arXiv:2601.18080v1 [math.FA] 26 Jan 2026

多通道信号加残差分析中算子缺陷恒等式的应用：通过迭代乘积与伸缩能量残差——及其在机器学习核中的应用

Palle E.T. Jorgensen、Myung-Sin Song 和 James Tian

摘要。我们提出一种新的算子理论框架，用于分析具有内在组件细分的复杂系统；这些组件通常呈现为“残差”，尤其是“伸缩能量残差”。我们证明了新的结果，由此得到可容许性／有效性以及关于能量残差的新的先验界。应用包括针对 λ_n-松弛变体的无限维 Kaczmarz 理论，以及 λ_n-有效性。我们还将该框架应用于广义机器学习算法和贪心核主成分分析（核主成分分析），证明了明确的收敛结果、残差能量分解，以及噪声下稳定性的判据。

目录

1. 引言 1 2. 多通道缺陷分解 2 3. 核插值 8 4. 随机丢弃（随机抽样） 12 5. 基于树分裂的核重新参数化 16 6. 贪心细化与核压缩 19 7. 与贪心核主成分分析的联系 22 参考文献 27

1. 引言

近期关于递归算法的论文，建立在有限算子乘积的极限之上（其因子为投影或更一般的算子），并使用单个参数 λ 来量化遗憾估计。受应用需求的启发，我们在此转向更广泛的情形。具体而言，我们不再使用 λ 的单一取值，而是提出一个更广泛的框架，改为使用

2020 年数学主题分类：15、46、47、70、37、40、37、55、49、22 主要分类号：N；次要分类号：N、L、Q、E。关键词与短语。算子理论、再生核希尔伯特空间、格拉姆矩阵、机器学习、算法、核的重新参数化、优化、数据自适应方向、能量残差、系统理论、内在细分、无限维卡茨马尔兹、多通道树分裂、主成分分析、贪心细化与核压缩、稳定性、噪声、随机抽样、核插值。

一个数列 λ_n，其中每个数对应递归构造中的一个算子因子。我们还将这一新设定融入核分析框架，并将其应用于机器学习（ML）的新方向。此外，我们还为机器学习提出了两项新的应用。

本文的主要结果如下：定理2.4 给出了一个新的望远镜恒等式和一个可和性判据，两者均建立在新的算子理论框架中。定理3.1则基于核分析，证明了 λ_n 动力学的精确能量平衡，并给出了采样点处残差平方求和的显式误差界，从而保证其有效性。

本文结构如下。我们在第2节介绍多通道缺陷分解的算子理论框架，并建立算子乘积的基本望远镜恒等式。构成后续分析基础的精确残差能量分解、可和性判据和收敛结果也将在本节讨论。

我们在第2 节建立的框架，并在第3 节将其推广到再生核希尔伯特空间，讨论核插值及相关的在线学习动力学。我们证明了精确的望远镜式能量平衡，并据此给出明确的收敛保证和有效性判据，这些结果以与采样点相关的秩一投影表示。

第4 节将确定性框架扩展到随机情形，包括随机采样和丢弃机制。本文分析了噪声下的稳定性与鲁棒性，并在温和的步长条件下推导出定量的偏差-方差权衡。

随后，我们在第5节利用多通道分裂结构，构建核的基于树的重参数化。这带来了多尺度特征表示和截断核。近似误差由舍弃的残差能量精确度量，并在核和格拉姆矩阵两个层面给出紧确界。

基于这一框架，第6 节介绍了一种用于核近似与压缩的贪心细化策略。在此过程中，根据经验能量贡献自适应地选择核分量。因此，该方法能够单调收敛至原始核，并提供实用的停止准则。

最后，第7 节将所提出的框架与贪心核主成分 l
分析联系起来，阐明其影响及望远镜式残差框架的优势——精确的能量分解、收敛保证，以及 d
噪声稳定性。

2. 多通道缺陷分解

我们的目标是提出一个数学框架，它源于各种涉及复杂系统的应用场景。这些系统被划分为若干组成部分，本文将其称为“残差”，或不同能级之间的差异。在技术文献中，“伸缩能量残差”一词指的是工程学、数理物理、结构工程以及计算流体力学（CFD）等领域所用的一个概念。在这些不同的应用中，它通常描述一种方法：将大型系统拆分为若干组成部分，从而以算法式、通常是“望远镜式”的方式将“残差”（即能级之间的差异）累加起来。“残差”通常反映观测值（或实际值）与模型预测值之间的差异。具体而言，“能量残差”指系统或模拟中未被计入的剩余能量。

我们从一类广泛的迭代分裂算法共有的一种机制入手：与压缩算子相关的缺陷恒等式。该恒等式在每一步都会给出一个典范的双通道“信号加残差”分解；经过迭代后，它会产生一个望远镜式能量-残差公式。经典的 Kaczmarz 型方法，以及 λ-松弛投影方案，都自然地适用于这一压缩算子框架。在这些情形下，通常会施加可容许性/有效性假设，以确保残差通道渐近趋于零（或等价地，能量残差收敛到正确的极限）；例如，参见[KM01、JST20]，其中讨论了无限维 Kaczmarz 理论；以及 [JJKS25]，其中讨论了 λ-松弛变体和 λ-有效性。

对于调和分析、有效系统以及广义机器学习算法中无限算子乘积的相关特殊应用，我们谨提请读者关注以下近期论文 [BGHJ26，JST25，JT25b，JT25a，BHJW25，HJW23，JS23]。

2.1. **缺陷分解**。对于我们当前的设定，采用希尔伯特空间这一一般框架会比较方便，记作 H，其范数和内积在此均已指定。 H 中的有界线性算子记为 $B(H)$。不过，为了对本文将要讨论的算法进行估计，采用压缩算子的情形会更有用。在下文第 3 节中，我们将进一步专门讨论 H 取为再生核希尔伯特空间（RKHS）的情形。

固定一个压缩算子 $A \in B(H)$，并定义其缺陷算子

$$D_A = (I - A^*A)^{1/2} \in B(H)_+ .$$

考虑映射

$$W_A : H \to H \oplus H, \qquad W_A x = (Ax, D_A x) .$$

那么 W_A 是等距映射。事实上，

$$\begin{aligned} \|W_A x\|^2 &= \|Ax\|^2 + \|D_A x\|^2 \\ &= \langle x, A^*Ax\rangle + \langle x, (I - A^*A)\, x\rangle = \|x\|^2 . \end{aligned}$$

等价地，

$$\|x\|^2 = \|Ax\|^2 + \|D_A x\|^2 , \qquad x \in H,$$

而用算子形式表示，

$$I = A^*A + D_A^2.$$

迭代这一恒等式很直接，但这是看清标准“能量残差”及其相关残差通道由来的最简洁方式。

命题 2.1。设 $A_1, A_2, \ldots$ 是 H 上的压缩算子。令 $T_0 = I$ 和 $T_n = A_n A_{n-1} \cdots A_1$（$n \geq 1$ 时）。那么，对于任意 $x \in H$ 和任意 $N \geq 1$，

$$\|x\|^2 = \|T_N x\|^2 + \sum_{n=1}^{N} \|D_{A_n} T_{n-1} x\|^2 .$$

等价地，对于每个 $N \geq 1$，

$$I - T_N^* T_N = \sum_{n=1}^{N} T_{n-1}^* D_{A_n}^2 T_{n-1},$$

其中求和项有限，因此在 $B(H)$ 中定义明确。

证明。 应用 $\|y\|^2 = \|A_n y\|^2 + \|D_{A_n} y\|^2$ ，其中 $y = T_{n-1}x$，并从 $n = 1$ 求和至 N。极化给出算子恒等式。 □

定义 2.2 （有效性/可容许性）。 设 $A_1, A_2, \ldots$ 是作用于 H上的压缩算子，并且

$$T_n = A_n A_{n-1} \cdots A_1.$$

如果序列 $\{A_n\}$ 满足 $\{T_n\}$ 在强算子拓扑（$T_n \xrightarrow{s} T_\infty$）下收敛，即存在 $T_\infty \in B$（H），使得对每个 $x \in H$都有$T_n x \to T_\infty x$ ，则称该序列有效（或可容许）。

2.2. **Kaczmarz-λP方案**。 现在，我们将命题2.1中的一般压缩算子望远镜式公式专门应用于支撑 Kaczmarz 型算法的投影驱动情形。其基本构件是如下形式的压缩算子 $I - \lambda P$ 其中 P是一个正交投影，且 $0 < \lambda < 2$。对这类因子进行迭代可得到乘积$\widetilde{T}_n$ 其对向量的作用编码了算法，而相关的能量$\widetilde{\ }$ $\widetilde{\ }$残差 $T_n^* T_n$ 则满足由命题 2.1所决定的自动单调衰减。于是，核心问题在于这种能量控制是否能进一步确保乘积本身收敛，即是否具有定义 2.2所述的有效性。下一个例子给出基本的缺陷恒等式，以及在λ-松弛投影情形下得到的望远镜式公式。

例 2.3 ([JJKS25])。 设 $P = P^* = P^2$为一个正交投影，并固定$0 < \lambda < 2$。定义 $A = I - \lambda P$。则 A 是一个压缩算子，因为

$$A^* A = A^2 = (I - \lambda P)^2 = I - \lambda (2 - \lambda) P \le I.$$

因此，

$$D_A^2 = I - A^* A = \lambda (2 - \lambda) P, \qquad D_A = \sqrt{\lambda (2 - \lambda)} P,$$

且缺陷恒等式变为

$$\|x\|^2 = \|(I - \lambda P) x\|^2 + \lambda (2 - \lambda) \|Px\|^2 .$$

现在取一个正交投影序列 $P_1, P_2, \ldots$ 并构造 λ-松弛乘积

$$\widetilde{T}_0 = I, \qquad \widetilde{T}_n = (I - \lambda P_n) \cdots (I - \lambda P_1) .$$

采用相应的归一化

$$\widetilde{Q}_n := \lambda P_n \widetilde{T}_{n-1},$$

命题2.1给出望远镜恒等式

$$I - \widetilde{T}_N^* \widetilde{T}_N = \lambda (2 - \lambda) \sum_{n=1}^{N} \widetilde{T}_{n-1}^* P_n \widetilde{T}_{n-1} = \frac{2 - \lambda}{\lambda} \sum_{n=1}^{N} \widetilde{Q}_n^* \widetilde{Q}_n.$$

在这种收缩驱动的情形下，一个核心的收敛性问题是$\widetilde{\ }\lambda$-松弛乘积 T_n 是否按照定义 2.2 的意义是有效的。相比之下，$\widetilde{\ }$ $\widetilde{\ }$相关的能量残差 $T_n^* T_n$ 总是强收敛：由于每个因子$I - \lambda P_n$ 都是压缩算子，因此有

$$0 \le \widetilde{T}_{n+1}^* \widetilde{T}_{n+1} = \widetilde{T}_n^* (I - \lambda P_{n+1})^* (I - \lambda P_{n+1}) \widetilde{T}_n \le \widetilde{T}_n^* \widetilde{T}_n,$$

~ ~因此 $T_n^* T_n \downarrow S_\infty$ 对于某个 $S_\infty \in B(H)_+$。上述望远镜恒等式通过确定每一步的耗散，对这种单调性作了进一步刻画：

$$\|x\|^2 - \|\widetilde{T}_N x\|^2 = \frac{2-\lambda}{\lambda} \sum_{n=1}^{N} \|\widetilde{Q}_n x\|^2.$$

被引用文献中的可容许性/有效性假设，是对投影序列（以及 λ）施加的额外结构条件，可将这种能量控制提升为对乘积本身的强收敛结论；参见~ T_n [KM01、JST20、JJKS25]。

引入松弛调度 $\{\lambda_n\}$ 会得到变步长 $\lambda_n P_n$ 方案；下一个定理给出相应的望远镜恒等式及一个保证有效性的可和性判据。

2.3. **望远镜式、可和性判据与有效性**。

定理 2.4。设 P_1、P_2、.....为 H上的正交投影，并设 $0 < \lambda_n < 2$为一个序列。令

$$A_n := I - \lambda_n P_n, \qquad T_0 = I, \qquad T_n := A_n A_{n-1} \cdots A_1.$$

则每个 A_n 都是压缩算子，并且有如下望远镜恒等式

$$I - T_N^* T_N = \sum_{n=1}^{N} \lambda_n \left(2 - \lambda_n\right) T_{n-1}^* P_n T_{n-1} = \sum_{n=1}^{N} Q_n^* Q_n, \tag{2.1}$$

