适用于小波变换并具有稳定可逆实现的基于网格的抽取

Nicki Holighaus、Günther Koliander、Clara Hollomey 和 Friedrich Pillichshamm er

arXiv:2301.01640v1 [eess.AS] 4 Jan 2023

摘要——**小波变换具有恒定的中心频率与带宽之比（Q因子），因此为音频数据提供了一种非常自然的表示方式。然而，可逆小波变换要么需要非均匀抽取——这会导致不规则数据结构，难以处理——要么需要过高的过采样率，从而带来无法接受的计算开销。本文提出一种新颖的小波变换抽取策略，可在过采样率接近 1 的情况下实现均匀抽取，并获得稳定的表示。具体而言，我们证明，所得表示的有限实现按照框架理论的意义是能量保持的。得到的小波系数可以存储在一个时频矩阵中，其中列自然对应时间帧，行对应频率通道。这种矩阵结构使我们能够直接使用大量已成功应用于时频音频处理的算法，而这些算法此前无法与小波变换结合使用。我们展示了该方法在基于非负矩阵分解的处理、起音检测和无相位重建中的应用。**

索引词—**小波变换、低差异序列、采样方法、信号重建、音频应用、平移不变系统、均匀抽取**

I. 引言

小波变换是最重要且研究最为深入的时频滤波器组之一，唯一可与之匹敌的是短时傅里叶变换。由于其中心频率与带宽之比恒定，即Q因子 [1], , 《小波》为音频提供了一种自然且具有物理意义的表示：每个滤波器的脉冲响应都包含其中心频率相同数量的波长。恒定的Q因子确保（西方）音阶上由按几何间隔排列的音调组成的相等区间都能得到同等良好的分辨，而与其绝对位置无关。最后，在高于500Hz的频率范围内，外周听觉系统在人类中可以很好地近似为常数Q滤波器组 [2]。简要讨论请参见，例如， [3]。因此，常数Q滤波器组总体而言 [4]–[6] ，尤其是小波变换 [7]–[10]，已广泛成功地用于语音和音频信号的分析。然而，在很长一段时间内，常数Q滤波器组无法用于，或至少不便于以足够高的保真度处理音频：小波基 [11]以及非抽取离散小波系统[12], [13] 具有非常低的Q因子，不适合处理音频。现代常数Q滤波器组实现支持调节Q因子和过采样率，但兼具计算效率、数值稳定性和可逆性的常数Q滤波器组直到最近才得以构建：在 [14] 中，这是通过有目的地修改用于快速小波变换的Mallat算法实现的；在 [15], [16] 中，则是借助数学框架理论 [17]实现的。为了实现完美重构，这些构造依赖于非均匀抽取，即选择足够小的抽取因子，使其与带宽或等价地与常数Q滤波器的中心频率成反比。[1]因此，每个通道都会以不同且通常不匹配的速率生成系数。另一种方法是完全不进行通道内抽取，但这样会导致在较大Q因子下出现高得令人难以接受的过采样。

尽管具有适当频率分辨率（即足够大的Q因子）的可逆表示的出现，是使常数-Q滤波器组更适用于音频处理的重要一步，但另一个根本问题仍然存在，此前已在 [3], 中连同一种部分解决方案一并讨论过：采用非均匀抽取后，滤波器组系数形成了一种不规则数据结构，无论从计算角度还是概念角度来看，都比矩阵更难处理。这类系数与任何依赖矩阵运算的方法都不兼容；要将其整合到现成的音频处理工具箱中，需要付出大量工作，例如 [18] 和 [19], 通常都是在系数具有矩阵结构的假设下实现的。这一影响在无处不在的短时傅里叶变换（STFT）算法中表现得最为明显；这类算法通常依赖于将短时傅里叶变换系数分割为时间帧，见 [20]和 [21]。这些时间帧会被分别（或以小组形式）处理，而且通常是在实时环境中进行的，但这种分割依赖于系数在时间上对齐这一假设。更准确地说，通道间系数关系被假定为与位置无关，

稿件收到日期：XXX；修订日期：XXX 年 8 月。N. Holighaus（通讯作者）和 C. Hollomey 任职于奥地利科学院声学研究所（ARI），地址：奥地利维也纳 1040，Wohllebengasse 12–14。G. Koliander 任职于 ARI 和维也纳大学数学学院，奥地利。F. Pillichshammer 任职于林茨约翰·开普勒大学金融数学与应用数论研究所，奥地利。电子邮件：{nicki.holighaus,guenther.koliander,clara.hollomey} @oeaw.ac.at, friedrich.pillichshammer@jku.ac.at。扩展结果、音频文件以及用于复现实验的代码可在以下网址获取：ltfat.org/notes/057。本研究得到奥地利科学基金（FWF）资助：I3067–N30（N.H.）、Y 1199“时频分析、随机性与采样”（G.K.），以及 F5509-N26（该项目属于“准蒙特卡罗方法：理论与应用”特别研究计划）（F.P.）。版权所有 (c) 2023 IEEE。允许个人使用本文材料。但如需将本文材料用于其他任何目的，必须向 IEEE 发送请求至 pubs-permissions@ieee.org，以取得许可。

[1]这种选择抽取因子的方法并非可逆常数Q变换所独有，而是大多数常数Q变换和小波变换实现所共有的。

而对于非均匀抽取而言，显然并非如此。因此，将针对短时傅里叶变换系数的成功处理方案适配到常数-Q设置仍然具有挑战性。非均匀抽取还会影响重建过程本身，见II-A节。最重要的是，即使是高效的重建算法 [22], [23] 也需要进行代价高昂的迭代，除非所有滤波器都是严格带限的 [24], ，而这会排除有限冲激响应滤波器的使用。

在这项工作中，我们提出一类基于连续小波变换准随机采样的小波滤波器组。这类滤波器组计算效率高，可通过不受限制地选择母小波来调节 Q 值，并且过采样率灵活。过采样率接近 1 时，所提出的滤波器组能够完美重构离散信号。独特之处在于，它们在实现这一目标的同时采用均匀抽取，从而克服了以往可逆恒定 Q 变换因不规则性而引发的困难。

关于时频表示随机与准随机采样的先前工作：本研究并非首个探讨时频表示或更一般的积分变换的（准）随机采样的工作。Levie等人在近期的一系列研究中，考察了时频积分算子的蒙特卡洛 [25], [26]和准蒙特卡洛 [27] 式随机采样。他们的工作关注如何利用蒙特卡洛积分和（准）随机时频样本，近似连续域中的时频处理。通过援引有关蒙特卡洛积分的既有结果，作者证明，在满足某项技术性结构条件时，近似误差可以得到控制。他们还表明，短时傅里叶变换、小波变换以及二者的一种定制融合（称为局部化时频变换）均满足这一条件。然而，他们并未研究采样后表示的可逆性或稳定性。在随机采样方面，相关文献通过相关采样研究这些性质；这一概念由 Bass 和 Gröchenig 针对带限函数提出 [28] ，随后被推广到包括时频表示在内的各种情形 [29]–[31], [32]。相关采样为有限体积域内局部化函数的稳定采样提供了概率框架，例如，能量在有限区间之外可以忽略不计的带限信号。我们的工作在多个方面有别于这些先前研究：我们提出的采样集并非完全（准）随机，而是对应于均匀时频网格，只是在每个小波通道中引入微小的准随机延迟。此外，我们考虑任意信号的完美重构，不作局部化假设。

贡献： 此前，人们认为基于网格的抽取策略不适用于连续小波变换，也不认为它能为离散信号提供高效且稳定可逆的表示。事实上，已知的构造方法在中等甚至较低的过采样率下都无法实现完美重构。本文提出了首个适用于具有可调 Q 因子的小波变换的基于网格的抽取策略，该策略在过采样率接近 1 时即可实现完美重构，并且具有出色的数值稳定性；在音频处理中常用于短时傅里叶变换的 2 至 8 倍中等过采样率下，它还能保持能量。为此，我们将平移不变系统 [33], [34] 与基于低差异序列的准随机采样思想相结合 [35]–[36]。与小波基和以往的可逆常 Q 实现类似，我们使用一组补偿滤波器来覆盖任意小的低频区域。我们通过计算有限域中抽取小波系统在一系列系统参数下的精确框架界比和累积频谱图，验证了该构造；这些参数涵盖母小波、频率通道数、抽取因子和过采样率的变化。作为概念验证，我们将所提出的小波抽取应用于若干音频应用。我们复现了一项基于 Itakura-Saito 非负矩阵分解（NMF）的信号增强实验，该实验由 Févotte 等人提出 [37]，并高度依赖于将表示系数自然地解释为时频矩阵。此外，我们还展示了如何通过对经典谱通量方法进行直接改编，将所提出的方案用于起音检测 [38, 第 3-A 节]。为说明所提出的方案在无需矩阵结构的任务中能够与成熟的非均匀常 Q 变换相媲美，我们还评估了所提出的方法与快速 Griffin-Lim 算法 [39]结合进行无相位重建的适用性。除标准的非规则小波抽取外，我们还将其与短时傅里叶变换进行了比较。

论文结构：我们先在第 II 节简要介绍小波系统和准随机序列，然后在第 III 节提出所述抽取方案。第 IV 节讨论实际实现的细节与复杂度，并对我们的构造进行数值评估。随后，我们在音频处理领域的三个示例实验中应用所述方案（第 V 节）。具体而言，我们研究了基于非负矩阵分解的信号分解、基于谱通量改编方法的起音检测，以及基于快速Griffin-Lim算法从时频系数进行的无相位重建。论文最后总结结果，并展望相关的未来工作（第 VI 节）。

II. 技术背景

在介绍我们新提出的抽取策略之前，我们先回顾小波系统和准随机序列的一些基础知识。

A. 小波变换

小波系统是一组函数（或向量），由单个原型函数——母小波——通过平移和膨胀生成。由于我们关注实值信号，尤其是音频信号，因此考虑一种母小波 ψ suchthatitsFouriertransformvanishes，使其傅里叶变换在负频率处为零，即 $\psi(\xi) = 0$ for $\xi \in (-\infty, 0]$。这类母小波通常称为解析，不过这一术语

