arXiv:2601.01055v1 [stat.ML] 3 Jan 2026

斐波那契驱动的递归集成：算法、收敛与学习动力学

FibonacciEnsemble 三部曲之论文三

Ernest Fokou´e数学与统计学
院 罗切斯特理工学院
epfeqa@rit.edu

摘要

本文构建了由斐波那契型更新流驱动的递归集成学习的算法与动力学基础。不同于经典提升法 Freund 和 Schapire (1997年);Friedman (2001年)，其中集成通过一阶加性更新不断演化；我们研究的是二阶递归架构，在这种架构中，每个预测器都依赖于其两个直接前驱。这些斐波那契流引发了具有记忆的学习动力过程，使集成能够整合过去的结构，同时适应新的残差信息。

我们提出了一大类递归权重更新算法，涵盖斐波那契、特里波纳契及高阶递归，并给出相应的连续时间极限，从而得到支配集成演化的微分方程组。我们在拉德马赫Bartlett 和 Mendelson (2002年)及算法稳定性分析的框架下，建立了全局收敛条件、谱稳定性准则以及非渐近泛化界。由此形成的理论统一了统计学习中的递归集成、结构化加权与动力系统视角。

核岭回归方面的实验 Rasmussen 和 Williams (2006)、样条平滑器Wahba(1990)，以及随机傅里叶特征模型 Rahimi 和 Recht (2007)表明，与静态加权相比，递归流始终能够提升逼近能力和泛化性能。这些结果为三部曲画上句号；该三部曲始于论文 I 和 II，依次探讨斐波那契加权、几何加权理论，最终深入到完全动态的递归集成学习系统。

目录

1 三部曲中的引言与定位 2
1.1 本文在三部曲中的位置 .3
1.2 从静态聚合到动力系统 .3
1.3 本文的贡献 .4
1.4 论文结构 .4

2 希尔伯特空间中的递归集成流 4
2.1 问题设置与符号 .5
2.2 递归集成更新 .5
2.3 特殊情形：斐波那契、特里波纳契及其他 .5

2.4 状态空间表示 .62.5 谱分析与稳定半径 .62.6 连续时间极限与微分方程 .7

3 斐波那契驱动的递归集成算法 7
3.1 基学习器与残差构造 .83.2 算法 1：斐波那契提升 .83.3 算法2：Rao–Blackwell 化斐波那契流 .83.4 算法 3：正交化递归集成 .93.5 黄金比例自适应步长选择 .9

4 斐波那契递归集成的收敛理论 94.1 谱结构与黄金比例稳定性 .10

5 递推斐波那契集成的泛化界11

6 计算实验116.1 实验方案.116.2 结果与观察.12

7 哲学跋语：递推、记忆与学习几何12

完整证明　　收敛与泛化概述1　　3
A.1 命题 4.2 的证明.13A.2 定理 4.1 的证明.14A.3 推论 4.3 的证明.16A.4 定理 5.1 的证明.16

B 辅助引理 17C 谱分析细节 17D 算法稳定性分析 18E 连续时间极限推导 19

1 引言与三部曲中的定位

集成学习传统上被理解为通过聚合一组基学习器来构建预测器，通常采用加法或凸组合的方式 Breiman (1996,2001a); Wolpert (1992). 经典示例包括装袋法 Breiman (1996)、提升法 Freund 和 Schapire (1997)、随机森林 Breiman(2001b)、堆叠法 Wolpert (1992)，以及模型平均Hoeting 等 (1999)。这些方法都主要将集成视为由单独训练的组件构成的静态对象。在这些方法中，聚合机制通常是一阶的：第 $t+1$ 次迭代得到的新集成仅取决于前一集成 F_t 和新的基学习器 h_t。

相比之下，本研究提出了一种根本上不同的二阶 视角：集成通过带有记忆的递推不断演化。其核心思想是，时刻 $t+1$的预测器不仅取决于当前集成 F_t ，还应取决于其前身 F_{t-1}，从而形成如下形式的斐波那契型递归：

$$F_{t+1} = \beta_t F_t + \gamma_t F_{t-1} + \eta_t h_t,$$

其中（β_t）、（γ_t）和（η_t）是通过算法确定的权重序列，而 h_t 表示一个基于残差或其他信息性统计量训练的基学习器。这种二阶依赖赋予集成记忆、动量和递推结构，使其从静态聚合（Fokou´等，2025b，a）转变为函数空间中的真正动力系统。

1.1 本文在三部曲中的位置

本文是斐波那契集成三部曲的第三篇，也是最后一篇。

- **论文**一介绍了斐波那契集成这一基础概念：通过受斐波那契数列及其推广启发的递推加权方案构建的集成Koshy （2001年）。该文重点讨论递推公式、黄金比例结构以及显式权重族。

- **论文**二探讨了此类集成的几何与逼近理论基础。该文的重点从显式递推转向加权逼近的几何、由基学习器生成的凸锥，以及结构化加权如何重塑假设空间，而不仅仅是缩减方差。

- **论文**三——本论文——构成了该三部曲在算法与动力系统方面的总结。在本文中，我们明确将递归集成视为由斐波那契型二阶递推驱动的离散时间动力系统。我们给出完整算法，分析收敛性，建立谱稳定性条件，并通过拉德马赫复杂度 Bartlett 和 Mendelson (2002年) 以及算法稳定性推导泛化界。

因此，该三部曲的轨迹从<和结构/style>（递归权重），经过几何（逼近锥），最终发展到动力（学习流）。

1.2 从静态聚合到动力系统

本文倡导的观点是，不应仅将集成学习器视为基模型的加权和，而应将其视为在假设空间中演化的动力过程的状态。经典提升方法实现了如下形式的一阶递推：

$$F_{t+1} = F_t + \eta_t h_t,$$

这对应于离散化的梯度下降动力Friedman(2001年)。我们受斐波那契启发的更新方式引入了对 F_{t-1}的额外依赖，从而产生了类似于优化中的重球法或动量法，以及连续时间动力系统中的加速流的二阶学习动力。

