arXiv:2601.01344v1 [stat.ME] 4 Jan 2026

约束路线对齐的自适应核回归：理论与迭代数据锐化

Shiyin Du[1], Yiting Chen[1], Wenzhi Yang[2,*]、Qiong Li[3,*]和Xiaoping Shi[1,*]

[1]不列颠哥伦比亚大学，计算机科学、数学、物理与统计系，基洛纳，V1V 1V7，加拿大[2]安徽大学大数据与统计学院，合肥，230601，中国[3]广东省跨学科研究与应用重点实验室；数据科学，北京师范大学-香港浸会大学联合国际学院，珠海，519087，中国[*]wzyang@ahu.edu.cn，qiongli@bnbu.edu.cn，xiaoping.shi@ubc.ca

摘要

测绘和交通工程中的路线线形设计通常涉及固定航路点约束，即路线必须精确通过特定坐标。现有文献主要依赖几何优化或控制理论样条框架，但在兼顾全局平滑度与精确点位贴合方面，系统性的统计建模方法仍然不足。本文提出一种用于解决固定航路点问题的自适应 Nadaraya-Watson（ANW）核回归估计量。通过引入航路点特定权重调节参数，ANW估计量将全局平滑与局部约束满足解耦，避免了朴素局部带宽收缩策略中常见的“锯齿状”伪影。为进一步提高估计精度，我们开发了一种迭代数据锐化算法，在保持核框架稳定性的同时系统地减小偏差。我们通过推导ANW估计量的渐近偏差和方差，并证明其在内部约束模型下的收敛性质，为该估计量建立了理论基础。一维和二维轨迹规划的数值案例研究表明，该方法能够有效平衡均方根误差（RMSE）与曲率平滑性。最后，我们通过将该框架应用于铁路和公路路线规划的实证研究，验证了其实际效用。总之，本研究为复杂约束路线线形设计问题提供了一种稳定、具有坚实理论基础且计算高效的解决方案。

关键词：自适应 Nadaraya–Watson、固定航路点约束、数据锐化、路线对齐、非参数回归、轨迹规划、渐近性质。

引言

在测绘与制图的路线线形规划中，约束通常是存在的。在实际的路线线形设计问题中，特定点的精确值通常是已知的，因此会产生要求路径精确通过这些给定点的约束，我们称之为固定航路点。此类问题最常见的例子之一，是对物理过程起点和终点的控制（Kano 等，2011）。固定航路点问题广泛存在于交通运输和测绘领域。在道路与公路设计中，Kang 等（2007）研究了“可行门”方法的应用：在道路建设期间经过指定的关键点（或区域）时，该方法可替代僵化的定点通过约束，使路线在满足几何要求的同时更具实用性。Maji（2017）采用路径规划器方法，自动生成穿过指定点并绕过限制区域的公路，在最大限度降低建设成本的同时，生成既短又平滑的路线。在铁路线路设计中，Wan 等（2024）将适合铁路建设的区域划分为“可行岛”，规定铁路必须穿过这些指定区域。随后在这些可行岛之间建立连接，最终确定可能的最短路线。在其他领域也存在一些固定航路点问题，例如海事与海洋学应用（Zhen 等，2023；Kytariolou 和 Themelis，2022）以及航空航天轨迹规划（Mellinger 和 Kumar，2011；Richter 等，2016）。

这些涉及航路点约束的路线规划研究，主要依赖计算优化和算法框架来满足“经过特定地点”的要求，而很少考虑统计建模方法。此外，现有研究从几何和优化的角度设计道路线形（Vázquez-Méndez，2023）。因此，亟需开发系统性的统计线形设计方法，以解决固定航路点问题。

在统计建模中，约束非参数回归旨在估计光滑函数，同时施加结构限制。在早期研究中，Hall 和 Huang（2001）研究了单调性约束下的核回归，而 Racine 等（2009）则通过二次规划将区间约束纳入等式约束，从而扩展了该方法。 Henderson 和 Parmeter（2009）进一步关注实际问题中约束的可计算性，将经济学中的结构约束转化为二次规划形式，从而实现了从理论到算法的过渡。随后，研究重点逐渐从一般形状约束转向固定航路点约束。 Xiong（2021）提出了重构方法，将固定航路点约束问题推广为非参数回归问题。在“重构”概念的基础上，Dontchev 等（2022）将固定航路点约束问题扩展为最优控制下的约束平滑问题。与基于核的方法的发展并行，基于样条的方法为在非参数回归中纳入等式和不等式约束提供了另一种实用途径（Saeidian 等，2021）。Sun 等（2000）在控制理论框架内构建了平滑样条，其中样条通过求解最优控制问题获得，该问题旨在最小化 L^2范数

控制输入，同时使系统输出接近预先指定的插值点。Kano 等（2003）确立了B样条与线性控制系统之间的紧密联系，并将插值样条和平滑样条表述为最优控制问题，通过最小化控制能量，使系统轨迹经过预先指定的数据点。Kano 等（2011）为带等式和不等式约束的样条建立了二次规划框架，并展示了该框架在轨迹规划中的应用，其中边界条件包含固定航路点约束。这些方法主要基于样条或控制模型开发，而非核回归。此外，大多数固定航路点约束所对应的是插值问题，而非真正对应现实测绘挑战的固定航路点问题。

现有研究仍缺乏针对满足固定航路点约束的非参数回归方法的系统且具体的解决方案。首先，在路线线形设计场景中，大多数研究侧重于纯优化，并未将固定点视为非参数回归问题。其次，在现有的核方法中，很少有方法能为固定航路点系统提供简单、易解释且易于实现的具体方案。此外，关于约束点收敛的理论讨论仍然有限。为此，我们提出一种自适应 Nadaraya–Watson 核回归估计量（ANW），通过航路点特定的固定权重调节参数（记为 λ_j）显式纳入航路点约束。该方法不仅保留了传统 Nadaraya–Watson 核回归一致的收敛性质，还能以可控且易于解释的方式将拟合曲线拉向固定航路点，因此更适用于线形设计任务，包括道路和铁路路线规划、交通轨迹建模以及其他空间路径设计问题。

本文其余部分安排如下。在建模部分，我们介绍内部约束模型，并构建朴素的带宽收缩估计量和所提出的 ANW 估计量，同时比较二者的局部行为。随后，我们详细说明 ANW 估计量的迭代数据锐化过程，并给出用于降低偏差的正式算法。在渐近性质部分，我们建立 ANW 估计量的理论框架并推导其渐近性质，所有正式的数学证明均置于附录。数值性质部分通过一维和二维轨迹规划的数值案例研究，评估该方法的性能与稳定性。实证应用部分则通过将该框架应用于铁路和公路路线规划，展示其实际效用。最后，本文总结研究发现，并讨论未来的研究方向。

建模

在本研究中，X 表示沿 s 路线的弧长 s。它代表实际线形设计中沿路线的位置。响应变量 Y 对应于沿路线变化的几何量，例如高程、横向偏移或坐标分量 $x(s)$，$y(s)$。令 (X_1, Y_1)，……，(X_n, Y_n) 为独立同分布的 $\mathbb{R}^{d+1}$ 值随机向量，且 Y 为实值。我们通常假设模型具有如下形式：$Y_i = m(X_i) + \varepsilon_i$，其中 $m(x)$ 是未知的光滑函数，且

ε_i 表示表示随机误差。考虑估计回归函数的问题 ，

$$m(x) = \mathbb{E}[Y|X = x],$$

使用 (X_1， Y_1)，……，(X_n， Y_n)，而不对其函数形式作出强假设。

核估计量（Nadaraya 1964年；Watson 1964年）提供了对m（x）的直观估计，它是 Y_i的局部加权平均，定义为

$$\hat{B}_{\mathrm{NW}}(x) = \hat{m}(x) = \frac{\sum_{i=1}^{n} w_i(x;h)Y_i}{\sum_{i=1}^{n} w_i(x;h)}$$

其中 $w_i(x;h) = K_h(x - X_i) = \frac{1}{h} K\left(\frac{x-X_i}{h}\right)$，$K$：$\mathbb{R}^d \to \mathbb{R}$是核函数，而$h = h(n) \in \mathbb{R}^+$ 是带宽。

固定航路点下的估计

我们考虑指标集 $I = \{1, 2, \ldots\ldots, n\}$ 上的内部约束模型。我们将该集合划分为两个子集：一个包含确定性固定航路点 $C \subset I$，其基数为 $|C| = m$；另一个是随机观测的补集 $D = I \setminus C$。对于指标 $i \in D$，观测值遵循标准的随机非参数回归模型：

$$Y_i = m(X_i) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i] = 0,$$

其中，$m(x) = \mathbb{E}[Y|X = x]$是目标回归函数。反之，对于指标$i \in C$（表示一组固定航路点），假设相应的观测值是确定性的，并且恰好位于回归曲线上，即 $Y_i = m(X_i)$。

朴素方法带宽收缩策略

纳入固定航路点的一种自然基线方法是调整局部平滑尺度。其直观思路是在固定航路点的邻域内收缩核带宽，从而迫使估计量更加贴近指定值。

令 $h > 0$ 为全局参考带宽，并令 $\gamma : (0, \infty) \to (0, 1]$ 为收缩因子。对于每个评估位置 x 和观测值 X_i，定义点依赖带宽

$$h_i(x) = \begin{cases} \gamma(h), & \text{if } |x - X_i| \le r_h \text{ for some } i \in C, \\ h, & \text{otherwise,} \end{cases}$$

其中，C 表示固定航路点的指标集，$r_h > 0$ 是指定的邻域半径。

由此得到的朴素核权重为

$$\tilde{w}_i(x;h) = \frac{1}{h_i(x)} K\left(\frac{x - X_i}{h_i(x)}\right),$$

从而得到该估计量

$$\hat{m}_{\mathrm{naive}}(x) = \frac{\sum_{i \in I} \tilde{w}_i(x;h)\, Y_i}{\sum_{i \in I} \tilde{w}_i(x;h)}. \tag{1}$$

尽管这种方法直观易懂，但它将航点遵循与局部平滑尺度紧密耦合。缩小$h_i(x)$会减小评估位置周围有效平均邻域的范围，这可能显著增大局部方差，并导致受约束区域附近出现锯齿伪影。此外，选择合适的收缩因子$\gamma(\cdot)$通常计算成本高昂，且缺乏可靠的理论指导。

采用固定航点调节参数的自适应纳达亚-沃森（ANW）

为克服这些局限，我们提出自适应纳达亚-沃森（ANW）估计量。ANW 不改变平滑尺度，而是保留全局带宽h，仅通过自适应重新加权纳入固定航路点的影响。定义如下：

$$\bar{w}_i(x;h) = \begin{cases} w_i(x;h), & i \in D, \\ \lambda_i w_i(x;h), & i \in C, \end{cases}$$

其中 $w_i(x;h) = \frac{1}{h} K\left(\frac{x-X_i}{h}\right)$，$\lambda_i \geq 1$是一个调优参数，用于控制第 i 个固定航路点的影响“强度”。原则上，航路点影响参数 $\{\lambda_i : i \in C\}$ 可以因固定航路点而异，从而施加不同程度的约束。然而，在本文中，我们重点关注一种实际相关的情形：所有航路点共用一个全局调优参数，并将其记为 $\lambda_i \equiv \lambda$（对所有 $i \in C$）。ANW 估计量是以下加权局部常数最小二乘问题的解：

$$\hat{m}_{\mathrm{ANW}}(x) = \arg\min_{B\in\mathbb{R}} \left[\sum_{i\in D} (Y_i - B)^2 w_i(x;h) + \sum_{i\in C} (Y_i - B)^2 \lambda w_i(x;h) \right], \tag{2}$$