[2]关于稳定性和能量保持的更多细节，请参见 II-A节。

解析小波变换过去至少有两种不同的用法 [40], [41]。连续小波系统通过对母小波进行膨胀（膨胀因子为 $s > 0$）和平移（平移量为 $x \in \mathbb{R}$）生成：

$$\psi_{(x,s)} := s^{-1/2}\psi\left(\frac{\bullet - x}{s}\right).$$

现在可以使用该系统对信号 f 进行滤波，从而得到连续小波变换。

$$\begin{aligned} W_\psi f(x,s) &= \frac{1}{\sqrt{s}}\int f(t)\overline{\psi\left(\frac{t-x}{s}\right)}\,\mathrm{d}t \\ &= \langle f, \psi_{(x,s)}\rangle_{L_2} \\ &= \left[f * \overline{\psi(-\bullet/s)}\right](x)\,. \end{aligned} \tag{1}$$

小波变换的数学研究通常通过变换系数的内积表示来考察 $W_\psi f$，但最后的等式证明了将其解释为滤波器组是合理的。重要的是，小波变换具有平移不变性，即延迟信号 $f(\bullet - y)$ 的小波变换等于 f 的延迟后（在第一个变量上）的变换。另请注意，尺度 s 与滤波器中心频率成反比：如果 ψ 的通带位于频率 ξ_1 附近，那么 $\psi(\bullet/s)$ 的通带位于 $\xi_s := \xi_1/s$ 附近。

在应用中，只能考虑连续小波系统的一个离散子集；该系统，或者等价地说，该变换，是抽取的。通常，平移-膨胀对的离散子集 $((x_{l,\,j},\ s_j))_{l\in\mathbb{Z},\ j\in I} \subset \mathbb{R}\times[0,\infty)$ 由特定的抽取规则生成。我们将相应的抽取小波系统记为 $\{\psi_{l,j}\}_{l\in\mathbb{Z},\ j\in I}$，其中 $\psi_{l,\,j} := \psi_{(x_{l,j},s_j)}$。关于抽取 $((x_{l,\,j},\ s_j))_{l\in\mathbb{Z},\ j\in I}$ 的选择，我们希望任何函数 f 都能从抽取变换中稳定恢复，并且 f 的能量能由其系数准确表示。从数学上说，这些性质等价于连续小波系统的离散子集 $(\psi_{l,\,j})_{l,\,j}$ 构成一个框架 [17], ，即满足能量等价关系

$$A\|f\|_2^2 \le \sum_{l,j}|\langle f, \psi_{(l,j)}\rangle|^2 \le B\|f\|_2^2, \tag{2}$$

对所有 f 以及某些常数 $0 < A \le B < \infty$均成立。隐含地，我们始终假设 A，B 是使（2）成立的最优常数。所谓上框架界（B）与下框架界（A）之比 $R_{\mathrm{FB}} := B/A$ 量化了抽取变换保持信号能量的程度，即其数值计算的稳定性。如果 $R_{\mathrm{FB}} \neq 1$，则从抽取变换中对 f 进行无误差重构需要一个对偶框架；在母小波和抽取 [15], [24], [42], 满足某些条件时，可以高效地预先计算该对偶框架，否则则可通过迭代方案实现 [22]，[23], [43]。如果变换系数发生修改，则合成函数 f 的能量仍可由修改后系数的能量的 A^{-1} 倍界定。请注意，小波框架的对偶框架不一定是小波框架，甚至不一定是滤波器组。事实上，这种情况才是例外，而非通常情形。

关于尺度 s_j 的选择，请注意，小波滤波器的带宽会随其中心频率 ξ_s 线性增加，而中心频率与其尺度 s 成反比。因此，按照固定的基尺度 $a > 1$ 的整数次幂，即 $s_j = a^{-j}$，对频率通道进行抽取是很自然的。相反，小波脉冲响应的宽度与尺度成正比，这表明应根据 $x_{l,\,j} = a^{-j}\cdot lb$ 进行通道内抽取，其中 $b > 0$ 为某个固定值，另请参见 [44,第 3] 章。直观地说，各个 $\psi_{l,\,j}$ 所覆盖的时频区域大致对应于面积恒定的椭圆；随着 j 增大，这些椭圆会变得越来越窄，或者更准确地说，对应于以 $(x_{l,\,j},\ \xi_1/s_j)$ 为中心、半径恒定的双曲圆，见图 1(a)。

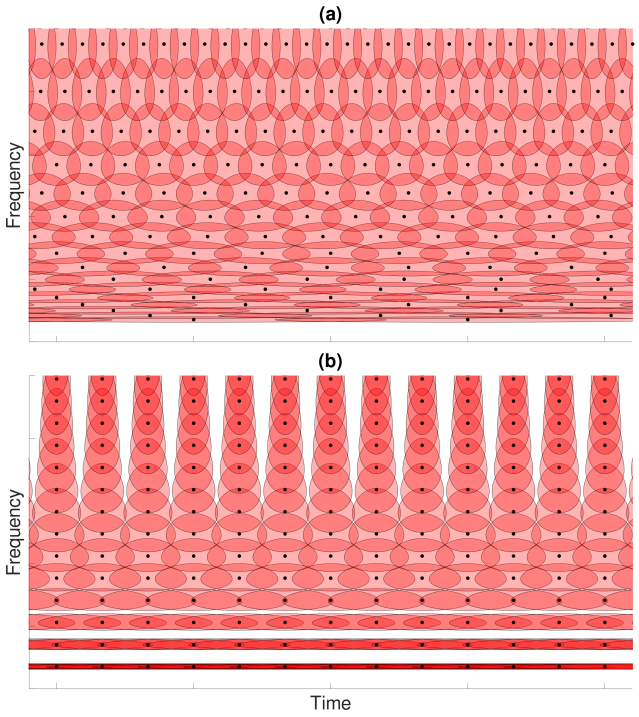

图1. 抽取小波系统的时频几何：(a) 经典非均匀小波抽取能够较为均匀地覆盖时频空间，但代价是引入非均匀抽取。(b) 在均匀网格上进行小波抽取时，离开一小段频率带后，覆盖会变得越来越不均匀。图中所示的双曲圆在两个面板中的半径相同。

这种抽取策略选择 $(x_{l,\,j},\ s_j) = (a^{-j}\cdot lb,\ a^{-j})$，采用几何频率间隔和依通道而异的抽取因子（跳步长度）$d_j = a^{-j}\cdot b$，是小波基、过完备小波系统和常数Q变换中常见且经过充分研究的做法，例如 [15], [44]。我们将这种约定称为经典小波抽取。如果 ψ 不是严格带限的，或者抽取因子过大，那么用于无误差重构的对偶框架只有在特殊情况下才会是滤波器组。此外，抽取因子d_j 依赖于尺度，导致滤波器组系数无法实现时间对齐；而且由于缺乏位置无关的关系，

[3]双曲圆包含所有与圆心的双曲距离小于给定半径的时间尺度对 $(x_{l,\,j},\ s_j)$。这里，$(x,\ s)$ 与 $(x',\ s')$ 之间的双曲距离按 2 arsinh $\sqrt{(x'-x)^2+s'-s^2}/2\sqrt{ss'}$ 计算。请注意，我们是在时频平面而非时间尺度平面中描绘这些圆。

不同频带系数之间无法表示为时频矩阵，也无法划分为结构相同的时间帧。在许多应用和实现中，这种不规则性会妨碍工作流程。与短时傅里叶变换不同，无论中心频率是线性还是对数间隔，上述小波变换的时频几何都不适用于在均匀矩形网格上进行抽取，见图1(b)。不难看出，而且事实上也很容易证明，这种抽取策略无法产生具有完美重构能力的数值稳定系统。鉴于短时傅里叶变换的许多成功处理方案都依赖于完美重构以及将系数表示为时频矩阵，这构成了在（音频）信号处理中采用小波的一大障碍。

B. 准随机序列

拟随机序列也称为低差异序列，是由数字或 k 维坐标构成的确定性序列，具有均匀分布随机数的某些特性，因此可在各种应用中替代后者，其中最重要的应用是用于数值积分的准蒙特卡洛方法 [35]。术语 *差异度* 指若干相关度量之一，用于量化一组点的均匀分布程度。这一性质很重要，例如，它有助于获得准蒙特卡洛积分的误差界。虽然大多数现代理论研究关注高维拟随机序列，但我们只需要一维序列。具体而言，低差异序列中任意 N 个连续元素在维度 $D = 1$ 上都能提供 N 个在单位区间内分布得特别均匀的数字。

小波的抽取策略可以从任何低差异序列推导得出。这里，我们重点介绍两个精心挑选的示例，分别基于两类重要的拟随机序列。

克罗内克序列： 一种看似简单的构造是所谓的克罗内克序列，其形式为 $(\{\alpha l\})_{l\geq 0} = (0,\ \{\alpha\},\ \{2\alpha\},\ \ldots\ldots)$，其中 α 是某个实数，而 $\{x\} := x - \lfloor x \rfloor$ 表示实数 x 的小数部分。众所周知，当 α 是一个*劣可逼近数*时，克罗内克序列的分布特别均匀。这类数是无理数，特别难以用有理数逼近。严格来说，若存在常数 $c > 0$，使得一个数 $\alpha \in \mathbb{R}$ 是劣可逼近数，则

$$|\alpha - p/q| > c/q^2,$$

对于所有非零整数 $p,\ q \in \mathbb{Z} \setminus \{0\}$，已知一个无理数是劣可逼近数，当且仅当其连分数展开的系数有界。在所有劣可逼近数中，黄金比例 $\phi = \frac{1+\sqrt{5}}{2}$ 使最优常数 c达到最大值。由于这一性质恰好也由所有形如 $\frac{a\phi+b}{c\phi+d}$ 的所谓等价数所共有，其中整数 $a,\ b,\ c,\ d$ 满足 $ad - bc = \pm 1$，因此我们同样可以考虑具有 $\alpha = 1 - 1/\phi$ 的克罗内克序列，即 $a,\ c = 1$、$b = -1$ 和 $d = 0$。有关更多克罗内克序列的信息，请参阅书籍[45], [46]。

数字 (0，1)-序列，定义在 $\mathbb{Z}_2$： 以 2 为底的 (0，1)-序列是单位区间 $[0, 1)$ 中的一个无限序列 $(x_l)_{l\geq 0} = (x_0,\ x_1,\ \ldots\ldots)$，具有如下性质：对于每个 $m \in \mathbb{N}_0$ 和每个 $k \in \{0, 1, \ldots, 2^m - 1\}$，基本区间 $[\frac{k}{2^m}, \frac{k+1}{2^m})$ 恰好包含点集中的一个元素

$$\{x_l\ :\ p2^m \leq l \leq (p+1)2^m - 1\} \quad \text{for every } p \in \mathbb{N}_0.$$

