Chapter 1

Introduction

John J. Benedetto
Paulo J. S. G. Ferreira

1.1 The Classical Sampling Theorem

Cauchy proved the following result in 1841 [Cau41].

Theorem 1.1. *[Cauchy's Theorem] Let*

$$f(t) = \sum_{|n| \leq M} c_n e^{2\pi i t n},$$

and set $N = 2M + 1$. *Then*

$$f(t) = \sin(\pi t N) \sum_{m=0}^{N-1} \frac{1}{N} f\left(\frac{m}{N}\right) \frac{(-1)^m}{\sin \pi \left(t - \frac{m}{N}\right)}. \tag{1.1}$$

Theorem 1.1 is a sampling theorem since (1.1) allows us to write f as a sum of its *sampled values* $f\left(\frac{m}{N}\right)$, $m = 0, 1, \ldots, N-1$, each multiplied by the corresponding translate m/N of the *sampling function* $\frac{1}{N} \frac{\sin \pi N t}{\sin \pi t}$. Further, it fits into the format of the Classical Sampling Theorem, Theorem 1.2, since the polynomial f can be thought of as an M-bandlimited function. Cauchy used Lagrange's interpolation formula in one of his proofs of (1.1).

Remark 1.1. *[Interpolation and Number Theory]*

a. Theorem 1.1 influenced some developments in interpolation theory and number theory at the end of the 19th century. For the former subject there is the work of Borel, Hadamard, and de la Vallée-Poussin during the period 1898–1908, e.g., [HM26, page 50], [Hig85, pages 29–50], and [Boa54].

b. Let $\pi(x)$ denote the number of primes less than or equal to x. The prime number theorem (PNT) asserts that $\pi(x) \sim x/\log x$, $x \to \infty$; and the PNT was proved independently in 1896 by Hadamard and de la Vallée-Poussin. Earlier, in 1894 (in a note in the Comptes Rendus Acad. Sci., Paris), von Koch constructed entire functions f having the form of nonperiodic versions on $\mathbb{R}$ of Cauchy's formula (1.1). Using such functions f as building blocks for more complicated expressions F, he was able to write $\pi(n)$ in terms of values $F(m)$ for $m \leq n$.

von Koch quoted the work of Hadamard and Poincaré on entire functions, and, although his approach plays a role in Steffensen's point of view [Ste14], one suspects it was not pursued after 1896.

It should be pointed out that the PNT is equivalent to the fact that

$$\forall t \in \mathbb{R}, \quad \zeta(1+it) \neq 0, \tag{1.2}$$

where ζ is the Riemann zeta function; and both Hadamard and de la Vallée-Poussin proved (1.2). A reasonable way to prove that (1.2) implies the PNT is to use Wiener's Tauberian Theorem [Ben75, pages 128–136] or the so-called Wiener-Ikehara Tauberian Theorem, cf., [Whi35, page 64] to see a role of the Tauberian Theorem in dealing with the study of the Newton-Gauss interpolation formula as developed by both Steffensen [Ste14] and E. T. Whittaker [Whi15].

c. In 1915, E. T. Whittaker [Whi15, page 187] introduced the terminology *cardinal function* in the context of interpolation theory, cf., the results of J. M. Whittaker [Whi29a].

For a given sequence $\{c_n\} \subseteq \mathbb{C}$ and a fixed value of $T > 0$, E. T. Whittaker considered a class $X(\mathbb{R})$ of functions on $\mathbb{R}$, each of whose elements f has the property that

$$\forall n \in \mathbb{Z}, \quad Tf(nT) = c_n. \tag{1.3}$$

He referred to $X(\mathbb{R})$ as a *cotabular set of functions*. For a given $f \in X(\mathbb{R})$ he then posed the problem of finding $f_c \in X(\mathbb{R})$ which is Ω-bandlimited, where $2T\Omega = 1$. His goal was to replace a "given function $f \in X(\mathbb{R})$ by a cotabular function in such a way as to remove all the rapid oscillations from it" [Whi15, page 184]. He referred to his solution

$$f_c(t) = (\sin 2\pi t\Omega) \sum c_n \frac{(-1)^n}{\pi(t-nT)}, \tag{1.4}$$

as the *cardinal function* of $X(\mathbb{R})$. (Whittaker's use of the adjective "cardinal" is in the sense of "primary" or "to hinge upon" from cardo, cardinis, the Latin noun for hinge.)