由此得到闭式表达式：

$$\hat{m}_{\mathrm{ANW}}(x) = \frac{\sum_{i\in D} w_i(x;h) Y_i + \sum_{i\in C} \lambda w_i(x;h) Y_i}{\sum_{i\in D} w_i(x;h) + \sum_{i\in C} \lambda w_i(x;h)}. \tag{3}$$

ANW的主要优势在于将平滑度与约束遵循解耦。由于 $\lambda \to \infty$，估计量无需缩小局部窗口即可通过固定航路点，从而保持由h所控制的全局平滑性。这种分离使ANW框架非常适合用于迭代数据锐化过程，以进一步实现偏差降低（Chen 等，2025年；Braun 和 Shi，2025年；Shi，2026年）。

图1对比了两种固定航路点约束策略的概念。朴素策略（蓝线）在固定航路点处局部减小带宽，使曲线趋向该约束，但会产生“尖峰”伪影并降低平滑度。相比之下，ANW估计量（绿线）在整个定义域内保持一致的全局平滑尺度h。通过仅调整权重调优参数 λ，ANW估计量便能使回归曲线平滑地经过固定航路点，而不会引入局部不规则性，也不会使约束遵循依赖于平均邻域的缩小。

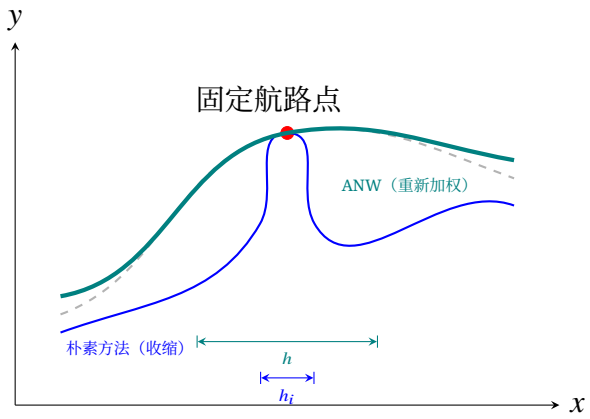


图1. 固定航路点约束的比较：朴素带宽收缩与 ANW 重新加权

在（h, λ）标准化尺度上的调优

我们使用交叉验证为每种方法选择最优带宽（Stone1974）。为使不同数据集之间的交叉验证误差和固定航路点偏差具有可比性，我们在拟合前将响应变量标准化，使其具有零均值和单位方差。具体而言，设

$$Y_i^\star = \frac{Y_i - \bar{Y}}{s_Y}, \qquad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \qquad s_Y^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2,$$

并令 $\hat{m}^\star_{\mathrm{ANW},\lambda,h}(x)$ 表示由标准化响应变量 $\{(X_i, Y_i^\star)\}_{i=1}^n$ 构造的 ANW 估计量（3）。

然后，我们通过最小化来选择 (λ, h)

$$\mathrm{Loss}(\lambda, h) = \mathrm{CV}^\star(\lambda, h) + \sum_{i \in C}\left(\hat{m}^\star_{\mathrm{ANW},\lambda,h}(X_i) - Y_i^\star\right)^2, \tag{4}$$

其中，$\mathrm{CV}^\star(\lambda, h)$ 表示在标准化尺度上计算的 K 折交叉验证误差。标准化公式隐含地赋予交叉验证误差和航路点偏差相同的权重。实际应用中，可以通过引入显式的权重参数来调整航点遵循与全局平滑性之间的相对重要性。此处我们将权重设为 1，以避免引入额外的调优参数，并专注于 (λ, h) 的表现。

为说明调优参数 λ的影响，我们在图 2 中展示了一个典型示例。随着 λ 增大，拟合曲线逐渐向固定航路点（红色圆圈）靠拢，对规定值的遵循也越来越严格。

ANW数据锐化（DS–ANW）

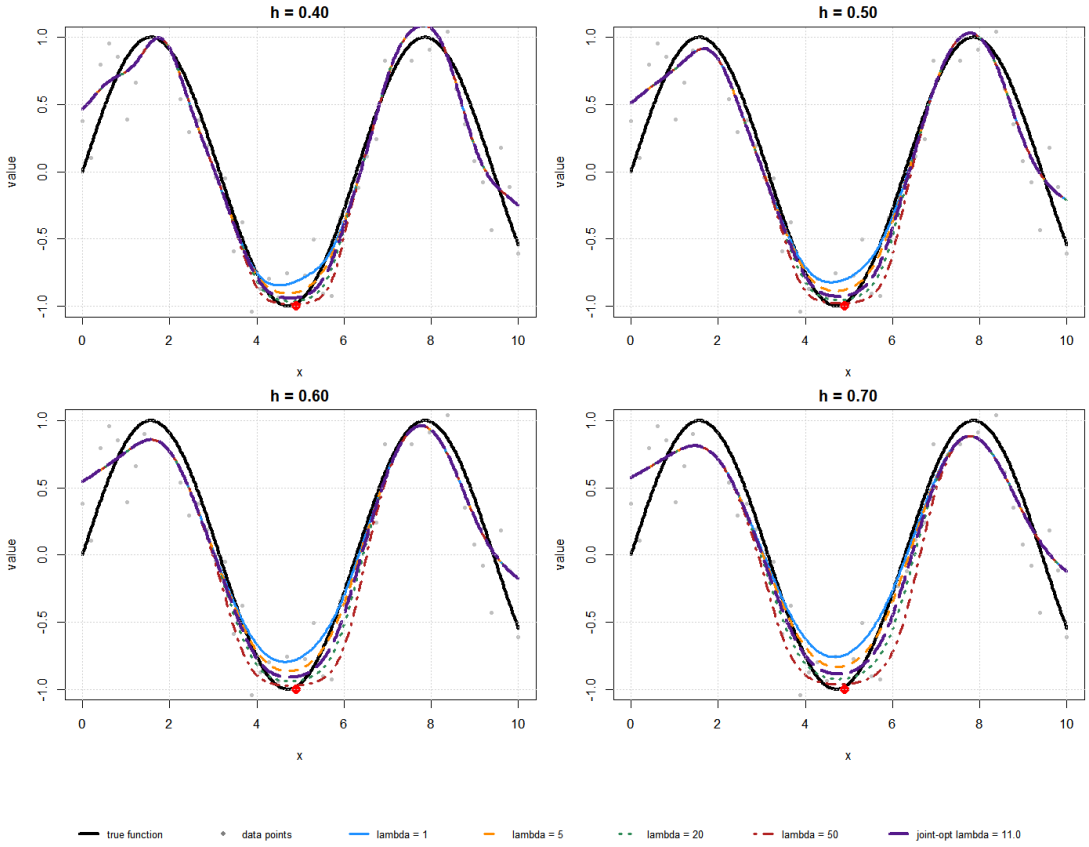


图 2. 单个航点约束下惩罚参数 λ 对拟合曲线的影响

动机：在不改变权重的情况下降低偏差

式 (3) 中的 ANW估计量是一种局部常数核回归（采用固定的航点调节参数），因此具有常见的平滑偏差。特别是，即使通过 λ控制航点遵循，内部点处的主导偏差项仍为 $O(h^2)$ 阶，与标准核回归相同。因此，我们采用一种偏差削减方法，不改变加权方案 $\bar{w}_i(x; h)$，从而保留固定航点机制。

数据锐化提供了这样一种策略：它不改变估计量形式，也不调整带宽，而是通过拟合残差迭代更新响应变量$\{Y_i\}_{i=1}^n$，然后再次应用相同的平滑算子。我们采用迭代数据锐化程序 IDS2，该程序旨在逐步降低偏差展开的各阶项，同时将方差膨胀控制在适度范围内（Chen 等，2025；Braun 和 Shi，2025；Shi，2026）。

ANW权重下的IDS2更新规则

令 $\hat{m}^{(0)}_{\mathrm{ANW}}(x)$ 表示由观测到的响应变量 $\{Y_i\}_{i=1}^n$ 构造的原始 ANW 估计量 (3)。定义锐化响应为 $Y_{(0)i} = Y_i$，并且对于$k = 0,1,\ldots, M-1$，更新

$$Y_i^{(k+1)} = Y_i + \left\{ Y_i^{(k)} - \hat{m}^{(k)}_{\mathrm{ANW}}(X_i) \right\}, \qquad i = 1,\ldots,n, \tag{5}$$

其中，$\hat{m}^{(k)}_{\mathrm{ANW}}(\cdot)$ 表示应用于锐化数据集 $\{(X_i, Y_{(k)i})\}_{i=1}^n$ 的 ANW 估计量；拟合值 $\hat{m}^{(k)}_{\mathrm{ANW}}(X_i)$ 是通过在 $x = X_i$ 处计算同一估计量得到的，整个过程中始终使用未改变的设计点 $\{X_i\}$。IDS2 的定义性特征在于，将第 k 次迭代的残差加回原始响应变量 Y_i，而不是加回到 $Y_i^{(k)}$，这对于降低迭代过程中的偏差至关重要。给定 $\{Y_i^{(k)}\}$，我们计算

$$\hat{m}^{(k)}_{\mathrm{ANW}}(x) = \frac{\sum_{i\in I} \bar{w}_i(x;h)\, Y_i^{(k)}}{\sum_{i\in I} \bar{w}_i(x;h)}, \tag{6}$$

使用相同的自适应权重

$$\bar{w}_i(x;h) = \begin{cases} w_i(x;h), & i \in D, \\ \lambda\, w_i(x;h), & i \in C, \end{cases} \qquad \lambda \geq 1.$$

经过 M 次迭代后，最终估计量

$$\hat{m}_{\mathrm{DS-ANW}}(x) := \hat{m}^{(M)}_{\mathrm{ANW}}(x)$$

称为数据锐化 ANW（DS–ANW）。

算法

公式 (5) 有一个简单的解释。在迭代次数 k 时，拟合值 $\hat{m}^{(k)}_{\mathrm{ANW}}(X_i)$ 表示 ANW 算子在 X_i 处施加的平滑程度。残差项 $Y_{(k)i} - \hat{m}^{(k)}_{\mathrm{ANW}}(X_i)$ 估计被平滑掉的部分。随后，IDS2 将该残差加回原始响应变量 Y_i，得到锐化后的响应，在保持相同加权机制的同时，部分恢复局部曲率。

在我们的设定中，锐化的一项关键优势是它不会改变固定航路点权重 λ，也不会改变 $\bar{w}_i(x;h)$ 内的带宽 h；因此，固定航点影响机制得以保留。此外，IDS2 的设计旨在通过迭代实现系统性的偏差降低，而早期的 IDS1 迭代（将残差加到当前的锐化响应上）则可能无法在第一步之后进一步降低偏差。

我们还要指出一个常见的注意事项：对于局部常数回归，边界区域会随着锐化步骤的增加而扩大，边界附近的偏差改善可能有限。在我们的应用中，通过将推断和比较集中在内部区域，或在端点附近进行足够密集的采样，可以缓解这一问题。

算法 1通过 IDS2 实现的数据锐化 ANW（DS–ANW）

1: 输入：数据 $\{(X_i, Y_i)\}_{i=1}^{n}$，固定航点集 C(其中 $D = I \setminus C$)，调优参数λ，带宽 h，迭代次数 M。 2: 初始化 $Y_i^{(0)} \leftarrow Y_i$，其中 $i = 1, \ldots, n$。 3: 对于 $k = 0$ =1, . . . , $M - 1$，执行 4: 使用 $\{(X_i, Y(k)\, i)\}$ 和权重 $\bar{w}_i(x; \mathrm{h})$，根据式 (6) 计算 $\hat{m}^{(k)}$ANW (x)。 5: 对于 $i = 1$=, . . . , n，执行 6: 更新 $Y(_{i}^{k+1} \leftarrow Y_i + \big(Y\big)\,(k)$ $_i$ $-\hat{m}^{(k)}$ANW(X_i)$\big)$。 7: 结束循环 8: 结束循环 9: 输出：$\hat{m}_{\mathrm{DS-ANW}}(x) \leftarrow \hat{}$ $m^{(M)}$ANW (x)。