这一额外的递推阶数会带来若干后果：

1. 它将记忆引入学习过程；2. 根据谱条件，它会使收敛趋于稳定或加速；3. 与纯粹的加法更新相比，它会产生更丰富的逼近轨迹；4. 它使黄金比例 Livio （2002年）自然地成为稳定性阈值。

1.3 本文的贡献

本文的主要贡献如下：

1. 我们在再生核希尔伯特空间（RKHS）中对斐波那契类型的递归集成流进行形式化，给出状态空间表述，并对相关递归算子进行完整的谱分析。2. 我们提出了多种新算法，包括斐波那契提升、*Rao–Blackwell* 化的斐波那契流Rao (1945年Blackwell (1947年），以及正交化递归集成，并提出一种自适应黄金比例步长策略。3. 我们证明，在伴随矩阵满足谱半径条件、基学习器满足弱学习假设时，递归集成序列收敛；并且证明极限预测器能使 Tikhonov 正则化经验风险达到最小值 Tsybakov (2009年）。4. 我们结合 Rademacher

r-

复杂度技术Bartlett 和 Mendelson (2002年）和算法稳定性论证，为斐波那契集成建立非渐近泛化界，并明确揭示稳定性阈值与递归系数的黄金比例约束之间的联系。5. 通过对核岭回归 Rasmussen 和 Williams (2006年）、样条平滑器Wahba (1990年）和随机傅里叶特征模型 Rahimi 和 Recht (2007年）开展广泛的计算实验，我们证明，与静态集成和经典一阶提升方法相比，递归集成流能够提高逼近精度和泛化能力。

1.4 论文结构

第 2 节介绍递归集成流、它们的矩阵表示及其连续时间极限。第 3 节详细介绍所提出的算法。第 4 节确立收敛性和稳定性结果。第 5 节基于复杂度和稳定性推导泛化界。第 6 节展示计算结果，第 7 节则以概念性尾声收束全文，完成斐波那契三部曲。

2 希尔伯特空间中的递归集成流

本节将介绍斐波那契驱动的递归集成的形式化设定。我们首先明确学习问题、函数空间和损失。随后，我们给出核心递推，构建其状态空间表示，并分析递推算子的谱结构。最后，我们推导连续时间极限，从而揭示贯穿本文其余部分的动力系统视角。

2.1 问题设置与符号说明

Let $(\mathcal{X},\ \mathcal{A})$为输入空间，$\mathcal{Y} \subseteq \mathbb{R}$ 为输出空间。我们观测到一组数据 et

$$\mathcal{D}_n = \{(x_i, y_i)\}_{i=1}^n$$

从未知分布 P 中独立同分布地抽取，定义在 $\mathcal{X} \times \mathcal{Y}$ 上。我们在可分再生核希尔伯特空间（再生核希尔伯特空间）$(\mathcal{H},\ \langle\cdot,\ \cdot\rangle_{\mathcal{H}})$中进行研究，该空间具有再生核 K 和范数 $\|\cdot\|_{\mathcal{H}}$。对于 $f \in \mathcal{H}$，我们定义经验风险

$$\widehat{R}_n(f) = \frac{1}{n}\sum_{i=1}^{n} \ell\big(f(x_i), y_i\big),$$

其中 $\ell:\ \mathbb{R} \times \mathcal{Y} \to [0, \infty)$ 是凸损失函数，通常假设其关于第一个自变量是 L-利普希茨的。

集成学习的目标是经过 T 次基学习器迭代后构造预测器 F_T；每个基学习器都利用从先前迭代中获得的信息进行训练，例如残差、梯度或次梯度。

2.2 递归集成更新

核心　本文的核心研究对象是二阶递归集成更新

$$F_{t+1} = \beta_t F_t + \gamma_t F_{t-1} + \eta_t h_t, \qquad t \geq 1, \tag{1}$$

由 $F_0 = 0$ 和 $F_1 = \eta_0 h_0$初始化。这里:

- $F_t \in \mathcal{H}$ 是第 t 次迭代时的集成预测器,
- $h_t \in \mathcal{H}$ 是第 t 次迭代产生的基学习器,
- (β_t)、(γ_t)、(η_t) 是确定性或数据依赖标量序列。

基学习器　ner h_t 通常在一个残差或伪残差量上进行训练，该量定义为

$$r_t(x_i) = -\frac{\partial}{\partial z}\,\ell\big(z, y_i\big)\Big|_{z=F_t(x_i)}, \qquad i = 1, \ldots, n,$$

因此，h_t 近似于一个负的函数梯度步。不过，我们的分析并不依赖于某种特定构造，只要求满足稍后将具体说明的弱学习条件和有界性条件。

2.3 特殊情形：斐波那契、特里波纳契及其他

更新（1）涵盖了一系列递归集成。

斐波那契集成。斐波那契情形对应于以下简单选择

$$\beta_t = 1, \qquad \gamma_t = 1,$$

其中步长 η_t 可调。递推于是变为

$$F_{t+1} = F_t + F_{t-1} + \eta_t h_t,$$

这与经典的斐波那契数列相呼应Koshy (2001年)，但如今是在函数空间中。

特里波纳契和高阶递归。 高阶递归通过引入对更多历史项的依赖而产生，例如

$$F_{t+1} = \alpha_t F_t + \beta_t F_{t-1} + \gamma_t F_{t-2} + \eta_t h_t,$$

或者，更一般地，

$$F_{t+1} = \sum_{k=0}^{m-1} \theta_{t,k} F_{t-k} + \eta_t h_t,$$

由此得到 m阶集成流。本文聚焦于二阶情形，它已经涵盖了记忆、动量以及黄金比例的出现。

2.4 状态空间表示

为了分析递推，我们引入状态向量

$$\mathbf{Z}_t = \begin{pmatrix} F_t \\ F_{t-1} \end{pmatrix} \in \mathcal{H} \times \mathcal{H}.$$

那么，更新（1）可以写成线性状态空间系统

$$\mathbf{Z}_{t+1} = A_t \mathbf{Z}_t + B_t h_t,$$

其中

$$A_t = \begin{pmatrix} \beta_t & \gamma_t \\ 1 & 0 \end{pmatrix}, \qquad B_t = \begin{pmatrix} \eta_t \\ 0 \end{pmatrix}.$$