这种序列的一个例子是著名的范德科普特序列 [47]（另见 [48]），它与位反转排列有关。通常，基数为 2 的 (0，1)-序列是通过所谓的数字方法，在阶数为二的有限域 $\mathbb{Z}_2$ 上构造的，即模 2 整数。为此，设 $\mathbf{C} = (c_{r,\ k})_{r,\ k\geq 1}$ 是一个定义在 $\mathbb{Z}_2$ 上的 $\infty \times \infty$-矩阵，即其元素 $c_{r,\ k}$ 取自 $\{0, 1\}$，并且对每个 $m \in \mathbb{N}$，左上角的 $m \times m$ 子矩阵都是非奇异的。然后，序列 $(x_l)_{l\geq 0}$ 定义为

$$x_l = \sum_{r=0}^{\infty} \frac{\eta_{l,r}}{2^{r+1}} \quad \text{where} \quad \eta_{l,r} = \sum_{k=0}^{\infty} c_{r+1,k+1} l_k \pmod 2,$$

其中 $l_k \in \{0, 1\}$ 是索引 l的二进制位，即 $l = l_0 + l_1 2 + l_2 2^2 + \cdots$（从某个位置开始显然都为 0），这是一个基数为 2 的 (0，1)-序列。在此背景下，取单位矩阵作为 $\mathbf{C}$，即可得到范德科普特序列。我们将使用由下列无限矩阵得到的特定数字 (0，1)-序列

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots \\ 1 & 1 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 1 & 1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

关于（数字）(0,1)-序列的更多信息，请参阅书籍 [35], [36] 或综述文章 [48,第 3 节2]。

III. 小波采样的新约定

由于 II-A节 中阐述的几何直觉，小波中心频率 $\xi_j \sim s_j^{-1}$ 呈线性间隔这一做法在文献中几乎未曾受到关注。出于同样的原因，均匀抽取，即对于某个固定的 $d > 0$，采用 $d_j = d$，通常也被忽略。现在我们将介绍一种遵循这两项惯例的抽取方案，可用于构造数值稳定、完美重构的小波变换，并且过采样率适中甚至极低。由于它是均匀滤波器组框架，标准结果 [33], [34] 保证其对偶框架同样是均匀滤波器组，且具有相同的抽取因子。然而，对偶滤波器组不一定是小波系统。特别是，这意味着所述对偶框架可以使用优化因式分解算法预先计算，因此，从变换系数进行合成的效率很高。所提出的方案得到一种近乎 时间对齐的抽取表示。更准确地说，各频带的系数之间存在与位置无关的固定关系，从而提供了一种有意义的

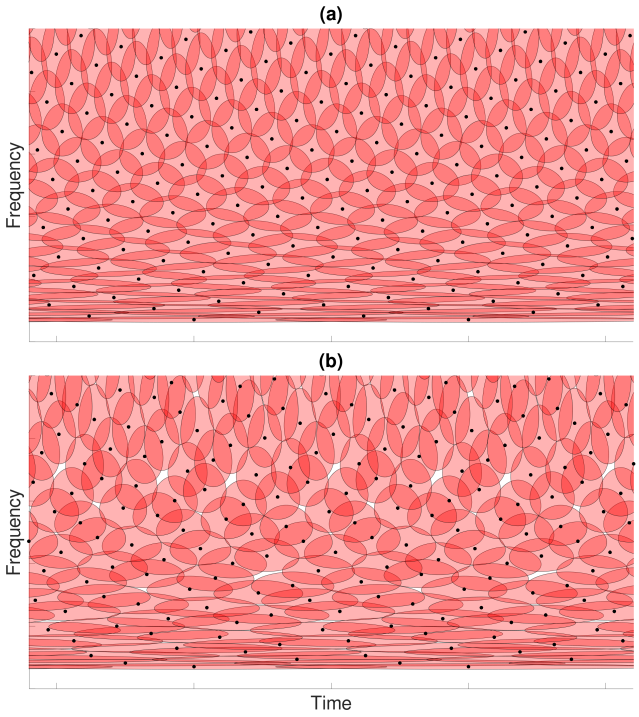

图2. 具有准随机延迟的抽取小波系统的时频几何：(a) 在均匀网格上进行抽取，并根据黄金数 克罗内克序列选取延迟，可获得与经典小波抽取（图1(a)）相当的均匀覆盖。(b) 若改为根据第II-B节所述的数字序列推导延迟，则得到的覆盖均匀性稍差。在这两种情况下，低于某一频率后，所得覆盖都会变得不均匀，类似于图1(b)的底部区域。请注意，为便于比较，图中双曲圆的半径与图1相同。只需略微增大半径，即可消除面板(b)中的盲区；而即使稍微减小半径，也会在面板(a)和图1(a)中引入类似的盲区。

时间帧概念，并提示了在时频矩阵中进行自然排列的方式。

给定感兴趣的最大尺度 $b > 0$ （或等价地，感兴趣的最低频率 $\xi_{\min} > 0$）、抽取因子$d > 0$、决定频率方向步长的第三个参数 $q > 0$ ，以及决定通道特定延迟的低差异序列（δ_0， δ_1， δ_2，.....），我们选取平移-膨胀对

$$(x_{l,j}, s_j) = \left(d(l+\delta_j), \frac{1}{b^{-1} + q^{-1} j} \right), \tag{3}$$

对于所有整数 $l \in \mathbb{Z}$ 和非负整数 $j \in \mathbb{N}_0$。换言之，我们考虑离散小波系统 $\{\psi_{l,j}\}_{l\in\mathbb{Z},\ j\in\mathbb{N}_0}$，其中

$$\psi_{l,j}(t) = \sqrt{\tfrac{1}{b} + \tfrac{j}{q}} \cdot \psi\left((\tfrac{1}{b} + \tfrac{j}{q}) \cdot (t - d(l+\delta_j)) \right). \tag{4}$$

与图1(b)所示的矩形网格不同，后者可通过上述构造生成，其中对所有 j，延迟均为 $\delta_j = 0$；若将 δ_j 选为拟随机序列中的元素，则得到的小波系统能够以出人意料的均匀程度覆盖时频平面，见图2。值得注意的是，若将 δ_j 选为克罗内克序列中的元素，则点 $\{(x_{l,\ j},\ \xi_j)\}_{l\in\mathbb{Z},\ j\in\mathbb{N}_0}$ 构成一个均匀（倾斜）的时频网格。

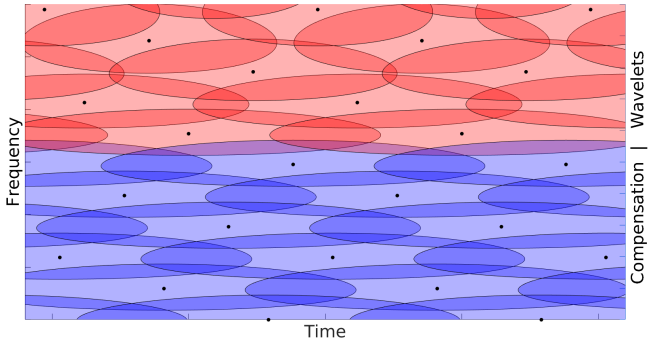

图3. 对低频区域的覆盖方案：对于采用准随机延迟的抽取小波系统，使用最大尺度小波滤波器的调制和延迟版本进行覆盖，准随机延迟根据黄金克罗内克序列选取。为便于观察，与图2(a)相比，频率轴经过了重新缩放。

A. 具有完美重构和均匀抽取的数值稳定小波变换

如图2所示，所提出的小波抽取能够以出人意料的均匀程度覆盖时频平面。然而，由于频率上的间距均匀，只有当小波的尺度足够小（相对于频率步长参数 q）时，这一结论才成立。由于小波带宽与其中心频率成正比，当基尺度 b 任意大时，我们发现低频区域的覆盖不足，与图1(b)类似。该区域的大小取决于所选小波和频率步长参数 q。为弥补覆盖不足，我们引入额外的补偿滤波器来覆盖低频区域

$$\psi_{l,j}(t) = \frac{1}{\sqrt{b}} \psi\left(\frac{t - d(l+\delta_j)}{b} \right) e^{2\pi i \xi_1 \cdot \frac{j(t - d(l+\delta_j))}{q}}, \tag{5}$$

对于负整数 $j \geq -q/b$。换言之，只要中心频率仍为非负值，我们就以均匀步长对基尺度 b 的小波进行解调。请注意，上述构造意味着准随机延迟序列的索引也相应发生了偏移，即我们将第j个元素记为 $\delta_{j-\lfloor q/b \rfloor}$ ，而不是 δ_j。图3展示了这一构造；为便于观察，我们对频率轴（但不对时间轴）相较于图2(a)进行了重新缩放。显然，这只是构造适用于覆盖低频区域的滤波器的多种可能方案之一；之所以在此采用该方案，是因为它保留了滤波器原型函数（即母小波），并保持系统在时间和频率上的均匀抽取。

完整系统 $\{\psi_{l,j}\}_{l\in\mathbb{Z},\ j\in\mathbb{Z}\cap[-q/b,\infty)}$ 仍采用均匀抽取，其性质，尤其是完美重构性质和数值稳定性，可通过均匀滤波器组的框架理论 [34] 或等价的平移不变系统 [33]，进行研究；这也为根据滤波器组系数进行重构提供了高效的算法。我们发现，在实际应用中，只要将抽取参数 d 和 $1/q$ 设得足够小，就足以获得可逆且数值稳定的滤波器组。在第IV-A节中，我们针对 $\mathbb{C}^L$中的有限小波系统验证了这一判断。

IV. 数值评估、复杂度与实现

尽管上一节按照 $b > 0$ 和 $q > 0$ 对$\{\psi_{l,j}\}_{l,\ j}$ 进行参数化的方式很自然，但不便于实际构造。因此，接下来我们采用以下参数化方式，并将范围限定在 $q/b \in \mathbb{N}$的情形。我们选择一个母小波 ψ、频率通道数 $M+1$，这些通道在 $[0, \xi_{\text{samp}}/2]$内以中心频率等间距分布，其中 ξ_{samp} 为采样率；此外，还选择补偿滤波器的数量 M_{C} 。它与上述参数化方式之间的关系由 $M_{\text{C}} = q/b$ 和 $\xi_1/q = \xi_{\text{samp}}/(2M)$ 给出。在所提供的实现中，小波会自动缩放，使得$\psi_{l,\ 0}$的中心频率等于 $\xi_{\text{samp}} \cdot M_{\text{C}}/(2M)$。随后，通过选择抽取因子 d来控制所得系统的冗余度，这与短时傅里叶变换的常见参数化方法类似。