模拟：DS–ANW 通过偏差降低改进 ANW

在实践中，M 保持较小（例如，$M = 1$，2、3），以平衡偏差降低与方差逐渐膨胀。IDS2 研究报告称，增加迭代次数可以降低估计误差，也可能降低有限样本下对带宽选择的敏感性（Chen 等，2025）。在我们的实验中，我们将 M 视为较小的调优参数，并将 DS–ANW 与未锐化 ANW 基线进行比较。

我们考虑回归模型下的一维对齐原型

$$Y_i = m(X_i) + \varepsilon_i, \qquad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2),$$

其中 $X_i \overset{iid}{\sim} U(0,\ 1)$、$n = 500$ 和 $\sigma = 0$。15。真实信号为

$$m(x) = \sin(2\pi x) + 0.3\cos(6\pi x).$$

为模拟固定航路点，我们随机选取两个观测位置。使用调优后的参数，我们比较基线 ANW估计量（$M = 0$）与其数据锐化扩展 DS–ANW；后者通过对 $M = 1$ 应用 IDS2 并进行 $M = 2$ 次迭代得到。

图 3 展示了连续的锐化步骤如何在保留固定航路点吸引机制的同时减少平滑偏差。

表 1. 不同锐化迭代次数下 ANW 和 DS–ANW 的性能比较

方法	h	λ	RMSE	航点误差	平滑度
ANW	0.030	100	0.127	0.080	0.245
DS–ANW ($M = 1$)	0.030	100	0.089	0.073	0.321
DS–ANW($M = 2$)	0.030	100	0.080	0.069	0.335
DS–ANW($M = 3$)	0.030	100	0.076	0.067	0.335

表 1 总结了 ANW 和 DS–ANW 在固定平滑配置下的性能，其中带宽h 和惩罚参数 λ 在所有锐化迭代中均保持不变

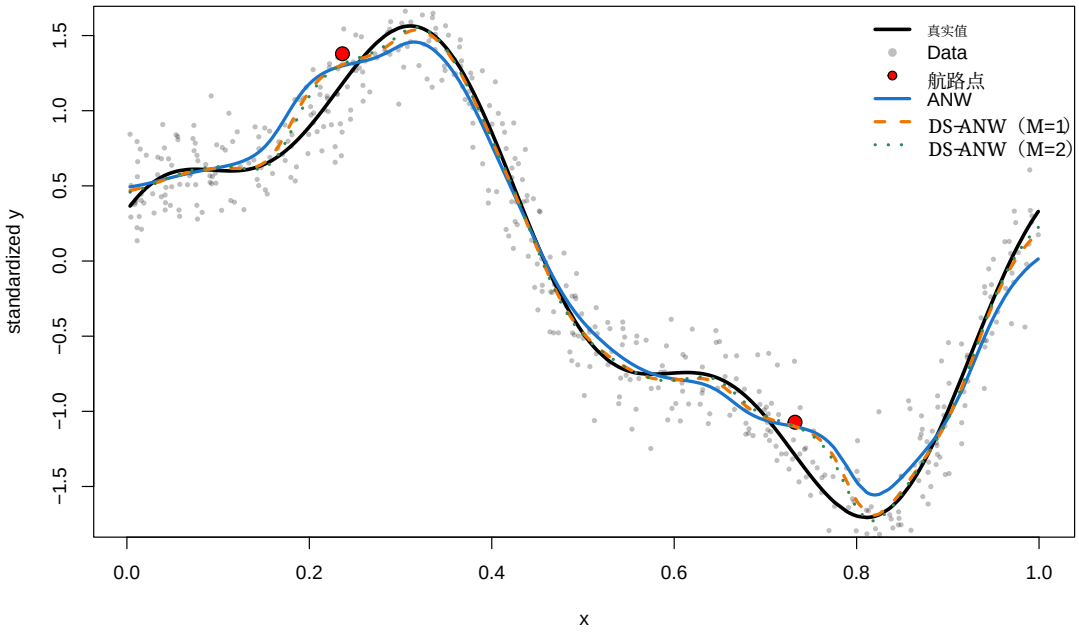


图 3. 通过迭代式 IDS2 精化比较 ANW 与 DS-ANW

结果展示了迭代数据锐化对全局准确度、固定航路点遵循度和平滑度的系统性影响。

随着锐化迭代次数增加，DS–ANW 相较于 ANW 显著降低了均方根误差，表明通过连续更新响应值可以有效进行偏差校正。同时，固定航路点误差从 ANW 到带有 $M = 3$ 的 DS–ANW 单调下降，说明在此配置下，适度锐化不会削弱对预设固定航路点的遵循程度，甚至可能略微改善航路点保真度。

然而，精度和固定航路点保真度的提升伴随着平滑度的明显损失：基于曲率的平滑度指标在首次锐化步骤后急剧上升，随后趋于稳定。尤其是，超过 $M = 2$ 后，均方根误差和固定航路点误差的边际改善相对有限，而平滑度惩罚则几乎没有进一步变化。

综合来看，这些结果表明，超过少量锐化迭代后，DS–ANW 的收益会逐渐减小。实践中，一到两次锐化步骤能够在偏差降低与形状规则性之间取得较好的折中；而增加迭代次数带来的益处有限，却会使曲线更加粗糙。

渐近性质

我们研究标准核平滑正则条件下 ANW估计量 $\hat{} B_{\mathrm{ANW}}(x)$ 的渐近行为。

假设 1（带宽）。 带宽 $h = h_n$ 满足

$$h \to 0 \quad and \quad nh \to \infty \qquad as\, n \to \infty.$$

假设2（核）。 核函数 K： $\mathbb{R} \to \mathbb{R}$ 是对称的，满足 s

$$\int K(u)\,du = 1, \qquad \int u^2 K(u)\,du = \sigma_K^2 < \infty, \qquad \int K^2(u)\,du = R(K) < \infty.$$

假设3 （平滑度）。 回归函数 m（x） $= \mathbb{E}[Y \mid X = x]$ 以及 X 的边际密度 f（x） 在 x的某个邻域内二阶连续可微，且满足 f（x） > 0。

假设4 （误差）。 $\varepsilon = Y - m$（ X） 的条件方差 σ^2（x） $=$Var（$Y \mid X =x$） 在 x 附近连续且有界。

假设5（固定航路点调优参数缩放）。 为避免内部点处的渐近支配（*Zhou* 等，*2024*），

$$\lambda \to \infty \quad and \quad \frac{\lambda}{nh^4} \to 0 \qquad (n \to \infty).$$

假设6 （约束数量）。 固定航路点的数量 q 是固定的，不会随 n增长。

固定航路点回归的一个核心要求是，估计量必须至少在渐近意义下，在约束位置精确恢复预先规定的值。尽管自适应 Nadaraya–Watson（ANW）估计量通过自适应重新加权纳入了固定航路点，但随着样本量增加，这种机制是否能够保证收敛到所施加的值，事先并不显而易见。

以下定理表明，在满足标准核正则条件且约束权重满足温和假设时，ANW估计量在固定航路点位置处逐点相合。

定理1 （**固定航点处的逐点收敛**）在假设*1*、*2*、*5*和*6*成立的条件下，

$$\hat{m}_{\mathrm{ANW}}(X_j) \xrightarrow{p} Y_j, \qquad j \in C. \tag{7}$$

定理1表明，ANW的自适应加权机制以强逐点意义强制满足固定航路点约束。具体而言，估计量不仅仅是保持在航路点取值附近，而是在每个受约束位置都依概率收敛到所施加的响应值。

这一结果为在航线设计问题中使用ANW作为满足约束的平滑器提供了理论依据；在这类问题中，必须精确遵循规定的航路点。

尽管定理1刻画了ANW估计量在固定航路点位置处的行为，但同样重要的是了解它在约束之外的渐近性质。尤其值得关注的是，为强制满足固定航路点而引入的自适应重新加权机制，是否会改变核回归经典的内部偏差—方差权衡。

以下定理表明，在不受约束的内部点处，ANW估计量与标准 Nadaraya–Watson 估计量具有相同的一阶渐近性质。

定理 2 (内部渐近性质)。 在假设 *1–5* 下，对于内部点 $x \notin \{X_j : j \in C\}$，*ANW*估计量满足

$$\mathrm{Bias}\{\hat{B}_{\mathrm{ANW}}(x)\} = \frac{1}{2}h^2\sigma_K^2\left(m''(x) + \frac{2m'(x)f'(x)}{f(x)}\right) + o(h^2), \tag{8}$$

$$\mathrm{Var}\{\hat{B}_{\mathrm{ANW}}(x)\} = \frac{\sigma^2(x)R(K)}{nhf(x)} + o\left(\frac{1}{nh}\right), \tag{9}$$

and

$$\sqrt{nh}\left(\hat{B}_{\mathrm{ANW}}(x) - m(x) - \mathrm{Bias}\{\hat{B}_{\mathrm{ANW}}(x)\}\right) \xrightarrow{d} N\left(0, \frac{\sigma^2(x)R(K)}{f(x)}\right). \tag{10}$$

ˆ因此， B_{ANW} 在内部点处与 Nadaraya–Watson 估计量渐近等价ˆm_{NW}（Fan和Gijbels 1996）。特别是，采用通常的选择$h_{\mathrm{opt}} \asymp n^{-1/5}$时，我们有 $nh^4 \asymp n^{1/5}$，因此 $\lambda \to \infty$ 而 $\lambda/(nh^4) \to 0$ 可以同时成立。因此，施加固定航路点约束不会渐近地增大内部偏差或方差，约束的影响仍局限于指定航路点附近的区域。

定理1和定理2的完整数学证明见附录。

数值性质

一维固定航点回归（一维）

首先，我们进行了简单的场景验证，以评估 ANW 和 DS–ANW 在兼顾固定航路点邻近性、平滑度和总体估计误差方面的表现。

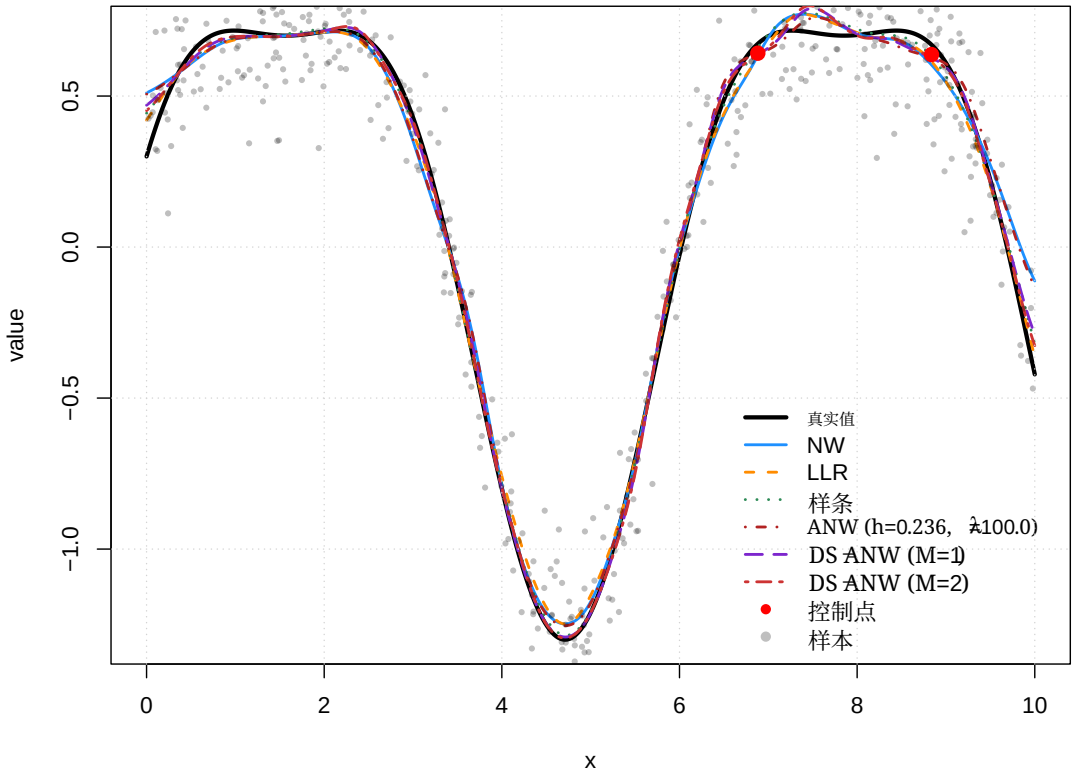
实验数据由一个已知测试函数生成，该函数同时呈现振荡行为和鲜明的结构特征：

$$Y_i = m(X_i) + \varepsilon_i, \quad m(x) = \sin x + 0.3\cos(2x), \quad x \in [0, 10], \tag{11}$$