We shall now state the Classical Sampling Theorem. To do so, we define the *Fourier transform* of $f \in L^1(\mathbb{R})$ to be the complex-valued function $\widehat{f} : \widehat{\mathbb{R}} \to \mathbb{C}$ defined by

$$\forall \gamma \in \widehat{\mathbb{R}}, \quad \widehat{f}(\gamma) = \int f(t) e^{-2\pi i t\gamma}\, dt, \tag{1.5}$$

where integration is over the real line $\mathbb{R}$, and $\widehat{\mathbb{R}}$ is $\mathbb{R}$ considered as a frequency domain. $L^1(\mathbb{R})$ is the space of complex-valued (Lebesgue) integrable functions on $\mathbb{R}$; and the Plancherel Theorem allows us to define $\widehat{f}$ for $f \in L^2(\mathbb{R})$, the space of square-integrable functions, i.e., $|f|^2 \in L^1(\mathbb{R})$. The L^2-norm of $f \in L^2(\mathbb{R})$ is $\|f\|_{L^2(\mathbb{R})} = (\int |f(t)|^2 \, dt)^{1/2}$. The *Paley–Wiener* space PW_Ω is

$$PW_\Omega = \{f \in L^2(\mathbb{R}) : \operatorname{supp} \widehat{f} \subseteq [-\Omega, \Omega]\},$$

where "supp $\widehat{f}$ " designates the support of $\widehat{f}$, i.e., the smallest closed set outside of which $\widehat{f}$ vanishes. For further reference, $F^\vee$ denotes the inverse Fourier transform of F, and $e_\gamma(t) = e^{2\pi i t\gamma}$.

Theorem 1.2. *[Classical Sampling Theorem] Let $T, \Omega > 0$ satisfy the condition that $0 < 2T\Omega \leq 1$, and let $s \in PW_{1/(2T)}$ satisfy the condition that $\widehat{s}$ is a bounded function on $\mathbb{R}$ which equals 1 on $[-\Omega, \Omega]$. Then*

$$\forall f \in PW_\Omega, \quad f = T \sum f(nT) \tau_{nT} s, \tag{1.6}$$

where the convergence of the sum is in L^2-norm and uniformly on $\mathbb{R}$, and where $(\tau_{nT} s)(t)$ designates the translation $s(t - nT)$.

In the sampling formula (1.6), the *sampling period* is T, the *sampling sequence* is $\{nT : n \in \mathbb{Z}\}$, the sequence of *sampled values* is $\{f(nT) : n \in \mathbb{Z}\}$, and the *sampling function* is s. Since the sampling sequence is equispaced, Theorem 1.2 is a *uniform* or *regular* sampling theorem.

The *sampling rate* is the number of samples taken per second. Because of Theorem 1.2, if $f \in PW_\Omega$ and $2T\Omega \leq 1$, and if f is sampled every T seconds, then f can be perfectly reconstructed in terms of these sampled values. In this case, with $2T\Omega \leq 1$, the *minimum* sampling rate for which we have reconstruction by (1.6) *for all* $f \in PW_\Omega$ is 2Ω samples per second. This minimum sampling rate, 2Ω, is the *Nyquist rate* [Nyq28].

Remark 1.2. *[History and Proof of the Classical Sampling Theorem]*

a. An important special case of Theorem 1.2 is for the Dirichlet sampling function

$$s(t) = \frac{\sin 2\pi\Omega t}{\pi t}, \tag{1.7}$$

in which case $2T\Omega = 1$ since $\widehat{s} = \mathbf{1}_{[-\Omega,\Omega)}$, the characteristic function of the frequency interval $[-\Omega, \Omega)$. (If $\Omega = 1/2$, s is the sinc function.) It should be pointed out that in the case $2T\Omega < 1$ and s is smooth, then the sampling formula converges faster than in the case of (1.7).

b. Theorem 1.2 is often called the Shannon Sampling Theorem, because of Shannon's use of (1.6) in his theory of communication, see [Sha48a], [Sha48b], [Sha49a], [Sha49b]. Shannon was in fact aware of the role of (1.6) in communication theory prior to his own work, and his application of Theorem 1.2

in this work should not in any way be confused with his profound contributions.

Theorem 1.2 is also sometimes named after E. T. Whittaker (1915) and Kotel'nikov (1933), but, as we indicated in Remark 1.1, the result was already known prior to their results. (A feature of this volume is Katsnelson's English translation of Kotel'nikov's famous paper [Kot33], which was written in Russian and has not been readily available.)

There is an extensive and important literature on uniform sampling formulas such as (1.6). Jerri's tutorial review [Jer77] (1977) has had an enormous influence, as have the fundamental papers of Butzer and his school, e.g., [But83b] and [BSS87], and the important historical survey by Higgins [Hig85], see also Higgins' [Hig96a] and Zayed's [Zay93] books.

c. The proof of Theorem 1.2 is elementary but depends on the beautiful idea of *periodization.* The essential calculation is

$$\begin{aligned} &\|f - T \sum_{|n|\leq N} f(nT)\tau_{nT} s\|_{L^2(\mathbb{R})} \\ &= \|\widehat{f}(\gamma) - T \sum_{|n|\leq N} f(nT) e^{-2\pi i n T\gamma}\widehat{s}(\gamma)\|_{L^2[-\frac{1}{2T},\frac{1}{2T}]} \\ &= \|G - S_N \widehat{s}\|_{L^2[\frac{1}{2T},\frac{1}{2T}]} = \|\widehat{s}(G - S_N)\|_{L^2[-\frac{1}{2T},\frac{1}{2T}]}, \end{aligned} \tag{1.8}$$

where G is the $1/T$-periodic function on $\widehat{\mathbb{R}}$ defined as $\widehat{f}\,\mathbf{1}_{(-\Omega,\Omega)}$ on $[-\frac{1}{2T},\frac{1}{2T})$ and where S_N is the N-th partial sum of the Fourier series of G. The right side of (1.2) tends to 0 as $N \to \infty$.