A. 数值评估

为了验证所提出的抽取策略确实能够得到可逆且数值稳定的小波系统，我们计算了采用该抽取策略的各种小波滤波器组的框架界比 R_{FB} 。这里，我们考虑作用于有限序列的滤波器组；在这种情况下，该比值等于实现滤波器组分析与合成复合过程的矩阵的条件数，因而可以直接计算。为此，我们借助大型时频分析工具箱（LTFAT，ltfat.org），该工具箱通过对上述矩阵进行高效因式分解来完成计算。请注意，当且仅当该滤波器组的 R_{FB} 有限时，它才是可逆的；当该值较小时，滤波器组稳定；若 $R_{\text{FB}} = 1$，则具有完美稳定性。本文所展示的评估可使用 ltfat.org/notes/057 上的代码复现。

在对所提评估方法进行预测试时，我们注意到，对于固定的 M，帧界比 R_{FB} 会随着 M_{C}的增加而单调下降，直到达到某个值 M_{C}；超过该值后，继续增加便不会带来任何益处。此外，我们发现该值 M_{C} 与 M的选择无关。增加 M_{C} 会使滤波器组的频率响应更加稳定；该响应等于 $\Psi = \sum_j |\widehat{\psi_{0,\ j}}|^2$，两者相差一个正的乘法常数，尤其是在低频区域。 Ψ 的剧烈波动会对 R_{FB}产生不利影响，而增加 M_{C} 可减小这些波动，这解释了观察到的第一个现象。之所以不依赖于 M ，是因为仅在 M 的选择上有所不同的两个此类滤波器组，其频率响应相差一个膨胀和一个正的乘法常数。总之，这些观察结果使我们能够大幅减少需要测试的配置数量。

考虑到上述因素，我们对框架界进行优化：首先，对于固定的母小波和过采样率，确定使 R_{FB} 达到最优的 M_{C} 最小值。随后通过穷举搜索，确定使 R_{FB} 达到最优的值 $M \in \{128, 256, 384, 512, 640, 768, 1024, 1280, 1536, 2048\}$，并额外限制 $M \geq 50 M_{\text{C}}$。最后，我们从已确定的最优值及其在 M_{C} 测试值序列中的两个相邻值出发，采用分治法进一步优化 M 的选择。

这一限制 $M_{\text{C}} \leq M/50$ 可确保 $\psi_{l,0}$ 的中心频率不超过 $\xi_{\text{samp}}/100$，例如，当采样率为 44.1 kHz 时，中心频率不超过 441 Hz。

结果：表 I 展示了过采样率为 1.2、2、4 和 8 时得到的最优框架界。作为母小波，我们考虑柯西小波 [49], , 其定义（相差一个归一化常数）为 $\widehat{\psi}(\xi) = \xi^{\frac{\alpha-1}{2}} e^{-2\pi\xi}$; 其中 α 分别取 100、300、900 和 2700，较大的 α 意味着较高的 Q因子。[4] 作为紧支撑小波的示例，我们考虑调制的四阶 B 样条，其定义（相差一个归一化常数）为 $\widehat{\psi}(\xi) = \sin(\pi(\xi - \xi_{\text{fm}}))^4/(\pi(\xi - \xi_{\text{fm}}))^4$,, 其中 ξ_{fm} 分别取 3、6 和 10，对应的品质因数大致相当于 α 分别为 257、1024 和 2842 的柯西小波[5], , 见 [50]。

正如预期的那样，最优框架界比会随着过采样率的提高而降低。 然而，在过采样率固定时，它对母小波和 Q 因子的依赖有限。具体而言，对于基于克罗内克序列的抽取，在几乎所有条件下，过采样率分别为 1，3，1.65 和 1.3 时，我们得到接近 R_{FB} 14、3、1.65 和 1.3 的值。对于冗余度为 4 及以下的情况，基于数字（0,1）序列的抽取所得结果稳定性略差，即 R_{FB}更大。有趣的是，当过采样率为 8 时，除一种情况外，采用这种抽取方案都能改善 R_{FB} ，这表明在使用较高过采样率时，基于数字（0,1）序列的抽取方案更有利。

请注意，只有当函数 M 中的 R_{FB} 为凸函数时，我们的分而治之细化过程才能保证找到其全局最小值。通常情况下并非如此，如图 4 所示：图中绘出了过采样率为 2 时，在上述穷举搜索中考虑的所有 M 选项以及所考虑的七种母小波对应的R_{FB} 值。因此，仍有可能进一步改进所报告的界。此外，这也解释了为何所报告的最优参数 M 并不随母小波的 Q因子单调变化；母小波的 Q因子会随着 α 或 ξ_{fm}取值增大而增大。

为了验证上文图1和图2所示的小波覆盖确实符合所提出的小波系统对时频域的良好覆盖，我们计算了这些系统在约为 2 的过采样率下的累积频谱图 [51], 见图5。具体而言，我们使用带高斯窗的短时傅里叶变换，计算系统中每个小波的谱图，并将它们全部相加。由于信号的谱图表示其时频能量分布，因此这一总和展示了不同小波系统能够良好表示的时频区域。我们的模拟基本证实了

[4]这些超参数对应的品质因数，是通过母小波的中心频率与其相对于最大值低 -3 dB 处的带宽之比估算得出的。它们分别为 2.9999、5.2053、9.0212 和 15.6282。5

[5]1600、8.3200 和 13.8675，而样条小波对应的数值分别为 4.8171、9.6229 和 16.0340。

图 3：B

表 I 所提出小波系统的优化框架界比。表中数值为 $R_{\mathrm{FB}}\,(M_{\mathrm{C}}, M)$，即通过 M_{C} 个补偿通道和 M 个总通道实现最优框架界比 R_{FB} 。

克罗内克序列延迟							
	柯西				B样条		
过采样	$\alpha = 100$	$\alpha = 300$	$\alpha = 900$	$\alpha = 2700$	$\xi_{\mathrm{fm}} = 3$	$\xi_{\mathrm{fm}} = 6$	$\xi_{\mathrm{fm}} = 10$
1.2	15.06 (2,102)	14.17 (4,202)	13.61 (6,307)	13.74 (11,550)	15.04 (4,202)	14.08 (7,449)	14.08 (12,620)
2	3.22 (3,448)	2.98 (5,253)	2.92 (9,501)	2。94 (15，764)	3.01 (5,253)	2.93 (9,511)	2.97 (15,764)
4	1.72 (4,768)	1.62 (7,350)	1.60 (12,768)	1.59 (20,1012)	1.60 (6,363)	1.59 (12,768)	1。59 (20，1024)
8	1.31 (4,214)	1。25 (8，404)	1.23 (14,702)	1.24 (26,1306)	1.25 (7,473)	1。24 (15，757)	1.25 (25,1250)
数字 (0,1)-序列延迟							
	柯西				B样条		
过采样	$\alpha = 100$	$\alpha = 300$	$\alpha = 900$	$\alpha = 2700$	$\xi_{\mathrm{fm}} = 3$	$\xi_{\mathrm{fm}} = 6$	$\xi_{\mathrm{fm}} = 10$
1.2	20.38 (2,127)	19.27 (4,255)	17.48 (6,384)	16.02 (11,1023)	20。67（4，256）	17.26 (7,384)	16.98 (12,640)
2	3.80 (3,260)	3.73 (5,384)	3.66 (8,656)	3.57 (14,771)	3.72 (5,388)	3.62 (9,771)	3.62 (14,771)
4	1.74 (4,255)	1.67 (7,383)	1。63 (12，639)	1.62 (21,1279)	1.78 (6,640)	1.62 (12,639)	1。60 (20，1023)
8	1.27 (5,256)	1.20 (9,511)	1。20 (16，1535)	1。21 (27,1791)	1。21（8，511）	1。27 (16，1408)	1.22 (27,1791)

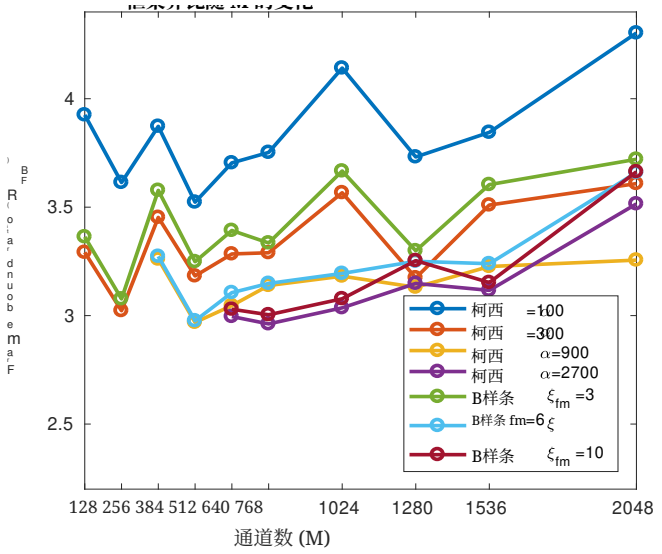

图4。展示了过采样率为2时， R_{FB} 对 M 的依赖关系示例，适用于克罗内克序列延迟。

图1和图2。图5(a)中的经典小波系统以及图5(c)中采用黄金克罗内克序列的延迟移位系统表现出最佳的均匀性。显然，图5(b)中没有延迟的均匀系统过度强调了某些时频区域，同时完全遗漏了其间的区域。图5(d)中基于数字(0，1)序列的带延迟均匀系统，覆盖区域的均匀程度稍逊。然而，由于此处展示的累积频谱图采用了双倍过采样，因此这一发现与计算得到的框架界一致：在低过采样率下，克罗内克序列延迟优于数字(0，1)序列延迟。

B. 分块处理与复杂度

所提出的小波抽取最直接且高效的实现采用正向和逆变换，例如FFT滤波器组 [52]，，与 [16]，[53]类似。与这些参考方法相比，所提出的这种抽取实现由于频率通道重叠数量较多，计算负荷略高。根据其构造，FFT滤波器组会一次性处理整个输入信号。不过，采用 [15]提出的切片方案，也可以实现分块处理。切片处理需要将输入数据流分割成较大的数据块，因此产生的延迟对于某些应用而言可能无法接受。短块长度的小波滤波器组实现通常相当复杂，尤其是因为它们通常采用非均匀抽取，参见例如 [54]。然而，在我们的设置中，通过采用具有有限支撑的母小波，利用简单的时域实现即可缩短块长度，从而减少延迟。尽管实时实现并非我们的主要目标，我们仍将概述此类实现的计算成本。