其中误差项相互独立且同分布， $\varepsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$。

样本量设为 500，这是一个中等规模的数据集。为展示 ANW 和 DS–ANW 接近固定航路点的能力，分别考虑了航路点偏差较小和较大的两种情况。我们比较了标准的 Nadaraya–Watson核回归（NW）、其自适应约束扩展（ANW）和数据锐化变体（DS–ANW），以及局部线性核回归（LLR）和基于样条的方法。我们使用四项指标评估估计量的性能。均方根误差（RMSE）是常用的全局拟合精度指标。为评估 ANW估计量满足点值约束的能力，我们计算了约束点处的误差（航路点误差）。此外，拟合曲线的平滑度基于二阶导数惩罚进行测量，该惩罚用于量化曲率（Green和Silverman 1993）。

我们首先考虑一种弱约束情景（情况1），其中两个固定航路点的选取使其相对于无约束基线拟合的残差较小。



Fi图 4. 带有两个固定航路点的一维约束回归（案例 1)

在这种情况下，施加的航路点约束接近自然数据趋势，因此不会与全局平滑性产生明显冲突。

如图 4 所示，当施加的航路点约束较为宽松时，大多数平滑方法的拟合结果在视觉上相似，全局准确度也相当。在这种情况下，固定航路点与整体数据结构并无明显冲突，约束方法与无约束方法之间的差异相对较小。

我们在图 5 中比较了六种具有代表性的非参数平滑方法与 ANW 和 DS-ANW。B样条回归（De Boor 1977）和平滑样条（Qiu1995）作为经典的全局平滑器纳入比较，而 P样条（Pya 和 Wood 2015）代表现代惩罚样条框架。局部加权散点平滑（Cleveland 和 Devlin 1988）是典型的局部回归方法，高斯过程回归（GPR）（Seeger 2004）代表一种基于概率的核方法，自适应薄板样条（TPS）（Wood 2003）则提供了一种自适应空间平滑基线。

为联合评估固定航路点遵循、全局准确度和平滑度，我们基于三个指标定义了综合评分：

$$\mathrm{CSS} = w_c\,\phi\!\left(\frac{\text{Waypoint Error}}{\tau_c}\right) + w_g\,\phi\!\left(\frac{\text{RMSE}}{\tau_g}\right) + w_s\,\phi\!\left(\frac{\text{Smoothness}}{\tau_s}\right), \tag{12}$$

其中，惩罚函数 $\phi(\cdot)$ 定义为

$$\phi(x) = \max(0,\, x - 1). \tag{13}$$

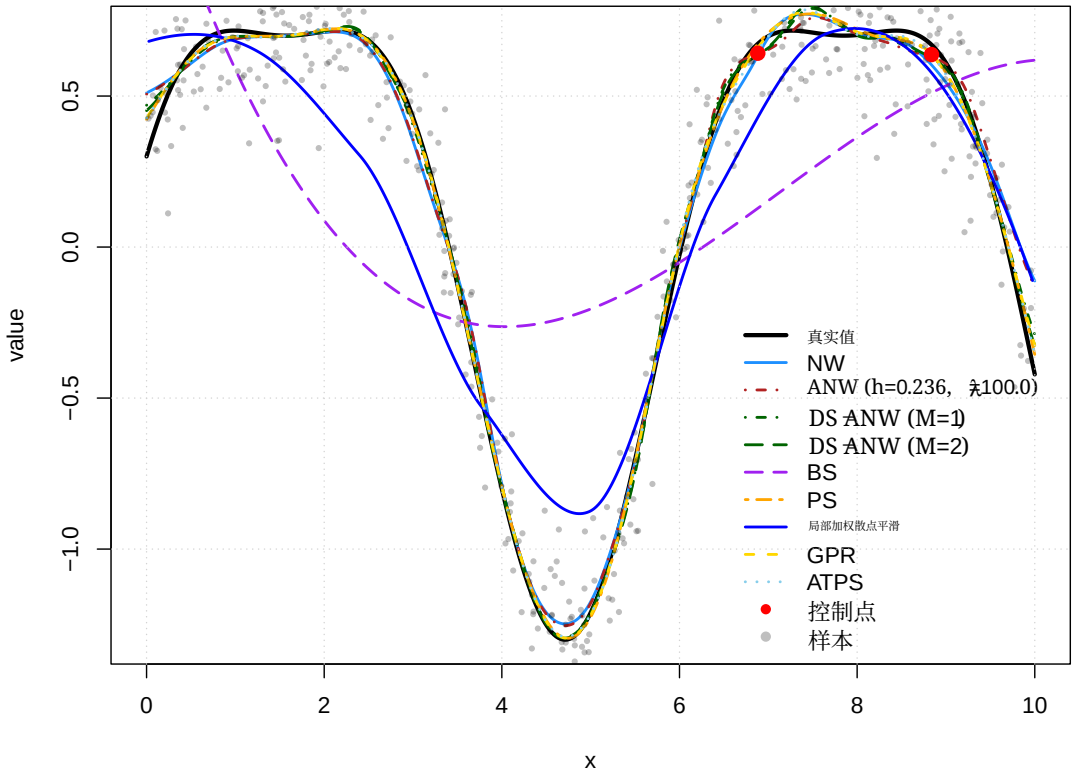


图5。 一维非参数平滑方法在固定航路点约束下的比较（情况1）。

每个容差参数 τ_c、 τ_g 和 τ_s 表示相应指标可接受的上限，超过该上限后将施加线性惩罚。在我们的实验中，航路点误差容差设为 $\tau_c = 0.10$，严格于噪声标准差（$\sigma = 0.15$），因此体现了测量噪声下较高的航路点要求。全局误差容差设为 $\tau_g = 0.05$。平滑度容差 τ_s 取为实验中所有候选拟合的平滑度中位数。权重固定为 $w_c = 0.5$、 $w_g = 0.3$ 和 $w_s = 0.2$。

在这种情况下，固定航路点的选择使其与基准拟合的偏差较小，对应于相对温和且易于处理的约束情景。因此，大多数方法都能实现较低的整体误差，并较好地满足固定航路点约束。

如表2所示，惩罚型和平滑概率平滑器，包括P样条、高斯过程回归（GPR）和自适应薄板样条（ATPS），均取得非常小的均方根误差值和接近于零的综合评分（CSS），这表明在施加的约束较弱时，它们具有较强的整体拟合效果。标准Nadaraya–Watson估计量在这一情景下的表现同样具有竞争力，约束偏差处于中等水平。

尽管在这种情况下，各方法之间的性能差异不那么明显，但 ANW 在基于核的估计量中取得了最小的航路点误差，同时保持了相当的全局准确度。这表明，自适应加权

表2. 一维仿真中估计器的性能比较（情况1）

模型	RMSE	航路点误差平滑度 CSS		
NW	0.060	0.025	4.43×10^{-7}	0.10
ANW	0.056	0.006	$6。11 \times 10^{-7}$	0.17
DS–ANW（$M = 1$）	0.034	0.008	6.22×10^{-7}	0.14
DS–ANW（$M = 2$）	0.032	0.009	$6。37 \times 10^{-7}$	0.15
BS	0.553	0.725	$1。03 \times 10^{-8}$	6.14
PS	0.026	0.021	3.63×10^{-7}	0.00
局部加权散点平滑	0.236	0.156	$1。31 \times 10^{-7}$	1.39
GPR	0.027	0.013	$3。66 \times 10^{-7}$	0.00
ATPS	0.028	0.006	3.86×10^{-7}	0.01

即使基线对齐效果已经很好，也能进一步增强航点遵循的稳定性。

数据锐化 ANW 估计量（DS–ANW， $M = 1$ 和 $M = 2$）相较于 NW 和 ANW 均进一步降低了均方根误差，表明迭代锐化能够有效降低偏差。尽管 DS–ANW 的航路点误差略高于 ANW，但偏差幅度较小，且在各次迭代中保持稳定，同时整体平滑度得以保留。这些结果表明，在低噪声、弱约束场景中，DS–ANW 能够在全局准确度与约束遵循之间实现较为理想的权衡。

接下来，我们考虑一个具有挑战性的约束情景（情况 2）：其中两个固定航路点经过刻意选择，使其相对于无约束基线拟合具有较大的残差。在这种情况下，施加的航路点要求与全局数据趋势严重冲突，因此严格实施约束要困难得多。

如图 6 所示，即使与全局结构存在强烈冲突，ANW 仍能够通过自适应重新加权，使拟合曲线显著靠近指定的航路点。相比之下，随着锐化迭代次数增加，数据锐化变体（DS–ANW）对航路点的遵循程度逐渐减弱，反映出在此情形下偏差降低与约束实施之间的权衡。

图7进一步将ANW和DS–ANW与几种具有代表性的非参数平滑方法进行了比较。未明确纳入航路点信息的方法对施加的约束基本不敏感，尽管在某些情况下取得了较低的全局均方根误差，但仍无法接近规定位置。

如表3所示，ANW通过在全局准确度、约束满足度和平滑度之间取得良好平衡，在所有竞争方法中获得了最低的综合评分（CSS）。在此情形下，固定航路点经过刻意选择，相对于基准拟合具有较大的残差，代表了一种具有挑战性的约束情景，其中逐点要求与全局平滑性存在强烈冲突。

在此类条件下，大多数无约束或弱约束平滑器，包括NW、P样条、高斯过程回归（GPR）、局部加权散点平滑（LOESS）和自适应薄板样条（ATPS），均表现出显著增大的航路点误差，尽管

