i$]）上不成立的假设矛盾。必然有 $s_i = s_i'$。因此，c_i 在（$-\infty$，δ_i]）上严格凸，这又导致另一矛盾：第一个条件在（$-\infty$，δ_i]）上确实成立。

最后，我们刚刚证明，无论哪种情况，c_i 要么是严格凸的，要么在 $(-\infty, \delta_i]$ 上等于 $s_i \mapsto \sqrt{1+x_i^2}|s_i - \delta_i|$。对称的论证可得，在 $[\delta_i, +\infty)$ 上也同样成立。因此，c_i 在 δ_i, $\sqrt{1+x_i^2}$ 处取到最小值，且为利普希茨函数，并满足两个区间 $(-\infty, \delta_i]$ 和 $[\delta_i, +\infty)$ 上的第一点。这直接意味着 c_i 在 $\mathbb{R}$ 上是凸函数。

现在还需证明两个区间上的第二点。我们来证明 $(-\infty, \delta_i]$ 上的情形：由连续性可知，只需证明 $(-\infty, \delta_i)$ 上的情形即可。考虑 $s_i < s_i' < \delta_i$ 和 $s_{i+1} \in [\delta_i, +\infty)$，使得

$$c_i(s_i) = g_{i+1}(s_{i+1}, s_i) + c_{i+1}(s_{i+1}).$$

根据 c_i 的定义，

$$c_i(s_i) - c_i(s_i') \geq g_{i+1}(s_{i+1}, s_i) - g_{i+1}(s_{i+1}, s_i').$$

经过直接计算可得，函数

$$h_2 : \begin{array}{c} (-\infty, \delta_i] \to \mathbb{R}_+ \\ s \mapsto g_{i+1}(s_{i+1}, s) \end{array}$$

是凸函数，且 $h_2'(\delta_i) = -\frac{1+x_i x_{i+1}}{\sqrt{1+x_{i+1}^2}}$。因此，$h_2' \leq -\frac{1+x_i x_{i+1}}{\sqrt{1+x_{i+1}^2}}$，这最终意味着

$$\begin{aligned} c_i(s_i) - c_i(s_i') &\geq h(s_i) - h(s_i') \\ &\geq \frac{1+x_i x_{i+1}}{\sqrt{1+x_{i+1}^2}}(s_i' - s_i). \end{aligned}$$

因此，第二点在 $(-\infty, \delta_i)$ 和 $(-\infty, \delta_i]$ 上由连续性得到验证。对称论证可得，在 $[\delta_i, +\infty)$ 上也有相同性质。

由归纳法可得引理 7 的第一部分。对称论证可得引理 7 的第二部分（针对 $i \leq i_0 - 2$）。 □

现在我们可以证明定理 2。仿照引理 2 的证明，方程（4）存在唯一极小化解，当且仅当以下问题存在唯一极小化解:

$$\min_{\mathbf{s} \in \mathcal{S}} \sum_{i=1}^{n-1} g_{i+1}(s_{i+1}, s_i). \tag{29}$$

根据引理 1，我们已经知道最小值能够取到。根据函数c_i的构造，方程（29）的任一最小值 ˜ 都满足

$$\tilde{s}_i \in \operatorname*{argmin}_{s_i \in S_i(\tilde{s}_{i+1})} g_{i+1}(\tilde{s}_{i+1}, s_i) + c_i(s_i) \quad \text{for any } i \in [i_0 - 2] \tag{30}$$

$$(\tilde{s}_{i_0-1}, \tilde{s}_{i_0}) \in \operatorname*{argmin}_{(s_{i_0-1}, s_{i_0}) \in \Lambda} g_{i_0}(s_{i_0}, s_{i_0-1}) + c_{i_0-1}(s_{i_0-1}) + c_{i_0}(s_{i_0}) \tag{31}$$

$$\tilde{s}_{i+1} \in \operatorname*{argmin}_{s_{i+1} \in S_i(\tilde{s}_i)} g_{i+1}(s_{i+1}, \tilde{s}_i) + c_{i+1}(s_{i+1}) \quad \text{for any } i \in \{i_0, \ldots, n-1\}$$