其中

$$Q_n := \sqrt{\lambda_n \left(2 - \lambda_n\right)}\, P_n T_{n-1}.$$

特别地，对于任意 $x \in H$ 和任意 $N \geq 1$,

$$\|x\|^2 - \|T_N x\|^2 = \sum_{n=1}^{N} \lambda_n \left(2 - \lambda_n\right) \|P_n T_{n-1} x\|^2 = \sum_{n=1}^{N} \|Q_n x\|^2 . \tag{2.2}$$

此外，假设

$$\sum_{n=1}^{\infty} \frac{\lambda_n}{2 - \lambda_n} < \infty. \tag{2.3}$$

则乘积 $\{T_n\}$ 按照定义 *2.2* 的意义是有效的，即 $T_n \xrightarrow{s} T_\infty$对于某个 $T_\infty \in B(H)$。

证明。由于 $P_n^2 = P_n$和 $P_n^* = P_n$，因此有

$$A_n^* A_n = A_n^2 = \left(I - \lambda_n P_n\right)^2 = I - \lambda_n \left(2 - \lambda_n\right) P_n \leq I,$$

因此 A_n 是一个压缩算子，且 $D_{A_n}^2 = I - A_n^* A_n = \lambda_n(2 - \lambda_n)P_n$。应用命题 2.1 以及这些 A_n 可得（2.1），从而得到（2.2）。

为确保有效性，固定 $x \in H$，并写出步差

$$T_n x - T_{n-1} x = \left(A_n - I\right) T_{n-1} x = -\lambda_n P_n T_{n-1} x.$$

根据柯西-施瓦茨不等式，

$$\begin{aligned}\sum_{n=1}^{\infty}\|T_nx-T_{n-1}x\| &= \sum_{n=1}^{\infty}\lambda_n\|P_nT_{n-1}x\| \\ &= \sum_{n=1}^{\infty}\sqrt{\frac{\lambda_n}{2-\lambda_n}}\,\sqrt{\lambda_n(2-\lambda_n)}\,\|P_nT_{n-1}x\| \\ &\le \left(\sum_{n=1}^{\infty}\frac{\lambda_n}{2-\lambda_n}\right)^{1/2}\left(\sum_{n=1}^{\infty}\lambda_n(2-\lambda_n)\|P_nT_{n-1}x\|^2\right)^{1/2}.\end{aligned}$$

第二个级数由（2.2）界定：

$$\sum_{n=1}^{\infty}\lambda_n(2-\lambda_n)\|P_nT_{n-1}x\|^2 \le \|x\|^2.$$

因此（2.3）推出 $\sum_n\|T_nx-T_{n-1}x\|<\infty$，所以 $\{T_nx\}$ 在 H中是柯西序列。由于 x 是任意的，$T_n \xrightarrow{s} T_\infty$。□

可求和条件（2.3）是对松弛调度 $\{\lambda_n\}$的一维约束；它为有效性提供了一个与投影无关的充分条件，而更精细的判据必然取决于投影的几何性质。

注记（变量松弛为何重要）。在卡茨马尔兹更新的学习式应用中，松弛参数充当步长。具体而言， [JJKS25] 中的分析 λ给出了遗憾界对 λ 的显式依赖，将“学习项”与“噪声底”区分开来，并促使人们在整个迭代过程中对 λ_n 进行噪声感知调优。因此，允许 λ_n 变化，便能适应标准的步长调度（例如，前期采取较激进的更新，随后逐渐减弱），而（2.3）则给出了一个简单且与投影无关的充分条件，以确保此类调度的有效性。

2.4. 多通道树分裂。设 $A_1,\dots,A_d\in B(H)$ 为有界算子，满足

$$\sum_{i=1}^{d}A_i^*A_i \le I.$$

等价地，$(A_1,\dots,A_d)$ 是列压缩。令

$$D := \left(I-\sum_{i=1}^{d}A_i^*A_i\right)^{1/2}.$$

基本观察是，这给出了范数的一种多通道分裂。

引理 2.5。对于每个 $x\in H$,

$$\|x\|^2 = \sum_{i=1}^{d}\|A_ix\|^2+\|Dx\|^2.$$

证明。我们有

$$\sum_{i=1}^{d}\|A_ix\|^2+\|Dx\|^2 = \left\langle x,\left(\sum\nolimits_{i=1}^{d}A_i^*A_i+D^2\right)x\right\rangle = \langle x,Ix\rangle = \|x\|^2.$$

□

现在，我们沿着一棵以 d 为分支数的根树迭代这一恒等式。

定义 2.6。令 $\mathcal{T}$ 为一棵以根节点为根的 d 叉树。我们用一个词标记每个节点

$$\alpha = i_1 \cdots i_n, \qquad i_k \in \{1, \dots, d\},$$

其根节点为 $\emptyset$（空词），深度为 $|\alpha| = n$。对于一个词 $\alpha = i_1 \cdots i_n$，我们定义

$$A_\alpha := A_{i_n} \cdots A_{i_1}, \qquad A_\emptyset := I.$$

对于 $x \in H$ 和一个节点 α，我们设

$$E_\alpha(x) := \|A_\alpha x\|^2, \qquad \Delta_\alpha(x) := \|DA_\alpha x\|^2.$$

对于每一层 $N \geq 0$，我们定义层能量

$$L_N(x) := \sum_{|\alpha|=N} E_\alpha(x) = \sum_{|\alpha|=N} \|A_\alpha x\|^2.$$

采用这一记号，多通道分裂给出了沿树的细化恒等式。

命题 2.7。对于每个 $x \in H$ 和每个整数 $N \geq 1$,

$$\|x\|^2 = \sum_{|\alpha|=N} \|A_\alpha x\|^2 + \sum_{n=0}^{N-1} \sum_{|\alpha|=n} \|DA_\alpha x\|^2. \tag{2.4}$$

证明。固定一个节点 α。将引理 2.5 应用于向量 $A_\alpha x$:

$$\|A_\alpha x\|^2 = \sum_{i=1}^{d} \|A_i A_\alpha x\|^2 + \|DA_\alpha x\|^2 = \sum_{i=1}^{d} \|A_{\alpha i} x\|^2 + \Delta_\alpha(x),$$

其中 αi 是通过追加 i 得到的 α 的子节点。现在对深度为 n 的所有节点求和。左侧变为 $L_n(x)$，右侧的第一项变为$L_{n+1}(x)$。因此

$$L_n(x) = L_{n+1}(x) + \sum_{|\alpha|=n} \Delta_\alpha(x)$$

对每个 $n \geq 0$ 都成立。将此恒等式从 $n = 0$ 求和至 $n = N - 1$，得到

$$L_0(x) = L_N(x) + \sum_{n=0}^{N-1} \sum_{|\alpha|=n} \Delta_\alpha(x).$$

由于 $L_0(x) = \|x\|^2$，命题得证。 □

下面将用到几个直接的推论。

推论 2.8（单调性和极限）。对于每个 $x \in H$，序列 $(L_N(x))_{N\geq 0}$都是递减的，并且极限

$$L_\infty(x) := \lim_{N\to\infty} L_N(x)$$

存在于 $\left[0, \|x\|^2\right]$ 中。此外，

$$\sum_{n\geq 0} \sum_{|\alpha|=n} \|DA_\alpha x\|^2 = \|x\|^2 - L_\infty(x).$$

证明。 根据命题 2.7的证明，我们有

$$L_{N+1}(x) = L_N(x) - \sum_{|\alpha|=N} \Delta_\alpha(x) \le L_N(x),$$

因此 $(L_N(x))$ 单调递减且以下界 0有界，因此收敛。在（ $N \to\infty$2.4）中代入即可得到最后一个结论。 □

推论 2.9 (列等距情形)。 若 $\sum_{i=1}^d A_i^* A_i = I$，则 $D = 0$，且对每个 $x \in H$ 和每个 $N \ge 0$,

$$\|x\|^2 = \sum_{|\alpha|=N} \|A_\alpha x\|^2 .$$

特别地，对所有 N， $L_N(x) = \|x\|^2$，且 $L_\infty(x) = \|x\|^2$。

证明。 若 $\sum A_i^* A_i = I$则 $D = 0$，因此所有 $\Delta_\alpha(x)$ 均为零，且 (2.4) 化简为所述恒等式。 □

推论 2.10。 假设存在常数 $c \in (0,\ 1)$ 使得

$$\sum_{i=1}^d A_i^* A_i \le cI.$$

那么对任意 $x \in H$ 和任意 $N \ge 0$,

$$\sum_{|\alpha|=N} \|A_\alpha x\|^2 \le c^N \|x\|^2 .$$

特别地 $L_\infty(x) = 0$，且

$$\sum_{n\ge 0} \sum_{|\alpha|=n} \|DA_\alpha x\|^2 = \|x\|^2 .$$

证明。 该假设意味着对每个 $y \in H$，都有 $\sum_{i=1}^d \|A_i y\|^2 \le c\|y\|^2$。固定 N，令 $y = A_\alpha x$ 并对所有长度为 N 的词 α 求和。于是

$$L_{N+1}(x) = \sum_{|\beta|=N+1} \|A_\beta x\|^2 = \sum_{|\alpha|=N} \sum_{i=1}^d \|A_i A_\alpha x\|^2 \le c \sum_{|\alpha|=N} \|A_\alpha x\|^2 = cL_N(x).$$

由归纳法可得 $L_N(x) \le c^N L_0(x) = c^N \|x\|^2$。其余结论由推论 2.8 得出。 □

3. 核插值

下一节将讨论核及其再生核希尔伯特空间（RKHSs）。为方便读者，现列出以下参考文献[JSY25， AJ21， AJV14， AJ22， PR16， FHR+23， Ste23， JP19]。

现在，我们将定理 $\lambda_n P_n$ 2.4 中的乘积 置于再生核希尔伯特空间（RKHS）的框架下，其中投影由点求值导出。设 X 为一个集合，设 H_k 为定义在 X 上的复值函数再生核希尔伯特空间，其再生核为 k。对于 $x \in X$，记 $k_x := k(\cdot,\ x) \in H_k$，于是

$$f(x) = \langle k_x, f\rangle, \qquad f \in H_k. \tag{3.1}$$

固定一个序列 $(x_n)_{n\geq 1}\subset X$，使得 $k(x_n,\ x_n)>0$ 对所有 n 均成立，并令

$$g_n:=\frac{k_{x_n}}{\|k_{x_n}\|}=\frac{k_{x_n}}{\sqrt{k(x_n,x_n)}}\in H_k.$$

令 $P_n=|g_n\rangle\langle g_n|$ 表示到 $\mathbb{C}g_n$上的正交投影，即

$$P_nf=\langle g_n,f\rangle g_n,\qquad f\in H_k.$$

在 x_n 处求值并利用再生性质（3.1），可知该投影

$$P_nf=\frac{f(x_n)}{k(x_n,x_n)}k_{x_n}.$$

现在固定 $0<\lambda_n<2$，令 $A_n:=I-\lambda_nP_n$，并令 $T_n=A_nA_{n-1}\cdots A_1$如定理 2.4 所述。由于 P_n 是正交投影，每个 A_n 都是压缩算子，因此（2.2）在 H_k中化为显式求值公式：

$$\begin{aligned}\|f\|^2-\|T_Nf\|^2&=\sum_{n=1}^{N}\lambda_n(2-\lambda_n)\|P_nT_{n-1}f\|^2\\&=\sum_{n=1}^{N}\lambda_n(2-\lambda_n)\frac{|(T_{n-1}f)(x_n)|^2}{k(x_n,x_n)}.\end{aligned}\tag{3.2}$$

特别地，由于 $N\to\infty$ 右侧的部分和单调递增且有 $\|f\|^2$作为上界，因此

$$\sum_{n=1}^{\infty}\lambda_n(2-\lambda_n)\frac{|(T_{n-1}f)(x_n)|^2}{k(x_n,x_n)}\leq\|f\|^2,\qquad f\in H_k.$$

为了将其与插值联系起来，我们转而采用通常的仿射更新。设 $(y_n)_{n\geq 1}\subset\mathbb{C}$ 为给定数据，并假设存在 $f^*\in H_k$ 使得

$$f^*(x_n)=y_n,\qquad n\geq 1.$$

从 $f_0=0$ 开始，递归定义 $(f_n)_{n\geq 0}\subset H_k$：

$$f_n=f_{n-1}+\lambda_n(y_n-f_{n-1}(x_n))\frac{k_{x_n}}{k(x_n,x_n)},\qquad n\geq 1.\tag{3.3}$$