The notions of periodization and the details for (1.2) are found in [Ben97].

d. A conceptually equivalent proof to that outlined in part *c* makes use of the Poisson Summation Formula (PSF),

$$T \sum f(t + nT) = \sum \widehat{f}\left(\frac{n}{T}\right) e^{2\pi i n t/T}, \tag{1.9}$$

which itself is essentially equivalent to (1.6). The details of proof of Theorem 1.2 by means of PSF are also found in [Ben97].

Periodization and the PSF are meaningful ideas in the setting of locally compact groups and their discrete subgroups; and there are applications as varied as the Euler-Maclaurin and Selberg trace formulas. It should be pointed out that the proof of the PSF can be a delicate exercise depending on the space of functions or distributions being considered, see [BZ97] for various proofs and the role of the notion of a *sampling multiplier.*

In the following result, we generalize Theorem 1.2 to the case of non-bandlimited functions. The hypotheses seem technical but are natural.

Theorem 1.3. *[Nonbandlimited Gabor Sampling Theorem] Let $T, \Omega > 0$ be constants for which $0 < 2T\Omega \leq 1$, and let $g \in PW_{1/(2T)}$ have the properties that $\widehat{g}$ is bounded on $\widehat{\mathbb{R}}$, $\widehat{g} = 1$ on $[-\Omega, \Omega]$, and, in case $2T\Omega < 1$, $\widehat{g}$ is continuous and*

$$|\widehat{g}| > 0 \text{ on } (-\frac{1}{2T}, -\Omega] \cup [\Omega, \frac{1}{2T}).$$

Set

$$G(\gamma) = \sum_{m \in \mathbb{Z}} |\widehat{g}(\gamma - mb)|^2 \text{ and } s(t) = \left(\frac{\widehat{g}}{G}\right)^{\vee}(t),$$

where $\Omega + 1/(2T) \leq b < 1/T$ in case $2T\Omega < 1$ and $\Omega + 1/(2T) = b$ if $2T\Omega = 1$. Then $s \in PW_{1/(2T)}$, $\widehat{s}$ is bounded on $\widehat{\mathbb{R}}$, $\widehat{s} = 1$ on $[-\Omega, \Omega]$,

$$\forall f \in L^2(\mathbb{R}), \quad f = T \sum_{m,n \in \mathbb{Z}} \langle \widehat{f}, e_{nT}\tau_{mb}\widehat{g}\rangle \tau_{-nT}(e_{mb}s), \tag{1.10}$$

and, as a consequence,

$$\forall f \in PW_{\Omega}, \quad f = T \sum f(nT)\tau_{nT}s, \tag{1.11}$$

where the convergence of each sum is in L^2-norm.

Remark 1.3. *[Gabor Systems and Aliasing]* Equation (1.11) is the Classical Sampling Theorem for bandlimited functions f and for the sampling function s. Note that if $f \in PW_{\Omega}$, then $\langle \widehat{f}, e_{nT}\tau_{mb}\widehat{g}\rangle = 0$ for all $m \neq 0$, and so equation (1.11) is a consequence of (1.10) since $\langle \widehat{f}, e_{nT}\widehat{g}\rangle = f(-nT)$.

Theorem 1.3 first appeared in [BH90] (1990), also see [Ben92], [Ben94]. The proof depends on the fact that the so-called *Gabor system* $\{e_{nT}\tau_{mb}\widehat{g}\}$ is a frame for $L^2(\widehat{\mathbb{R}})$. *Frames* will be defined in Section 1.2. There are analogues of Theorem 1.3 in terms of *wavelet systems* $\{2^{m/2}\,\psi(2^m t - n)\}$, e.g., [Mey90], [Dau92], and there are many classical expansions of nonbandlimited functions, e.g., [BS92].

Criteria for linear spans of Gabor systems to be dense in $L^2(\widehat{\mathbb{R}})$ go back to von Neumann [vN55, pages 405 ff.], cf., [BGZ75]. Signal decompositions in terms of Gabor systems, leading to Gabor frames, were initiated by Gabor [Gab46] in 1946, see [BF94] (Chapters 3 and 7) as well as [FS98].