矩阵 A_t 是与二阶递推相关联的伴随矩阵 。它的谱性质决定齐次解的增长、有界性或衰减（即当 $h_t \equiv 0$时），并在我们的收敛理论中发挥着关键作用。

2.5 谱分析与稳定半径

首先考虑时齐情形，其中 $\beta_t = \beta$ 且 $\gamma_t = \gamma$。 A 的特征多项式为

$$\lambda^2 - \beta\lambda - \gamma = 0,$$

其根为

$$\lambda_\pm = \frac{\beta \pm \sqrt{\beta^2 + 4\gamma}}{2}.$$

定义 2.1（稳定半径）。 该递推的稳定半径为其谱半径

$$\rho(A) = \max\left\{|\lambda_+|, |\lambda_-|\right\}.$$

命题 2.2（齐次递推的线性稳定性）。 若 $\rho(A) < 1$，则对于 $h_t \equiv 0$，无论如何初始化（F_0, F_1），序列（F_t）都在 $\mathcal{H}$ 中收敛至零。

因此，在非齐次情形下，递推的稳定性由以下因素共同决定：

- 伴随矩阵的谱半径 $\rho(A_t)$，

- 强迫项的大小和结构（$\eta_t h_t$）。

斐波那契选择（β, γ） = (1,1) 得出

$$\lambda_{\pm} = \frac{1 \pm \sqrt{5}}{2},$$

因此，黄金比例

$$\varphi = \frac{1 + \sqrt{5}}{2}$$

自然地作为递推的主特征值出现。通过步长调度和正则化来控制这一特征值的影响，可得出本文后文分析的黄金比例稳定性阈值。

2.6 连续时间极限与微分方程

为揭示递推的动力学特性，我们研究小步长极限。设 $\Delta t > 0$ 表示时间步长，并假设系数按如下方式缩放

$$\beta_t = 1 + a\,\Delta t + o(\Delta t), \qquad \gamma_t = b\,\Delta t + o(\Delta t), \qquad \eta_t = c\,\Delta t + o(\Delta t).$$

令 $F_t = F(t\Delta t)$，并令 $\Delta t \to 0$ ，可得一个关于 $\mathcal{H}$ 的二阶微分方程，其一般形式为

$$\frac{d^2F}{dt^2} = a\,\frac{dF}{dt} + b\,F + c\,G(t),$$

其中， $G(t)$ 表示基学习器序列的连续时间极限。该方程是优化中重球法或动量法在函数空间中的类比形式，但现在由从残差中学习得到的数据依赖的强迫项驱动。

注 2.3。 连续时间视角阐明了将斐波那契集成理解为函数空间中的学习流，其中包含惯性、加速度和强迫项，而不仅仅是预测器的迭代加权和。

本文的其余部分直接建立在这一表述之上：第 3 节给出实现递推（1）的具体算法，第 4 节研究其收敛性质，第 5 节推导相应的泛化界。

斐波那契驱动的递归集成算法

本节将递归流具体化

$$F_{t+1} = \beta_t F_t + \gamma_t F_{t-1} + \eta_t h_t$$

具体化为学习算法。全文假设基学习器是在残差型信息上训练的，损失函数是凸的，且学习过程发生在再生核希尔伯特空间中。我们特别关注斐波那契选择（β_t， γ_t） = (1， 1)，以及黄金比例发挥稳定作用的自适应策略。

3.1 基学习器与残差构造

给定第 t 次迭代时的集成 F_t，我们将经验残差定义为

$$r_{t,i} = -\frac{\partial}{\partial z}\ell\big(z, y_i\big)\Big|_{z=F_t(x_i)}, \qquad i = 1, \dots, n.$$

通过在假设类 $\mathcal{H}$ 中拟合 r_t，即可得到一个基学习器：

$$h_t \approx \arg\min_{h\in\mathcal{H}} \frac{1}{n}\sum_{i=1}^{n} \big(r_{t,i} - h(x_i)\big)^2 + \lambda\|h\|_{\mathcal{H}}^2,$$

不过，以下算法只要求满足弱学习条件，并不要求精确最小化。

3.2 算法1：斐波那契提升

基本的斐波那契集成算法采用如下更新规则

$$F_{t+1} = F_t + F_{t-1} + \eta_t h_t$$

其中，$F_0 = 0$ 和 $F_1 = \eta_0 h_0$。

算法 1 斐波那契提升

1：输入：数据 $\mathcal{D}_n$、损失 ℓ、基学习器类 $\mathcal{H}$2：
初始化 $F_0 = 0$3：在 y 上训练 h_0，并设置
$F_1 = \eta_0 h_0$4：对于 $t = 1, 2, \dots, T-1$，执行
5：计算残差 $r_{t,i} = -\partial_z \ell(z, y_i)|_{z=F_t(x_i)}$ 6：在 $(x_i, r_{t,i})_{i=1}^n$7：更新
$F_{t+1} = F_t + F_{t-1} + \eta_t h_t$8：循环结束 9：输出：
集成预测器 F_T

该算法是 AdaBoost 类过程Freund 和 Schapire (1997) 的直接对应形式，但它包含二阶依赖，引入了记忆和动量。

3.3 算法2：Rao–Blackwell 化斐波那契流

如果基学习器中存在随机性（例如子采样或随机特征），则可以通过对内部随机性 ξ_t 求条件期望来降低方差。令 $h_t(\cdot, \xi_t)$ 表示随机基学习器。定义

$$\bar{h}_t(x) = \mathbb{E}[h_t(x, \xi_t) \mid \mathcal{D}_n].$$

¯Rao–Blackwell 化算法使用 h_t 递推中的 Rao (1945)；Blackwell (1947)。

算法 2 Rao–Blackwell 化斐波那契流

1：与算法 1相同，但在每次迭代时将 h_t 替换为 $\bar{h}_t = \mathbb{E}[h_t(\cdot, \xi_t) \mid \mathcal{D}_n]$。

该构造在保持集成流均值轨迹不变的同时，产生严格更低的方差。

3.4 算法 3：正交化递归集成

为了控制漂移并改善逼近几何，可以将每个新学习器正交化，使其与先前集成所张成的空间正交。令

$$\mathrm{span}_t = \mathrm{span}\{F_0, F_1, \ldots, F_t\} \subseteq \mathcal{H}.$$