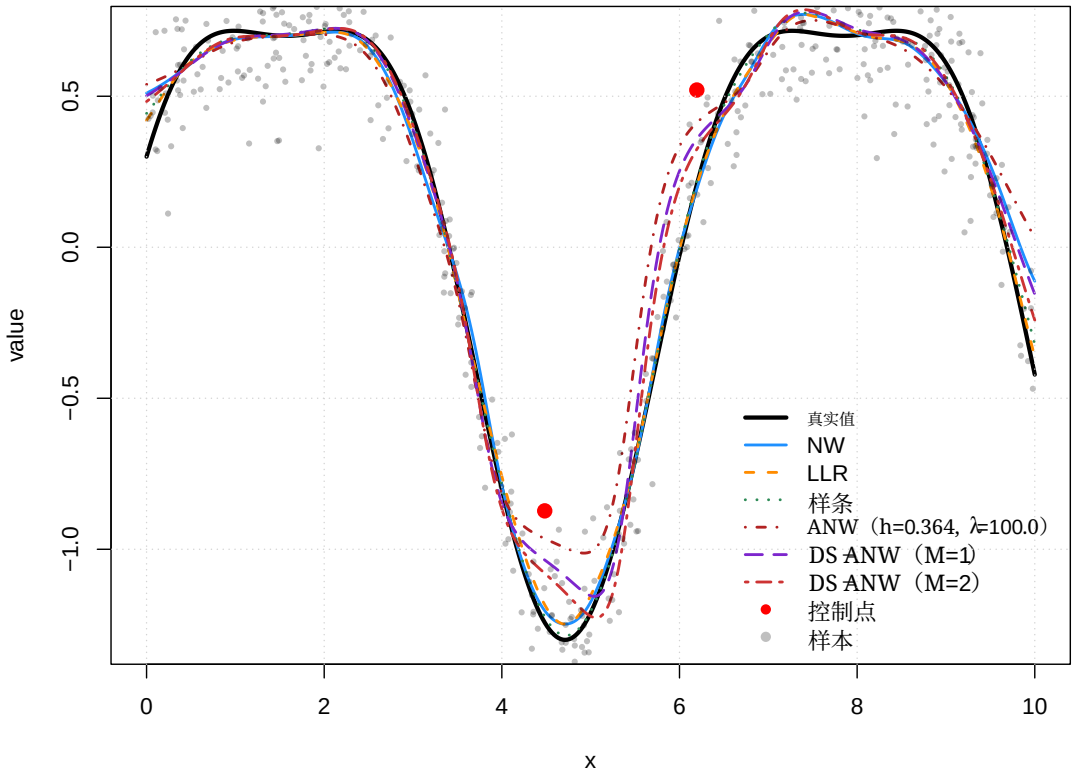


图 6. 具有两个固定航路点的一维约束回归（情况 2）

表3. 一维仿真中估计器的性能比较（情况 2）

模型	RMSE	航路点误差平滑度		CSS
NW	0.060	0.329	4.43×10^{-7}	1.25
ANW	0.159	0.100	9.46×10^{-7}	0.97
DS–ANW ($M = 1$)	0.103	0.162	$1。20 \times 10^{-6}$	1.08
DS–ANW ($M = 2$)	0.082	0.202	$1。24 \times 10^{-6}$	1.18
BS	0.553	1.691	$1。03 \times 10^{-8}$	10.97
PS	0.026	0.328	$3。63 \times 10^{-7}$	1.14
局部加权散点平滑	0.236	0.345	$1。31 \times 10^{-7}$	2.34
GPR	0.027	0.324	$3。66 \times 10^{-7}$	1.12
ATPS	0.028	0.320	$3。86 \times 10^{-7}$	1.11

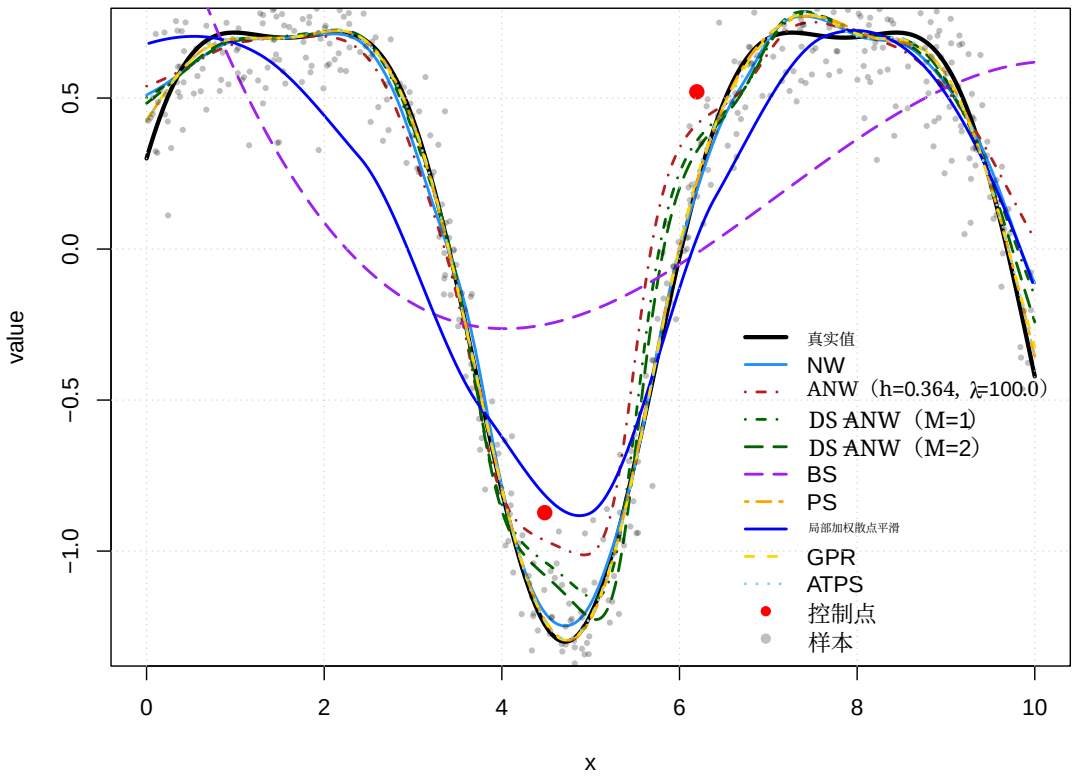


图7。 固定航路点约束下的一维非参数平滑方法比较（情况 2）

保持较低的均方根误差值。这表明，仅凭全局拟合质量不足以满足相距较远的固定航路点要求。相比之下，ANW 通过自适应惩罚机制明确地对固定航路点重新赋权，使航路点误差显著低于所有其他方法，因此 CSS 值也最小。

数据锐化 ANW 估计量（DS–ANW，使用 $M = 1$ 和 $M = 2$）相较于 ANW 大幅降低了均方根误差，体现了通过迭代锐化有效进行偏差校正。然而，随着锐化迭代次数增加，全局准确度的提升伴随着航路点误差增大、平滑度降低。因此，在这种约束很强的情形下，DS–ANW 的 CSS 值高于 ANW，这表明激进的偏差降低可能会在一定程度上抵消对相距较远的固定航路点进行预期约束的效果。

总的来说，这些结果清楚地揭示了 ANW 与 DS–ANW 作用上的区别：ANW 更适用于由强约束或相互冲突的固定航路点主导的情形；而 DS–ANW 提供了偏差-方差权衡，有利于提高全局准确度，但当指定的固定航路点与基线结构偏离较大时，可能会损害对约束的严格遵循。

表4报告了 ANW 对航路点惩罚参数 λ 和固定航路点数量 q（情况 2）的敏感性分析。对于每个固定的 q，增大 λ 会加强约束实施，使航路点

表4. ANW 关于 q 和 λ 的敏感性分析（情况 2）

q	λ	RMSE	航路点误差
1	10	0.06	0.33
1	100	0.10	0.10
1	1000	0.14	**0.01**
2	10	0.06	0.30
2	100	0.12	0.09
2	1000	0.17	**0.01**
3	10	0.07	0.30
3	100	0.14	0.09
3	1000	0.20	**0.01**

误差单调下降。局部贴合度的提升伴随着全局均方根误差的增加，这表明拟合精度与精确满足航路点之间存在预期的权衡。此外，对于固定的 λ，使用更多航路点（更大的 q）通常会使问题约束更严格，并倾向于增大均方根误差；而不同 q 下的航路点误差则基本相当，最终主要取决于 λ的选择。

二维约束轨迹拟合（二维）

为模拟空间场景，我们生成由路径变量 $(x(s)$、 $y(s))$参数化的二维轨迹 $s \in [0,1]$ (Mellinger 和 Kumar， 2011)。随后，将沿路线变化的标量响应建模为

$$z(s) = m(s) + \varepsilon(s), \qquad \varepsilon(s) \overset{iid}{\sim} \mathcal{N}(0, \sigma^2).$$

尽管数据位于二维空间中，回归分析仍以 s 作为变量，因为它自然地表示沿轨迹的位置（Ramsay和Silverman（2005））。在设定三个固定航路点（$q = 3$）的条件下，我们比较了标准 Nadaraya–Watson估计量与 ANW 和 DS–ANW。在二维设置中，ANW 和每个 DS–ANW 变体的带宽与惩罚参数均通过交叉验证分别选取，以反映空间轨迹的复杂性增加。

图8展示了具有多个指定固定航路点的二维轨迹几何，这些航路点作为空间约束施加。在图9中，展示了自适应 Nadaraya–Watson（ANW）机制在二维空间场景中的应用，其中拟合曲面在局部被拉向指定的固定航路点，以满足施加的约束。数据锐化变体（DS–ANW）在保持相同约束执行机制的同时，通过迭代偏差降低进一步改进拟合。