所提出的均匀抽取可通过固定块长度来实现，其中每个块对应一个时间帧，且每个通道包含一个滤波器组系数。每个时间帧中所有系数的相对时间位置都相同。这种实现的计算过程简单直接，可通过直接计算输入信号与小波原子之间的内积来完成，见式（1）。由于小波变换具有恒定的Q因子，这种计算的成本仅随通道数量呈对数增长。具体而言，对于 $M + 1$ 个等间隔频率通道，我们以抽取因子d 计算小波系数，其中前 M_{C} 个通道为补偿滤波器。假设在所考虑的最大尺度下，小波脉冲响应的长度为 L_{W} 个样本。由于延迟位于区间 $[0, d)$内，我们可以将输入划分为长度为 $L_{\mathrm{B}} = L_{\mathrm{W}} + d$的时间帧（或块），相邻帧之间重叠 L_{W} 个样本。在每个时间帧中，我们为每个通道计算一个系数。计算补偿滤波器时，每个时间帧的总成本分别为 $M_{\mathrm{C}} \cdot L_{\mathrm{W}}$ 次乘法和加法。对于

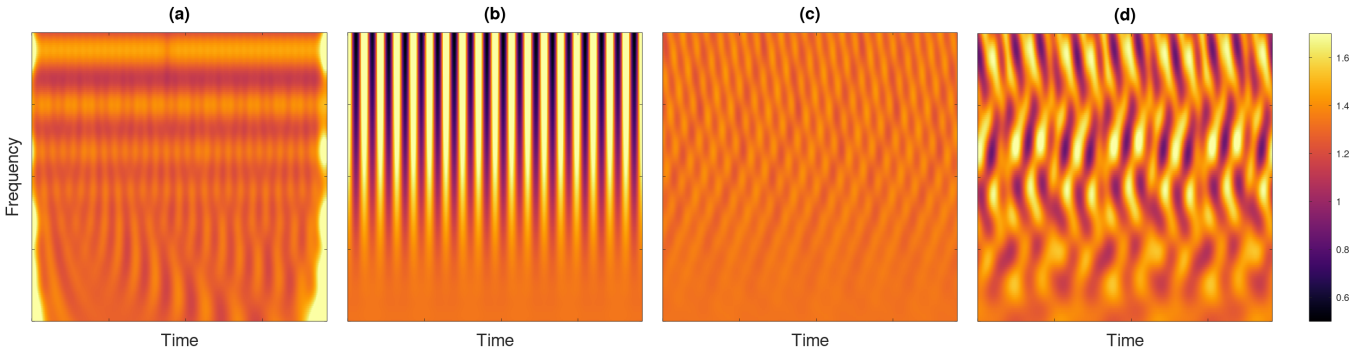

图5. 抽取小波系统的累积频谱图：(a) 经典非均匀抽取，(b) 均匀网格，(c) 延迟根据黄金 克罗内克序列选择的均匀网格，(d) 根据第II-B节所述数字序列导出的延迟。

其余滤波器，脉冲响应长度会随着尺度减小而缩短，因此很容易看出

$$L_{\mathrm{W}} \cdot \sum_{j=M_{\mathrm{C}}}^{M} \frac{M_{\mathrm{C}}}{j} \leq M_{\mathrm{C}} L_{\mathrm{W}} \cdot \int_{M_{\mathrm{C}}-1}^{M} s^{-1}\, ds = M_{\mathrm{C}} L_{\mathrm{W}} \cdot \ln(M/(M_{\mathrm{C}}-1)) \tag{6}$$

每个时间帧需要进行乘法和加法运算。这里，我们采用了调和级数部分和的标准积分估计。请注意，上述计算忽略了将脉冲响应长度舍入到下一个整数的情况，并假设 $M_{\mathrm{C}} \geq 2$。总的来说，正变换的直接实现大约需要

$$M_{\mathrm{C}} L_{\mathrm{W}} \cdot (1 + \ln(M/(M_{\mathrm{C}}-1)) \tag{7}$$

每个时间帧需要进行乘法和加法运算，并引入 L_{B} 个样本的延迟。采用低通滤波并随后进行抽取的更复杂实现（其思路与 [3] 类似），可以显著优化系数计算，但会产生少量误差。

通过对偶滤波器计算具有完美重构的反向变换（参见 II-A 节）。对于本文中的小波配置，对偶滤波器的有效脉冲响应长度与相应原始滤波器的脉冲响应长度处于同一量级。这意味着，反向变换的计算成本和延迟与正变换相近。不过，由于对偶脉冲响应预计不具有有限支撑，因此必须对其进行近似，这会在精度与延迟之间引入权衡。

C. 大型时频分析工具箱中的小波

第IV-A节和第V节中的所有计算均依赖于大型时频分析工具箱（LTFAT）中的小波滤波器组实现。该工具箱在发布 LTFAT 2.5.0 时进行了更新，以支持所提出的小波抽取。具体而言，函数 waveletfilters 根据给定的输入参数，提供小波滤波器的频率响应，以及一组抽取因子和滤波器的中心频率。默认情况下，waveletfilters 接受滤波器组长度 L 和一个小波尺度向量作为输入参数，其中单位尺度 $s = 1$ 对应于一个中心频率 $\xi_{\mathrm{samp}}/20$。此外，还可以使用一系列逐步细化的可选输入参数来定制小波滤波器，以适应大量应用场景，包括 [40], [50] 中的场景以及本文所述场景。除了直接指定小波尺度之外，waveletfilters 还支持在指定频率范围内自动分配滤波器；滤波器间隔可以按几何方式排列（每倍频程设定固定数量的频点），也可以按线性方式排列（设定固定数量的通道）。

对于本文而言，最重要的是，waveletfilters 提供了控制母小波和抽取设置以及过采样率的选项。默认情况下，waveletfilters 会选择柯西型母小波，并采用适配小波带宽的非均匀整数抽取因子。目前已实现适用于正尺度和负尺度的柯西、莫尔斯、莫莱小波、频率B样条、解析样条和复样条小波。单个滤波器由独立函数 freqwavelet 生成，从而便于今后添加更多小波原型。waveletfilters 支持均匀抽取和非均匀抽取，抽取因子可以是整数或分数。为构建完美重构滤波器组，waveletfilters 提供了两种方案，用于覆盖从 0 Hz 到最大小波尺度的中心频率这一频率范围：使用单个低通滤波器，或使用与最大小波尺度对应滤波器的多个频移副本，如第 III-A 节所述。

例如，调用 [g,a,fc]=waveletfilters(Ls,'linear', 2,MC/M,1,M−MC+1,{'cauchy',900}, 'uniform','redtar', 4,'repeat','delay',@dly) 可为长度为 Ls 且频率间隔为线性的信号提供小波滤波器组。假设采样率为 2，则 M−MC+1 个小波尺度分布在频率 MC/M 和 1（奈奎斯特频率）之间。该滤波器组使用参数为 $\alpha = 900$ 的柯西小波、均匀抽取以及为 4 的目标过采样率。最后，repeat 标志会在频率 0 与 MC/M 之间添加 MC 个补偿滤波器，从而得到 M+1 个总通道。函数 @dly 用于生成所需的延迟序列，并将在内部调用；调用时以所需序列元素的数量 M+1 和抽取因子向量 a 作为输入参数。输出

[g,a,fc] 由滤波器频率响应的元胞数组 g、抽取因子向量 a 和中心频率向量 fc 组成。

LTFAT 的滤波器组模块提供了大量函数，用于处理和分析由此构造的滤波器组。函数 filterbank 通过基于快速傅里叶变换的卷积计算滤波器组系数。其输入包括目标信号，以及由 waveletfilters 提供的小波滤波器及其对应的抽取因子。滤波器组系数的合成由 ifilterbank 实现，可通过 plotfilterbank 将其可视化。若需要具有完美重构能力的分析—合成滤波器组对，可对 waveletfilters 的输出应用 filterbankdual 来获得对偶滤波器组；对于仅覆盖正频率的滤波器组，则可应用 filterbankrealdual。请注意，这仅适用于均匀滤波器组，或满足严格条件的非均匀滤波器组。一般而言，只要分析滤波器组可逆，就可以通过迭代方法实现完美重构，该方法由 ifilterbankiter 实现。滤波器组的框架界可分别通过 filterbankbounds 和 filterbankrealbounds 获取，例如用于验证可逆性。LTFAT 滤波器组模块还提供了一些精选方法以实现高级功能，例如无相位重建 [39], [59]和时频重分配 [60]。

最后，waveletfilters 的输出与 LTFAT 中的分块处理框架兼容，因此可以通过实现 [15] 中提出的一种切片处理变体，开展实验性的实时应用。

本研究使用的代码可在 ltfat.org/notes/057 找到，其中展示了多个小波滤波器组的构建过程，以及上述部分函数的用法。更详细的演示将集成到下一个 LTFAT版本中。

五、音频处理实验

我们展示了所提出的小波抽取方案的三种应用。在第一种应用中，我们复现了先前基于 NMF 的信号分解研究 [37]，中的实验；这些研究最初采用短时傅里叶变换。我们只需将短时傅里叶变换替换为小波变换，以即插即用的方式进行处理，同时保持其他所有参数不变。我们只调整了小波变换的参数，即母小波ψ、通道数 M 和补偿通道 M_C。鉴于所考虑的处理方案最初是为短时傅里叶变换设计的，并未作任何适配，因此我们并不预期采用小波变换会优于原有的短时傅里叶变换方法。相反，我们旨在说明，即便采用简单的即插即用方法，也能取得可比的结果。

在第二个应用中，我们考虑音乐信号中的起音检测。利用谱通量这一思想 [38, 第 3 节 A] ，针对所提出的抽取方案中的小波系数，而不是通常使用的短时傅里叶变换系数，我们计算了用于检测起始点的检测函数。小型标注测试集的结果 [61]与通过使用短时傅里叶变换系数的相同方法检测到的起始点，以及现成可用的起音检测器 [19]的输出进行了比较。