Equation (1.10) can be thought of in terms of quantifying aliasing error. In fact, we know that in dealing with high frequency time series it is necessary to sample closely in order to capture all of the fluctuations. Thus, in the case of very high frequency information f, thought of as "infinite frequencies" and hence nonbandlimited, we can not reconstruct f with a discrete set of samples. In this context, the terms of the sum in (1.10) for $n \in \mathbb{Z}$ and $m \neq 0$ can be thought of as dealing with the "infinite frequencies" associated with an arbitrary signal $f \in L^2(\mathbb{R})$.

Besides the extension Theorem 1.3 of Theorem 1.2 to the case of non-bandlimited functions, we can also ask about d-dimensional versions of Theorem 1.2. To this end we give the following definition.

Definition 1.1. *[Lattices]* A *lattice* $H \subseteq \mathbb{R}^d$ is the image of $\mathbb{Z}^d$ under some nonsingular linear transformation, i.e., H is a discrete subgroup of Euclidean space $\mathbb{R}^d$ consisting of integral linear combinations of elements $v_1, v_2, \ldots, v_d \in \mathbb{R}^d$, which form a basis for $\mathbb{R}^d$. The *reciprocal lattice* $\Lambda \subseteq \widehat{\mathbb{R}}^d$ of H is the lattice consisting of all $\gamma \in \widehat{\mathbb{R}}^d$ with the property that the inner product $\langle x, \gamma \rangle$ is an integer n_x for each $x \in H$.

A *unit cell* of a lattice Λ is a set $E \subseteq \widehat{\mathbb{R}}^d$, not necessarily connected, such that $\mathcal{T} = \{\gamma + E : \gamma \in \Lambda\}$ is a *tiling* or *partition* of $\widehat{\mathbb{R}}^d$, i.e., the elements of $\mathcal{T}$ are pairwise disjoint and

$$\bigcup_{\gamma \in \Lambda} (\gamma + E) = \widehat{\mathbb{R}}^d. \tag{1.12}$$

There are many possible choices for the unit cell of a given lattice. For example, the Voronoi cell or Brillouin zone is the unit cell of Λ defined as the set of all points in $\widehat{\mathbb{R}}^d$ closer to the origin than to any other lattice point.

Theorem 1.4. *[A d-dimensional Uniform Sampling Theorem] Let $H \subseteq \mathbb{R}^d$ be a lattice and let $E \subseteq \widehat{\mathbb{R}}^d$ be a unit cell of the reciprocal lattice Λ. Define the sampling function*

$$\forall x \in \mathbb{R}^d, \quad s_E(x) = \frac{1}{|E|} \int_E e^{2\pi i \langle x, \gamma \rangle}\, d\gamma,$$

where $|E|$ is the Lebesgue measure of E, and let $f \in L^2(\mathbb{R}^d)$ have the property that $\widehat{f} = 0$ a.e. off of E.

a. There is a continuous function f_c on $\mathbb{R}^d$ such that $f = f_c$ a.e.

b. If f is continuous on $\mathbb{R}^d$, then

$$f = \sum_{y \in H} f(y) \tau_y s_E,$$

where the convergence of the sum is in L^2-norm and uniformly on $\mathbb{R}^d$.

Remark 1.4. *[Early Applications Motivating d-Dimensional Uniform Sampling]*

In the mid-1950s, Brillouin [Bri56] (1956) discussed 3-dimensional uniform sampling with regard to some crystallographic problems (see the terminology "Brillouin zone" prior to Theorem 1.4), and Bracewell [Bra56] (1956) used 2-dimensional uniform sampling with regard to issues in radio astronomy. Miyakawa's basic formulation [Miy59] (1959), in terms of the "Nyquist relationships" between lattice and unit cell, was in the context of multivariate stochastic processes, see Sasakawa's applications of Miyakawa's Theorem in [Sas] (1960-61). Petersen and Middleton [PM62b] (1962) completed Miyakawa's approach, both theoretically and with many important examples, cf., Prosser's sampling theorem and analysis of truncation error in [Pro66]. Although not "early" relative to the

1950s, we also mention Dubois' application [Dub85] (1985) to video systems and Mersereau's work on hexagonal sampling [Mer79] (1979).

Motivating-applications from mathematics go back to J. M. Whittaker's uniform sampling theorem [Whi35] (1935) for entire functions f of order less than two. The sampled values in Whittaker's formula are $f(m+in)$, $m,n \in \mathbb{Z}$; and, as noted by Pólya, the formula yielded a positive solution to Littlewood's conjecture that if $\{f(m+in)\}$ is bounded then f is a constant. There are many other papers on d-dimensional uniform sampling, and Higgins' exposition [Hig85] is a reasonable place to start (but not to finish!).