图10对比展示了 ANW 与前文介绍的其他平滑方法，并着重突出了它们在存在固定航路点时的不同行为。

表5总结了二维空间仿真中具有代表性的估计器的性能。全局平滑器和概率方法，包括样条回归，

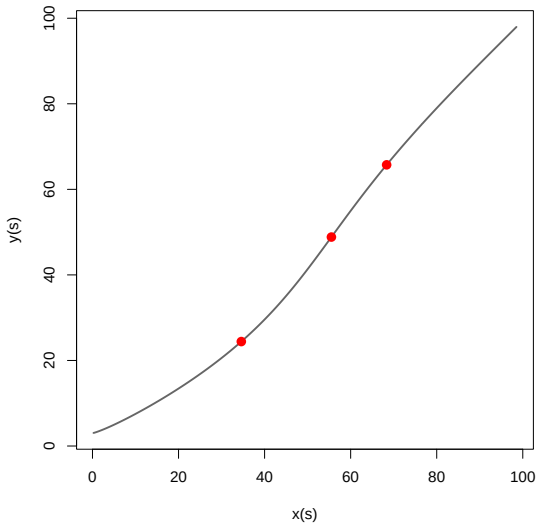


图 8. 带固定航路点的轨迹几何

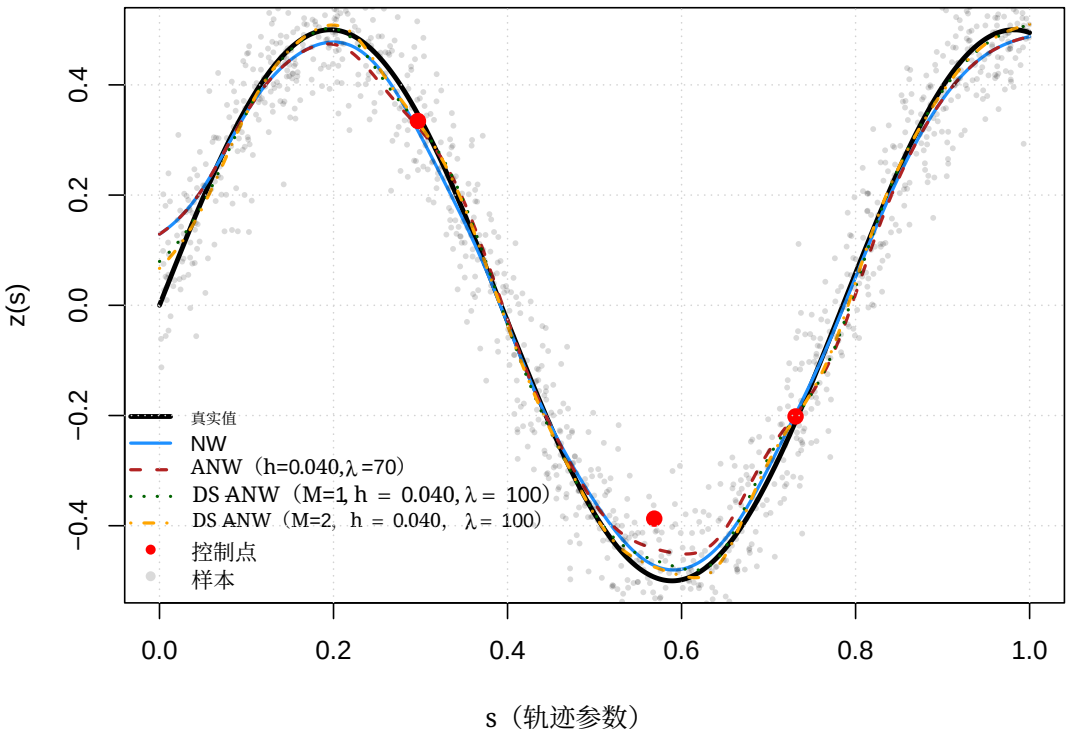


图9。 带固定航路点的二维约束路径拟合

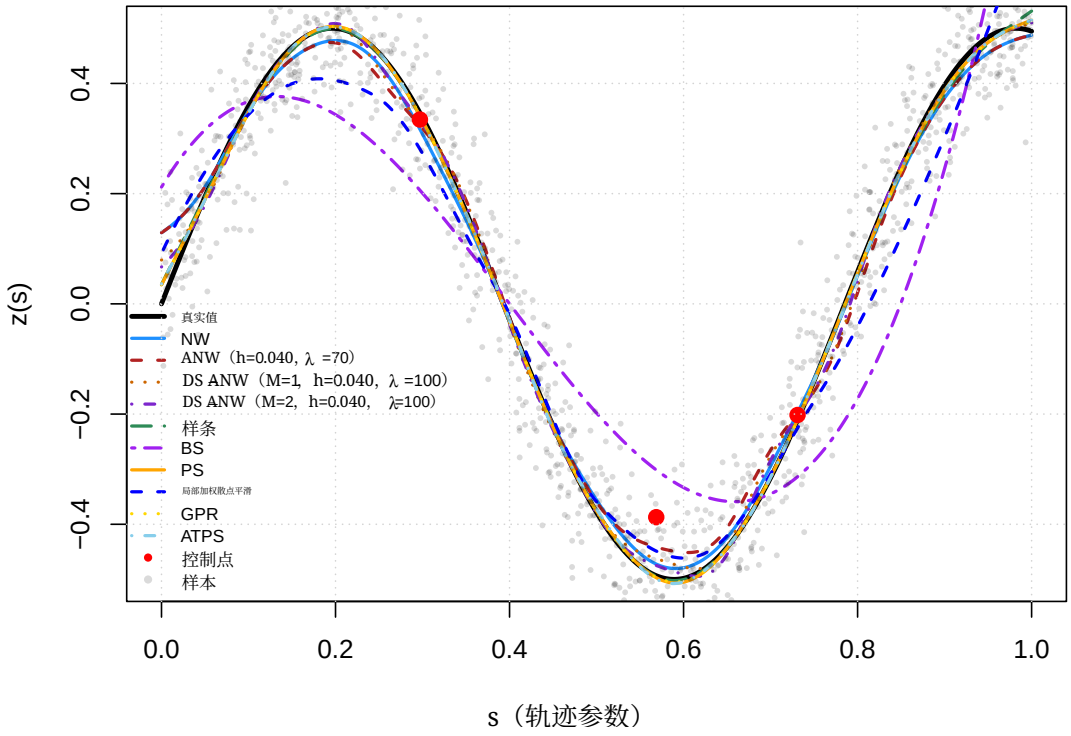


图10。 固定固定航路点的二维约束路径拟合

表5。 二维空间仿真中估计器的性能比较

模型	RMSE	航路点误差平滑度 CSS		
NW	0.02	0.05	8.19×10^{-9}	0.01
ANW	0.03	0.03	1.31×10^{-8}	0.14
DS–ANW ($M = 1$)	0.02	0.04	1.76×10^{-8}	0.25
DS–ANW ($M = 2$)	0.01	0.05	$1。68 \times 10^{-8}$	0.23
样条	0.01	0.06	7.10×10^{-9}	0.00
BS	0.16	0.43	$5。55 \times 10^{-9}$	2.27
PS	0.01	0.06	7.56×10^{-9}	0.00
局部加权散点平滑	0.07	0.05	$6。89 \times 10^{-9}$	0.09
GPR	0.01	0.06	7.61×10^{-9}	0.00
ATPS	0.01	0.07	7.88×10^{-9}	0.00

P样条、高斯过程回归和 ATPS 的均方根误差值最小，表明仅根据预测误差评估时，它们具有较强的全局拟合精度。然而，这些方法通常会产生较大的航点误差，说明它们在适应外部施加的固定航路点方面灵活性有限。

与无约束平滑器相比，ANW 大幅降低了航路点误差，体现了其自适应加权机制在确保航点遵循方面的有效性。与表现最佳的全局平滑器相比，这一改进伴随着均方根误差的增加，展现了二维情境下全局准确度与约束满足度之间的权衡。

数据锐化变体 DS–ANW 进一步降低了相较于 ANW 的均方根误差，尤其是在锐化迭代次数较少时。与此同时，随着锐化程度提高，航路点精度会有所下降，这凸显了数据锐化带来的偏差—方差—约束权衡。总体而言，ANW 和 DS–ANW 在全局拟合与航路点约束之间实现了可控的折衷；当实际应用特别重视精确遵循约束时，它们能取得颇具竞争力的综合评分。

实证应用

在路线对齐应用中，外部施加的固定航路点（例如要求铁路经过班夫）通过将其视为确定性固定航点而纳入我们的框架。尽管这些位置并非源自观测轨迹数据，但它们会被补充到观测集中，并纳入约束集 C。通过为其分配较大的惩罚权重λ，估计曲线实际上会被迫插值经过这些指定位置。重要的是，由于补充点被视为普通观测值，同样的约束信息也会提供给所有竞争方法，从而确保公平且一致的比较。在这一表述下，外部几何要求可作为所提出模型中的内部固定航路点得到自然处理。

铁路线路设计

在这些应用中，可用数据仅由地理坐标组成（纬，度、经度）。为了将这些观测值嵌入一维非参数回归框架，我们采用经度作为协变量

$$X := \text{lon},$$

并将纬度作为响应变量

$$Y := \text{lat}.$$

尽管这种构造不同于理论上按弧长进行的参数化 $X = s$，但它在保留路线相对几何形状的同时，为数据提供了一维的实际排序。

基线是从温哥华到多伦多的真实铁路路线，名为“The canadaian”。该基线的轨迹点来自 VIA Rail， https://www.viarail.ca/en/developer-resources其使用许可为加拿大开放政府许可协议第2版， https://open.canada.ca/en/open-g

overnment-licence-canada。现有铁路网并未经过风景价值极高的班夫地区。为展示所提方法的灵活性，我们构建了一条假想的铁路线路，并将班夫设为固定航路点。

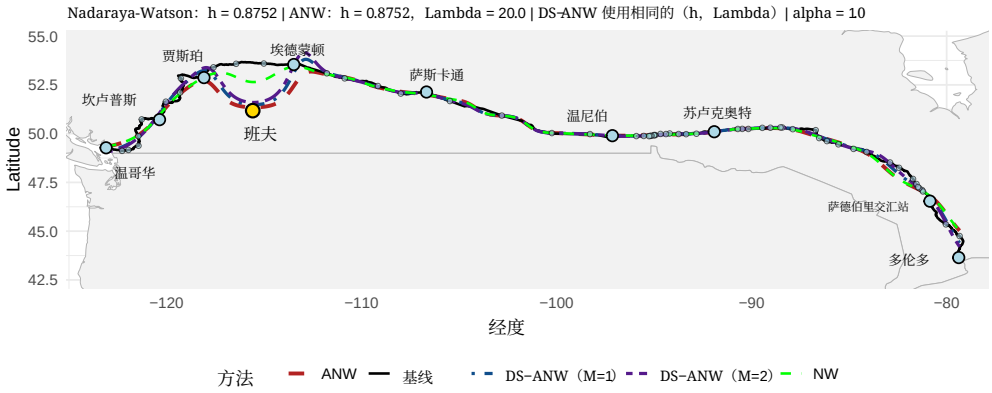


图11. 固定航路点条件下的一维铁路线路对齐中 Nadaraya-Watson、ANW 和 DS–ANW 的比较

在图11中，ANW及其数据锐化变体精确地满足了班夫处的固定航路点。远离约束点时，ANW生成的轨迹总体上与NW相当，同时仍能将拟合结果拉向指定航路点。我们还报告了DS–ANW（其中 $M = 1$和 $M = 2$），相较于ANW，它进一步锐化了局部特征。然而，在埃德蒙顿附近，由于停靠位置变得非常稀疏，DS–ANW可能出现明显的局部过冲：附近可用数据有限，降低了基于核的估计的稳定性，而锐化更新可能会放大这种局部不稳定性。

在图12中，我们将ANW及其数据锐化变体与六种具有代表性的非参数平滑方法进行了比较。ANW和DS–ANW能够将拟合轨迹拉向班夫处的固定航路点，而其他方法基本不受这一约束的影响，无法靠近指定位置。

表 6报告了铁路线路应用中的性能指标。与无约束平滑器相比，标准Nadaraya–Watson估计量的均方根误差相对较低，但航点误差非常大，表明其与指定固定航路点存在显著偏差。惩罚型平滑器和概率型平滑器也表现出类似的情况，包括P样条、平滑样条、高斯过程回归（GPR）和自适应薄板样条（ATPS）；这些方法优先考虑全局平滑性，代价是航点遵循性较差。

相比之下，ANW 显著降低了航路点误差，从而在所有竞争方法中取得了最小的综合评分（CSS）。尽管与纯粹的全局平滑器相比，ANW 的均方根误差（RMSE）更高、平滑度更低，但其自适应重加权机制能够有效约束固定航路点；这在铁路场景中至关重要，因为固定航路点代表的是强制性的设计

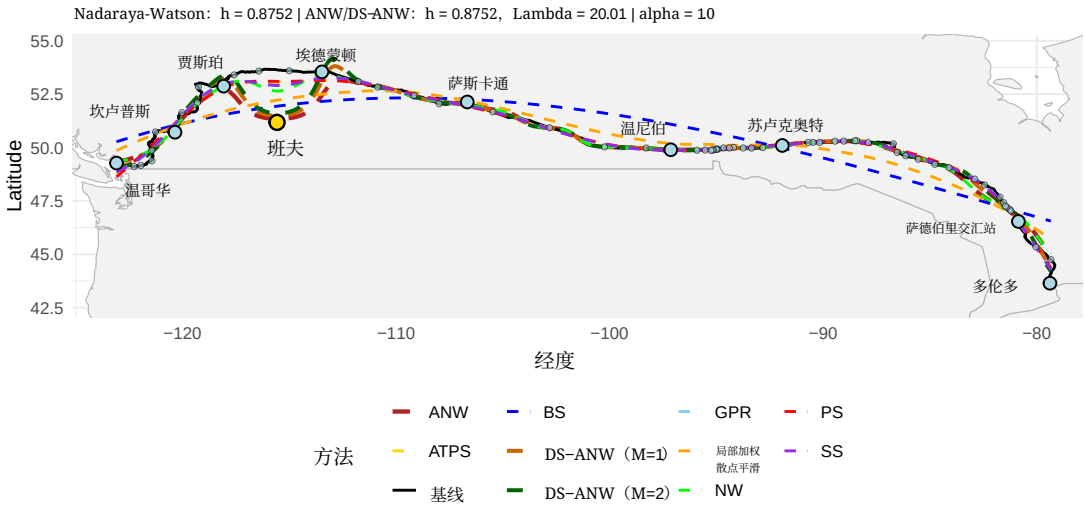


图12. 固定航点约束下的一维铁路线路非参数回归方法比较

表 6. 铁路应用的性能指标

模型	RMSE	航路点误差	平滑度	CSS
NW	0.29	1.47	9.57×10^{-8}	8.00
ANW	0.65	0.17	$2。01 \times 10^{-7}$	3.29
DS–ANW ($M = 1$)	0.55	0.30	$3。69 \times 10^{-7}$	3.55
DS–ANW ($M = 2$)	0.51	0.42	$4。65 \times 10^{-7}$	3.98
B样条	1.07	0.91	$1。86 \times 10^{-10}$	9.10
P样条	0.22	1.92	$1。02 \times 10^{-8}$	9.92
局部加权散点平滑	0.70	1.04	$1。65 \times 10^{-9}$	7.87
平滑样条	0.22	1.73	$2。20 \times 10^{-8}$	8.93
GPR	0.22	1.73	$2。39 \times 10^{-8}$	8.98
ATPS	0.22	1.86	$1。34 \times 10^{-8}$	9.62