现在还需证明所有这些问题都存在唯一极小化解。首先假设公式 (31) 存在不同的极小化解 (s_{i_0-1}, s_{i_0}) 和 (s_{i_0-1}', s_{i_0}')。注意以下 h_{i_0-1}：$(s, s') \mapsto g_{i_0}(s, s') + c_{i_0-1}(s') + c_{i_0}(s)$。根据极小化原理以及三个函数 g_{i_0}，c_{i_0-1}，c_{i_0} 的凸性，对于任意 $t \in (0, 1)$:

$$h_{i_0-1}(t(s_{i_0-1}, s_{i_0}) + (1-t)(s_{i_0-1}', s_{i_0}')) \leq t h_{i_0-1}(s_{i_0-1}, s_{i_0}) + (1-t) h_{i_0-1}(s_{i_0-1}', s_{i_0}') \tag{32}$$

$$= h_{i_0-1}(s_{i_0-1}, s_{i_0}). \tag{33}$$

于是，连接 (s_{i_0-1}, s_{i_0}) 和 (s_{i_0-1}', s_{i_0}') 的整条线段都是极小化解。不失一般性，我们可以假设 s_{i_0} 和 s_{i_0-1}' 都位于 δ_{i_0-1} 的同一侧（例如小于δ_{i_0-1}），且 s_{i_0} 和 s_{i_0}' 都位于 δ_{i_0}的同一侧。

此外，公式（33）意味着在 g_{i_0} 上存在等式，从而导致向量（$s_{i_0-1} - \delta_{i_0-1}$，$s_{i_0} - \delta_{i_0-1}$）和（$s_{i_0-1}' - \delta_{i_0-1}$，$s_{i_0}' - \delta_{i_0-1}$）共线。特别地，$s_{i_0} \neq s_{i_0}'$ 两者均

$s_{i_0-1} \neq s'_{i_0-1}$。此外，我们在 c_{i_0-1} 和 c_{i_0} 上均有等式成立，根据引理 7 的第一点

$$\begin{aligned} |c_{i_0-1}(s_{i_0-1}) - c_{i_0-1}(s'_{i_0-1})| &= \sqrt{1+x_{i_0-1}^2}|s_{i_0-1} - s_{i_0-1}| \\ |c_{i_0}(s_{i_0}) - c_{i_0}(s'_{i_0})| &= \sqrt{1+x_{i_0}^2}|s_{i_0} - s_{i_0}|. \end{aligned} \tag{34}$$

对于 $u = (x_{i_0-1},\ 1)$ 和 $v = (x_{i_0},\ 1)$，根据 $(s_{i_0-1} - \delta_{i_0-1},\ \ s_{i_0} - \delta_{i_0-1})$ 与 $(s'_{i_0-1} - \delta_{i_0-1},\ s'_{i_0} - \delta_{i_0-1})$ 的（正）共线性：

$$\begin{aligned} &|g_{i_0}(s_{i_0},\ s_{i_0-1}) - g_{i_0}(s'_{i_0},\ s'_{i_0-1})| = \big|\|(s_{i_0} - \delta_{i_0-1})v - (s_{i_0-1} - \delta_{i_0-1})u\| - \|(s'_{i_0} - \delta_{i_0-1})v - (s'_{i_0-1} - \delta_{i_0-1})u\|\big| \\ &= \|(s_{i_0} - s'_{i_0})v - (s_{i_0-1} - s'_{i_0-1})u\|. \end{aligned}$$