The proofs of Theorem 1.4 and its uniform variants (sic!) are conceptually similar to that of Theorem 1.2. They depend essentially on periodization in the guise of the proper PSF or of the canonical Fourier expansions associated with $\mathbb{R}^d$ and its discrete subgroups. The Nyquist hypothesis $2T\Omega \leq 1$ and bandlimited hypothesis $f \in PW_\Omega$ of Theorem 1.2 are replaced in Theorem 1.4 by the pairing H, Λ (where Λ is essential for choosing some unit cell $E \subseteq \widehat{\mathbb{R}}^d$) and the hypothesis that $\widehat{f} = 0$ off of E, respectively.

Our final topic in this section is Igor Kluvánek's uniform sampling theorem [Klu65] (1965) for locally compact abelian groups (LCAGs). We are stating it, not only for its own sake, but because this more abstract context allows us to see clearly the role of congruence in conceiving uniform sampling formulas, see Remark 1.11. Further, the relation between congruence and uniform sampling has implications in wavelet theory, see Remark 1.13, and provides an understanding of "saving bandwidth" in some applications, see Example 1.12.

Let G be a LCAG with dual group $\widehat{G}$, and let $H \subseteq G$ be a discrete subgroup. $H^\perp$ denotes the annihilator subgroup

$$H^\perp = \{\gamma \in \widehat{G} : \forall x \in H, \quad (x,\gamma) = 1\}.$$

Haar measure on a locally compact abelian group X is denoted by m_X. In our case, we adjust the Haar measures on $\widehat{G}$, $H^\perp$, and $\widehat{G}/H^\perp$ so that $m_{\widehat{G}/H^\perp}(\widehat{G}/H^\perp) = 1$, $m_H(\{x\}) = 1 = m_{H^\perp}(\{\gamma\})$ for $x \in H$ and $\gamma \in H^\perp$, and

$$\int_{\widehat{G}} = \int_{\widehat{G}/H^\perp} \int_{H^\perp} .$$

Discrete annihilator subgroups of $\widehat{\mathbb{R}}^d$ are lattices Λ considered in Theorem 1.4. The PSF for LCAGs has been known from early-on, e.g., [Loo53, page 153] and [Rei68]. It is

$$\int_H f dm_H = \int_{H^\perp} \widehat{f} dm_{H^\perp}, \tag{1.13}$$

and it can be used to prove Theorem 1.5.

Theorem 1.5. *[Kluvánek's Uniform Sampling Theorem for LCAGs] Let $H \subseteq G$ be a discrete subgroup of a LCAG G, with discrete annihilator subgroup $H^{\perp} \subseteq \widehat{G}$. Let $E \subseteq \widehat{G}$ be any subset of finite Haar measure for which the canonical surjective map*

$$h : \widehat{G} \longrightarrow \widehat{G}/H^{\perp} \tag{1.14}$$

restricted to E is a bijection; and define the sampling function

$$\forall x \in G, \quad s_E(x) = \int_E (x,\gamma) dm_{\widehat{G}}(\gamma), \tag{1.15}$$

where $m_{\widehat{G}}$ is adjusted to have the property that $m_{\widehat{G}}(E) = 1$. Let $f \in L^2(G)$, and assume $\widehat{f} = 0$ a.e. off of E.

a. *There is a continuous function f_c on G such that $f = f_c$ a.e.*

b. *If f is continuous on G, then*

$$f = \sum_{y \in H} f(y) \tau_y s_E,$$

where the convergence of the sum is in L^2-norm and uniformly on G. Further, the "Gaussian quadrature" formula

$$\|f\|^2_{L^2(G)} = \sum_{y \in H} |f(y)|^2$$

is valid.

Remark 1.5. *[The Nyquist Rate and the Role of Congruence]*

a. The Nyquist hypothesis $2T\Omega \leq 1$ and the bandlimited hypothesis $f \in PW_{\Omega}$ of Theorem 1.2 are replaced in Theorem 1.5 by the pairing $H, H^{\perp}$ (where, because of (1.12), $H^{\perp}$ is essential for choosing some set $E \subseteq \widehat{G}$) and the hypothesis that $\widehat{f} = 0$ off of E, respectively.

   Clearly, the sampling function s_E defined in Equation (1.15) is the analogue of the Dirichlet sampling function defined in Equation (1.7).

b. Of course, in the case of sampling functions $s \in L^2(\mathbb{R})$ for which $\widehat{s}$ is the characteristic function $\mathbf{1}_E$ of some set E, the set E can be more complicated than the interval $[-\Omega, \Omega) \subseteq \widehat{\mathbb{R}}$. This is precisely the point of our formulation of Kluvánek's theorem in terms of the bijectivity hypothesis associated with (1.14). For example, in the case of $G = \mathbb{R}$, $E = [-\Omega, \Omega)$, and $H^{\perp} = 2\Omega\mathbb{Z}$ we have a tiling $\{2\Omega n + E : n \in \mathbb{Z}\}$ of $\widehat{\mathbb{R}}$ as in (1.12). On the other hand, there are many other such tilings for this lattice $H^{\perp}$. We could take $E = [-2\Omega, -\Omega) \cup [\Omega, 2\Omega)$, just as one does in defining the Shannon (or Littlewood-Paley or sinc) wavelet, see more on this relationship in Remark 1.6.