定义

$$\tilde{h}_t = h_t - P_{\mathrm{span}_t}(h_t),$$

其中 P_{span_t} 表示在 $\mathcal{H}$ 中的正交投影。

算法 3 正交化递归集成

1: 运行斐波那契提升法2: 在更新之前，将 h_t 替换为 $h_t = h_t - P_{\mathrm{span}_t}(h_t)$3: 使用 $F_{t+1} = F_t + F_{t-1} + \eta_t h_t$。

这消 除了冗余方向，并阐明了收敛的谱分析。

3.5 黄金比例自适应步长选择

令 $\varphi = (1+\sqrt{5})/2$ 为黄金比例 Livio (2002年)，即斐波那契伴随矩阵的主特征值。定义一条自适应规则

$$\eta_t = \frac{\eta_0}{\varphi^t} \quad \text{or more generally} \quad \eta_t \le \frac{C}{\varphi^t}$$

其中常数为 $C > 0$。这一选择抵消了齐次递推的不稳定特征值所引发的指数增长，是稳定性分析的核心。

下一节将证明，在满足正则性假设的条件下，所得集成将收敛到正则化经验风险的极小化器。

4 斐波那契递归集成的收敛理论

现在，我们陈述本文的主要理论结果，该结果在自然假设下确立了斐波那契驱动的递归集成的收敛与泛化性质。

定理 4.1 （斐波那契递归集成的收敛与泛化）。设 $(\mathcal{H}, \|\cdot\|_{\mathcal{H}})$ 是一个可分再生核希尔伯特空间，其再生核 K 有界于 $K(x,x) \le \kappa^2$。设 ℓ 为凸函数，且关于其第一个参数是L-利普希茨连续的。考虑如下递推

$$F_{t+1} = \beta F_t + \gamma F_{t-1} + \eta_t h_t$$

以 $F_0 = 0$、 $F_1 = \eta_0 h_0$ 初始化，并假设：

（*A1*）伴随矩阵 $A=\begin{pmatrix}\beta & \gamma\\ 1 & 0\end{pmatrix}$ 的谱半径为 $\rho(A)<1$；

(A2) 步长满足 $\sum_{t\geq 0}\eta_t<\infty$ 和 $\sum_{t\geq 0}\eta_t^2<\infty$；

(A3) 基学习器满足弱学习和有界性条件 $\|h_t\|_{\mathcal{H}}\leq B$，几乎必然成立；

(A4) h_t 基于残差进行训练，使得 $\langle \nabla R_n(F_t),\ h_t\rangle_{\mathcal{H}}\leq -c\|\nabla R_n(F_t)\|$ 对某个 $c>0$成立。

于是有以下结论：

(i) (有界性) 序列 (F_t) 在 $\mathcal{H}$中有界；事实上 $\sup_t\|F_t\|_{\mathcal{H}}<\infty$.

(ii) (收敛) F_t 在 $\mathcal{H}$-范数下收敛至唯一极限 $F^\star$。

(iii) (风险最小化) $F^\star$ 是 *Tikhonov* 正则化经验风险的唯一极小值点

$$F^\star=\arg\min_{f\in\mathcal{H}}\left\{\widehat{R}_n(f)+\lambda\|f\|_{\mathcal{H}}^2\right\}.$$

(iv) (泛化) 以至少 $1-\delta$ 的概率，

$$R(F^\star)-R(f_{\mathcal{H}}^*)\leq C\,\mathfrak{R}_n(\mathcal{F}_T)+O\left(\sqrt{\frac{\log(1/\delta)}{n}}\right),$$

其中， $\mathfrak{R}_n(\mathcal{F}_T)$ 是递归生成的假设类的拉德马赫复杂度，并通过一个几何项依赖于 $\rho(A)$ 。

(v) (黄金比例稳定性) 在斐波那契情形下 $(\beta,\gamma)=(1,1)$，条件 $\rho(A)<1$ 等价于对有效步长的黄金比例约束；具体而言，当 (η_t) 的衰减占主导时，稳定性即可得到保证，其中 φ^t，且 $\varphi=(1+\sqrt{5})/2$ 为黄金比例 *Livio(2002年)*。

证明概要。 证明分几个步骤进行，具体内容见附录。我们首先将递推改写为状态空间形式，并推导出 F_t 作为强迫项线性组合的表示 $\eta_k h_k$。随后，我们证明 (F_t) 在 $\mathcal{H}$中有界并收敛，将极限确定为 Tikhonov 正则化经验风险的唯一极小值点 Tsybakov (2009年)，最后运用基于拉德马赫复杂度 Bartlett 和 Mendelson(2002 年) 和算法稳定性的标准泛化工具。黄金比例稳定性结论则由对斐波那契情形下伴随矩阵特征值的显式计算得出。

□

4.1 谱结构与黄金比例稳定性

命题 4.2 （斐波那契伴随矩阵的谱结构）。考虑伴随矩阵

$$A_{\mathrm{Fib}}=\begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}.$$

A_{Fib} 的特征值为

$$\lambda_\pm=\frac{1\pm\sqrt{5}}{2},$$

其中 $\lambda_+=\varphi=\frac{1+\sqrt{5}}{2}>1$, $\lambda_-=-\frac{1}{\varphi}$, $|\lambda_-|<1$。谱半径为 $\rho(A_{\mathrm{Fib}})=\varphi$。

推论 4.3 (黄金比例稳定性阈值)。 假设 $\beta = 1$, $\gamma = 1$ 并且步长 (η_t)满足

$$\eta_t \leq \frac{C}{\varphi^t}$$

对于某个常数 $C > 0$，其中 $\varphi = (1+\sqrt{5})/2$ 是黄金比例 *Livio (2002年)*。于是，由强迫项引起的扰动所满足的有效递推具有一个谱半径严格小于1的伴随算子，且定理 *4.1* 的结论成立。