由于 $s_{i_0} \neq s'_{i_0}$ 和 $s_{i_0-1} \neq s'_{i_0-1}$，根据三角不等式可得两个严格不等式

$$\begin{aligned} &\left|\sqrt{1+x_{i_0}^2}|s_{i_0} - s'_{i_0}| - \sqrt{1+x_{i_0-1}^2}|s_{i_0-1} - s'_{i_0-1}|\right| < |g_{i_0}(s_{i_0}, s_{i_0-1}) - g_{i_0}(s'_{i_0}, s'_{i_0-1})|, \\ &\sqrt{1+x_{i_0}^2}|s_{i_0} - s'_{i_0}| + \sqrt{1+x_{i_0-1}^2}|s_{i_0-1} - s'_{i_0-1}| > |g_{i_0}(s_{i_0}, s_{i_0-1}) - g_{i_0}(s'_{i_0}, s'_{i_0-1})|. \end{aligned}$$

将此结果与公式 (34)结合，可得

$$g_{i_0}(s_{i_0}, s_{i_0-1}) - g_{i_0}(s'_{i_0}, s'_{i_0-1}) \neq c_{i_0}(s_{i_0}) - c_{i_0}(s'_{i_0}) + c_{i_0-1}(s_{i_0-1}) - c_{i_0-1}(s'_{i_0-1}).$$

Thi这与 s_{i_0-1}, s_{i_0} 和 s_{i_0-1}, s_{i_0} 均为公式 (31)的极小化解这一事实相矛盾。因此 e, Equ式 (31)存在唯一极小化解。

此外，最小化问题

$$\min_{s_{i+1} \in S_i(\tilde{s}_i)} g_{i+1}(s_{i+1}, \tilde{s}_i) + c_{i+1}(s_{i+1})$$

对于任意 $i \in \{i_0, \ldots, n-1\}$，都存在唯一极小化解。事实上，要么 $\tilde{s}_i = \delta_i$，此时约束集是单元素集；要么函数 $s_{i+1} \mapsto g_{i+1}(s_{i+1}, \tilde{s}_i)$ 对于 $\tilde{s}_i \neq \delta_i$ 是严格凸的。对于公式 (28)的最小化问题，也有类似的论证。因此，定理 2 的证明得证。

E.2 引理 3 的证明

令 $i \in \{i_0, \ldots, n-1\}$。回顾一下，

$$s^*_{i+1} = \operatorname*{argmin}_{s_{i+1} \in S_i(s^*_i)} g_{i+1}(s_{i+1}, s^*_i) + c_{i+1}(s_{i+1}).$$

如果 $i = n-1$，显然当 $s^*_n = \delta_{n-1}$ 时目标函数达到最小值，因为 x_{n-1} 和 x_n 均为正数。否则，不妨假设 $s^*_i > \delta_i$。根据引理 7，目标函数在 $(\quad -\infty \quad \delta_i\ \delta_{i+1}\ ]$ $\quad S_i\ s^*_i\ = [\delta_i, +\infty$ $\quad s^*_{i+1} \in [\quad \delta_i\ \delta_{i+1}\quad \delta_i]$ $\subset$，$[\quad \delta_i\ \delta_{i+1}\ ,\quad \delta_i\ \delta_{i+1}\ ]$ $\quad s^*_i = \delta_i$ $\quad s^*_i < \delta_i$ min(,) 和 ( )) 上递减，由此可得 min(,) min(,) max(,)。此情形显然成立，类似的论证也适用于其他情况。

现在考虑 $i = i_0 - 1$。首先假设 $s^*_{i_0-1} > \delta_{i_0-1}$，则

$$s^*_{i_0} = \operatorname*{argmin}_{s_{i_0} < \delta_{i_0-1}} g_{i_0}(s_{i_0}, s^*_{i_0-1}) + c_{i_0}(s_{i_0}).$$

得益于引理 7 的最后一点

$$c_{i_0}(s_{i_0}) - c_{i_0}(s'_{i_0}) \geq \frac{1 + x_{i_0}x_{i_0+1}}{\sqrt{1+x_{i_0+1}^2}}(s_{i_0} - s'_{i_0}) \qquad \text{for any } s_{i_0} < s'_{i_0} \leq \delta_{i_0}. \tag{35}$$