c. We note that since $H^\perp$ is discrete, then $\widehat{G}/H^\perp$ is compact because it can be identified with the dual group of H, e.g., [Rud62, Chapter 2.1], thereby allowing us to make the normalization $m_{\widehat{G}/H^\perp}(\widehat{G}/H^\perp) = 1$ prior to Theorem 1.5. This, along with the other aforementioned normalizations, establishes fixed Haar measures on $\widehat{G}/H^\perp$, $H, H^\perp$, and $\widehat{G}$. Finally, we can normalize Haar measure on G so that if the Fourier transform is formally defined as

$$\widehat{f}(\gamma) = \int_G f(x)(-x,\gamma)dm_G(x),$$

then the formal inversion formula is

$$f(x) = \int_{\widehat{G}} \widehat{f}(\gamma)(x,\gamma)dm_{\widehat{G}}(\gamma).$$

All of these normalizations are required to state the PSF (1.13) and to prove Theorem 1.5.

Example 1.1. [Efficient Sampling] Let s be the Dirichlet sampling function in (1.7). Suppose $f \in PW_{5\Omega}$ and $2T(5\Omega) = 1$. Then the Classical Sampling Theorem asserts that

$$f = \frac{1}{10\Omega} \sum f\left(\frac{n}{10\Omega}\right) \tau_{n/10\Omega} s; \tag{1.16}$$

and, in particular, sampling takes place on the set $\{n/10\Omega : n \in \mathbb{Z}\}$. If f has the further property that $\widehat{f}$ vanishes off of $E = [-5\Omega, -4\Omega) \cup [4\Omega, 5\Omega) \subseteq \widehat{\mathbb{R}}$, then (1.16) still requires sampling on the set $\{n/10\Omega\}$. On the other hand, we can consider $E \subseteq \widehat{\mathbb{R}}$ as the domain of the function h of (1.14) with the property that $h(E) = \widehat{\mathbb{R}}/H^\perp$ so that $H^\perp = \{2\Omega n : n \in \mathbb{Z}\}$ and $H = \{n/2\Omega : n \in \mathbb{Z}\}$. Thus, in this case, Theorem 1.5 allows us to write

$$f = \frac{1}{2\Omega} \sum f\left(\frac{n}{2\Omega}\right) \tau_{n/2\Omega} s_E;$$

and, in particular, we need only sample from the set $\{n/2\Omega\}$.

Remark 1.6. *[Wavelets and Tiling]* The notion of congruence inherent in the canonical function h of (1.14) and our choice of E plays a role in many areas of mathematics including several closely related to the present topic.

a. In 1990 Albert Cohen used such congruence criteria to establish orthonormality of scaling functions associated with wavelet multiresolution analysis (MRA), for a given quadrature mirror filter (QMF), e.g., [Dau92, pages 182–186].

b. Congruence of the type described by (1.14) and in Remark 1.5 also plays a fundamental role in the recent deep work on self-similar tilings of $\mathbb{R}^d$ by K. Gröchenig, A. Haas, J. C. Lagarias, W. Madych, Yang B. Wang, et al., e.g., [GM92], [LW97].

c. Until recently, it had been thought that integer dilates and multi-integer translates of more than one function were required to generate a wavelet orthonormal basis for $\mathbb{R}^d$, $d \geq 2$. This is the case when such bases are generated from a MRA. However, in 1992, using the notion of congruence and other methods, David Larson and Xingde Dai showed the existence of (non-MRA) dyadic wavelet orthonormal bases for $\mathbb{R}^d$, $d \geq 2$, from a single function ψ [DL97], [XDS97], cf., [HWW97]. These bases are of the form

$$\psi_{m,n}(x) = 2^{md/2}\psi(2^m x - n), \quad m \in \mathbb{Z}, \quad n \in \mathbb{Z}^d,$$

cf., the 1-dimensional form in Remark 1.3.

The first specific constructions of single dyadic wavelets are due to Soardi and Weiland [SW98] and Zakharov [Zak96]. A general and implementable method of construction was introduced by Benedetto and Leon [BL], [BL00] in 1998. Their method depends essentially on the function h of (1.14) and, in particular, on the underlying properties of the d-dimensional version of the Shannon wavelet mentioned in Remark 1.5. These properties are similar to those used by Leonardo da Vinci and Maurits C. Escher in some of their drawings.

1.2 Non-Uniform Sampling and Frames

This section serves as an introductory exposition of the subject of non-uniform sampling, hopefully providing some motivation and historical background. As with Section 1.1, it is constrained, not only by available space, but also by the authors' perception of the subject and their limitations.