令 $e_n:=f^*-f_n$。简单计算表明，误差也由相同的λ_nP_n 因子演化：

$$e_n=e_{n-1}-\lambda_nP_ne_{n-1}=(I-\lambda_nP_n)e_{n-1}=A_ne_{n-1},$$

因此 $e_n=T_ne_0=T_nf^*$。将 $f=f^*$代入（3.2）得到一个恒等式，其中耗散项恰好是插值残差。

定理 3.1。假设插值约束 $f^*(x_n)=y_n$存在解$f^*\in H_k$，并令 (f_n) 由 (3.3) 定义。则对于每个 $N\geq 1$，

$$\|f^*-f_N\|^2=\|f^*\|^2-\sum_{n=1}^{N}\lambda_n(2-\lambda_n)\frac{|y_n-f_{n-1}(x_n)|^2}{k(x_n,x_n)}.\tag{3.4}$$

特别是，

$$\sum_{n=1}^{\infty}\lambda_n(2-\lambda_n)\frac{|y_n-f_{n-1}(x_n)|^2}{k(x_n,x_n)}\leq\|f^*\|^2.$$

证明。 我们有 $e_n = f^* - f_n = T_n f^*$。此外，

$$(T_{n-1}f^*)(x_n) = \langle k_{x_n}, T_{n-1}f^* \rangle = \langle k_{x_n}, e_{n-1} \rangle = y_n - f_{n-1}(x_n),$$

由于 $e_{n-1} = f^* - f_{n-1}$且 $f^*(x_n) = y_n$。将 $f = f^*$代入（3.2）得到（3.4）。□

备注 3.2。恒等式（3.4）为 $\lambda_n P_n$ 动力学给出了精确的能量平衡：误差范数通过对采样点处的残差平方求和而下降，权重为 $\lambda_n(2-\lambda_n)/k(x_n,x_n)$。要实现收敛，此外还需要满足定义 2.2 意义下的有效性假设。如果 $T_n \xrightarrow{s} T_\infty$，则 $e_n = T_n f^* \to T_\infty f^*$在 H_k中成立，从而有 $f_n \to f^* - T_\infty f^*$。特别地，如果 $T_\infty f^* = 0$ （例如，如果 $T_n \xrightarrow{s} 0$），则 f_n 在 H_k 中收敛到插值函数f^*。根据定理 2.4，可求和条件（2.3）是一个与投影无关的充分假设，可确保 $\{T_n\}$的有效性；而在当前核设置下，更精确的判据取决于核截面 $\{k_{x_n}\}$的几何结构。

接下来我们考察加性噪声的影响。更新的几何结构不变，但精确恒等式（3.4）被期望误差的分解所取代，该分解由耗散部分和噪声项组成。

现在假设数据具有如下形式

$$y_n = f^*(x_n) + \varepsilon_n,$$

其中 ε_n 是复值随机变量。我们采用通常的平方可积、零均值假设

$$\mathbb{E}(\varepsilon_n \mid x_n) = 0, \qquad \mathbb{E}\left(|\varepsilon_n|^2 \mid x_n\right) \le \sigma^2 \tag{3.5}$$

对于某个 $\sigma \ge 0$，并假设噪声 $\{\varepsilon_n\}$ 在 n之间相互独立。松弛更新（3.3）保持不变，并继续记 $e_n = f^* - f_n$。

命题 3.3. 设 $\{f_n\}$ 由 (3.3) 定义，其中 $y_n = f^*(x_n) + \varepsilon_n$，其中$f^* \in H_k$，且噪声 $\{\varepsilon_n\}$满足 (3.5)。令 $e_n = f^* - f_n$。那么对于每个$N \ge 1$,

$$\mathbb{E}\|e_N\|^2 = \|f^*\|^2 - \sum_{n=1}^{N} \lambda_n(2-\lambda_n)\,\mathbb{E}\|P_n e_{n-1}\|^2 + \sum_{n=1}^{N} \lambda_n^2 \mathbb{E}\left(\frac{|\varepsilon_n|^2}{k(x_n,x_n)}\right). \tag{3.6}$$

证明。 在步骤 n 有

$$y_n - f_{n-1}(x_n) = (f^*(x_n) - f_{n-1}(x_n)) + \varepsilon_n = e_{n-1}(x_n) + \varepsilon_n,$$

因此（3.3）可得

$$\begin{aligned} e_n &= e_{n-1} - \lambda_n (e_{n-1}(x_n) + \varepsilon_n) \frac{k_{x_n}}{k(x_n,x_n)} \\ &= (I - \lambda_n P_n) e_{n-1} - \lambda_n \varepsilon_n \frac{k_{x_n}}{k(x_n,x_n)}. \end{aligned}$$

引入

$$u_n := (I - \lambda_n P_n) e_{n-1}, \qquad v_n := \lambda_n \varepsilon_n \frac{k_{x_n}}{k(x_n,x_n)},$$

因此 $e_n = u_n - v_n$。与确定性情形一样，

$$\|u_n\|^2 = \|e_{n-1}\|^2 - \lambda_n(2-\lambda_n)\|P_n e_{n-1}\|^2,$$

and

$$\|v_n\|^2 = \lambda_n^2 |\varepsilon_n|^2 \frac{\|k_{x_n}\|^2}{k(x_n,x_n)^2} = \lambda_n^2 \frac{|\varepsilon_n|^2}{k(x_n,x_n)}.$$

我们展开

$$\|e_n\|^2 = \|u_n - v_n\|^2 = \|u_n\|^2 + \|v_n\|^2 - \langle u_n, v_n\rangle - \langle v_n, u_n\rangle .$$

在 e_{n-1} 和 x_n的条件下，向量 u_n 是固定的，而 v_n 关于 ε_n是线性的。由（3.5），

$$\mathbb{E}(\varepsilon_n \mid e_{n-1}, x_n) = \mathbb{E}(\varepsilon_n \mid x_n) = 0,$$

类似地，$\mathbb{E}(\varepsilon_n \mid e_{n-1}, x_n) = 0$，因此两个混合项的条件均值都为零。由此取条件期望可得

$$\mathbb{E}\left(\|e_n\|^2 \mid e_{n-1}, x_n\right) = \|e_{n-1}\|^2 - \lambda_n(2-\lambda_n)\|P_n e_{n-1}\|^2 + \lambda_n^2 \mathbb{E}\left(\frac{|\varepsilon_n|^2}{k(x_n,x_n)} \mid x_n\right).$$

将这些关系对 $n = 1, \ldots, N$ 求和，并取完全期望，可得

$$\mathbb{E}\|e_N\|^2 = \mathbb{E}\|e_0\|^2 - \sum_{n=1}^{N} \lambda_n(2-\lambda_n)\mathbb{E}\|P_n e_{n-1}\|^2 + \sum_{n=1}^{N} \lambda_n^2 \mathbb{E}\left(\frac{|\varepsilon_n|^2}{k(x_n,x_n)}\right).$$

由于 $e_0 = f^*$，这正是（3.6）。 □

（3.6）中的第一项求和与（3.4）中的耗散项相同，只是现在对数据中的随机性取了平均；而第二项求和则衡量累积的噪声。对核对角元作一个简单的统一下界，即可将（3.6）转化为期望误差的界。

推论 3.4。 除命题 *3.3*的假设外，设存在常数 $\kappa > 0$ 使得

$$k(x_n,x_n) \geq \kappa, \qquad n \geq 1. \tag{3.7}$$

那么，对于每个 $N \geq 1$,

$$\mathbb{E}\|e_N\|^2 \leq \|f^*\|^2 + \frac{\sigma^2}{\kappa} \sum_{n=1}^{N} \lambda_n^2. \tag{3.8}$$

特别地，若 $\sum_{n=1}^{\infty} \lambda_n^2 < \infty$，则 上确界$_{N\geq 1}\, \mathbb{E}\|e_N\|^2 < \infty$。

证明。 从（3.6）出发，舍去耗散项可得

$$\mathbb{E}\|e_N\|^2 \leq \|f^*\|^2 + \sum_{n=1}^{N} \lambda_n^2 \mathbb{E}\left(\frac{|\varepsilon_n|^2}{k(x_n,x_n)}\right).$$

由（3.5）和（3.7），

$$\mathbb{E}\left(\frac{|\varepsilon_n|^2}{k(x_n,x_n)}\right) \leq \frac{1}{\kappa}\mathbb{E}|\varepsilon_n|^2 \leq \frac{\sigma^2}{\kappa},$$

因此

$$\sum_{n=1}^{N} \lambda_n^2 \mathbb{E}\left(\frac{|\varepsilon_n|^2}{k(x_n,x_n)}\right) \leq \frac{\sigma^2}{\kappa} \sum_{n=1}^{N} \lambda_n^2.$$

由此得到（3.8）。最后一个结论可由假设立即推出$\sum_n \lambda_n^2 < \infty$。□

4. 随机丢弃（随机抽样）

在许多应用中，更新并不会在每一步都执行：样本可能缺失，传感器可能无法传输，或者人们可能有意跳过更新，将其作为一种正则化方式。一种简单的建模方法是引入一个伯努利随机变量，随机开启或关闭更新。关键在于，基本能量恒等式沿每条样本路径仍然成立，因此只需稍加处理，便可得到简洁的超鞅和稳定性结论。

设 H 是一个希尔伯特空间，并设 $\{P_n\}_{n\geq 1}$ 是 H 上一列正交投影。取 $\lambda \in (0,\ 2)$，并设 $\{\eta_n\}_{n\geq 1}$ 为 $\{0,1\}$-值随机变量。我们考虑间歇松弛

$$x_n = x_{n-1} - \lambda \eta_n P_n x_{n-1}, \qquad n \geq 1.$$

等价地，对于 $A_n = I - \lambda \eta_n P_n$，我们有 $x_n = A_n x_{n-1}$，对所有 $n \geq 1$ 均成立。

推论 4.1. 对于 $\{\eta_n\}$ 的每一种实现以及每个 $x \in H$，

$$\|A_n x\|^2 = \|x\|^2 - \eta_n \lambda (2-\lambda) \|P_n x\|^2 .$$

特别地，$\|x_n\|$ 单调不增，且

$$\|x_0\|^2 - \|x_N\|^2 = \sum_{n=1}^{N} \eta_n \lambda (2-\lambda) \|P_n x_{n-1}\|^2$$

对于每个 $N \geq 1$ 均成立。

证明. 这直接由例 2.3 得出，因为 $\eta_n \in \{0,1\}$ 且

$$(I - \lambda \eta_n P_n)^2 = I - \eta_n \lambda (2-\lambda) P_n.$$

□

这些恒等式是确定性的，也就是说，它们对每条样本路径都成立。为了获得收敛信息，将迭代过程置于一个滤过中会更方便。设 $\{\mathcal{F}_n\}_{n\geq 0}$ 是一个滤过，使得 x_{n-1} 是 $\mathcal{F}_{n-1}$-可测的，且 (P_n, η_n) 是 $\mathcal{F}_n$-可测的。令

$$p_n = \mathbb{E}\left[\eta_n \mid \mathcal{F}_{n-1}\right].$$

命题 4.2。沿用上述记号，

$$\mathbb{E}\left[\|x_n\|^2 \mid \mathcal{F}_{n-1}\right] = \|x_{n-1}\|^2 - \lambda (2-\lambda)\, \mathbb{E}\left[\eta_n \|P_n x_{n-1}\|^2 \mid \mathcal{F}_{n-1}\right].$$

特别地，$\left\{\|x_n\|^2\right\}$ 是一个非负超鞅，并且

$$\sum_{n\geq 1} \eta_n \lambda (2-\lambda) \|P_n x_{n-1}\|^2 < \infty$$

几乎必然成立。

证明。对第一个推论中的逐路径恒等式取条件期望，并利用 x_{n-1} 是 $\mathcal{F}_{n-1}$-可测的这一事实。超鞅性质显而易见。由望远镜恒等式以及 $\|x_n\|$ 的有界性可得几乎必然可求和。□