注意，函数

$$h: \begin{array}{c} (-\infty, \delta_{i_0-1}] \to \mathbb{R}_+ \\ s \mapsto g_{i_0}(s, s^*_{i_0-1}) \end{array}$$

是凸的，且满足 $h'(\delta_{i_0-1}) = -\frac{1+x_{i_0-1}x_{i_0}}{\sqrt{1+x_{i_0-1}^2}}$。由于 $x_{i_0-1} < 0 \leq x_{i_0+1}$，可得

$$-\frac{1 + x_{i_0-1}x_{i_0}}{\sqrt{1+x_{i_0-1}^2}} \leq x_{i_0} \leq \frac{1 + x_{i_0}x_{i_0+1}}{\sqrt{1+x_{i_0+1}^2}}.$$

由式 (35) 可知，函数 $s_{i_0} \mapsto g_{i_0}(s_{i_0}, s^*_{i_0-1}) + c_{i_0}(s_{i_0})$ 在 $(-\infty, \min(\delta_{i_0-1}, \delta_{i_0})]$ 上递减。因此，如上所述，这意味着 $s^*_{i_0} \in [\min(\delta_{i_0-1}, \delta_{i_0}), \max(\delta_{i_0-1}, \delta_{i_0})]$。$s^*_{i_0-1} = \delta_{i_0-1}$ 的情形显然成立；若 $s^*_{i_0-1} < \delta_{i_0-1}$，类似的论证同样适用。

我们证明了，对于任意 $i \in \{i_0, \dots, n\}$，都有 $s^*_{i+1} \in [\min(\delta_i, \delta_{i+1}), \max(\delta_i, \delta_{i+1})]$。对于 $i \in [i_0 - 1]$，对称的论证同样适用。至此，引理 3 的证明完成。

E.3 引理 4 的证明

我们先证明，任意最稀疏插值器 f 至少具有右侧求和给出的数量的折点。为此，我们实际上证明：对于 $k \geq 1$，在任意区间 $(x_{n_k-1}, x_{n_{k+1}})$ 上，只要满足 $\delta_{n_k-1} \neq \delta_{n_k}$，$f$，就至少有 $\left\lceil \frac{n_{k+1}-n_k}{2} \right\rceil$ 个折点，其符号由 $\operatorname{sign}(\delta_{n_k} - \delta_{n_k-1})$ 给出。考虑任意满足 $\delta_{n_k-1} \neq \delta_{n_k}$ 的 $k \geq 1$。不失一般性，假设 $\delta_{n_k-1} < \delta_{n_k}$。根据公式（10）的定义：

$$\delta_j > \delta_{j-1} \quad \text{for any } j \in \{n_k, \dots, n_{k+1} - 1\}.$$

显然，对于任意 $n_k \leq j \leq n_{k+1} - 1$，$f$ 在每个形如[14] (x_{j-1}, x_{j+1}) 的区间上都必须至少计入一个正拐点。注意，我们可以构造 $\left\lceil \frac{n_{k+1}-n_k}{2} \right\rceil$ 个互不相交的此类区间。因此，f 在 $(x_{n_k-1}, x_{n_{k+1}})$ 上至少有 $\left\lceil \frac{n_{k+1}-n_k}{2} \right\rceil$ 个正拐点。

根据定义，形如 $(x_{n_k-1}, x_{n_{k+1}})$ 且满足 $\delta_{n_k-1} < \delta_{n_k}$ 的区间互不相交。因此，f 的正拐点总数至少为

$$\sum_{k \geq 1} \left\lceil \frac{n_{k+1} - n_k}{2} \right\rceil \mathbb{1}_{\delta_{n_k-1} < \delta_{n_k}}.$$

类似地，f 的负拐点总数至少为

$$\sum_{k \geq 1} \left\lceil \frac{n_{k+1} - n_k}{2} \right\rceil \mathbb{1}_{\delta_{n_k-1} > \delta_{n_k}},$$