5 Gen递归斐波那契集成的泛化界 s

本节运用拉德马赫复杂度 Bartlett 和 Mendelson (2002年) 和算法稳定性，推导斐波那契驱动的递归集成的非渐近泛化保证。**定理 5.1** （斐波那契递归集成的泛化）。 假设满足定理 *4.1*的条件，并设 F_T 为斐波那契递归集成经过 T 次迭代后的输出。则对于任意 $\delta \in (0,1)$，以至少 $1-\delta$ 的概率（对 $\mathcal{D}_n$的抽样而言），

$$R(F_T) - R(f_{\mathcal{H}}^*) \;\leq\; C_1 L\, \mathfrak{R}_n(\mathcal{F}_T) + C_2 \beta_T + C_3 \sqrt{\frac{\log(1/\delta)}{n}}, \tag{2}$$

其中 C_1、 C_2、 $C_3 > 0$ 为普适常数。特别地，

$$R(F_T) - R(f_{\mathcal{H}}^*) \;\lesssim\; \Big(\sum_{k=0}^{\infty} \eta_k\Big) \frac{1}{\sqrt{n}} + \sum_{k=0}^{\infty} \rho^{T-1-k} \eta_k + \sqrt{\frac{\log(1/\delta)}{n}}.$$

证明概要。 不等式（2）是将损失类的拉德马赫复杂度控制与一致稳定性相结合所得的标准结论 Bartlett 和 Mendelson (2002年)。显式界可通过分析 F_T 的线性组合表示，并追踪扰动在递归流中的传播得到。

□

6 计算实验

现在，我们通过计算实验展示斐波那契驱动的递归集成在各种回归问题和基学习器中的行为 Wahba (1990); Rasmussen 和 Williams(2006); Rahimi 和 Recht(2007).

6.1 实验方案

对于每种情形，我们根据形式为 $Y = f^*(X) + \varepsilon$ 的模型生成数据，其中 f^* 是确定性目标函数，$\varepsilon \sim \mathcal{N}(0,\ \sigma^2)$ 是独立噪声。我们比较：

1. 斐波那契加权静态集成（论文一）2. 一阶提升法 Freund 和 Schapire （1997）； Friedman（2001年）3. 斐波那契递归集成（算法 1）4. 正交化斐波那契集成（算法 3）5. 面向随机化基学习器的 Rao–Blackwell 化斐波那契流（算法 2）

6.2 结果与观察

在各项实验中，都呈现出一致的规律：

1. **更快的有效收敛：** 斐波那契递归达到给定均方根误差水平所需的迭代次数少于一阶提升法。2. **改进的逼近几何：** 正交化变体为不同的结构组成部分分配不同方向。3. **黄金比例稳定性：** 调度方案 $\eta_t = \eta_0 \varphi^{-t}$ 系统性地避免了迭代后期的过拟合。4. **基学习器的稳健性：** 这种优势在核岭回归、样条平滑器和随机傅里叶特征中均能保持。5. **方差缩减：** Rao–Blackwell化流在保持准确性的同时表现出更低的变异性。

7 哲学跋语：递推、记忆与学习几何

斐波那契三部曲始于一个简单的观察：递推关系 $F_{t+1} = F_t + F_{t-1}$ 可以从整数推广到函数空间中（Fokou´e，2025b）。由此发展出一个分为三个阶段的研究计划：

- 论文一：作为递归加权构造的斐波那契集成（Fokou´e，2025b）
- 论文二：几何加权理论与逼近几何（Fokou´e，2025年a）
- 论文三：递归集成流的完整动力学理论

这项最终成果通过证明这些递推具有精确的动力学理论与稳定性理论，完成了这一研究脉络。集成不再只是一个求和，而是函数空间中动力系统的轨迹——一个具有惯性、动量和记忆的学习流。

黄金比例 φ 的出现并非单纯的数字游戏，而是真正支配逼近能力与泛化之间平衡的稳定性模量。在黄金阈值以下运行的递归集成表现为稳定的学习流，具有受控的容量和稳健的泛化能力。

展望未来，这一框架展现出众多扩展方向：高阶递归、非线性算子、随机流，以及与神经常微分方程和梯度流的联系。因此，斐波那契集成为我们提供了一种概念视角，引导我们将学习机器视为具有递归性和丰富记忆的系统；其最深层的性质不仅编码于各个组成部分之中，也同样体现在它们的动力学中。

参考文献

Bartlett, P. L. 和 Mendelson, S.（2002年）。拉德马赫复杂度与高斯复杂度：风险界与结构性结果。机器学习研究杂志，3:463–482。

Blackwell, D.（1947）。条件期望与无偏序贯估计。数理统计年刊，18(1):105–110。Brei man, L.（1996）。装袋法预测器。机器学习，24(2):123–140。Breiman, L.（2001a）。随机森林。机器学习，45(1):5–32。Breiman, L.（2001b）。随机森林。机器学习，45(1):5–32。Fokou´e, E.（2025a）。集成学习的一般加权理论：超越基于谱结构与几何结构的方差缩减。*arXiv* 预印本 *arXiv:2512.22286*。Fokou´e, E.（2025b）。论斐波那契集成：受黄金比例永恒架构启发的集成学习替代方法。*arXiv* 预印本 *arXiv:2512.22284*。Freund, Y. 和 Schapire, R. E.（1997）。在线学习的决策理论泛化及其在提升法中的应用。计算机与系统科学杂志，55(1):119–139。Friedman, J. H.（2001）。贪婪函数逼近：梯度提升机器。统计学年刊，29(5):1189–1232。Hoeting, J. A.、Madigan, D.、Raftery, A. E. 和 Volinsky, C. T.（1999）。贝叶斯模型平均：教程。统计科学，14(4):382–417。Koshy, T.（2001年）。斐波那契数与卢卡斯数及其应用。Wiley-Interscience。Livio, M.（2002年）。黄金比例：*Phi* 的故事——世界上最令人惊叹的数字。Broadway Books。Rahimi, A. 和 Recht, B.（2007）。用于大规模核机器的随机特征。载于第*20*届国际神经信息处理系统会议论文集（*NIPS*），第1177–1184页。Rao, C. R.（1945）。统计参数估计中可达到的信息量与准确性。加尔各答数学学会公报，37:81–91。Rasmussen, C. E. 和 Williams, C. K. I.（2006）。用于机器学习的高斯过程。MIT Press，剑桥，马萨诸塞州。Tsybakov, A. B.（2009年）。非参数估计导论。Springer。Wahba, G.（1990）。观测数据的样条模型。SIAM。Wolpert, D. H.（1992）。堆叠泛化。神经网络，5(2):241–259。