为确定极限点，我们对投影序列施加一个几何假设。我们固定一个确定性的公共不动点子空间：

假设 4.3. 存在一个固定的闭子空间 $M \subseteq H$，使得

$$P_n m = 0 \qquad for\ all\ m \in M \tag{4.1}$$

几乎必然地，对每个 $n \geq 1$均成立。

我们用 P_M 表示到 M 上的正交投影，用 $M^{\perp}$ 表示其正交补。

定理 4.4。 假设 *4.3*。假设存在 $\beta > 0$ 使得对每个 $x \in M^{\perp}$,

$$\langle \mathbb{E}\left[\eta_n P_n \mid \mathcal{F}_{n-1}\right] x, x\rangle \geq \beta \left\|x\right\|^2$$

对所有 $n \geq 1$几乎必然成立。则对每个初始值 $x_0 \in H$,

$$\mathbb{E}\left[\left\|x_n - P_M x_0\right\|^2\right] \leq \left(1 - \beta\lambda\left(2-\lambda\right)\right)^n \left\|x_0 - P_M x_0\right\|^2,$$

因此 $x_n \to P_M x_0$ 属于 L^2。此外,

$$x_n \to P_M x_0 \qquad almost\ surely.$$

证明。 记 $e_n = x_n - P_M x_0$。由于对所有 $m \in M$， $P_M x_0 \in M$且 $P_n m = 0$，我们有 $P_n e_{n-1} = P_n x_{n-1}$。此外，对所有 n， $e_n \in M^{\perp}$，因为 $x_n - P_M x_0$与 x_n 相差一个属于 M 的元素，并且每次更新都会减去一个属于 $\mathrm{ran} P_n \subseteq M^{\perp}$的向量。事实上，若 $y = P_n z$ 且 $m \in M$，则

$$\langle y, m\rangle = \langle P_n z, m\rangle = \langle z, P_n m\rangle = 0,$$

因此 $\mathrm{ran} P_n \subseteq M^{\perp}$。

条件期望恒等式给出

$$\mathbb{E}\left[\left\|e_n\right\|^2 \mid \mathcal{F}_{n-1}\right] = \left\|e_{n-1}\right\|^2 - \lambda\left(2-\lambda\right)\mathbb{E}\left[\eta_n \left\|P_n e_{n-1}\right\|^2 \mid \mathcal{F}_{n-1}\right].$$

由于 $\left\|P_n e_{n-1}\right\|^2 = \langle P_n e_{n-1},\ e_{n-1}\rangle$ 且 e_{n-1} 是 $\mathcal{F}_{n-1}$-可测的，我们有

$$\mathbb{E}\left[\eta_n \left\|P_n e_{n-1}\right\|^2 \mid \mathcal{F}_{n-1}\right] = \langle \mathbb{E}\left[\eta_n P_n \mid \mathcal{F}_{n-1}\right] e_{n-1}, e_{n-1}\rangle.$$

利用强制性假设可得

$$\mathbb{E}\left[\left\|e_n\right\|^2 \mid \mathcal{F}_{n-1}\right] \leq \left(1 - \beta\lambda\left(2-\lambda\right)\right)\left\|e_{n-1}\right\|^2.$$

取期望并迭代可得

$$\mathbb{E}\left\|e_n\right\|^2 \leq \left(1 - \beta\lambda\left(2-\lambda\right)\right)^n \left\|e_0\right\|^2,$$

这就是所述的 L^2 界。

对于几乎必然收敛，令 $c = \beta\lambda(2-\lambda) \in (0,\ 1)$。上述不等式表明， $\left\{(1-c)^{-n}\left\|e_n\right\|^2\right\}$ 是一个非负超鞅，因此它几乎必然收敛到一个有限的随机极限。特别地， $\left\|e_n\right\|^2$ 几乎必然收敛。另一方面,

$$\mathbb{E}\left\|e_n\right\|^2 \leq (1-c)^n \left\|e_0\right\|^2$$

因此 $\sum_{n\geq 1}\mathbb{E}\left\|e_n\right\|^2 < \infty$。根据马尔可夫不等式和博雷尔-坎特利引理, $\left\|e_n\right\| \to 0$ 几乎必然成立。因此 $x_n \to P_M x_0$ 几乎必然成立。□

注释 4.5。 一种常见情形是，给定 $\mathcal{F}_{n-1}$ 时， η_n 和 P_n 条件独立，且 $p_n = \mathbb{E}[\eta_n \mid \mathcal{F}_{n-1}] \geq p > 0$ 几乎必然成立，同时存在 $\alpha > 0$，使得

$$\langle \mathbb{E}\left[P_n \mid \mathcal{F}_{n-1}\right] x, x\rangle \geq \alpha \left\|x\right\|^2, \qquad x \in M^{\perp}.$$

Then

$$\mathbb{E}\left[\eta_n P_n \mid \mathcal{F}_{n-1}\right] = \mathbb{E}\left[\eta_n \mid \mathcal{F}_{n-1}\right] \mathbb{E}\left[P_n \mid \mathcal{F}_{n-1}\right]$$

并且上述假设意味着

$$\langle \mathbb{E}\left[\eta_n P_n \mid \mathcal{F}_{n-1}\right] x, x\rangle \geq p\alpha \left\|x\right\|^2, \qquad x \in M^{\perp}.$$

因此定理 4.4 适用于 $\beta = p\alpha$。

备注 4.6。 间歇更新不会改变基本的耗散机制；它们通过条件平均算子 $\mathbb{E}\left[\eta_n P_n \mid \mathcal{F}_{n-1}\right]$ 在 $M^{\perp}$上的作用来调整有效收缩率。特别地，定理 4.4 表明，对这些条件平均算子的一致强制性界可导致几何 L^2 收敛，并且几乎必然收敛到 $P_M x_0$。一个更接近循环 Kaczmarz 精神的确定性类比，是假设族 $\{P_n\}$满足窗口下帧界。此处暂不讨论这种情形。

记录一个将丢弃与可预测松弛调度相结合的变步长版本会很方便。

定理 4.7。 令 $\{\lambda_n\}_{n\geq 1}$ 为可预测序列，即对于每个 n， λ_n 都是 $\mathcal{F}_{n-1}$-可测的，并假设 $0 < \lambda_n < 2$ 几乎处处成立。考虑间歇更新

$$x_n = x_{n-1} - \lambda_n \eta_n P_n x_{n-1}, \qquad n \geq 1.$$

则逐路径恒等式成立：

$$\left\|x_n\right\|^2 = \left\|x_{n-1}\right\|^2 - \eta_n \lambda_n \left(2 - \lambda_n\right) \left\|P_n x_{n-1}\right\|^2.$$

此外，假设存在 $\gamma > 0$，使得对于每个 $x \in M^{\perp}$，

$$\langle \mathbb{E}\left[\eta_n \lambda_n \left(2 - \lambda_n\right) P_n \mid \mathcal{F}_{n-1}\right] x, x\rangle \geq \gamma \left\|x\right\|^2$$

对所有 n 几乎处处成立。则对于每个 $x_0 \in H$，

$$\mathbb{E}\left[\left\|x_n - P_M x_0\right\|^2\right] \leq (1-\gamma)^n \left\|x_0 - P_M x_0\right\|^2,$$

并且 $x_n \to P_M x_0$ 几乎处处成立。

证明。 逐路径恒等式可通过应用例2.3并取步长$\lambda_n\eta_n$得到。对于收缩估计，令 $e_n = x_n - P_M x_0$。如上所述，

$$\mathbb{E}\left[\left\|e_n\right\|^2 \mid \mathcal{F}_{n-1}\right] = \left\|e_{n-1}\right\|^2 - \mathbb{E}\left[\eta_n \lambda_n \left(2 - \lambda_n\right) \left\|P_n e_{n-1}\right\|^2 \mid \mathcal{F}_{n-1}\right],$$

最后一项等于

$$\langle \mathbb{E}\left[\eta_n \lambda_n \left(2 - \lambda_n\right) P_n \mid \mathcal{F}_{n-1}\right] e_{n-1}, e_{n-1}\rangle.$$

利用强制性假设可得

$$\mathbb{E}\left[\left\|e_n\right\|^2 \mid \mathcal{F}_{n-1}\right] \leq (1-\gamma) \left\|e_{n-1}\right\|^2.$$

迭代可得 L^2 界。几乎必然收敛的证明与定理 4.4相同。

□

注 4.8。 在标量再生核希尔伯特空间 Kaczmarz 设置中（见第 3 节），取一个再生核希尔伯特空间 H_k，其核截面为 $k_x \in H_k$ 并令

$$g_n = \frac{k_{x_n}}{\|k_{x_n}\|}, \qquad P_n = |g_n\rangle \langle g_n| .$$

给定数据 y_n，定义间歇更新

$$f_{n+1} = f_n + \lambda \eta_n \frac{y_n - f_n(x_n)}{\|k_{x_n}\|^2} k_{x_n}.$$

若 f^* 是目标函数，记作 $e_n = f_n - f^*$，那么在无噪声插值情形下 $y_n = f^*(x_n)$，有

$$e_{n+1} = (I - \lambda \eta_n P_n)\, e_n,$$

因此，同一逐路径恒等式成立：

$$\|e_{n+1}\|_{H_k}^2 = \|e_n\|_{H_k}^2 - \eta_n \lambda (2-\lambda) \, |\langle g_n, e_n \rangle|^2 .$$

在含噪模型 $y_n = f^*(x_n) + \xi_n$ 中，丢弃因子 η_n 同时乘以相应条件期望递推中的耗散项和噪声注入项，因此，间歇更新以一种直观明了的方式，按照通常的早停图景缩放偏差与方差贡献。

推论 4.9。 设 H_k 是可分度量空间 X上的再生核希尔伯特空间，并假设核 k 连续，因此每个 $f \in H_k$ 都是连续的。设 ρ是 X上的博雷尔概率测度，并假设$(x_n)_{n\geq 1}$ 服从 ρ 分布且独立同分布。假设 η_n 是独立同分布的 $\{0,1\}$值随机变量，满足 $\mathbb{P}(\eta_n = 1) = p \in (0,1)$，且与$(x_n)$独立。定义

$$g_n = \frac{k_{x_n}}{\|k_{x_n}\|}, \qquad P_n = |g_n\rangle \langle g_n| .$$

令 $S = \operatorname{supp}(\rho)$，并令

$$M = \{ f \in H_k : \ f(x) = 0 \ \textit{for all}\ x \in S \}.$$

假设存在 $\alpha > 0$，使得对每个 $f \in M^{\perp}$，

$$\int \frac{|f(x)|^2}{\|k_x\|^2} d\rho(x) \geq \alpha \|f\|_{H_k}^2 .$$

令 $f^* \in H_k$，并考虑带有间歇更新的无噪声插值模型 $y_n = f^*(x_n)$

$$f_{n+1} = f_n + \lambda \eta_n \frac{y_n - f_n(x_n)}{\|k_{x_n}\|^2} k_{x_n}, \qquad 0 < \lambda < 2.$$

那么，假设 $f_0 \in M^{\perp}$（例如 $f_0 = 0$），我们有

$$\mathbb{E}\left[\|f_n - P_{M^{\perp}} f^*\|_{H_k}^2 \right] \leq (1 - p\alpha\lambda(2-\lambda))^n \, \|f_0 - P_{M^{\perp}} f^*\|_{H_k}^2 ,$$

因此 $f_n \to P_{M^{\perp}} f^*$ 属于 L^2。

证明。 对于每个 n，核截面满足 $\langle k_{x_n}, f \rangle = f(x_n)$，因此

$$P_n f = \langle g_n, f \rangle \, g_n = \frac{f(x_n)}{\|k_{x_n}\|^2} k_{x_n}.$$

特别地，$\ker P_n = \{ f : \ f(x_n) = 0 \}$。由于 k 连续且 X 可分，独立同分布样本 $\{x_n\}$ 几乎必然在 $S =$ 支撑(ρ)中稠密。因此，若 $f \in$

$\bigcap_{n\geq 1}\ker P_n$，则对所有 n，$f(x_n)=0$；由连续性可知，f 在 $\{x_n\}$的闭包上消失，因此也在 S上消失。这表明几乎必然

$$\bigcap_{n\geq 1}\ker P_n = M.$$

令 $\mathcal{F}_n$ 为由 $\{(x_j,\ \eta_j)\}_{1\leq j\leq n}$ 生成的自然滤过。由于 (x_n) 独立同分布，P_n 与 $\mathcal{F}_{n-1}$ 独立，且与 P_1 同分布，因此

$$\mathbb{E}\left[P_n \mid \mathcal{F}_{n-1}\right] = \mathbb{E}\left[P_1\right]$$