要求，而非可灵活调整的偏好。

数据锐化 ANW 估计量（DS–ANW，采用 $M = 1$ 和 $M = 2$）相较于 ANW 进一步降低了均方根误差，反映出通过迭代锐化实现了部分偏差校正。然而，全局准确度的提升伴随着航路点误差显著增加，并且随着锐化迭代次数增加，平滑度逐渐下降。因此，在此应用中，DS–ANW 得出的 CSS 值高于 ANW。

这种现象可归因于铁路车站的空间分布不规则，尤其是在局部信息有限的稀疏区域。在这些区域，迭代锐化可能会放大局部外推效应，导致形状失真并降低航路点保真度。总体而言，这些结果表明，尽管

DS–ANW 可以改善全局拟合，但在实际铁路线路问题中，ANW 仍是严格约束固定航路点最可靠的选择。

在此应用中，我们将班夫设为固定航路点，以展示所提出方法在实际应用中的灵活性。然而，现实中，VIA Rail 运营的主要跨大陆客运铁路并不直接服务班夫。相反，长途路线会绕开该地区，选择更偏北、人口密度和运营需求更高的通道。班夫的铁路直达服务主要限于季节性或豪华旅游列车，而非常规城际运输。这反映了景观价值与经济可行性之间的现实权衡，也解释了为何此处构建的假设线形偏离现有铁路网。

高速公路线形设计

我们选取萨斯喀彻温省的11号公路，这是一条贯穿萨斯喀彻温省南北的主要公路。数据来自萨斯喀彻温省政府， https://geohub.saskatchewan.ca/datasets/da074f6eb1814ef1b033b7a090c93cd3/about，依据 GOS 标准无限制使用数据许可协议 v2.0。该数据集包含沿11号公路的55个主要交叉路口，数据取自维基百科， https://en.wikipedia.org/wiki/Saskatchewan_Highway_11。为便于建模，我们旋转坐标系，使公路的主要走向与水平轴对齐；随后再将拟合曲线旋转回原始方向。霍沃登和马滕斯维尔被视为固定航路点，因为它们是位于11号公路走廊沿线、人口较多的社区。

在 11号公路的应用中，原始地理坐标（x， y） = (经度，纬度)构成了一条近乎南北走向的路线。在原始坐标系中，这会在（x， y）平面上形成一条近乎垂直的曲线，无法表示为单值函数 $x = f(y)$ 或 $y = g(x)$。因此，无法直接进行非参数回归。为了获得适用于核回归的一维表示，我们对坐标系进行旋转。令

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \qquad \theta = 90^\circ.$$

这一变换将原来的南北走向路线映射为近似东西走向的曲线。在旋转后的坐标系中，变换后的路线几乎呈水平，因此 y' 近似为关于 x'的单值函数：

$$y' \approx f(x').$$

这种几何重新参数化大大简化了建模：可将协变量X 取为水平坐标 x'，将响应变量 Y 取为垂直坐标 y'。

图 13 和图 14 展示了旋转坐标系中的拟合路线，其中水平轴作为协变量，垂直轴作为响应变量。这种表示方式可以直观地评估不同平滑策略下的航点遵循情况和全局平滑性。

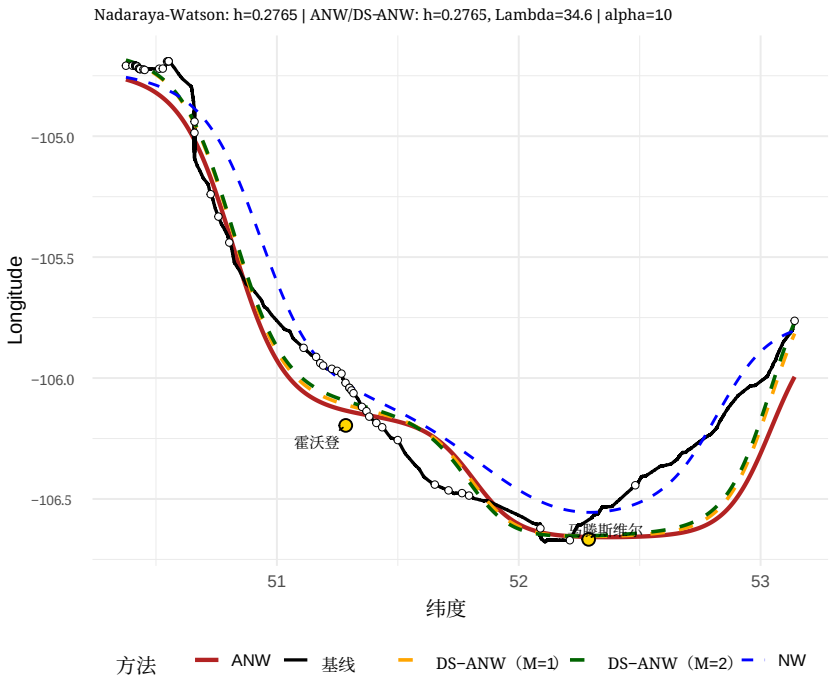


图 13 固定航路点条件下，11号公路旋转坐标系中的 ANW 和 DS–ANW 拟合

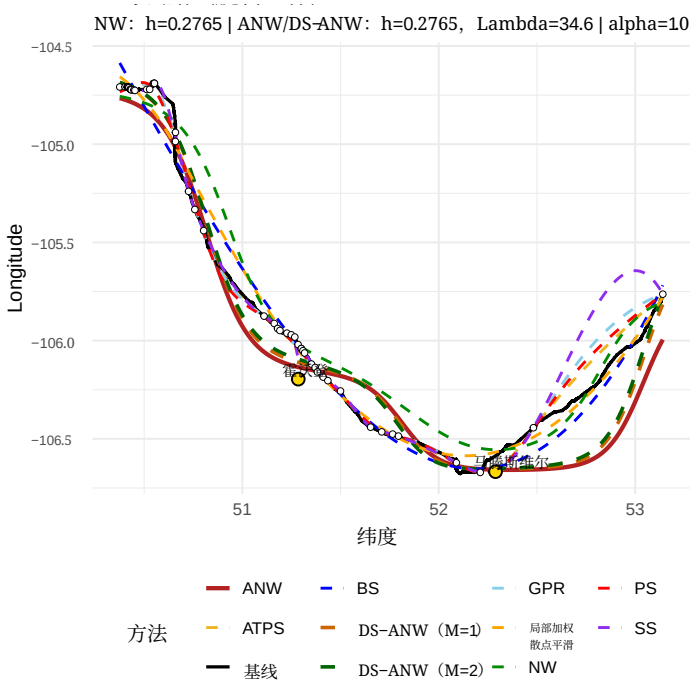


图 14 固定航路点约束下，11号公路旋转坐标系中非参数平滑方法的比较

图14比较了 ANW 和 DS–ANW 与采用相同航路点约束的几种代表性非参数平滑方法。未明确纳入航点信息的方法，在指定位置附近往往会出现比基于 ANW 的方法更大的偏差；而基于 ANW 的方法则能更好地平衡航点遵循与全局平滑性。

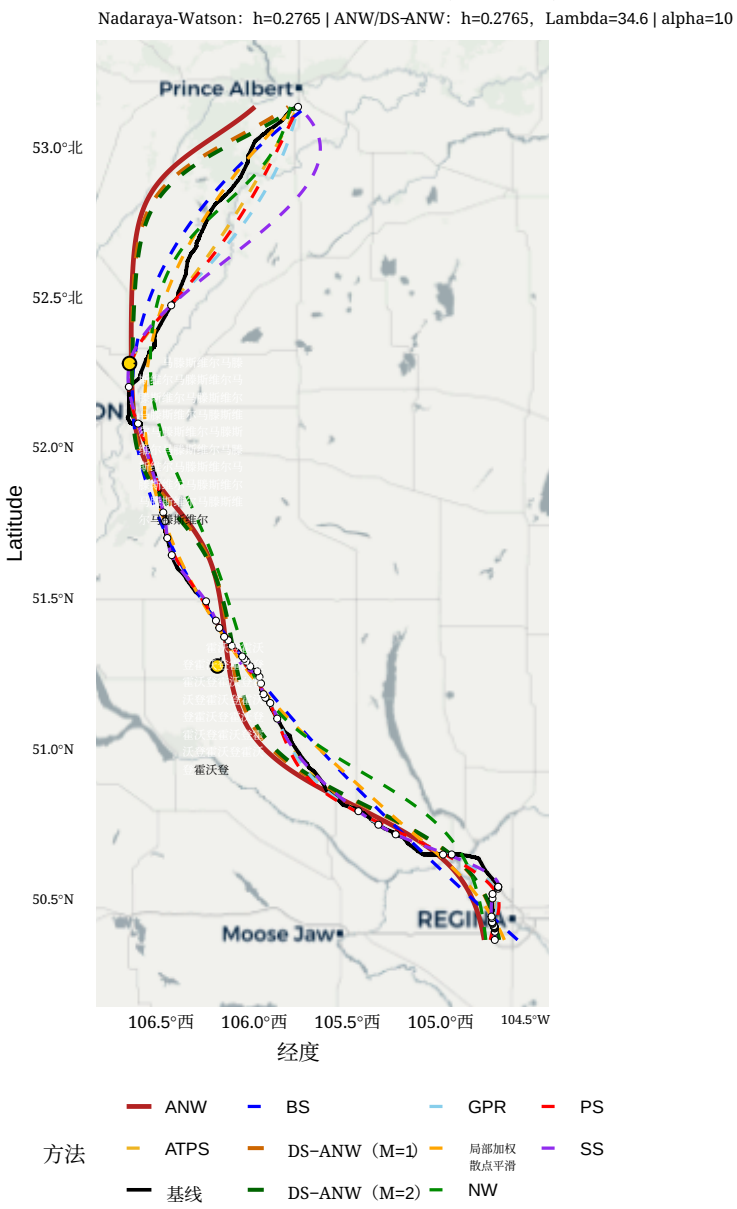


图15. 映射回原始地理坐标的 11号公路拟合线形

图15将拟合后的对齐结果映射回原始地理坐标系。拟合曲线贴合高速公路的整体形状，同时保持靠近固定航路点。在某一局部纬度范围内出现的明显偏差主要

是由于沿路线的采样密度高度不均匀：该区域内可用的观测值很少，因此 NW 型平滑器和航路点正则化变体都缺乏足够的数据支持。结果是，局部拟合变得更加敏感，可能偏离基线。