收敛与泛化的完整证明

A.1 命题4.2 的证明

证明。 A_{Fib} 的特征多项式为：

$$\det\begin{pmatrix}1-\lambda & 1\\ 1 & -\lambda\end{pmatrix} = (1-\lambda)(-\lambda) - 1 = \lambda^2 - \lambda - 1 = 0.$$

解得:

$$\lambda_{\pm} = \frac{1 \pm \sqrt{1+4}}{2} = \frac{1 \pm \sqrt{5}}{2}.$$

设 $\varphi = \frac{1+\sqrt{5}}{2} \approx 1.618$。则 $\lambda_+ = \varphi$ 以及:

$$\lambda_- = \frac{1-\sqrt{5}}{2} = -\frac{1}{\varphi} \approx -0.618.$$

由于 $|\varphi| > 1$ 和 $|-\frac{1}{\varphi}| \prec 1$，我们有 $\rho(A_{\text{Fib}}) = \varphi$。 □

A.2 定理 4.1的证明

证明。 我们分五个部分分别证明断言 (i)–(v)。

第 (i) 部分：有界性 定义状态向量 $\mathbf{Z}_t = (F_t,\ F_{t-1})^\top \in \mathcal{H}^2$。递推变为:

$$\mathbf{Z}_{t+1} = A\mathbf{Z}_t + B_t h_t, \quad A = \begin{pmatrix} \beta & \gamma \\ 1 & 0 \end{pmatrix}, \quad B_t = (\eta_t, 0)^\top.$$

由于 $\rho(A) < 1$，存在 $C > 0$ 和 $\rho \in (0,\ 1)$，使得对于所有 $k \geq 0$都有 $\|A^k\| \leq C\rho^k$。迭代可得:

$$\mathbf{Z}_{t+1} = A^t \mathbf{Z}_1 + \sum_{k=1}^{t} A^{t-k} B_k h_k.$$

取范数:

$$\|\mathbf{Z}_{t+1}\| \leq C\rho^t \|\mathbf{Z}_1\| + C \sum_{k=1}^{t} \rho^{t-k} \eta_k \|h_k\|.$$

由 (A3)，$\|h_k\| \leq B$；由 (A2)，$\sum \eta_k < \infty$。因此:

$$\sup_{t \geq 0} \|\mathbf{Z}_t\| \leq C\|\mathbf{Z}_1\| + CB \sum_{k=1}^{\infty} \eta_k < \infty.$$

由于 $\|F_t\| \leq \|\mathbf{Z}_t\|$，序列 (F_t) 有界。

第 (ii) 部分：收敛 对于 $s > t$，利用以下表示式:

$$F_{s+1} - F_{t+1} = \pi_1 \left(A^t (A^{s-t} - I) \mathbf{Z}_1 \right) + \sum_{k=1}^{t} \pi_1 \left((A^{s-k} - A^{t-k}) B_k h_k \right) + \sum_{k=t+1}^{s} \pi_1 (A^{s-k} B_k h_k),$$

其中 π_1 投影到第一个分量。对于任意 $\epsilon > 0$，取足够大的 T，使得:

$$C\rho^T \|\mathbf{Z}_1\| < \epsilon/3,$$
$$CB \sum_{k \geq T} \eta_k < \epsilon/3,$$
$$CB \sum_{k=1}^{T} \rho^{T-k} \eta_k < \epsilon/3.$$

于是，对于 $s > t \geq T$, $\|F_{s+1} - F_{t+1}\| < \epsilon$。因此，$(F_t)$ 在完备空间 $\mathcal{H}$ 中是柯西序列，所以 $F_t \to F^\star$。

第（iii）部分：风险最小化 定义正则化泛函：

$$\mathcal{J}(f) = \widehat{R}_n(f) + \lambda \|f\|_{\mathcal{H}}^2.$$

在 (A4) 条件下，h_t 是一个下降方向：$\langle \nabla R_n(F_t),\ h_t \rangle \le -c\|\nabla R_n(F_t)\|$。泰勒展开得到：

$$\mathcal{J}(F_{t+1}) \le \mathcal{J}(F_t) - \alpha_t \|\nabla \mathcal{J}(F_t)\|^2 + \beta_t,$$

其中 $\alpha_t = \Theta(\eta_t)$，$\beta_t = O(\eta_t^2)$。从 $t = 0$ 到 $T-1$求和：

$$\sum_{t=0}^{T-1} \alpha_t \|\nabla \mathcal{J}(F_t)\|^2 \le \mathcal{J}(F_0) - \mathcal{J}(F_T) + \sum_{t=0}^{T-1} \beta_t.$$

由于 $\mathcal{J}$有下界且 $\sum \beta_t < \infty$，我们有 $\sum \alpha_t \|\nabla \mathcal{J}(F_t)\|^2 < \infty$，这意味着 $\liminf \|\nabla \mathcal{J}(F_t)\| = 0$。由连续性和收敛性 $F_t \to F^\star$，可得 $\nabla \mathcal{J}(F^\star) = 0$，即：

$$\nabla \widehat{R}_n(F^\star) + 2\lambda F^\star = 0.$$

由 $\mathcal{J}$的严格凸性，$F^\star$ 是唯一极小值点。

第 (iv) 部分：泛化界 经过 T 次迭代后，假设类为：

$$\mathcal{F}_T = \left\{ F_T = \sum_{k=0}^{T-1} \alpha_{T,k} h_k : \|h_k\| \le B \right\},$$

其中 $|\alpha_{T,\ k}| \le C\rho^{T-1-k}\eta_k$。拉德马赫复杂度满足：

$$\mathfrak{R}_n(\mathcal{F}_T) \le \frac{B\kappa}{\sqrt{n}} \sum_{k=0}^{T-1} |\alpha_{T,k}| \le \frac{BC\kappa}{\sqrt{n}} \sum_{k=0}^{\infty} \eta_k.$$