对所有 n 均成立。此外，η_n 与 P_n 及 $\mathcal{F}_{n-1}$ 独立，因此 $p_n=\mathbb{E}\left[\eta_n \mid \mathcal{F}_{n-1}\right]=p$。直接计算可得，对于 $f\in H_k$,

$$\left\langle \mathbb{E}\left[P_1\right] f, f\right\rangle = \mathbb{E}\left[\left|\langle g_1, f\rangle\right|^2\right] = \int \frac{|f(x)|^2}{\|k_x\|^2} d\rho(x).$$

因此，关于 ρ 的假设恰好是定理 4.4 中的强制性条件，其中$\beta=p\alpha$。将定理 4.4 应用于误差 $e_n=f_n-f^*$可得 $e_n\to P_M e_0=P_M f_0-P_M f^*$，因此

$$f_n = f^* + e_n \to P_M f_0 + P_{M^\perp} f^*.$$

在假设 $f_0\in M^\perp$下，有 $P_M f_0=0$，这就是在 L^2中收敛到 $P_{M^\perp} f^*$，如所述。□

5. 基于树分裂的核重新参数化

我们沿用第 2.4 节的多通道树分裂设置。特别地，对于每个深度 $N\geq 1$，分裂恒等式（2.4）给出一个典范坐标映射W_N，将其映射到希尔伯特直和 H_N，其中

$$H_N := \left(\bigoplus\nolimits_{|\alpha|=N} H\right) \oplus \left(\bigoplus\nolimits_{|\alpha|<N} H\right),$$

配备标准的 ℓ^2 直和内积。

接着我们定义

$$W_N x := \left((A_\alpha x)_{|\alpha|=N}, (DA_\alpha x)_{|\alpha|<N}\right), \quad x\in H. \tag{5.1}$$

采用 H_N上的这一范数，对于所有 $x\in H$,

$$\|W_N x\|^2 = \sum_{|\alpha|=N}\|A_\alpha x\|^2 + \sum_{|\alpha|<N}\|DA_\alpha x\|^2 = \|x\|^2. \tag{5.2}$$

因此，W_N 是一个等距映射，因此对所有 $x,\ y\in H$,

$$\langle x, y\rangle = \langle W_N x, W_N y\rangle. \tag{5.3}$$

这一观察（5.2）-（5.3是下文构造的出发点。

设 X 为任意集合。设 k_0 为定义在 $X\times X$ 上的正定核，其特征映射为 $\Phi_0\colon X\to H$，使得

$$k_0(x,y) := \langle \Phi_0(x), \Phi_0(y)\rangle. \tag{5.4}$$

对于每个 $N\geq 1$，我们构造多尺度特征映射 $\Phi_N := W_N\circ\Phi_0$，即（5.1）与 Φ_0。由于 W_N 是等距映射，利用（5.4）可得恒等式

$$k_0(x,y) = \langle \Phi_N(x), \Phi_N(y)\rangle, \quad x,y\in X. \tag{5.5}$$

因此，树坐标为由 Φ_0诱导的同一核提供了一种结构化表示，而不改变 k_0。

核的截断。舍弃（或加权） W_N的某些坐标后，会得到一族新的核。我们给出一种基本的截断方式，足以满足后续使用。

固定整数 $0 \le M \le N$，并通过仅保留深度不超过 $M-1$的缺陷块来定义截断特征映射，

$$\widetilde{\Phi}_M(x) := \left((DA_\alpha \Phi_0(x))_{|\alpha|<M} \right), \quad x \in X, \tag{5.6}$$

将其视为 H的相应多个副本的直和中的一个元素。我们将关联的截断核定义为

$$\begin{aligned} \widetilde{k}_M(x,y) &:= \left\langle \widetilde{\Phi}_M(x), \widetilde{\Phi}_M(y) \right\rangle \\ &= \sum_{|\alpha|<M} \langle DA_\alpha \Phi_0(x), DA_\alpha \Phi_0(y) \rangle . \end{aligned} \tag{5.7}$$

~ ~由于 k_M 是 Φ_M的格拉姆核，因此它是正定的。

为了量化 k_M 与 k_0 之间的差异，我们引入截断级别 M之后深度为 N 的丢弃能量，

$$R_{N,M}(x) := \sum_{|\alpha|=N} \|A_\alpha \Phi_0(x)\|^2 + \sum_{M \le |\alpha| < N} \|DA_\alpha \Phi_0(x)\|^2, \quad x \in X. \tag{5.8}$$

定理 5.1。对于所有 $x \in X$，

$$k_0(x,x) - \widetilde{k}_M(x,x) = R_{N,M}(x) \ge 0. \tag{5.9}$$

此外，对于所有 x、$y \in X$，

$$\left| k_0(x,y) - \widetilde{k}_M(x,y) \right| \le \left(R_{N,M}(x) R_{N,M}(y) \right)^{1/2}. \tag{5.10}$$

证明。 固定 $N \ge 1$ 和 $0 \le M \le N$。回顾一下，

$$W_N \Phi_0(x) = \left((A_\alpha \Phi_0(x))_{|\alpha|=N}, (DA_\alpha \Phi_0(x))_{|\alpha|<N} \right) \in H_N,$$

~并且 $\Phi_M(x)$ 来自 (5.6由缺陷块组成，且 $|\alpha| < M$。我们将~$\Phi_M(x)$ 视为 H_N中的一个元素，方法是将其嵌入为 $W_N\Phi_0(x)$ 支撑在坐标 $\{\alpha$ 上的子向量： $|\alpha| < M\}$ 在缺陷部分取值，其他位置为零。

定义被丢弃的坐标向量如下：

$$\Psi_{N,M}(x) := \left((A_\alpha \Phi_0(x))_{|\alpha|=N}, (DA_\alpha \Phi_0(x))_{M \le |\alpha| < N} \right) \in H_N.$$

在直和 H_N 中，不同的坐标块彼此正交，因此，支撑在互不相交的指标集上的向量彼此正交。根据构造，~中 $\Phi_M(x)$ 和 $\Psi_{N、M}(x)$ 的支集互不相交；因此我们有如下正交分解：

$$W_N \Phi_0(x) = \widetilde{\Phi}_M(x) \oplus \Psi_{N,M}(x).$$

对于对角恒等式，取范数并使用毕达哥拉斯定理：

$$\|W_N \Phi_0(x)\|^2 = \|\widetilde{\Phi}_M(x)\|^2 + \|\Psi_{N,M}(x)\|^2.$$

由于 W_N 是等距映射，$\|W_N\Phi_0(x)\|^2 = \|\Phi_0(x)\|^2 = k_0(x,\ x)$，而根据定义 $\|\widetilde{\Phi}_M(x)\|^2 = \widetilde{k}_M(x,\ x)$。最后，展开 $\Psi_{N,\ M}(x)$ 的直和范数，得到

$$\|\Psi_{N,M}(x)\|^2 = \sum_{|\alpha|=N} \|A_\alpha \Phi_0(x)\|^2 + \sum_{M\le|\alpha|<N} \|DA_\alpha \Phi_0(x)\|^2 = R_{N,M}(x).$$

代入这些恒等式可得（5.9），特别是 $R_{N,\ M}(x)\ \ge 0$。

对于非对角线界，使用相同的分解来表示内积：

$$\begin{aligned} k_0(x,y) &= \langle \Phi_0(x), \Phi_0(y)\rangle = \langle W_N\Phi_0(x), W_N\Phi_0(y)\rangle \\ &= \left\langle \widetilde{\Phi}_M(x), \widetilde{\Phi}_M(y)\right\rangle + \langle \Psi_{N,M}(x), \Psi_{N,M}(y)\rangle \\ &= \widetilde{k}_M(x,y) + \langle \Psi_{N,M}(x), \Psi_{N,M}(y)\rangle\,. \end{aligned}$$

因此

$$k_0(x,y) - \widetilde{k}_M(x,y) = \langle \Psi_{N,M}(x), \Psi_{N,M}(y)\rangle\,,$$

并且由柯西-施瓦茨不等式 H_N 可得

$$\left|k_0(x,y) - \widetilde{k}_M(x,y)\right| \le \|\Psi_{N,M}(x)\|\,\|\Psi_{N,M}(y)\| = \left(R_{N,M}(x)R_{N,M}(y)\right)^{1/2}.$$

□

特别地，上一节中关于树能量的任何估计（例如推论2.10得出的几何衰减结论）都会立即给出对 $R_{N,M}(x)$ 的定量控制，进而控制 $k_0 - k_M$，而无需额外论证。

推论 5.2. 假设存在常数 $c \in (0,1)$ 使得

$$\sum_{i=1}^{d} A_i^* A_i \le cI.$$

那么对于每个 $M \ge 0$、每个 $N \ge M$ 和每个 $x \in X$，都有

$$R_{N,M}(x) = \sum_{|\alpha|=M} \|A_\alpha \Phi_0(x)\|^2 \le c^M k_0(x,x)\,. \tag{5.11}$$

因此，

$$0 \le k_0(x,x) - \widetilde{k}_M(x,x) \le c^M k_0(x,x)\,, \tag{5.12}$$

并且对于所有 $x, y \in X$，

$$\left|k_0(x,y) - \widetilde{k}_M(x,y)\right| \le c^M \left(k_0(x,x)\,k_0(y,y)\right)^{1/2}.$$

证明。固定 $x \in X$ 并令 $v = \Phi_0(x)\ \in H$。按照定义 2.6中的记号，层级能量满足

$$L_n(v) = L_{n+1}(v) + \sum_{|\alpha|=n} \|DA_\alpha v\|^2$$

如命题 2.7 的证明中所示。将该恒等式从 $n = M$ 到$n = N-1$ 求和可得

$$L_M(v) = L_N(v) + \sum_{M\le|\alpha|<N} \|DA_\alpha v\|^2.$$

根据 L_N 的定义和$R_{N\text{的定义，}M}$，右侧为

$$\sum_{|\alpha|=N} \|A_\alpha v\|^2 + \sum_{M\le|\alpha|<N} \|DA_\alpha v\|^2 = R_{N,M}(x).$$

因此$R_{N,\,M}(x) = L_M(v) = \sum_{|\alpha|=M} \|A_\alpha v\|^2$，从而证明了式（5.11）。在一致收缩假设下，推论 2.10 给出 $L_M(v) \le c^M \|v\|^2$。□

我们还给出了格拉姆矩阵的样本级不等式，该不等式由~相同的特征表示得出。给定点 $x_1, \dots, x_m \in X$，令 K_0和 K_M为格拉姆矩阵

$$(K_0)_{j\ell} = k_0(x_j, x_\ell), \qquad (\widetilde{K}_M)_{j\ell} = \widetilde{k}_M(x_j, x_\ell).$$

~**推论 5.3**。有 $0 \le K_0 - K_M$，

$$\operatorname{tr}\left(K_0 - \widetilde{K}_M\right) = \sum_{j=1}^m R_{N,M}(x_j), \tag{5.13}$$

$$K_0 - \widetilde{K}_M \le \left(\sum_{j=1}^m R_{N,M}(x_j)\right) I. \tag{5.14}$$

~证明。 $K_0 - K_M$是丢弃坐标的格拉姆矩阵，因此是半正定的。迹恒等式由上面的（5.9）得出。最后一个界是由以下初等推论得到的：$0 \le M \le \operatorname{tr}(M)\, I$对于半正定矩阵 M。□

注 5.4。尽管这一构造在截断之前不会改变基础核 k_0，但仍然很有用。在许多情况下，人们希望使用比 Φ_0更简单或成本更低的特征映射，同时仍保持与~原始核的行为相近。多尺度映射 Φ_N 及其截断 Φ_M 按深度组织 Φ_0 中的信息，而舍弃量 $R_{N、M}$（x）则精确衡量了在超过深度 M的层级 N 舍弃了多少 Φ_0（x）。本节中的恒等式~表明，$R_{N、M}$（x）是 k_0（x、x）与 $-k_M$（x、x）之差，并且~对于所有 x、y，它都能界定 $|k_0$（x、y）与 $-k_M$（x、y）$|$。对于有限样本 $\{x_j\}$，格拉姆~矩阵满足（5.14），因此只要舍弃能量$R_{N、M}$（x_j）足够小，K_M 的特征值就会与K_0 的特征值始终保持接近。在实践中，这为选择截断深度 M（或选择保留哪些节点）提供了一种有原则的方法：在降低复杂度的同时，仍能定量控制截断核与原始核之间的差异。