在第三项应用中，我们比较了在所提出的抽取方案和经典抽取方案下，无相位重建与快速 Griffin-Lim 算法（FGLA） [39], [62] 之间的性能。由于 FGLA 不依赖于将变换系数划分为时间帧，因此我们可以将本方法的处理性能与经典小波抽取进行比较。此外，还提供了基于短时傅里叶变换系数、使用 FGLA 进行无相位重建的结果作为参考。请注意，Griffin-Lim算法的实时变体确实依赖逐帧处理，例如 [63], ，并且可以适配所提出的抽取方案。

随附的音频文件及用于复现所述实验的代码可在此处获取：ltfat.org/notes/057。

A. 使用非负矩阵分解进行信号分解

在 [37], 中，Févotte 等人表明，采用 Itakura-Saito 代价函数的非负矩阵分解，能够将音频数据的短时傅里叶变换频谱图有意义地分解为 $S = W \cdot H$，其中 W 和 H 分别为非负的分量矩阵和激活矩阵。作者对路易斯·阿姆斯特朗与 His Hot Five 演奏的《My Heart (Will Always Lead Me Back to You)》这段 20 世纪 20 年代录音进行了分解、去噪和上混。根据 [37],中的主观分析，这段录音包含小号、单簧管、长号、低音提琴和钢琴声部，同时还存在明显的嘶嘶噪声和爆裂声。所用信号为一段时长 108 秒的截取片段，采样率为 $\xi_{\text{samp}} = 11\,025$ 赫兹，总长度为 $L = 1\,191\,735$ 个样本。在原始研究中，研究者基于过采样率为 2 的 STFT 频谱图计算了包含 10 个分量的 Itakura-Saito 非负矩阵分解，频率范围从零到奈奎斯特频率，共有 $M = 129$ 个通道。研究者使用逆伽马先验对分量激活矩阵 H进行正则化。

在我们的实验中，我们采用 Févotte [64],后续工作的代码，该代码使用了不同的平滑先验，其他方面则完全相同。与该工作一样，正则化参数设为 $\lambda = 25$。我们使用带有柯西小波 [40], [49],的小波变换，其超参数为 $\alpha = 450$。抽取基于克罗内克序列，过采样率设为 2。参数按照第IV-A节进行优化[6] ，最终得到 $M + 1 = 449$ 个通道和 $M_C = 6$ 个补偿通道，即 $\psi_{l\text{的中心频率}0}$ 为 74Hz。为达到所需的过采样率，我们将抽取因子设为 $d = 448$。所得的非负矩阵分解分量和信号分解如图 6 所示。

结果： 与 [37], 中展示的结果相比，我们可以看到，使用小波变换时， W 中各个分量的谐波结构不那么明显，但仍然存在。考虑到较大的

[6]由于内存限制，我们将优化限制为最多选择 769 个通道。

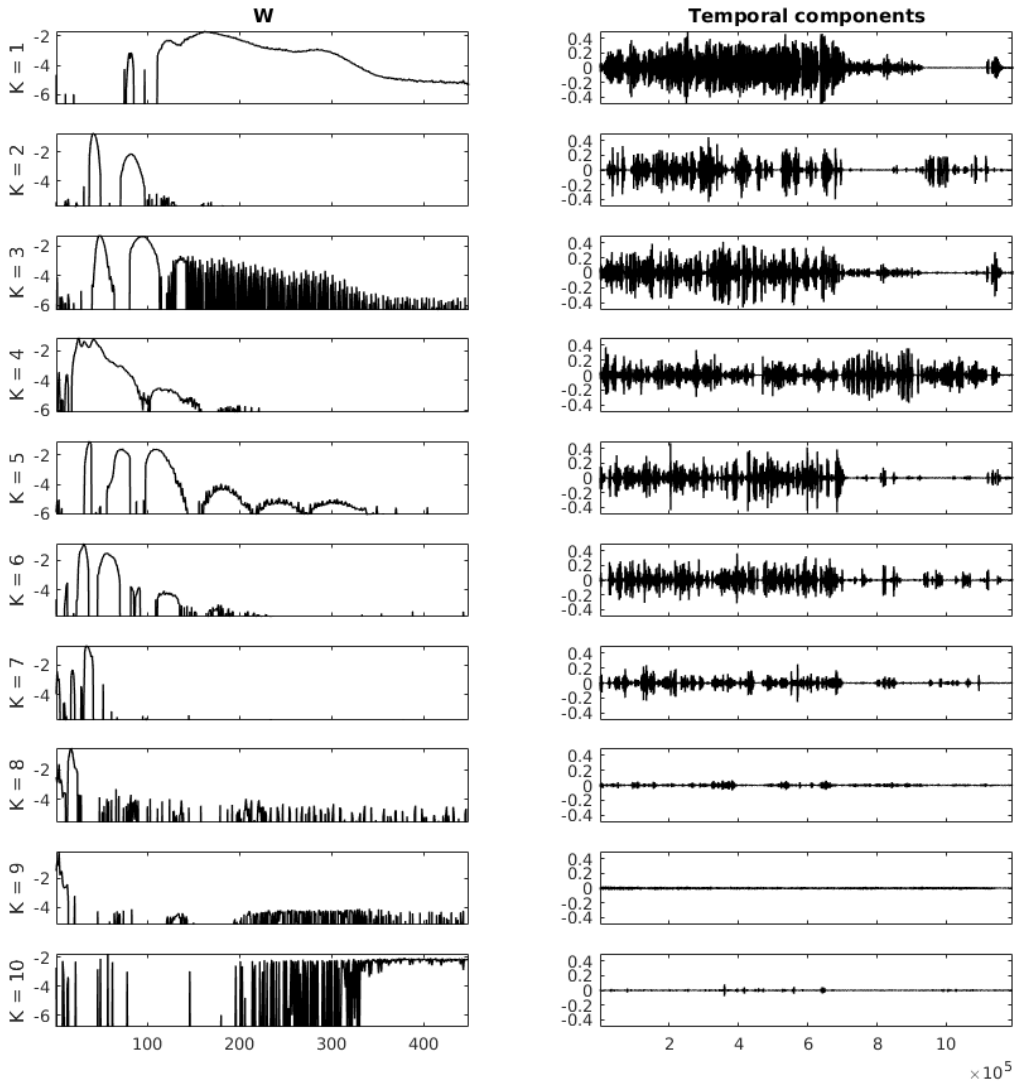

图 6. 音乐片段的非负矩阵分解。（左） W 的分贝缩放列，即非负矩阵分解分量。（右）重构信号分量。分量 4 和 6–8 捕捉了钢琴和低音提琴音轨的大部分内容。分量 9、10 表征了嘶声和噼啪噪声的大部分内容。其余分量包含小号和单簧管主旋律音轨的大部分内容，以及长号。

高频小波带宽以及基于克罗内克序列的抽取所采用的非对齐相空间覆盖，这并不完全令人意外。尽管如此，该分解仍实现了主旋律、伴奏和噪声的分离，其效果与 [37], [64] 中展示的短时傅里叶变换结果相当。主要区别在于，长号声轨未能清晰分离，而是大部分与主旋律声轨混合在一起。网站上提供了音频示例，包括各个独立分量，以及原始信号的去噪版本和升混版本。

B. 音乐起始检测

在第二个实验中，我们使用所提出的小波抽取来检测音频信号中的起始点。更具体地说，我们采用一种谱通量方法，例如[38, 第 3-A 节] 中所述的方法，但用所提出的小波系统替代短时傅里叶变换。谱通量衡量不同频带中幅值或能量的增加量（减少量设为 0），并对每个时间帧的这些增量求和。由此得到的时变谱通量函数随后用作基于以下假设的峰值拾取过程的依据：谱通量的局部极大值是新音乐事件的起始点。

在我们的实验中，我们使用所提出的小波系统的通道，替代短时傅里叶变换的频率通道。信号 f 对应的谱通量定义为

$$S(l) = \sum_j H\big(|W_\psi f(x_{l,j}, s_j)| - |W_\psi f(x_{l-1,j}, s_j)|\big) \quad (8)$$

其中 $H(x) = (x + |x|)/2$ 是一个整流器，求和仅针对与我们的新型抽取方案相对应的通道进行，并忽略 M_C 补偿通道。我们选择过采样因子为 4 的系统，使用$\alpha = 2700$ 的柯西小波，并选择能获得最佳帧界的通道数，即 $M = 1012$和 $M_\mathrm{C} = 20$ （见表 I）。为避免出现虚假的局部极大值，我们额外采用了时变阈值处理步骤。这里使用局部中位数的倍数 λ，这是一种在 [38,第 4-B 节]中推荐的稳健选择。

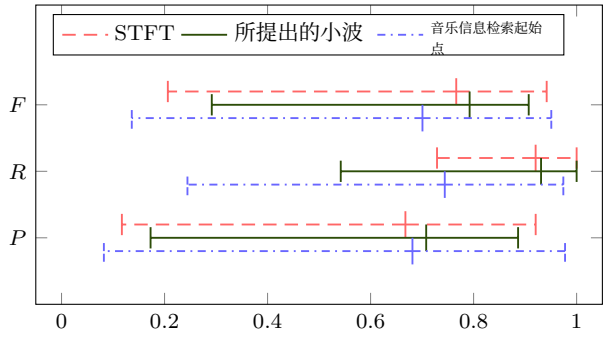

图 7. 须状图分别显示了 19 个信号和三种起音检测器的最小值、中位数和最大值：F-测度 F、召回率 R 和精确率 P。前两种起音检测器分别基于针对所提出的小波变换或短时傅里叶变换计算的谱通量。第三种方法是来自 MIRtoolbox 的标准方法，基于振幅包络。

在实验中，我们使用阿利坎特大学模式识别与人工智能组（PRAIg-UA）提供的带标注起音检测数据库 [61]。在性能评估中，如果估计的起始点与标注的起始点相差不超过 50 毫秒，我们就认为该起始点检测正确（这是文献中常用的度量方式[38, 第 5-A 节]）。随后，我们计算精确率 P，即正确检测到的起始点数与估计起始点总数之比；并计算召回率 R，即正确检测到的起始点数与标注起始点总数之比。作为单一的性能度量，我们还计算 F度量 $F = 2PR/(P + R)$。我们将本方法与基于短时傅里叶变换和汉宁窗的谱通量进行比较，即将式 (8) 中的小波系数替换为短时傅里叶变换系数。此处，我们采用与小波情形相同的抽取因子，并调整通道数和窗长，以获得过采样因子为 4 的紧框架。通过针对 [61], 中的第一个音频样本优化 F度量，我们将短时傅里叶变换的阈值因子λ 设为 1.24，将小波情形的阈值因子设为 1.34。以下结果中不包括该样本。我们还将结果与 MIRtoolbox [19]中实现的基础起音检测算法进行比较，该算法基于振幅包络。