由此得到引理 4 的第一部分

$$\min_{\substack{f \\ \forall i, f(x_i) = y_i}} \|f''\|_0 \geq \sum_{k \geq 1} \left\lceil \frac{n_{k+1} - n_k}{2} \right\rceil \mathbb{1}_{\delta_{n_k-1} \neq \delta_{n_k}}.$$

现在我们构造一个恰好具有所需数量折点的插值函数。注意，引理 4 中考虑的问题具有平移不变性（方程（4）则不具有这一性质）。因此，不失一般性，我们可以假设 $x_1 \geq 0$。这样可以简化以下斜率序列$\mathbf{s} \in S$的定义：

$$s_1 = \delta_1$$
$$\text{and for any } i \in \{2, \dots, n\}, \quad s_i = \begin{cases} \delta_{i-1} \text{ if } (s_{i-1} = \delta_{i-1} \text{ or } i = n_k \text{ for some } k \geq 1) \\ s_i = \delta_i \text{ otherwise.} \end{cases}$$

容易验证$\mathbf{s} \in S$。现在考虑由引理 6 的映射根据斜率序列得到的函数 f，以及满足 $\delta_{n_k-1} \neq \delta_{n_k}$ 的区间 $[x_{n_k-1}, x_{n_{k+1}})$。根据定义，对于任意满足 $n_k + 1 \leq n_k + 1 + 2p < n_{k+1}$ 的 p，都有 $s_{n_k+1+2p} = \delta_{n_k+1+2p}$。这意味着 f在区间 $[x_{n_k+1+2p}, x_{n_k+2+2p})$ 内没有折点。由此，经过简单计算可知，f 在 $[x_{n_k}, x_{n_{k+1}})$ 上至多有$\left\lceil \frac{n_{k+1}-n_k}{2} \right\rceil$ 个折点。此外，当 $i = n_k$ 时，$s_i = \delta_{i-1}$，因此 f 在 $\delta_{n_k-1} = \delta_{n_k}$ 时于区间 $[x_{n_k}, x_{n_{k+1}})$ 上没有折点。因此，f是一个至多具有

$$\sum_{k \geq 1} \left\lceil \frac{n_{k+1} - n_k}{2} \right\rceil \mathbb{1}_{\delta_{n_k-1} \neq \delta_{n_k}},$$

折点，这就完成了引理 4 的证明。

E.4 定理 3 的证明

设 f 为方程（4）的极小化解。定理 3 的证明分别表明，f 在每个 $(x_{n_k-1}, x_{n_{k+1}})$ 上恰有 $\left\lceil \frac{n_{k+1}-n_k}{2} \right\rceil \mathbb{1}_{\delta_{n_k-1} \neq \delta_{n_k}}$ 个折点。以下固定 $k \geq 0$。

[14]否则，导数在该区间上将弱单调递减，这与插值相矛盾。

首先假设 $\delta_{n_k-1}=\delta_{n_k}$。那么，根据引理 3 以及 n_k 和 n_{k+1} 的定义，可直接推出，对于任意 $i=n_k,\dots,n_{k+1}-1$，都有 $s_i^*=\delta_{n_k-1}$。这意味着相关插值器，即 f 在 $(x_{n_k-1},x_{n_{k+1}})$ 上没有折点。

现在假设 $\delta_{n_k-1}\neq\delta_{n_k}$。不失一般性，假设 $\delta_{n_k-1}<\delta_{n_k}$。根据公式（10）的定义：

$$\delta_j>\delta_{j-1}\quad \text{for any } j\in\{n_k,\dots,n_{k+1}-1\}.$$

此外，根据 n_k的定义，我们有

$$\begin{cases}\text{either } n_k=1\\ \text{or } \delta_{n_k-1}\le\delta_{n_k-2}\end{cases}\qquad\text{and}\qquad\begin{cases}\text{either } n_{k+1}=n\\ \text{or } \delta_{n_{k+1}}\le\delta_{n_{k+1}-1}\end{cases}$$