6. 贪心细化与核压缩

我们继续采用第 5 节的设定。固定一个深度 $N \ge 1$，并始终保持不变。记

$$\mathcal{A}_{<N} := \{\alpha : |\alpha| < N\} \tag{6.1}$$

表示严格位于第 N 层之上的节点集合（包括空词 $\emptyset$）；对于每个 $\alpha \in \mathcal{A}_{<N}$，记 K_α为由 α- 缺陷块诱导的格拉姆核，

$$k_\alpha(x,y) := \langle DA_\alpha\Phi_0(x), DA_\alpha\Phi_0(y)\rangle, \quad x,y \in X.$$

因此 k_α 是正定的，对于任意有限样本 $\{x_1, \dots, x_m\}$，其格拉姆矩阵

$$(K_\alpha)_{j\ell} = k_\alpha(x_j, x_\ell)$$

是半正定的。

与第5节中的情况一样，深度为M的截断对应于保留前缀集 $\{\alpha$ 中的缺陷块: $|\alpha|<M\}$。对于自适应细化，允许更一般的前缀集会更方便。

定义 6.1。若子集 $S\subseteq\mathcal{A}_{<N}$ 满足： $\alpha\in S$ 且 $\alpha\neq\emptyset$蕴含 $\alpha^{-}\in S$，则称其为前缀闭，其中 α^{-} 表示 α 的父节点（删除最后一个符号）。

给定这样的 S，定义截断特征映射

$$\Phi_S(x):=\left(\left(DA_\alpha\Phi_0(x)\right)_{\alpha\in S}\right),\qquad x\in X, \tag{6.2}$$

将其视为由 S索引的 H 副本的希尔伯特直和中的元素，并定义其格拉姆核

$$\begin{aligned}k_S(x,y)&:=\langle\Phi_S(x),\Phi_S(y)\rangle\\&=\sum_{\alpha\in S}\langle DA_\alpha\Phi_0(x),DA_\alpha\Phi_0(y)\rangle.\end{aligned} \tag{6.3}$$

由于 k_S 是格拉姆核，因此它是正定的。深度为 N 时，超出 S 的残差能量定义为

$$R_{N,S}(x):=\sum_{|\alpha|=N}\|A_\alpha\Phi_0(x)\|^2+\sum_{\alpha\in\mathcal{A}_{<N}\setminus S}\|DA_\alpha\Phi_0(x)\|^2,\qquad x\in X. \tag{6.4}$$

当 $S=\{\alpha:\ |\alpha|<M\}$ 这恰好就是第 R_N 节$_M$ 中的量 5。与第 5 节相同的正交分解论证给出了以下推广。

推论 6.2。对于所有 $x\in X$，均有

$$k_0(x,x)-k_S(x,x)=R_{N,S}(x)\geq 0. \tag{6.5}$$

此外，对于所有 x、$y\in X$，

$$|k_0(x,y)-k_S(x,y)|\leq\left(R_{N,S}(x)R_{N,S}(y)\right)^{1/2}. \tag{6.6}$$

对于有限样本 $\{x_1,\ldots,x_m\}$，记 K_S 为 k_S 的格拉姆矩阵，记 K_0 为 k_0 的格拉姆矩阵。那么 $K_S=\sum_{\alpha\in S}K_\alpha$，并且与之前完全相同， K_0-K_S 是丢弃坐标的格拉姆矩阵，因此是半正定的。特别地，

$$\begin{aligned}0\leq K_0-K_S&\leq\operatorname{tr}(K_0-K_S)\,I\\\operatorname{tr}(K_0-K_S)&=\sum_{j=1}^{m}R_{N,S}(x_j).\end{aligned}$$

这个迹量将作为下文贪心压缩方案中的主要误差度量。

接下来，我们介绍树上的一个简单细化规则。固定一个样本 $\{x_1,\ldots,x_m\}\subset X$。对每个节点 $\alpha\in\mathcal{A}_{<N}$，定义其样本能量

$$E(\alpha):=\operatorname{tr}(K_\alpha)=\sum_{j=1}^{m}\|DA_\alpha\Phi_0(x_j)\|^2.$$

如果 $S\subseteq\mathcal{A}_{<N}$ 是前缀闭的，其 边界 为

$$\partial S:=\left\{\beta\in\mathcal{A}_{<N}\setminus S:\beta^{-}\in S\right\}.$$

贪心规则是:

算法。从一个前缀闭集 S_0 开始（例如 $S_0=\{\emptyset\}$）。给定 S_t，选择

$$\alpha_t \in \arg\max_{\beta\in\partial S_t} E(\beta),$$

并设 $S_{t+1} := S_t \cup \{\alpha_t\}$。

每一步都会增加一个缺陷块，并保持前缀闭性。下一个命题给出了基本的单调性，并确定了迹误差的精确下降量。

命题 6.3。对于每个 t都有

$$K_{S_t} \le K_{S_{t+1}} \le K_0,$$

and

$$\operatorname{tr}\left(K_0 - K_{S_{t+1}}\right) = \operatorname{tr}\left(K_0 - K_{S_t}\right) - E(\alpha_t).$$

特别地，$\operatorname{tr}(K_0 - K_{S_t})$ 在 t 中单调不增，并提供一种事后停止准则。

证明。由于 $K_{S_{t+1}} = K_{S_t} + K_{\alpha_t}$ 和 $K_{\alpha_t} \ge 0$，可得 $K_{S_t} \le K_{S_{t+1}}$。此外，$K_0 - K_{S_t}$ 是丢弃坐标的格拉姆矩阵，因此是半正定的，所以 $K_{S_t} \le K_0$。取迹可得

$$\operatorname{tr}\left(K_0 - K_{S_{t+1}}\right) = \operatorname{tr}\left(K_0 - K_{S_t} - K_{\alpha_t}\right) = \operatorname{tr}\left(K_0 - K_{S_t}\right) - \operatorname{tr}\left(K_{\alpha_t}\right),$$

根据定义，$\operatorname{tr}(K_{\alpha_t}) = E(\alpha_t)$。 □

核学习。接下来，我们讨论与标准核学习最直接相关的推论。设 $y \in \mathbb{R}^m$ 是样本上的标签，并固定 $\lambda > 0$。对于格拉姆矩阵 K，定义样本上的核岭预测器如下：

$$\widehat{y}(K) := K\left(K + \lambda I\right)^{-1} y \in \mathbb{R}^m.$$

下一个命题表明，迹界控制了以贪心截断替代完整核所产生的误差。

定理 6.4。对于每个前缀闭的 $S \subseteq \mathcal{A}_{<N}$，有

$$\|\widehat{y}(K_0) - \widehat{y}(K_S)\| \le \frac{1}{\lambda}\|K_0 - K_S\|\,\|y\| \le \frac{1}{\lambda}\operatorname{tr}\left(K_0 - K_S\right)\|y\|. \tag{6.7}$$

特别地，对于贪心序列 (S_t)，有

$$\|\widehat{y}(K_0) - \widehat{y}(K_{S_t})\| \le \frac{1}{\lambda}\operatorname{tr}\left(K_0 - K_{S_t}\right)\|y\| = \frac{1}{\lambda}\sum_{j=1}^{m} R_{N,S_t}(x_j)\,\|y\|. \tag{6.8}$$

证明。利用 $\widehat{y}(K) = (I - \lambda(K+\lambda I)^{-1})y$，可得

$$\widehat{y}(K_0) - \widehat{y}(K_S) = \lambda\left(\left(K_S + \lambda I\right)^{-1} - \left(K_0 + \lambda I\right)^{-1}\right) y.$$

预解式恒等式给出

$$\left(K_S + \lambda I\right)^{-1} - \left(K_0 + \lambda I\right)^{-1} = \left(K_S + \lambda I\right)^{-1}\left(K_0 - K_S\right)\left(K_0 + \lambda I\right)^{-1}.$$

由于 $K_0 \ge 0$ 和 $K_S \ge 0$，因此有

$$\|\left(K_0 + \lambda I\right)^{-1}\| \le \frac{1}{\lambda}, \qquad \|\left(K_S + \lambda I\right)^{-1}\| \le \frac{1}{\lambda},$$

因此

$$\|\widehat{y}(K_0) - \widehat{y}(K_S)\| \le \lambda \cdot \frac{1}{\lambda}\|K_0 - K_S\| \cdot \frac{1}{\lambda}\|y\| = \frac{1}{\lambda}\|K_0 - K_S\|\,\|y\|.$$

最后，由于 $K_0 - K_S \ge 0$，初等推论 $0 \le M \le \mathrm{tr}\,(M)\ I$ 给出

$$\|K_0 - K_S\| \le \mathrm{tr}\,(K_0 - K_S),$$

并且，由 $R_{N,\ S}$ 的定义可得所述迹公式。 □

注 6.5. 定理 6.4提示了一条具体的停止规则。若指定容差 $\varepsilon > 0$ 并选择 t 使得

$$\sum_{j=1}^{m} R_{N,S_t}(x_j) \le \varepsilon\lambda,$$

则样本级岭预测器满足

$$\|\widehat{y}(K_0) - \widehat{y}(K_{S_t})\| \le \varepsilon\|y\|.$$

从这个意义上说，舍弃的树能量为压缩核在诱导学习映射 $y \mapsto \widehat{y}(K)$ 层面上的偏离程度提供了定量界。

备注 6.6. 还可以考虑其他贪心评分方案。上述方案由迹界驱动，因为它可直接通过块能量$\|DA_\alpha\Phi_0(x_j)\|^2$ 计算，并能得出简洁的稳定性估计。在希望优先考虑条件数或信息增益而非迹误差的应用中，也可以采用相同的前缀闭合探索，但根据候选项 $\beta \in \partial S$ 的更新 K_β的矩阵泛函进行评分，例如

$$S \mapsto \log\det\left(I + \frac{1}{\lambda}K_S\right).$$

只要给定样本上的块矩阵K_α 具有额外结构（例如，秩一致地较低，或其因式分解可以高效维护），就可以分析这类选择。本文不探讨这一方向，因为基于迹的方案已经能根据舍弃能量$R_{N,S}(x_j)$给出明确且可验证的界。

7. 与贪心核主成分分析的联系

下面我们说明上述结果对主成分分析（PCA）的意义，以扩展我们先前的结果。PCA 是核分析当前一项重要的应用。我们回顾一下，PCA 提供了强大的降维工具，有助于将高维数据变换为一组较小的互不相关变量，即主成分（PC）。这尤其有用，因为它可以在保留最大方差（信息）的同时完成，也就是说，选择主成分。因此，通过确定方差最大的正交轴，PCA 可以实现：1）降低噪声；2）处理多重共线性；3）促进特征选择与可视化。为方便读者，列出以下参考文献[BN03, WYLZ23, GGB18, JHZW19, GJ06, Bis06, Bis13, ZBB04, AFS18, VD16]。近年来，核主成分分析及其应用得到了广泛研究，并在多个方向上取得了进展[LHN18, Raj18, WGLP19, RS99, GGB18, DWGC18, THH19, LLY+19, MMP19, JHZW19, CLS+19, GK19]。

在此，我们将经典主成分分析（PCA）与前文各节的结果联系起来，应用于贪心核主成分分析。请参见 [JS07、JST22] ，了解经典主成分分析；并参见 [OB05、WSH21、LY09、DTX10、FH06] ，了解贪心核主成分分析。

GKPCA 中的“贪心”一词，指的是采用“贪心”方法设计算法技术，从训练数据中选取一个具有代表性的小型子集，而非使用整个数据集，将其用于当前问题的特定核模型；因此，贪心算法能够缩减特征空间计算中所用核矩阵的规模。其应用广泛多样，包括人脸识别、说话人识别（噪声环境下的非线性特征）、生物医学信号处理和分类（用于提高支持向量机（SVM）等分类器的训练效率）。

上文第 3 节给出了插值结果，同时也展示了再生核希尔伯特空间中的压缩算子流，其能量核算精确，且残差分解由秩一投影给出。这正是主成分分析和核主成分分析的基础结构。PCA 是一种特殊情形，其中投影的选择旨在最大化方差；而本框架固定投影，并对残差能量进行优化。我们在第2节、第 3 节和第 5节提出了收敛条件、残差能量分解以及噪声下的稳定性。命题 3.3 比经典 PCA/KPCA 更强 [OB05，WSH21，LY09，DTX10，FH06]。