结果： 我们的结果如图 7 所示。基于谱通量的方法平均而言略优于基于振幅包络的参考方法。然而，没有任何一种方法在所有信号上都始终最好或最差。基于短时傅里叶变换的小波谱通量与基于小波的谱通量之间的差异平均而言很小，但针对特定信号时，差异可能相当显著。这表明，针对不同信号类别，某一种方法可能更具优势；深入分析各自的优缺点，是一个值得进一步研究的方向。最后，在两种谱通量的情况下，召回率相对较高而精确率较低，这表明阈值因子 λ 实际上设得过低，无法在误检和漏检之间取得良好平衡。

C. 使用快速*Griffin-Lim*算法进行无相位重建

在本实验中，我们评估无相位重建，即在所提出的小波抽取方案中，仅根据幅度时频系数重建音频信号。Griffin-Lim算法 [62] （GLA）仍然是从短时傅里叶变换或一般时频谱图进行无相位重建时最常用的迭代方法。这里，我们考虑 Perraudin 等人提出的快速Griffin-Lim算法（FGLA）变体 [39], ，该变体引入了 Nesterov式加速项。重建误差以相对频谱误差衡量，通常称为频谱收敛，其定义为

$$\mathbf{err_{MS}}(f, f_r) = 10 \log_{10} \frac{\| |W_\psi f_r| - |W_\psi f| \|^2}{\|W_\psi f\|^2}, \qquad (9)$$

目标信号 f 和重建信号 f_r。在文献中，关于量$\mathbf{err_{MS}}$ 是基于哪种时频表示计算的，往往并不明确。由于我们要比较不同表示和参数选择下的结果，因此按照 [65], 固定用于计算$\mathbf{err_{MS}}$的表示。这里，我们选择一种过采样率很高的小波变换作为参考表示，该变换使用柯西型母小波，且$\alpha = 1000$，具有 $M+1 = 181$ 个按几何间隔排列的频率通道， $d = 7$，即约 50 倍过采样。

实验设置与 [50, 第 4.2]节类似：我们考虑来自 EBU SQAM 数据集[7]的相同 15 个信号，并在过采样率[8] 为 3、5 和 10 时，测试参数为 $\alpha = 1000$ 的柯西型母小波。对于所提出的方法，我们分别为低、中、高过采样选择 $M = 750$、968 和 1369。我们还使用 11、13 和 17 个补偿通道。补偿通道的数量按照第 IV-A 节中的方法确定。在 $M = 750$ 的过采样率为 3 时取得令人鼓舞的结果后，通过对采样网格进行各向同性缩放来选择更高过采样率下的参数，即乘积 dM 保持不变。[9]作为参照，我们首先考虑过采样率相同但采用几何频率间隔的小波变换。由于 [50, 第 4.2]节中使用的均匀抽取方案在低和中过采样率下会导致系统不稳定，因此我们还采用了 II-A 节中讨论的通道相关抽取。低、中、高过采样率下的通道数分别设为 $M = 90$、125 和 180。此外，我们还考虑短时傅里叶变换，使用 1536 个采样点的汉宁窗，并在相同冗余度下采用 $M = 1536$、1920 和 2880 个通道进行均匀抽取。为避免初始化不佳带来的问题，我们先针对初始相位 0 和五个随机均匀分布的相位初始化，分别执行 20 次快速 Griffin-Lim 算法迭代。然后从这六个候选结果中选出最佳结果，用于

[7]信号最初的 5 01、02、04、14、15、16、27、39、49、50、51、52、53、54、70 秒以及 .[8]请注意，考虑到系数为复数， [50, 第 4 节2] 中考虑的值 M/a 大约对应于过采样率的一半。[9]请注意，要求 d, $M \in \mathbb{N}$ 会导致舍入误差。此外，我们使用的实现会选择满足 $M/(2d)$ 至少达到所需过采样率的最大值 d。因此，实际过采样率可能略高。

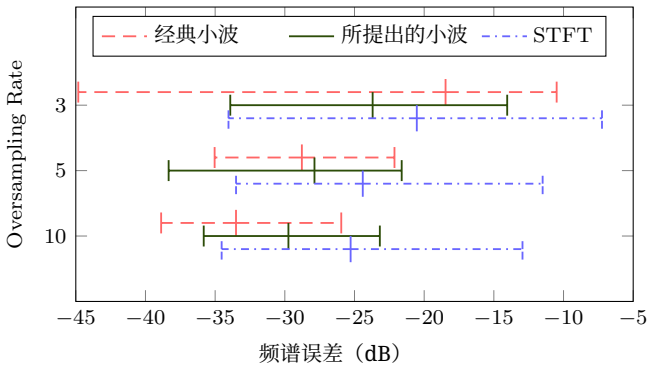

图8. 须状图展示了对 15 个信号和三种变换进行无相位重建后的最小、中位数和最大频谱误差。这三种变换分别为经典小波变换、采用基于克罗内克序列的抽取的所提出的小波变换，以及短时傅里叶变换。不同的过采样率按垂直方向排列。

再执行 130 次快速 Griffin-Lim 算法迭代，以计算最终解。

结果：图8总结了我们的结果。总体而言，我们发现过采样率提高时性能明显改善，这与 [65] 中得到的结果一致。平均而言，在中等和高过采样率下，经典小波抽取优于短时傅里叶变换；在高过采样率下，也优于所提出的方法。在中等过采样率下，两种小波方法大致相当。在低过采样率下，所提出的方法优于两种参考方法。我们在相关网站上提供了音频示例，以及在计算**误差**$_{\mathrm{MS}}$ 时以参考短时傅里叶变换为基准得到的额外结果。正如 [65],**误差**$_{\mathrm{MS}}$ 所示，该误差偏向于与参考相似的表示；我们发现，与图8相比，以短时傅里叶变换为参考计算**误差**$_{\mathrm{MS}}$ 时，总体上更有利于短时傅里叶变换而非小波变换，尤其更有利于克罗内克抽取而非经典小波抽取。不过，整体发现相似，因此我们不在此列出这些结果。

VI. 结论与展望

我们提出了一种新颖的均匀小波抽取方案，并引入了准随机延迟。通过概念分析和数值评估，我们证明，即使过采样率接近 1，该方案也适用于构造具有完美重建性质的小波系统。在采用 Itakura-Saito 非负矩阵分解的音频分解应用中，我们展示了：只需将滤波器组系数视为时频矩阵，先前为短时傅里叶变换提出的处理方案便可轻松适配我们的设置，这与经典的非均匀小波抽取形成对比。此外，我们通过一组小规模音频样本测试集说明，只需对最初为短时傅里叶变换定义的频谱通量方法进行简单调整，我们的方法便可用于起音检测。最后，我们观察到，该方法在多种音频信号的无相位重建中表现出令人鼓舞的性能。

我们预计，所提出的抽取策略将成为未来在不同应用领域开展工作的宝贵资源。无论是在音频领域还是其他领域。此外，尽管本文对所提出的方法的分析仅限于有限系统，但我们预计，所提出的构造在连续域中同样能够产生可逆小波变换。目前正在结合框架理论和函数空间理论，对其在该情形下的形式化数学性质进行严格研究。

致谢

我们感谢 Cédric Févotte 慷慨提供 Itakura-Saito 非负矩阵分解的代码，并指导我们如何使用。我们还感谢 Georg Tauböck 就所提出方法的潜在应用进行了富有成效的讨论。

参考文献

[1] J. C. Brown 和 M. S. Puckette，“恒定Q变换的高效计算算法”，《美国声学学会期刊》，第92卷，第5期，第2698–2701页，1992年。[2] B. C. Moore，《听觉心理学导论》。Brill，2012年。[3] C. Schörkhuber 和 A. Klapuri，“用于音乐处理的恒定Q变换工具箱”，载于第七届声音与音乐计算会议，西班牙巴塞罗那，2010年，第3–64页。[4] C. Schörkhuber、A. Klapuri 和 A. Sontacchi，“使用恒定Q变换进行音频移调”，音频工程学会期刊，第61卷，第7/8期，第562–572页，2013年。 [在线]。可访问：http: //www.aes.org/e-lib/browse.cfm?elib=16071

[5] B. Fuentes、A. Liutkus、R. Badeau 和 G. Richard，“使用恒定Q变换提取主旋律的概率模型”，载于2012年IEEE国际声学、语音与信号处理会议（ICASSP），2012年，第5357–5360页。[6] M. Todisco、H. Delgado 和 N. Evans，“恒定Q倒谱系数：自动说话人验证中的欺骗攻击对策”，《计算机语音与语言》，第45卷，第516–535页，2017年。 [在线]。可访问：https://www.sciencedirect.com/science/ article/pii/S0885230816303114

[7] C. O. Sakar、G. Serbes、A. Gunduz、H. C. Tunc、H. Nizam、B. E. Sakar、M. Tutuncu、T. Aydin、M. E. Isenkul 和 H. Apaydin，“用于帕金森病分类的语音信号处理算法比较分析及可调Q因子小波变换的应用”，《应用软计算》，第74卷，第255–263页，2019年。 [在线]。可访问：https://www.sciencedirect.com/science/article/pii/S1568494618305799

[8] S. Kadambe 和 G. Boudreaux-Bartels，“小波变换在语音信号音高检测中的应用”，《IEEE信息论汇刊》，第38卷，第2期，第917–924页，1992年。[9] G. Tzanetakis 和 P. Cook，“音频信号的音乐流派分类”，《IEEE语音与音频处理汇刊》，第10卷，第5期，第293–302页，2002年。[10] C.-C. Lin、S.-H. Chen、T.-K. Truong 和 Y. Chang，“基于小波和支持向量机的音频分类与类别划分”，《IEEE语音与音频处理汇刊》，第13卷，第5期，第644–651页，2005年。[11] S. Mallat，《信号处理小波之旅》（第三版）。波士顿：Academic Press，2009年。[12] M. Holschneider、R. Kronland-Martinet、J. Morlet 和 P. Tchamitchian，“借助小波变换进行信号分析的实时算法”，载于《小波》，J.-M. Combes、A. Grossmann 和 P. Tchamitchian 编。柏林、海德堡：Springer Berlin Heidelberg，1990年，第286–297页。[13] P. Dutilleux，“用于计算小波变换的‘空洞算法’实现”，载于《小波》，J.-M. Combes、A. Grossmann 和 P. Tchamitchian 编。柏林、海德堡：Springer Berlin Heidelberg，1990年，第298–304页。[14] I. W. Selesnick，“具有可调Q因子的小波变换”，《IEEE信号处理汇刊》，第59卷，第8期，第3560–3575页，2011年。[15] N. Holighaus、M. Dörfler、G. A. Velasco 和 T. Grill，“可逆实时恒定Q变换框架”，《IEEE音频、语音与语言处理汇刊》，第21卷，第4期，第775–785页，2013年4月。[16] C. Schörkhuber、A. Klapuri、N. Holighaus 和 M. Dörfler，“用于具有对数频率分辨率的高效完美重构时频变换的Matlab工具箱”，载于《第53届国际音频工程学会会议：语义音频》论文集，2014年1月。[在线]。可访问：http://www.aes.org/e-lib/browse.cfm?elib=17112