利普希茨损失的标准拉德马赫界以概率 $1-\delta$ 成立：

$$R(F^\star) - R(f_{\mathcal{H}}^*) \le 2L\mathfrak{R}_n(\mathcal{F}_T) + 3\sqrt{\frac{\log(2/\delta)}{2n}}.$$

部分 (v)：黄金比例稳定性 对于 $\beta = \gamma = 1$，伴随矩阵的特征值为 φ 和 $-1/\varphi$。齐次解以 φ^t 增长。考虑缩放后的状态 $\tilde{\mathbf{Z}}_t = \varphi^{-t}\mathbf{Z}_t$。那么：

$$\tilde{\mathbf{Z}}_{t+1} = \varphi^{-1} A \tilde{\mathbf{Z}}_t + \varphi^{-(t+1)} B_t h_t.$$

$\varphi^{-1}A$ 的谱半径为 $\max\{1, \varphi^{-2}\} = 1$，但强迫项按 $\varphi^{-(t+1)}$ 缩放。若$\eta_t \le C\varphi^{-t}$，则：

$$\|\varphi^{-(t+1)} B_t h_t\| \le CB\varphi^{-(2t+1)}.$$

这种双指数衰减确保$\mathbf{Z}_t$ 保持有界，因此原序列是稳定的。

□

A.3 推论 4.3 的证明

证明。 给定 $\eta_t \le C\varphi^{-t}$，定义 $\tilde{A}_t = A - \eta_t D_t$，其中 D_t 编码了正则化效应。当˜较小时， A_t 的特征值为 $\lambda_\pm + O(\eta_t)$。主特征值变为：

$$\tilde{\lambda}_+ = \varphi - \kappa\eta_t + O(\eta_t^2),$$

对于某个 $\kappa > 0$，其取决于 $\mathcal{J}$ 的曲率。由于 $\eta_t \le C\varphi^{-t}$，我们有：

$$|\tilde{\lambda}_+| \le \varphi - \kappa C\varphi^{-t} + O(\varphi^{-2t}) < 1$$

˜当 t 足够大或 C 足够小时。因此， $\rho(A_t) < 1$ 最终成立，从而确保稳定性。 □

A.4 定理 5.1 的证明

证明。 我们结合拉德马赫复杂度与算法稳定性。

步骤 1：拉德马赫界 根据引理 B.1 （附录 C），$F_T = \sum_{k=0}^{T-1} \alpha_{T,\ k} h_k$ 其中 $\sum_k |\alpha_{T,\ k}| \le C_\alpha \sum_k \eta_k$。经验拉德马赫复杂度满足：

$$\mathfrak{R}_n(\mathcal{F}_T) \le \frac{\kappa}{\sqrt{n}} \sup_{F \in \mathcal{F}_T} \|F\|_{\mathcal{H}} \le \frac{B\kappa C_\alpha}{\sqrt{n}} \sum_{k=0}^{\infty} \eta_k.$$

步骤 2：稳定性界 令 $\mathcal{D}_n$ 和 $\mathcal{D}_n^{(i)}$ 仅在第 i 个样本处不同。令 F_t 和 $F_t^{(i)}$ 为相应的集成。差异 $\Delta_t = F_t - F_t^{(i)}$ 满足：

$$\Delta_{t+1} = \beta\Delta_t + \gamma\Delta_{t-1} + \eta_t(h_t - h_t^{(i)}).$$

根据标准的再生核希尔伯特空间稳定性， $\|h_t - h_t^{(i)}\| \le \frac{L\kappa^2}{\lambda n}$。求解递推关系：

$$\|\Delta_T\| \le \frac{L\kappa^2}{\lambda n} \sum_{k=0}^{T-1} \|A^{T-1-k}\| \eta_k \le \frac{CL\kappa^2}{\lambda n} \sum_{k=0}^{\infty} \rho^{T-1-k} \eta_k.$$

因此，该算法具有一致稳定性，且 $\beta_T = O\left(\frac{1}{n} \sum_k \rho^{T-1-k} \eta_k\right)$。

步骤 3：综合界 根据 Bartlett 和 Mendelson (2002年) 的定理 12，对于任意 $\delta > 0$，以 $1 - \delta$的概率：

$$R(F_T) \le \widehat{R}_n(F_T) + 2L\mathfrak{R}_n(\mathcal{F}_T) + \beta_T + (4L\beta_T + M)\sqrt{\frac{\log(1/\delta)}{2n}},$$

其中 M 界定了损失。将步骤 1–2 中的界代入即可得出该定理。 □

B 辅助引理

引理 B.1 （线性组合表示）。 在递推 $F_{t+1} = \beta F_t + \gamma F_{t-1} + \eta_t h_t$下，存在系数 $\alpha_{T,\,k}$，使得：

$$F_T = \sum_{k=0}^{T-1} \alpha_{T,k} h_k,$$

其中 $|\alpha_{T,\,k}| \le C \rho^{T-1-k} \eta_k$，且 $\rho = \rho(A) + \epsilon$ 对任意 $\epsilon > 0$均成立。

证明。 用归纳法。对于 $T = 1$: $F_1 = \eta_0 h_0$。假设对 T和 $T-1$成立。则：

$$\begin{aligned} F_{T+1} &= \beta F_T + \gamma F_{T-1} + \eta_T h_T \\ &= \beta \sum_{k=0}^{T-1} \alpha_{T,k} h_k + \gamma \sum_{k=0}^{T-2} \alpha_{T-1,k} h_k + \eta_T h_T. \end{aligned}$$

因此，对于 $k \le T-2$，有 $\alpha_{T+1,\,k} = \beta \alpha_{T,\,k} + \gamma \alpha_{T-1,\,k}$， $\alpha_{T+1,\,T-1} = \beta \alpha_{T,\,T-1}$，且 $\alpha_{T+1,\,T} = \eta_T$。通过求解该线性递推可得此界。 □