我们首先定义贪心核主成分分析（KPCA）。

定义 7.1。 [OB05] 设 $X = \{x_1 \cdots, x_m\}$ 为输入空间中的一个训练集。设 H_k 为定义在 X 上的复值函数再生核希尔伯特空间，其再生核为 k。对于 $x_n \in X$，在 x_n 处的核截面为 $k_{x_n} := k(\cdot, x_n) \in H_k$，其中 $g_n := \frac{k_{x_n}}{\sqrt{k(x_n,x_n)}}$是归一化核特征。秩一再生核希尔伯特空间投影定义为

$$P_n f = \frac{f(x_n)}{k(x_n, x_n)} k_{x_n}.$$

算法 7.2。 [OB05] （贪心子集选择）贪心核主成分分析会构建一个字典 $D = \{d_1 \cdots, d_l\} \subseteq X$ 其中包含通过贪心方式选出的训练点；这些训练点依据数据模型能否由已选字典元素的张成空间良好近似来确定。

在迭代 n 中，给定字典 $D_{n-1} = \{d_1 \cdots, d_{n-1}\}$，针对下一个训练点 x_n计算以下两项：

1. 将 k_{x_n} 投影到字典特征$\{\phi(d_i)\}$张成的空间上，

2. 特征空间中的残差范数：

$$\delta_n := \min_{a \in R^{n-1}} \|\phi(x_n) - \sum_{i=1}^{n-1} a_i \phi(d_i)\|^2$$

其中 $\phi(x_n) = k_{x_n}$是再生核希尔伯特空间（*RKHS*）的特征映射。

如果 $\delta_n > \epsilon$，贪心规则会将下一个训练点 x_n 纳入字典；然后 $d_n = x_n$，否则不添加 x_n，即跳过。

该准则通过贪心方式扩展再生核希尔伯特空间中的张成空间，以最佳近似整个数据集的特征映射；它选择在特征空间中与当前子空间的残差范数距离最远的点。

算法 7.3。 [OB05] （贪心核主成分分析子空间近似）根据上述词典 $D = \{d_1 \cdots, d_n\} \subseteq X$ 构建：

1. 构造字典格拉姆矩阵 $G \in \mathbb{R}^{n\times n}$，其中

$$G_{ij} = k(d_j, d_j).$$

2. 计算格拉姆矩阵的特征值 $\lambda_1 \geq \cdots \geq \lambda_p$，以及对应的特征向量 G $v_1 \cdots v_p$。

3. 构造近似核主成分分析分量：

$$u_j^{(D)} = \sum_{i=1}^{n} (v_j)_i \, k_{d_i}, \quad where \quad j = 1, \cdots, p.$$

这里，$u_j^{(D)} \in H_k$ 是再生核希尔伯特空间中与字典 $D = \{d_1 \cdots, d_n\}$相关联的近似主方向，它们近似于通过 X上的完整格拉姆矩阵得到的真实核主成分分析特征函数。

定义 7.4。 [OB05] 在上述条件下，对于任意新点 $x \in X$，其在近似子空间上的贪心核主成分分析投影/嵌入为

$$f_j^{(D)}(x) = \langle u_j^{(D)}, g_x \rangle = \sum_{i=1}^{n} (v_j)_i \, \frac{k(d_i, x)}{\sqrt{k\,(x, x)}}, \qquad for \quad j = 1, \cdots, p.$$

备注 7.5。 在前文的框架中，字典 D 张成一个子空间

$$H_D = \operatorname{span}\{k_{d_i} : d_i \in D\} \subseteq H_k.$$

核主成分分析分量是限制在 H_D上的经验协方差算子的主特征函数。贪心距离准则对应于选择x_n，使得投影到 H_D 上的投影算子在再生核希尔伯特空间中留下最大的残差。

我们将理论与 [OB05] 中的贪心核主成分分析联系起来；这是一种用于近似特征函数估计的子集选择方法，其中选择字典点以最大化特征空间中未解释的方差。随后，在缩减后的字典张成空间中应用经典核主成分分析。有关核主成分分析的详细信息，请参阅 [JS07，JST22] 。

注释 7.6（与主成分分析的算子级联系）。 在经典主成分分析的算子形式中，我们设 H 上存在一个协方差算子，而主成分分析给出如下投影：

$$P_n = |g_n\rangle\langle g_n|$$

其中，g_n 是 G [JS07，JST22]。在第 3节中，以动态望远镜流替代静态主成分分析

$$T_n = (1 - \lambda_1 P_1) \cdots (1 - \lambda_n P_n)$$

利用恒等式

$$\|f\|^2 - \|T_N f\|^2 = \sum_{n=1}^{N} \lambda_n \, (2 - \lambda_n) \, \|P_n T_{n-1} f\|^2$$

每一步都会移除方向 P_n 上的一部分能量，这种移除过程可控、单调且测量精确。此外，无需进行特征分解。

此外，在第3节中，

$$g_n := \frac{k_{x_n}}{\|k_{x_n}\|}$$

更新在特征空间中进行，且不会构造协方差算子。该更新

$$f_n = f_{n-1} + \lambda_n \left(y_n - f_{n-1}\left(x_n\right)\right) \frac{k_{x_n}}{k\left(x_n, x_n\right)}, \qquad n \geq 1. \tag{7.1}$$

等价于增量核主成分分析

经典核主成分分析通常只能给出渐近结果或期望界，因此无法实现精确的望远镜式能量平衡。但根据定理 3.1 我们的新框架能够实现望远镜式能量平衡。

在经典主成分分析框架中，其中 $y_n = 0$, P_n 被选为主导经验方差方向，且 $\lambda_n = 1$。因此， $T_n = \prod_{k=1}^{n}(I - P_k)$ 这就是消去式主成分分析。我们的结果通过命题 3.3 给出了收敛条件、残余能量分解以及噪声下的稳定性。

第 3 节与精确的贪心能量分解相联系，并由此得到以下定理。

定理7.7。 （精确贪心能量分解）设 H_k 是核为k的再生核希尔伯特空间。设 $(x_n)_{n\geq 1} \subset X$ 且令，$0 < \lambda_n < 2$。定义秩一投影

$$P_n f = \frac{f\left(x_n\right)}{k\left(x_n, x_n\right)} k_{x_n},$$

以及贪心残差算子

$$T_n = \prod_{j=1}^{n}(I - \lambda_j P_j), \quad T_0 = I.$$

那么，对于每个 $f \in H_k$ 以及每个 $N \geq 1$，望远镜级数(7.2) 成立

$$\begin{aligned} \|f\|^2 - \|T_N f\|^2 &= \sum_{n=1}^{N} \lambda_n \left(2 - \lambda_n\right) \|P_n T_{n-1} f\|^2 \\ &= \sum_{n=1}^{N} \lambda_n \left(2 - \lambda_n\right) \frac{\left|\left(T_{n-1} f\right)\left(x_n\right)\right|^2}{k\left(x_n, x_n\right)}. \end{aligned} \tag{7.2}$$

证明。 每个 P_{x_n} 都是正交投影，因此 $P_{x_n} = P_{x_n}^* = P_{x_n}^2$。对于任意 $g \in H_k$,$\|($ $I - \lambda_n P_{x_n})g\|^2 = \|g\|^2 - 2\lambda_n\|P_{x_n}g\|^2 + \lambda_n^2\|P_{x_n}g\|^2 = \|g\|^2 - \lambda_n(2-\lambda_n)\|P_{x_n}g\|^2$.令 $g = T_{n-1}f$ 并从 $n = 1$到 N 求和，即得望远镜级数（7.2）。□

该恒等式将 f 的总方差分解为残差方差 $\|T_N f\|^2$ 和每个贪心步骤所解释的累积方差。与经典（核）主成分分析不同，该分解在每次有限迭代中都精确成立，且不依赖正交性或谱假设。

下述定理由（2.4）和（5.1）得出。贪心表述的一项核心优势在于，收敛完全可以通过（2.4）中的步长序列来刻画。

定理 7.8。 （贪心核主成分分析的强收敛）设步长满足（2.3）

$$\sum_{n=1}^{\infty} \frac{\lambda_n}{2-\lambda_n} < \infty. \tag{7.3}$$

则算子序列 (T_n) 强收敛到一个有界算子T_∞。特别地，对于每个 $f \in H_k$,

$$T_n f \to T_\infty f$$

在范数意义下，且残差能量 $\|T_n f\|^2$ 单调递减至 $\|T_\infty f\|^2$。

证明。 每个因子 $A_n = I - \lambda_n P_{x_n}$ 都是一个压缩算子。此外

$$||T_{n-1}f - T_n f||^2 = \lambda_n^2 ||P_{x_n} T_{n-1}||^2 \le \frac{\lambda_n}{2-\lambda_n}(\|T_{n-1}f\|^2 - \|T_n f\|^2),$$

其中，不等式由（5.1）得出。可求和性假设意味着 $(T_n f)$对每个 f 都是柯西序列，因此在范数意义下收敛。一致有界性保证 T_n 强收敛到有界极限 T_∞。 □

这提供了与经典主成分分析中完全方差捕获直接对应的结果，且无需进行特征分解。

以下定理直接由命题 3.3得出

定理 7.9。 （噪声稳定性与偏差—方差分解）假设观测值受到噪声污染，噪声按上述命题 *3.3*中的方式添加，

$$y_n = f^*(x_n) + \varepsilon_n,$$

其中 (3.5)

$$\mathbb{E}(\varepsilon_n \mid x_n) = 0, \qquad \mathbb{E}\left(|\varepsilon_n|^2 \mid x_n\right) \le \sigma^2$$

对于某个 $\sigma \ge 0$，并假设噪声 $\{\varepsilon_n\}$ 在各个 n之间相互独立。进一步假设

$$k(x_n, x_n) \ge \kappa, \qquad \sum_{n=1}^{\infty} \lambda_n^2 < \infty.$$

那么，贪心核主成分分析的迭代满足以下条件

$$\mathbb{E}\,\|f^* - f_N\|^2 \le \|f^*\|^2 - \sum_{n=1}^{N} \lambda_n (2-\lambda) \mathbb{E}\,\|P_{x_n} e_{n-1}\|^2 + \sum_{n=1}^{N} \lambda_n^2 \mathbb{E}\left(\frac{|\varepsilon_n|^2}{k(x_n, x_n)}\right),$$

并且所添加的噪声有界于

$$\frac{\sigma^2}{\kappa} \sum_{n=1}^{N} \lambda_n^2 < \infty.$$

证明。 将每次更新分解为由 f^* 驱动的确定性分量，以及由 ϵ_n. 驱动的随机分量。 P_{x_n} 的正交性和零均值假设消除了期望中的交叉项。现在，方差项可由下式得出

$$\|\lambda_n P_{x_n} \epsilon_n\|^2 = \lambda_n^2 \frac{|\varepsilon_n|^2}{k(x_n, x_n)},$$

并且，根据所述假设可知该级数可求和。 □

这为鲁棒性提供了一个透明且定量的判据，这与经典核主成分分析形成对比；在经典方法中，噪声通常仅通过截断或正则化间接处理。

结论：贪心核主成分分析以压缩算子驱动的算子流取代全局谱分解，沿着数据自适应方向逐步消除方差。前文各节整合而成的框架提供了：(1) 精确的有限步能量分解；(2) 在简单步长条件下的强算子收敛；以及 (3) 存在噪声时明确的偏差-方差权衡和稳定性保证。

综上，这些结果确立了贪心核主成分分析作为一种有理论依据、可在线运行的经典核主成分分析替代方法，尤其适用于大规模、流式处理以及考虑噪声的学习问题。

关于各种应用，请参见 [TS06，SWH24，TTLdS07，LY09，HBW+24]。

致谢。第一位署名作者谨向许多同事致谢，感谢他们进行了许多有益的讨论，其中包括 Daniel Alpay、Chad Berner、Dorin Dutkay、John E. Herr、Halyun Jeong 和 Eric S. Weber。