由于根据假设 1，$n_{k+1}\le n_k+3$，下面的引理 8 指出，在所有情况下，f 在（x_{n_k-1}，$x_{n_{k+1}}$）上恰有$\left\lceil\frac{n_{k+1}-n_k}{2}\right\rceil$个折点。

如果 $\delta_{n_k-1}>\delta_{n_k}$，则对称的论证同样成立。总之，$f$在每个（$x_{n_k-1}$，$x_{n_{k+1}}$）上恰有 $\left\lceil\frac{n_{k+1}-n_k}{2}\right\rceil\mathbb{1}_{\delta_{n_k-1}\neq\delta_{n_k}}$ 个折点。这意味着 f 至多有

$$\sum_{k\ge1}\left\lceil\frac{n_{k+1}-n_k}{2}\right\rceil\mathbb{1}_{\delta_{n_k-1}\neq\delta_{n_k}}$$

个折点。结合引理 4，即完成定理 3 的证明。

引理 8。对于任意 $k\ge0$，如果 $\delta_{n_k-1}<\delta_{n_k}$，那么公式（4）$f$ 的极小化解具有

1. 1 *kink on* $(x_{n_k-1},x_{n_{k+1}})$ *if* $n_{k+1}=n_k+1$*;*
2. 1 *kink on* $(x_{n_k-1},x_{n_{k+1}})$ *if* $n_{k+1}=n_k+2$*;*
3. 2 *kinks on* $(x_{n_k-1},x_{n_{k+1}})$ *if* $n_{k+1}=n_k+3$.

引理 8 以这种非紧凑的方式表述，是因为其证明分别展示了三种情形（论证类似）。

证明。1) 考虑 $n_{k+1}=n_k+1$。首先假设 $x_{n_k}\ge0$。引理 3 意味着 $s^*_{n_k+1}\in[\delta_{n_k+1},\delta_{n_k}]$以及 $s^*_{n_k}\in[\delta_{n_k-1},\delta_{n_k}]$。特别地，$s^*_{n_k}\le\delta_{n_k}$，这意味着 $s^*_{n_k+1}=\delta_{n_k}$。类似地，$s^*_{n_k-1}\ge\delta_{n_k-1}$，这意味着 $s^*_{n_k}=\delta_{n_k-1}$。利用引理 6 中的映射，$s^*_{n_k}$ 和 $s^*_{n_k+1}$ 这两个值都表明，相关函数 f 在（x_{n_k-1}，x_{n_k+1}）上恰好有一个折点，该折点位于 x_{n_k}。如果 $x_{n_k}<0$，则类似的论证同样成立。

2) 现在考虑 $n_{k+1}=n_k+2$。

首先假设 $x_{n_k+2}<0$。根据引理 3，我们可以类似于情形 1) 证明 $s^*_{n_k+1}=\delta_{n_k+1}$。

现在假设 $x_{n_k+2}\ge0$。类似于情形 1)，$s^*_{n_k+2}=\delta_{n_k+1}$。对于 $i=n_k$，斜率的极小化问题在 s^*_{i+1} 上变为：

$$s^*_{i+1}=\operatorname*{argmin}_{s\in\tilde S}g_{i+1}(s,s^*_i)+g_{i+2}(\delta_{i+1},s),$$

其中，当 $x_{i+1}\ge0$ 时为 $\tilde S=S_i(s^*_i)$，否则为 $\tilde S=\{\delta_{i+1}\}$。注意，$g_{i+1}(s,s^*_i)$ 关于其第一个自变量是 $\sqrt{1+x^2_{i+1}}$-利普希茨的，而 $g_{i+2}(\delta_{i+1},s)=\sqrt{1+x^2_{i+1}}|s-\delta_{i+1}|$。此外，$s^*_{n_k}\in[\delta_{n_k-1},\delta_{n_k}]$。因此，要么 $x_{n_k+1}\ge0$ 且 $s^*_{n_k}=\delta_{n_k}=s^*_{n_k+1}$；要么 $s^*_{n_k+1}=\delta_{n_k+1}$。对称地，当考察点 x_{n_k-1},x_{n_k} 时：

- either $s^*_{n_k}=\delta_{n_k-1}$;
- or ($x_{n_k}<0$ and $s^*_{n_k}=\delta_{n_k}=s^*_{n_k+1}$).