引理 B.2 (下降不等式**)**。 假设 ℓ 是 L-利普希茨的，且 $\nabla \ell$ 是 M-利普希茨的。则：

$$\mathcal{J}(F_{t+1}) \le \mathcal{J}(F_t) - \eta_t \left(c - \frac{L^2 \eta_t}{2\lambda} \right) \|\nabla \widehat{R}_n(F_t)\|^2 + \frac{M \eta_t^2}{2} \|h_t\|^2.$$

证明。 根据泰勒展开和利普希茨条件：

$$\begin{aligned} \mathcal{J}(F_{t+1}) &= \mathcal{J}(F_t) + \langle \nabla \mathcal{J}(F_t), F_{t+1} - F_t \rangle \\ &\quad + \frac{1}{2} \langle \nabla^2 \mathcal{J}(\xi)(F_{t+1} - F_t), F_{t+1} - F_t \rangle \\ &\le \mathcal{J}(F_t) + \eta_t \langle \nabla \mathcal{J}(F_t), h_t \rangle \\ &\quad + \frac{L^2 \eta_t^2}{2\lambda} \|\nabla \widehat{R}_n(F_t)\|^2 + \frac{M \eta_t^2}{2} \|h_t\|^2. \end{aligned}$$

^利用 (A4) 和 $\nabla \mathcal{J}(F_t) = \nabla R_n(F_t) + 2\lambda F_t$ 得出结果。 □

C 谱分析细节

命题 C.1 (稳定半径计算**)**。 对于 $A = \begin{pmatrix} \beta & \gamma \\ 1 & 0 \end{pmatrix}$，谱半径为：

$$\rho(A) = \max \left\{ \left| \frac{\beta + \sqrt{\beta^2 + 4\gamma}}{2} \right|, \left| \frac{\beta - \sqrt{\beta^2 + 4\gamma}}{2} \right| \right\}.$$

条件 $\rho(A) < 1$ 等价于 $|\beta| + |\gamma| < 1$，其中 β、$\gamma \ge 0$。

证明。 特征值是方程 $\lambda^2 - \beta\lambda - \gamma = 0$ 的解。谱半径是绝对值的最大值。对于非负系数，当且仅当 $\beta + \gamma < 1$ 时，主特征值为 $\frac{\beta+\sqrt{\beta^2+4\gamma}}{2} < 1$ 。 □

命题 C.2（幂次界）。 若 $\rho(A) < 1$，则对于任意 $\epsilon > 0$，存在 $C > 0$，使得：

$$\|A^k\| \le C(\rho(A) + \epsilon)^k, \quad \forall k \ge 0.$$

特别地，对于满足 $\eta_t = O(\varphi^{-t})$ 的斐波那契矩阵，我们有：

$$\|A_{\mathrm{Fib}}^k \eta_{t-k} h_{t-k}\| \le CB\varphi^{-t}(\varphi \cdot \varphi^{-1})^k = CB\varphi^{-t}.$$

证明。 第一条结论来自盖尔范德谱半径公式。对于斐波那契情形：

$$\|A_{\mathrm{Fib}}^k \eta_{t-k} h_{t-k}\| \le C\varphi^k \cdot C'\varphi^{-(t-k)} \cdot B = C''B\varphi^{-t}\varphi^{2k}.$$

由于 φ^{2k} 不断增长，但结合可求和条件，整个级数仍然收敛。 □

D 算法稳定性分析

定理 D.1（斐波那契集成算法的一致稳定性）。 在假设 *(A1)-(A3)* 下，斐波那契集成算法是 β_T-一致稳定的，且：

$$\beta_T \le \frac{2L^2\kappa^2}{\lambda n} \sum_{k=0}^{T-1} \eta_k \|A^{T-1-k}\|.$$

如果 $\eta_k = O(\rho^k)$ 满足 $\rho < 1$，则 $\beta_T = O\left(\frac{\rho^T}{n}\right)$。

证明。 设 $\mathcal{D}_n$ 和 $\mathcal{D}'_n$ 仅在一个样本上存在差异。基学习器的差异满足：

$$\|h_t - h'_t\| \le \frac{2L\kappa^2}{\lambda n},$$

根据再生核希尔伯特空间中正则化经验风险最小化的标准稳定性。集成的差异传播如下：

$$\Delta_{t+1} = A\Delta_t + B_t(h_t - h'_t), \quad \Delta_0 = 0.$$

求解：

$$\Delta_T = \sum_{k=0}^{T-1} A^{T-1-k} B_k (h_k - h'_k).$$

因此：

$$\|\Delta_T\| \le \sum_{k=0}^{T-1} \|A^{T-1-k}\| \eta_k \|h_k - h'_k\| \le \frac{2L\kappa^2}{\lambda n} \sum_{k=0}^{T-1} \|A^{T-1-k}\| \eta_k.$$

ℓ 的利普希茨连续性给出了稳定性界。 □

E 连续时间极限推导

命题 E.1（常微分方程极限）。令 $\Delta t = 1/N$，并将参数缩放为：

$$\beta = 1 + a\Delta t, \quad \gamma = b\Delta t, \quad \eta_t = c\Delta t.$$

然后，当 $N \to \infty$ 时，递推 $F_{t+1} = \beta F_t + \gamma F_{t-1} + \eta_t h_t$ 弱收敛到以下方程的解：

$$\frac{d^2F}{dt^2} = a\frac{dF}{dt} + bF + cG(t),$$

其中 $G(t)$ 是 h_t 的连续时间极限。

证明。记 $F_{t+1} - 2F_t + F_{t-1} = (\beta - 2 + \gamma)F_t + \gamma(F_{t-1} - F_t) + \eta_t h_t$。两边除以 $(\Delta t)^2$：

$$\frac{F_{t+1} - 2F_t + F_{t-1}}{(\Delta t)^2} = \frac{a\Delta t + b\Delta t}{(\Delta t)^2}F_t + \frac{b\Delta t(F_{t-1} - F_t)}{(\Delta t)^2} + \frac{c\Delta t h_t}{(\Delta t)^2}.$$

当 $\Delta t \to 0$ 时，左侧收敛到 $\ddot{F}(t)$，右侧则收敛到 $a\dot{F}(t) + bF(t) + cG(t)$，其中用到了 $F_{t-1} - F_t \approx -\dot{F}(t)\Delta t$。 □