Our point of departure is not non-uniform sampling, but rather the topic of closed spans of complex exponentials. Throughout this section, we let $\Lambda = \{\lambda_k : k \in \mathbb{Z}, \lambda_{-1} < 0 \leq \lambda_0, \text{ and } \lim_{k\to\pm\infty} \lambda_k = \pm\infty\} \subseteq \widehat{\mathbb{R}}$ be a strictly increasing sequence, which is *uniformly discrete* or *separated* in the sense that

$$\exists \delta > 0 \text{ such that } \forall k \in \mathbb{Z}, \quad \lambda_{k+1} - \lambda_k \geq \delta.$$

Further, for each $R > 0$, we let

$$X_R = \overline{\text{span}}\{e_\lambda : \lambda \in \Lambda\}$$

denote the closed linear span of $\{e_\lambda : \lambda \in \Lambda\}$ in $L^2[-R, R]$. Paley and Wiener [PW34, Chapter VI] refer to X_R as the "closure" of the set $\{e_\lambda : \lambda \in \Lambda\}$ of complex exponential functions. If $X_R = L^2[-R, R]$, then $\{e_\lambda : \lambda \in \Lambda\}$ is *complete* in $L^2[-R, R]$.

Definition 1.2. *[The Closure of Sets of Complex Exponential Functions]*

a. The *radius of completeness* $R_c(\Lambda)$ of Λ is

$$R_c(\Lambda) = \sup\{R \geq 0 : X_R = L^2[-R, R]\}.$$

$R_c(\Lambda)$ is well-defined since it is clear that if $R_1 < R_2$ and $X_{R_2} = L^2[-R_2, R_2]$, then $X_{R_1} = L^2[-R_1, R_1]$.

b. An essential *problem* is to compute $R_c(\Lambda)$, and in particular to find an intrinsic property of Λ so as to conclude that $X_R = L^2[-R, R]$ for a given R. Note that $X_R = L^2[-R, R]$ for each $R < R_c(\Lambda)$ and $X_R \neq L^2[-R, R]$ for each $R > R_c(\Lambda)$. $R_c(\Lambda)$ is equal to the radii of completeness for the L^p-spaces $L^p[-R, R]$, $1 \leq p < \infty$, as well as the space $C[-R, R]$ of continuous functions on $[-R, R]$, e.g., [Sch43].

The following theorem is an early, fundamental, and deep result due to Paley and Wiener [PW34, Section 26].

Theorem 1.6. *[A Paley–Wiener Completeness Theorem] Assume that Λ has the property that $\lambda_0 = 0$ and $\lambda_{-k} = -\lambda_k$ for $k \geq 1$. For each $\gamma > 0$ let $n(\gamma)$ be the cardinality of $\{\lambda_k : k \geq 1 \ \text{ and } \ \lambda_k \leq \gamma\}$. If*

$$\limsup_{\gamma \to \infty} \frac{n(\gamma)}{\gamma} > 2R, \tag{1.17}$$

then $X_R = L^2[-R, R]$.

Remark 1.7. *[Completeness and Density]*

a. The "$\overline{\lim}$" on the left side of (1.17) is a density condition, and such conditions are essential hypotheses, not only for completeness theorems such as Theorem 1.6 and Equation (1.23) below, but also for non-uniform sampling formulas such as (1.31) and (1.32).

   In engineering terms and in the context of non-uniform sampling, we can expect completeness if the number of samples per unit time exceeds *on average* twice the largest frequency in the given signal, i.e., if the average sampling rate exceeds the Nyquist rate. This sampling criteria is a density condition, and accurately quantifying the correct density to obtain completeness is difficult, e.g., see Definition 1.3 and Remark 1.8.

   Further, there are genuine engineering applications of some of these completeness theorems in uniquely determining signals from their non-uniformly spaced samples, e.g., Beutler's work [Beu66b] using results of Levinson.

b. Paley and Wiener's book [PW34] (1934) is the progenitor and driving force for an extensive and deep theory relating refinements of Theorem 1.6 with various notions of density, see Definition 1.17. Some of the many notable works since then are due to Levinson [Lev40] (1940), Duffin-Eachus [DE42]

(1942) and Duffin-Schaeffer [DS52] (1952), Kahane [Kah62] (1962), Beurling and Landau, e.g., [Lan67a] (1967), and Beurling and Malliavin [BM62] (1962), [BM67] (1967).

There are also world class expositions due to Koosis [Koo79] (1970), [Koo96] (1996) and Redheffer [Red77] (1977) reflecting the authors' profound understanding of the problems and their own seminal contributions from the 1960s, cf., [Boa54]. A more recent and justly influential exposition is due to Young [You80].

c. Paley and Wiener's formulation and vision relating density and completeness also has an important history prior to their work.