因此，最终有两种可能的情况：

- either ($s^*_{n_k} = \delta_{n_k-1}$ and $s^*_{n_k+1} = \delta_{n_k+1}$);
- or ($s^*_{n_k} = \delta_{n_k} = s^*_{n_k+1}$ and $x_{n_k} < 0 \le x_{n_k+1}$).

当 $x_{n_k} < 0 \le x_{n_k+1}$ 时，我们还有 $s^*_{n_k-1} = \delta_{n_k-1}$ 和 $s^*_{n_k+2} = \delta_{n_k+1}$。经过直接计算可得，对于函数 g_i，选择斜率$s^*_{n_k} = \delta_{n_k-1}$ 和 $s^*_{n_k+1} = \delta_{n_k+1}$时，代价更小。

因此，在任何情况下，$s^*_{n_k} = \delta_{n_k-1}$ 和 $s^*_{n_k+1} = \delta_{n_k+1}$。随后，引理 6 中的映射表明，$f$ 在（x_{n_k-1}，x_{n_k+2}）上恰有一个折点，位于（x_{n_k}，x_{n_k+1}）内。事实上，我们要么有 $a_{n_k-1} = 0$，要么有 $\tau_{n_k-1} = x_{n_k-1}$；类似地，要么有 $a_{n_k+1} = 0$，要么有 $\tau_{n_k+1} = x_{n_k+2}$。

3）现在考虑 $n_{k+1} = n_k + 3$。与情况 2）类似，我们同时有

$$\begin{cases}\text{either } s^*_{n_k+2} = \delta_{n_k+2}\\ \text{or } (s^*_{n_k+1} = \delta_{n_k+1} = s^*_{n_k+2} \text{ and } x^*_{n_k+2} \ge 0)\end{cases}$$

$$\text{and} \qquad \begin{cases}\text{either } s^*_{n_k} = \delta_{n_k-1}\\ \text{or } (s^*_{n_k} = \delta_{n_k} = s^*_{n_k+1} \text{ and } x_{n_k} < 0).\end{cases}$$

考虑所有可能的情况后，引理 6 中的映射表明，f 在（x_{n_k-1}，x_{n_k+3}）上恰有两个折点，它们位于 $[x_{n_k}, x_{n_k+2}]$ 中。 □

E.5 对 $x_1 \ge 0$ 情形的调整分析

本节说明如何将本节的分析调整到一个更简单的情形，即所有 x 均为正数。在该情形下，引理 7 在完全相同的条件下仍然成立（但其第二部分没有用处）。在此基础上，定理 2 的证明只需证明对于任意 $\tilde{\mathbf{s}} \in \mathcal{S}$，这些最小化问题具有唯一性：

$$\min_{s_{i_0} \in \mathbb{R}} c_{i_0}(s_{i_0})$$

$$\min_{s_{i+1} \in S_i(\tilde{s}_i)} g_{i+1}(s_{i+1}, \tilde{s}_i) + c_{i+1}(s_{i+1}) \quad \text{for any } i \in \{i_0, \dots, n-1\}.$$

第一个问题的唯一解是 δ_{i_0}，由引理 7 可知；而对于第二个问题，同附录 E.1 中相同的论证同样适用。

对于引理 3 的证明，附录 E.2 中完全相同的论证对任意 $i \ge i_0 + 1$ 都成立。对于 $i = i_0 = 1$，在该情况下显然有 $s^*_1 = \delta_1$，从而得到引理 3。