[17] O. Christensen，《框架与Riesz基导论》，应用与数值调和分析丛书。沙姆：施普林格国际出版公司，2016年。[18] M. S. Puckette 等，“Pure Data”，载于国际计算机音乐会议，1997年。[19] O. Lartillot、P. Toiviainen 和 T. Eerola，“用于音乐信息检索的Matlab工具箱”，载于《数据分析、机器学习与应用》，C. Preisach、H. Burkhardt、L. Schmidt-Thieme 和 R. Decker 编。柏林，德国：施普林格，2008年，第261–268页。[20] “信息技术——用于数字存储媒体、速率最高约为1.5 Mbit/s的活动图像及相关音频编码——第3部分：音频”，国际标准化组织，日内瓦，瑞士，标准，1993年8月。[21] W. G. Gardner，“无输入/输出延迟的高效卷积”，载于第97届音频工程学会年会。音频工程学会，1994年。[22] T. Necciari、N. Holighaus、P. Balazs、Z. Prˇuša、P. Majdak 和 O. Derrien，“Audlet滤波器组：使用听觉频率尺度的多用途分析/合成框架”，《应用科学》，第8卷，第1期：96，2018年。[23] T. Necciari、P. Balazs、N. Holighaus 和 P. L. Søndergaard，“erblet变换：一种基于听觉的具有完美重构的时频表示”，载于2013年IEEE国际声学、语音与信号处理会议，2013年，第498–502页。[24] P. Balazs、M. Dörfler、F. Jaillet、N. Holighaus 和 G. Velasco，“非平稳Gabor框架的理论、实现与应用”，《计算与应用数学杂志》，第236卷，第6期，第1481–1496页，2011年10月。[25] R. Levie 和 H. Avron，“使用连续框架的随机信号处理”，《傅里叶分析与应用杂志》，第28卷，第5期，2021年。[26] ——，“时频分析中的随机连续框架”，arXiv预印本 arXiv:2009.*10525 [*math.NA*]*, 2021。[27] R. Levie、H. Avron 和 G. Kutyniok，“准蒙特卡洛时频分析”，arXiv预印本 arXiv:2011.*02025 [*math.NA*]*, 2021。[28] R. F. Bass 和 K. Gröchenig，“带限函数的相关采样”，《伊利诺伊数学杂志》，第 57, 1, 43 58, 卷，第 2013 [ ] 10 1215 57, 1, 43 58, no. pp. – 2013 [ ] 10 1215 。在线。可访问：https://doi.org/ . /ijm/ 1403534485 [29] H. Führ 和 J. Xian，“有限生成移位不变空间中的相关采样”，《逼近理论杂志》，第240卷，15，2019年 1pp. – 。[30] D. Patel 和 S. Sampath，“$l^p(\mathbb{Z}^n)$再生核子空间中的随机采样”，《数学分析与应用杂志》，第491卷，第1期，第124270页，2020年。[31] P. Goyal、D. Patel 和 S. Sampath，“混合勒贝格空间再生核子空间中的随机采样”，2021年。[32] G. A. Velasco，“时频局部化函数的短时傅里叶变换的相关采样”，2017年。[33] A. J. E. M. Janssen，《Weyl-Heisenberg框架的对偶条件》。波士顿，马萨诸塞州：Birkhäuser Boston，1998年，第33–84页。[34] H. Bölcskei、F. Hlawatsch 和 H. Feichtinger，“过采样滤波器组的框架理论分析”，《IEEE信号处理汇刊》，第46卷，第12期，第3256–3268页，1998年。[35] H. Niederreiter，《随机数生成与准蒙特卡洛方法》。工业与应用数学学会，1992年。 [在线]。可访问：https://epubs.siam.org/doi/abs/10.1137/1.9781611970081[36] J. Dick 和 F. Pillichshammer，《数字网与序列：差异理论与准蒙特卡洛积分》。剑桥大学出版社，2010年。[37] C. Févotte、N. Bertin 和 J.-L. Durrieu，“具有Itakura-Saito散度的非负矩阵分解：在音乐分析中的应用”，《神经计算》，第21卷，第3期，第793–830页，2009年。[38] J. P. Bello、L. Daudet、S. Abdallah、C. Duxbury、M. Davies 和 M. B. Sandler，“音乐信号起音检测教程”，《IEEE语音与音频处理》，第13卷，第5期，第1035–1047页，2005年。[39] N. Perraudin、P. Balazs 和 P. L. Sondergaard，“ 快速 Griffin-Lim 算法 ”，载于 IEEE 应用信号处理、音频与声学会议论文集，美国纽约州纽帕尔茨，2013年10月。[40] N. Holighaus、G. Koliander、Z. Prˇuša 和 L. D. Abreu，“解析小波变换的刻画与一种新的无相位重构算法”，《IEEE信号处理汇刊》，第67卷，第15期，第3894–3908页，2019年。[41] J. M. Lilly 和 S. C. Olhede，“关于解析小波变换”，《IEEE信息论汇刊》，第56卷，4156，2010年8月，4135 no. pp. – 。[42] I. Daubechies 、 A. Grossmann 和 Y. Meyer，“ 简便的非正交展开 ”，《数学物理学杂志》，第27卷，第5期，第1271–1283页，1986年。[43] K. Grochenig，“框架算法的加速”，《IEEE信号处理汇刊》，第41卷，第12期，第3331–3340页，1993年。[44] I. Daubechies，《小波十讲》。SIAM，1992年。[45] M. Dr mota 和 R. F. Tichy，《序列、差异与应用》。Springer Berlin Heidelberg，1997年。[46] L. Kuipers 和 H. Niederreiter，《序列的均匀分布》。John Wiley，1974年。[47] J. G. van der Corput，“分布函数 I-II”，《阿姆斯特丹皇家科学院学报》，第 38 卷，第 813–821、1058–1066 页，1935年。[48] H. Faure、P. Kritzer 和 F. Pillichshammer，“从 van der Corput 到准蒙特卡罗求积规则的现代序列构造”，《数学研究》，第 26 卷，第 5 期，第 760–822 页，2015年，谨以此文纪念 J.G. van der Corput（1890–1975）。[49] I. Daubechies 和 T. Paul，“时频局部化算子——几何相空间方法：II 膨胀的应用”，《逆问题》，第 4 卷，第 3 期，第 661–680 页，1988年8月。[50] N. Holighaus、G. Koliander、Z. Prˇuša 和 L. Abreu，“从小波变换幅度进行非迭代无相位重建”，载于《2019年国际数字音频效果会议》（DAFx19），2019年9月。 [在线]。网址：http://dafx.de/paper-archive/2019/DAFx2019_paper_23.pdf[51] L. D. Abreu、K. Gröchenig 和 J. L. Romero，“关于累积频谱图”，《美国数学会汇刊》，第 368 卷，第 3629–3649 页，2016年1月。[52] J. O. Smith，“音频 FFT 滤波器组”，《第12届国际数字音频效果会议论文集》（DAFx-09），科莫，2009年。[53] G. A. Velasco、N. Holighaus、M. Dörfler 和 T. Grill，“利用非平稳 Gabor 框架构造可逆常数 Q 变换”，DAFX11 论文集，巴黎，第 33 卷，2011年。[54] Z. Prˇuša，“分段离散小波变换”，博士学位论文，布尔诺理工大学，布尔诺，2012年。[55] I. Daubechies 和 T. Paul，“时频局部化算子——几何相空间方法：II 膨胀的应用”，《逆问题》，第 4 卷，第 3 期，第 661—680 页，1988年。[56] S. C. Olhede 和 A. T. Walden，“广义 Morse 小波”，《IEEE 信号处理汇刊》，第 50 卷，第 11 期，第 2661–2670 页，2002年11月。[57] R. X. Gao 和 R. Yan，《小波：制造业中的理论与应用》。Springer Science & Business Media，2010年。[58] K. N. Chaudhury 和 M. Unser，“小波基希尔伯特变换对及类 Gabor 变换的构造”，《IEEE 信号处理汇刊》，第 57 卷，第 9 期，第 3411—3425 页，2009年。[59] Z. Prˇuša 和 N. Holighaus，“非迭代滤波器组相位（重）构建”，载于《第25届欧洲信号处理会议论文集》（EUSIPCO–2017），2017年8月，第 952–956 页。[60] N. Holighaus、Z. Prˇuša 和 P. L. S. ndergaard，“通用时频滤波器组的重分配与同步压缩、子采样及处理”，《信号处理》，第 125 卷，增刊 C，第 1–8 页，2016年。[61] “起始点检测数据库”，由阿利坎特大学模式识别与人工智能组（PRAIg-UA）提供。 [在线]。网址：https://grfia.dlsi.ua.es/cm/projects/prosemus/ database.php [62] D. Griffin 和 J. Lim，“从修正短时傅里叶变换进行信号估计”，《IEEE 声学、语音与信号处理汇刊》，第 32 卷，第 2 期，第 236–243 页，1984年4月。[63] X. Zhu、G. Beauregard 和 L. Wyse，“从修正短时傅里叶变换幅度谱进行实时信号估计”，《IEEE 音频、语音与语言处理汇刊》，第 15 卷，第 1645–1653 页，2007年7月。[64] C. Févotte，“用于平滑 Itakura-Saito 非负矩阵分解的主化-最小化算法”，载于《2011年 IEEE 国际声学、语音与信号处理会议》（ ICASSP ），2011 年，第 1980–1983 页。[65] A. Marafioti、N. Holighaus 和 P. Majdak，“音频时频相位恢复——变换参数的影响”，《IEEE 信号处理汇刊》，第 69 卷，第 3585–3596 页，2021年。