表 7. 高速公路应用的性能指标

模型	RMSE	航路点误差	平滑度	CSS
NW	0.14	0.15	$1。92 \times 10^{-10}$	0.58
ANW	0.19	0.04	$3。28 \times 10^{-10}$	0.48
DS–ANW ($M = 1$)	0.16	0.06	$3。85 \times 10^{-10}$	0.52
DS–ANW ($M = 2$)	0.15	0.07	$3。84 \times 10^{-10}$	0.54
B样条（B样条）	0.09	1.32	2.87×10^{-11}	0.67
P样条	0.07	0.11	$5。14 \times 10^{-10}$	0.54
局部加权散点平滑	0.07	0.12	$4。65 \times 10^{-11}$	0.48
平滑样条	0.15	0.08	2.56×10^{-8}	0.64
GPR	0.08	0.11	$5。68 \times 10^{-10}$	0.57
ATPS	0.06	0.11	$6。03 \times 10^{-10}$	0.54

表 7 总结了高速公路路线对齐应用的性能指标。与无约束平滑器相比，标准 Nadaraya–Watson估计量具有较好的全局准确度，但约束偏差明显更大。多种惩罚平滑器和概率平滑器，包括 B样条、P样条、局部加权散点平滑、高斯过程回归（GPR）和自适应薄板样条（ATPS），均取得了较低的均方根误差值，但通常无法满足规定的固定航路点要求，这一点从相对较大的航路点误差中可见一斑。

相比之下，ANW 在所有方法中实现了最低的综合评分（CSS），显著降低航路点误差的同时，保持了可接受的全局准确度和平滑度。这表明，对固定航路点进行自适应重新加权在高速公路场景中尤为有效，因为该场景中的航路点约束强度适中，且沿路线空间分布均匀。

数据锐化 ANW 估计量（DS–ANW，使用 $M = 1$ 和 $M = 2$）进一步降低了均方根误差，相较于 ANW 有所改善，反映出迭代锐化带来的偏差降低。然而，全局准确度的提升伴随着航路点误差略有增加，因此综合评分值仍略高于 ANW。值得注意的是，DS–ANW 的平滑度与 ANW 相当，这表明轻度锐化在此应用中不会引入严重振荡。

结论

我们提出了一种用于固定航路点路线对齐的自适应 Nadaraya–Watson 估计量（ANW）。通过引入航路点特定调节参数 λ，同时保持全局带宽 h，ANW 将约束执行与平滑解耦，并避免了在固定航路点附近采用朴素带宽缩减时常产生的局部不规则性。我们进一步将 ANW 与迭代数据锐化（IDS2）相结合

得到 DS–ANW，在不改变航路点加权机制的情况下减少了平滑偏差。

我们的理论证明，在适当的λ缩放条件下，固定航路点处可实现逐点收敛，并表明在内部点处，ANW 渐近等价于经典 Nadaraya-Watson 估计量。一维和二维数值研究表明，ANW 能在全局准确度、平滑度和航点遵循之间实现稳定的平衡，尤其是在固定航路点与基线几何形态相冲突时。DS–ANW 通过少量锐化步骤（通常为 $M = 1$–2 步）还能进一步降低均方根误差，但收益可能递减，并且在采样稀疏区域的稳定性可能较差。铁路和公路线形应用表明，将外部设计要求作为确定性航路点纳入模型，是一种将工程约束整合到统计平滑流程中的简单、透明的方法。

讨论：稳健性考量与可能的扩展

所提出的自适应 Nadaraya–Watson 估计量（ANW）及其数据锐化变体 DS–ANW，旨在经典平方误差回归框架下兼顾平滑度与严格的航点遵循。与大多数基于核的平滑器一样，该估计量在轻尾噪声假设下具有最优性，但可能对离群值或局部污染较为敏感，尤其是在采样稀疏的区域。

从实际的线形设计角度来看，此类污染可能源自测量误差、数字化伪影，或沿路线进行数据采集时的不规则性。在这些情况下，自然可以考虑采用稳健性增强措施，在保留自适应加权所引入的航路点吸引机制的同时，限制极端观测值的影响。

一个可行的方向是将分布鲁棒优化（DRO）的思想引入 ANW框架（Blanchet 等，2019）。从概念上说，可以用带惩罚项或经过稳健化处理的局部准则，替代单纯的平方误差目标函数，将极端局部均值收缩至基线值。重要的是，这些修改可以在不改变固定航点调节参数的情况下实现 λ，从而保持航路点约束不变。

然而，稳健性通常会以增加偏差为代价，尤其是在局部曲率较大或几何形状快速变化的区域附近。这种权衡在对齐设计问题中尤为重要，因为几何保真度和航路点精度都至关重要。尽管稳健化可以抑制离群值的影响，但也可能削弱轨迹中合理的局部特征。

数据锐化在此情境下提供了一种互补机制。作为一种独立于底层估计量运行的偏差降低封装方法，它原则上可以应用于标准 ANW 变体和稳健化 ANW 变体之上。这表明可以构建一个灵活的模块化框架，根据数据特征和设计要求，以不同方式组合航路点约束、稳健性和偏差降低。

A 对分布鲁棒 ANW 估计器及其 r

与数据锐化之间相互作用的系统研究超出了本文的范围，将留待未来工作。这种扩展需要对由此产生的偏差–方差权衡进行细致的理论分析，并在重尾或受污染的噪声模型下开展广泛的实证评估。

数据可用性

可通过发送电子邮件至 xiaoping.shi@ubc.ca 申请获取该数据集。

致谢

作者 Qiong Li 获得了国家自然科学基金（项目编号：12271047）以及广东与香港高校“1+1+1”联合研究合作计划（2025A0505000010）的资助。作者 Xiaoping Shi 获得了加拿大自然科学与工程研究委员会探索基金 RGPIN 2022-03264、加拿大自然科学与工程研究委员会联盟国际催化基金 ALLRP 590341-23 的资助，并得到不列颠哥伦比亚大学奥肯那根分校（UBC-O）副校长研究项目的支持；该项目与 UBC-O 欧文·K·巴伯理学院合作开展。

作者贡献

S.D. 和 X.S. 设计了本研究。S.D. 开展了本研究并撰写了论文主体。Y.C. 协助进行了比较分析。W.Y.、Q.L. 和 X.S. 审阅并修订了论文。所有作者均审阅了论文。

参考文献

Blanchet, J., Kang, Y., and Murthy, K. (2019).“稳健的Wasserstein轮廓推断及其在机器学习中的应用。”《应用概率杂志》, 56, 174–195. Braun, W. J. and Shi, X. (2025).“用于惩罚回归的数据锐化及其在变点估计中的应用。”《统计与计算》, 修订中。Chen, H., Braun, W. J., and Shi, X. (2025).“迭代数据锐化。”《计算与图形统计学杂志》, 34(1), 162–174. Cleveland, W. S. and Devlin, S. J. (1988).“局部加权回归：一种通过局部拟合进行回归分析的方法。”《美国统计学会杂志》, 83(403), 596–610. De Boor, C. (1977).“用于B样条计算的软件包。”《*SIAM*数值分析杂志》, 14(3), 441–472. Dontchev, A. L., Kolmanovsky, I. V., and Tran, T. B. (2022).“通过最优控制进行约束数据平滑。”《最优控制应用与方法》, 43(4), 1257–1269. Fan, J. and Gijbels, I. (1996).《局部多项式建模及其应用》, 第66卷, 《统计与应用概率专著丛书》。Chapman and Hall/CRC，博卡拉顿，佛罗里达州。Green, P. J. and Silverman, B. W. (1993).《非参数回归与广义线性模型：粗糙度惩罚方法》。Chapman and Hall/CRC，第1版。Hall, P. and Huang, L.-S. (2001).“受单调性约束的非参数核回归。”《统计年刊》, 29(3), 624–647。

Henderson, D. J. and Parmeter, C. F. (2009).“在非参数回归中施加经济约束：理论、实现与扩展。”《非参数计量经济学方法》，Emerald Group Publishing Limited, 433–469. Kang, M. W., Schonfeld, P., and Jong, J.-C. (2007).“通过可行门优化公路线形。”《先进交通杂志》, 41(2), 115–144. Kano, H., Egerstedt, M., Nakata, H., and Martin, C. F. (2003).“B样条与控制理论。”《应用数学与计算》, 145(2-3), 263–288. Kano, H., Fujioka, H., and Martin, C. F. (2011).“带约束的最优平滑与插值样条。”《应用数学与计算》, 218(5), 1831–1844. Kytariolou, A. and Themelis, N. (2022).“基于预测天气数据并考虑安全标准的船舶航线优化。”《航海杂志》, 75(6), 1310–1331. Maji, A. (2017).“使用路径规划器方法优化公路平面线形。” *WIT Trans Built Environ*, 176, 81–92. Mellinger, D. and Kumar, V. (2011).“四旋翼飞行器的最小snap轨迹生成与控制。” *2011*年*IEEE*国际机器人与自动化会议, IEEE, 2520–2525. Nadaraya, E. A. (1964).“论回归估计。”《概率论及其应用》, 9(1), 141–142. Pya, N. and Wood, S. N. (2015).“ 形状约束加性模型。”《统计与计算》, 25(3), 543–559. Qiu, C. (1995).“非参数回归与广义线性模型：粗糙度惩罚方法。” *Technometrics*, 37(3), 342–343. Racine, J. S., Parmeter, C. F., and Du, P. (2009).“约束非参数核回归：估计与推断。” 报告编号, Virginia Tech AAEC与麦克马斯特大学经济学系。工作论文。获取地址：https://www.tse-fr.eu/sites/default/files/medias/stories/SEMINAIRES_TSE_2008_09/ECONOMETRIE/racine.pdf。Ramsay, J. O. and Silverman, B. W. (2005). 函数型数据分析。Springer，第2版。Richter, C., Bry, A., and Roy, N. (2016).“面向密集室内环境中四旋翼飞行器激进飞行的多项式轨迹规划。”《机器人研究：第*16*届国际研讨会*ISRR*》, Springer, 649–666. Saeidian, J., Sarfraz, M., Azizi, A., and Jalilian, S. (2021).“基于三次Hermite样条的约束插值新方法。”《数学杂志》, 2021(1), 5925163. Seeger, M. (2004).“用于机器学习的高斯过程。”《国际神经系统杂志》, 14(02), 69–106. Shi, X. (2026).“用于降低变换回归学习偏差的数据锐化。未发表手稿。Stone, M. (1974).“统计预测的交叉验证选择与评估。”《英国皇家统计学会杂志：*B*辑（方法学）》, 36(2), 111–133. Sun, S., Egerstedt, M. B., and Martin, C. F. (2000).“控制理论平滑

样条。” *IEEE* 自动控制汇刊, 45(12), 2271–2279. Vázquez-Méndez, M. E. (2023). “道路立体交叉平面线形的自动设计：以喇叭形立体交叉为例。” 测量工程期刊, 149(2), 04023002. Wan, X., Pu, H., Schonfeld, P., Song, T., Li, W., and Peng, L. (2024). “基于语义拓扑地图的密集分布障碍物区域铁路线形优化。” 集成计算机辅助工程, 31(4), 421–437. Watson, G. S. (1964年). “平滑回归分析。” *Sankhy¯a:* 印度统计学杂志 *A*辑, 359–372. Wood, S. N. (2003年). “薄板回归样条。” 英国皇家统计学会杂志 *B*辑：统计方法学, 65(1), 95–114. Xiong, S. (2021年). “重构方法：从插值到回归。”*Technometrics*, 63(2), 225–235. Zhen, R., Gu, Q., Shi, Z., and Suo, Y. (2023年). “一种考虑海流、水深和船舶交通分隔规则的改进 A* 船舶路径规划算法。”海洋科学与工程杂志, 11(7), 1439. Zhou, Y., Gao, M., Chen, Y., and Shi, X. (2024年). “马尔可夫链的自适应惩罚似然方法。” *arXiv* 预印本 *arXiv:2406.00322*。