最后，当 $x_1 \ge 0$ 时，定理 3 的证明遵循相同的思路。

E.6 推论 1 的证明

推论 1 的证明。对于分类，自然可以定义如下划分：

$$\begin{aligned} &n_1 = 1 \text{ and for any } k \ge 0 \text{ such that } n_k < n+1,\\ &n_{k+1} = \min\{j \in \{n_k+1, \dots, n\} \mid y_{n_k} \ne y_j\} \cup \{n+1\}. \end{aligned} \tag{36}$$

该划分将数据分开，使得对于任意 k, y_i，$i \in \{n_k, n_{k+1}-1\}$ 的值都相同。记 K 为公式（36）中定义的 n_k 数量，即 $n_K = n+1$。由此，只需注意到，任何间隔分类器在 $k \in [K-2]$ 的 $[x_{n_k}, x_{n_{k+1}})$ 中至少有一个折点：

$$\min_{\substack{f \\ \forall i \in [n], y_i f(x_i) \ge 1}} \|f''\|_0 = K - 2.$$

与引理 1 的证明类似，我们可以先证明最小值的存在。[15] 现在考虑 f 的一个极小化解

$$\min_{\substack{f \\ \forall i\in[n], y_i f(x_i)\geq 1}} \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}}. \tag{37}$$

定义集合

$$S = \left\{ n_k \mid k \in \{2,\dots,K-1\} \right\} \cup \left\{ n_k - 1 \mid k \in \{2,\dots,K-1\} \right\}.$$

根据 f 的连续性，我们可以选择另一个训练集 $(\tilde{x}_i,\ \tilde{y}_i)$，使其满足：

$$\tilde{x}_i \in [x_{n_k-1}, x_{n_k}] \quad \text{for any } i \in \{n_k - 1, n_k\},$$
$$y_i = f(\tilde{x}_i) \quad \text{for any } i \in S.$$

那么，直接应用定理 2 可得，最小化问题

$$\min_{\substack{\tilde{f} \\ \forall i\in S, y_i = \tilde{f}(\tilde{x}_i)}} \left\| \sqrt{1+x^2} \tilde{f}'' \right\|_{\mathrm{TV}}, \tag{38}$$

存在唯一极小化解，记为 f_{reg}。但还需注意，根据引理 3，这个唯一极小化解也属于方程（37）的约束集，因此

$$\left\| \sqrt{1+x^2} f''_{\mathrm{reg}} \right\|_{\mathrm{TV}} \geq \left\| \sqrt{1+x^2} f'' \right\|_{\mathrm{TV}}.$$

然而，由于 f 属于方程（38）的约束集，实际上两者相等；又由于方程（38）的极小化解唯一，

$$f_{\mathrm{reg}} = f.$$

此外，容易验证，对于数据 $(\tilde{x}_i,\ y_i)_{i\in S}$，假设 1 成立，其中 $n_{k+1} = n_k + 2$。因此，定理 3 蕴含方程（38）的极小化解是集合 $(\tilde{x}_i,\ y_i)_{i\in S}$ 的最稀疏插值器之一，即它恰好有 $K-2$ 个折点。这进而意味着 $\|f''\|_0 = K-2$，因此

$$\operatorname*{argmin}_{\substack{f \\ \forall i\in[n], y_i f(x_i)\geq 1}} \overline{R}_1(f) \subset \operatorname*{argmin}_{\substack{f \\ \forall i\in[n], y_i f(x_i)\geq 1}} \|f''\|_0. \tag{39}$$

□

[15] 要证明序列 (a_j, b_j) 有界，这里稍微更棘手一些。数据要么是线性可分的，此时最小值为 0；要么不是线性可分的。若数据不是线性可分的，则 (a_j, b_j) 必然有界，因为 (μ_j, a_j, b_j) 在 (a_j, b_j) 任意大时都会表现为一个线性分类器。