参考文献

[AFS18] Carlos M. Alaíz、Michaël Fanuel 和 Johan A. K. Suykens，核主成分分析的凸优化表述及其在半监督学习中的应用，《IEEE 神经网络与学习系统汇刊》 **29** 2018 8, 3863 3869 3854652 ( )，第 – 期。MR AJ Daniel Alpay 和 [21] Palle E. T. Jorgensen，再生核希尔伯特空间的新刻画及其在度量几何中的应用，《Opuscula Math.》41 3, 283 2021 300 4302453 ( )，第 – 期。MR [AJ22] Daniel Alpay 和 Palle Jorgensen，μ-布朗运动、对偶性、扩散、变换与再生核希尔伯特空间，《J. Theoret. Probab.》35 4，2757 2022 2783 4509085 ( )，第 – 期。MR [AJV14] Daniel Alpay、Palle Jorgensen 和 Dan Volok，相对再生核希尔伯特空间，《Proc. Amer. Math. Soc.》 **142** 2014 11, 3889 3895 3251728 ( )，第 – 期。MR BGHJ [26] Chad M. Berner、Noah G. Giddings、John E. Herr 和 Palle E. T. Jorgensen，二维奇异纤维测度的傅里叶级数，《J. Math. Anal. Appl.》 **553** 2026 2, 129930, 21 4941845 ( )，论文编号 。MR BHJW [25] Chad Berner、John E. Herr、Palle E. T. Jorgensen 和 Eric S. Weber，高维奇异测度的傅里叶级数，《J. Fourier Anal. Appl.》31 (2025)，第 1 期，论文编号 1，33。MR 4835160[Bis06] Christopher M. Bishop，《模式识别与机器学习》，信息科学与统计丛书，Springer，纽约，2006。MR 2247587[Bis 13] ，《基于模型的机器学习》，《Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.》371 (2013)，第 1984 期，20120222，17。MR 3005666[BN03] Mikhail Belkin 和 Partha Niyogi，用于降维和数据表示的拉普拉斯特征映射，《Neural Computation》15 (2003)，第 6 期，1373–1396。[CLS+19] Kayla D. Coleman、Allison Lewis、Ralph C. Smith、Brian Williams、Max Morris 和 Bassam Khuwaileh，复杂模型中用于降维的主动子空间的无梯度构建及其在中子学中的应用，《SIAM/ASA J. Uncertain. Quantif.》7 (2019)，第 1 期，117–142。MR 3900802[DTX10] Mingtao Ding、Zheng Tian 和 Haixia Xu，自适应核主成分分析，《 Signal Processing 》 90 (2010)，第 5 期，1542–1553，统计信号与阵列处理专题。[DWGC18] Jianghu J. Dong、Liangliang Wang、Jagbir Gill 和 Jiguo Cao，肾移植后肾小球滤过率曲线的函数型主成分分析，《医学研究统计方法》 **27** 2018 12, 3785 3796 3878657 ( )，第 – 期。MR FH [06] Vojtěch Franc 和 Václav Hlaváč，贪心核主成分分析，第 87–105 页，Springer Berlin Heidelberg，柏林，海德堡，2006。

[FHR+23] Christian Fiedler、Michael Herty、Michael Rom、Chiara Segala 和 Sebastian Trimpe，平均场极限下的再生核希尔伯特空间，《运动学与相关模型》16 (2023)，第 6 期，850–870。MR 4619928[GGB18] Alon Gonen 和 Ran Gilad-Bachrach，基于平滑敏感度的差分隐私主成分分析方法，《算法学习理论》2018，《机器学习研究会议论文集》（PMLR），第 83 卷，《机器学习研究论文集》PMLR，[ ], 2018, 13 3857315 出版地未标明，第 页。[06]
MR GJ M. R. Gupta 和 N. P. Jacobson，小波主成分分析及其在高光谱图像中的应用，2006 年国际图像处理会议，2006 年 10 月，第 1585–1588 页。[GK19] Constantin Grigo 和 Phaedon-Stelios Koutsourelakis，不确定性传播的贝叶斯模型与降维：随机介质中的应用，《SIAM/ASA J. Uncertain. Quantif.》7 (2019)，第 1 期，292–323。MR 3922239[HBW+24] Robin Herkert、Patrick Buchfink、Tizian Wenzel、Bernard Haasdonk、Pavel Toktaliev 和 Oleg Iliev，用于根据三维多孔几何数据近似反应流突破曲线的贪心核方法，《Mathematics》12 (2024)，第 13 期。[HJW23] John E. Herr、Palle E. T. Jorgensen 和 Eric S. Weber，二维分形的傅里叶级数，《从经典分析到分形分析》。第 1 卷。献给 Robert Strichartz 的致敬文集，《应用数值调和分析丛书》，Birkhäuser/Springer，Cham， [2023] © 2023, 183 229 4676388 第 – 页。MR JHZW J. Jin、T. Huang、J. L. Zheng 和 P. [19]
H. Wen，采用映射与直接积分算法的降维分析，《Eng. Anal. Bound. Elem.》99 (2019)，122–130。MR 3883202[JJKS25] Halyun Jeong、Palle E. T. Jorgensen、Hyun-Kyoung Kwon 和 Myung-Sin Song，无限维算子／分块 Kaczmarz 算法：遗憾界与 λ-有效性，arXiv.2511.07604 (2025)。[JP19] Palle E. T. Jorgensen 和 Erin P. J. Pearse，连续网络与离散网络、图拉普拉斯算子及再生核希尔伯特空间，《J. Math. Anal. Appl.》469 (2019 2, 765 807 3860446)，第 – 期。MR JS Palle E. [07]
T. Jorgensen 和 Myung-Sin Song，熵编码、希尔伯特空间与 Karhunen-Loève 变换，《J. Math. Phys.》48 (2007)，第 10 期，103503。[JS23] Palle E. T. Jorgensen 和 Myung-Sin Song，《多层系统数学——数据、尺度变换、图像、信号与分形》，《当代数学及其应用：专著、综述与讲义》，第 8 卷，World Scientific Publishing Co. Pte. Ltd.，Hackensack，新泽西州， [2023] ©2023。MR 4619155[JST20] Palle Jorgensen、Myung-Sin Song 和 James Tian，一种适用于投影序列、无穷乘积及 IFS L^2 空间中框架应用的 Kaczmarz 算法，《Adv. Oper. Theory》 **5** 2020 3, 1100 1131 4126821 ( )，第 – 期。MR JST Palle [22]
E. T. Jorgensen、Myung-Sin Song 和 James Tian，无限维随机变换与再生核希尔伯特空间，《Sampling Theory, Signal Processing, and Data Analysis》21 (2022)。[JST25] Palle E. T. Jorgensen、Myung-Sin Song 和 James Tian，算子理论、核与前馈神经网络，《Complex Anal. Oper. Theory》19 (2025)，第 7 期，论文编号 179，25。MR 4957888[JSY25] Sanghun Jeong、Sanghun Shin 和 Hojin Yang，再生核希尔伯特空间中用于高维生存数据的特征筛选滤波器，《Comm. Statist. Theory Methods》 **54** 2025 13, 4101 4120 4905634 ( )，第 – 期。MR JT a Palle [25]
E. T. Jorgensen 和 James Tian，通过核分析探索特征空间选择的新对偶性，《Sampl. Theory Signal Process. Data Anal.》23 (2025)，第 1 期，论文编号 5，32。MR [JT25b] ，算子值核、机器学习与动力系统，《Phys. D》476 (2025)，论文编号 134657，6。MR 4887444[KM01] Stanisł aw Kwapień 和 Jan Mycielski，关于无限维空间中的近似 Kaczmarz 算法，《 Studia Math. 》 **148** 2001 1, 75 86 1881441 ( )，[18]
第 – 期。MR LHN Yeejin Lee、Keigo Hirakawa 和 Truong Q. Nguyen，用于原始图像传感器数据压缩的相机感知多分辨率分析，《IEEE Trans. Image Process.》27 (2018)，第 6 期，2806–2817。MR 3780557

[LLY+19] Xiao Lin、Ruosha Li、Fangrong Yan、Tao Lu 和 Xuelin Huang，《基于纵向数据函数型主成分分析的分位数残余寿命回归及动态预测》，《医学研究统计方法》28 (2019)，第 4 期，1216–1229。MR 3934645[LY09] Xiaofang Liu 和 Chun Yang，《用于分类中训练数据缩减和非线性特征提取的贪心核主成分分析》，MIPPR 2009：自动目标识别与图像分析（Tianxu Zhang、Bruce Hirsch、Zhiguo Cao 和 Hanqing Lu 编），国际光学工程学会（SPIE）会议系列，第 7495 卷，2009 年 10 月，第 749530 页。[MMP19] Shahar Mendelson、Emanuel Milman 和 Grigoris Paouris，《通过降维实现广义对偶苏达科夫下界——一项计划》，Studia Math. 244 2，159 2019 202 3850675 ( )，第 – 期。MR [OB05] Marie Ouimet 和 Yoshua Bengio，《贪心谱嵌入》，第十届人工智能与统计国际研讨会论文集（Robert G. Cowell 和 Zoubin Ghahramani 编），机器学习研究论文集，第 R5 卷，PMLR，2005 年 1 月 6–8 日，PMLR 于 2021 年 3 月 30 日重新发布，第 253–260 页。[PR16] Vern I. Paulsen 和 Mrinal Raghupathi，《再生核希尔伯特空间理论导论》，剑桥高等数学研究丛书，第 152 卷，剑桥大学出版社，剑桥，

2016 3526117 。MR Raj S. P. Raja，《利用 [18] Bandelet 多尺度变换和 DES 加密技术实现安全医学图像压缩》，《国际小波、多分辨率与信息处理杂志》16 (2018)，第 4 期，1850028，33。MR 3820672[RS99] B. Ya. Ryabko 和 M. P. Sharova，《低熵源的快速编码》，《信息传输问题》35 (1999)，第 1 期，49–60。MR 1720704[Ste23] George Stepaniants，《在再生核希尔伯特空间中学习偏微分方程》，《机器学习研究杂志》24 (2023)，论文编号 [86], 72。MR 4582508[SWH24] Gabriele Santin、Tizian Wenzel 和 Bernard Haasdonk，《目标数据依赖型核贪心插值在索博列夫再生核希尔伯特空间中的最优性》，《美国工业与应用数学学会数值分析杂志》62 (2024)，第 5 期，2249–2275。[THH19] I-Ping Tu、Su-Yun Huang 和 Dai-Ni Hsieh，《多线性主成分分析的广义自由度》，《多元分析杂志》173 (2019)，26–37。MR 3913046[TS06] T. Tangkuampien 和 D. Suter，《通过贪心核主成分分析滤波实现人体运动去噪》，第十八届国际模式识别会议（ICPR'06），第 3 卷，2006 年，第 457–460 页。[TTLdS07] A.M. Tome、A.R. Teixeira、E.W. Lang 和 A. Martins da Silva，《将贪心核主成分分析应用于单通道脑电记录》，2007 年 IEEE 医学与生物学工程学会第 29 届年度国际会议，2007 年，第 5441–5444 页。[VD16] Hai X. Vo 和 Louis J. Durlofsky，《用于复杂地质模型高效参数化的正则化核主成分分析》，《计算物理学杂志》322 (2016)，859–881。MR 3534893[WGLP19] Ning Wang、Shuangkui Ge、Baobin Li 和 Lizhong Peng，《基于多小波的多描述图像压缩》，《国际小波、多分辨率与信息处理杂志》17 (2019)，第 1 期，1850063，22。MR 3911884[WSH21] Tizian Wenzel、Gabriele Santin 和 Bernard Haasdonk，《一类新型稳定化贪心核近似算法：收敛性、稳定性与均匀点分布》，《逼近理论杂志》262 (2021)，105508。[WYLZ23] Hao Wang、Chunyu Yang、Xiaomin Liu 和 Linna Zhou，《具有全状态约束的不确定多输入多输出奇异摄动系统的基于神经网络的自适应控制》，《IEEE 神经网络与学习系统汇刊》34 (2023)，第 7 期，3764–3774。MR 4616401[ZBB04] Laurent Zwald、Olivier Bousquet 和 Gilles Blanchard，《核主成分分析的统计性质》，《学习理论》，计算机科学讲义，第 3120 卷，Springer，柏林，2004 年，第 594–608 页。MR 2177937

(Palle E.T. Jorgensen) 数学系，爱荷华大学，爱荷华城，IA 52242-1419，美国电子邮件地址: palle-jorgensen@uiowa.edu (Myung-Sin Song) 数学与统计系，南伊利诺伊大学爱德华兹维尔分校，爱德华兹维尔，IL 62026，美国电子邮件地址: msong@siue.edu (James Tian) 《数学评论》，535 W. William St, Suite 210, 安娜堡，MI 48103，美国电子邮件地址: james.ftian@gmail.com