Theorem 1.6 is a significant extension of a completeness theorem due to Pólya and Szász (Jahresbericht der Deutschen Mathematiker Vereinigung, 43 (1933), 20) whose density condition is (1.17) with "$\underline{\lim}$" instead of "$\overline{\lim}$". The added depth of Theorem 1.6 is due to the use of Paley and Wiener's fundamental theorem on quasi-analytic functions, viz., Theorem XII of [PW34].

Wiener, himself, proved a completeness theorem in 1927 [Wie27] assuming a uniform density criterion, see Definition 1.3. Such a criterion allows a comparison between the sequences $\{e_n : n \in \mathbb{Z}\}$ and $\{e_\lambda : \lambda \in \Lambda\}$ in the sense that the completeness of $\{e_\lambda : \lambda \in \Lambda\}$ can be deduced from the completeness of $\{e_n : n \in \mathbb{Z}\}$.

Wiener was influenced by Birkhoff [Bir17] and Walsh [Wal21] for this point of view, and there was related work by Carleman [Car22] (1922) and Pólya [Pól29] (1929), among others. There was also important work by Dini [Din54] (1917), who not only obtained completeness theorems, but actually provided non-harmonic Fourier series expansions $\sum c_n \, e_{\lambda_n}$ for sequences $\{\lambda_n\}$ obtained from some transcendental equations from mathematical physics. Dini's formulation of non-harmonic Fourier series (see Definition 1.4) in this context goes back to his 1880 classic book on Fourier series.

d. The material in parts *b* and *c* is expanded as background in a forthcoming tutorial on multidimensional non-uniform sampling [BWa].

Definition 1.3. *[Density Criteria]* Let $\Lambda = \{\lambda_k : k \in \mathbb{Z},\ \lambda_{-1} < 0 \leq \lambda_0, \text{ and } \lim_{k\to\pm\infty} \lambda_k = \pm\infty\} \subseteq \widehat{\mathbb{R}}$ be a strictly increasing uniformly discrete sequence, and define the function

$$n_\Lambda = \sum k\mathbf{1}_{[\lambda_\mathbf{k},\lambda_\mathbf{k+1})}$$

whose distributional derivative is $n_\Lambda' = \sum_{\lambda\in\Lambda} \delta_\lambda$. Clearly, if $n : (0,\infty) \to \{0,1,\dots\}$ is defined by $n(\gamma) = \text{card}\{\lambda_k : |\lambda_k| \leq \gamma\}$, where "card" is cardinality, then $n(\gamma) = n_\Lambda(\gamma) - n_\Lambda(-\gamma)$.

a. A reasonable definition of the *density* of Λ is

$$\lim_{\gamma\to\infty} \frac{n(\gamma)}{2\gamma}$$

when this limit exists. As such, and since $n_\Lambda(\gamma_k) = k$, we shall define the *natural density* of Λ as

$$D_n(\Lambda) = \lim_{|k|\to\infty} \frac{k}{\lambda_k} \tag{1.18}$$

when the limit in (1.18) exists. Otherwise, we consider the *upper*, resp., *lower*, *natural densities*

$$D_n^+(\Lambda) = \limsup_{|k|\to\infty} \frac{k}{\lambda_k}, \quad \text{resp.,} \quad D_n^-(\Lambda) = \liminf_{|k|\to\infty} \frac{k}{\lambda_k}.$$

b. Λ has *uniform density* $D_u(\Lambda) > 0$ if

$$\exists C > 0 \text{ such that } \forall |\gamma| > 0,\ |n_\Lambda(\gamma) - D_u(\Lambda)\gamma| \le C. \tag{1.19}$$

Since $n_\Lambda(\lambda_k) = k$, the uniform density inequality (1.19) is equivalent to the condition that

$$\exists C > 0 \ \text{ such that } \ \forall k, \quad |\frac{k}{\lambda_k} - D_u(\Lambda)| \le \frac{C}{|\lambda_k|},$$

or, equivalently,

$$|\lambda_k - \frac{k}{D_u(\Lambda)}| \le \frac{C}{D_u(\Lambda)} = L.$$

In particular, if $D_u(\Lambda)$ exists, then $D_n(\Lambda)$ exists and equals $D_u(\Lambda)$, cf., Proposition 1.18. In fact, uniform density can be viewed as natural density constrained by a convergence rate of $1/|\lambda_k|$.

c. For each $\gamma > 0$ and each interval $I \subseteq \widehat{\mathbb{R}}$ of length γ, let $n_I(\gamma) = \text{card}\,\{\lambda_k \in I\}$. Define

$$n^-(\gamma) = \inf_I n_I(\gamma) \quad \text{and} \quad n^+(\gamma) = \sup_I n_I(\gamma).$$

The *lower* and *upper Beurling densities* of Λ are

$$D_b^-(\Lambda) = \lim_{\gamma\to\infty} \frac{n^-(\gamma)}{\gamma} \quad \text{and} \quad D_b^+(\Lambda) = \lim_{\gamma\to\infty} \frac{n^+(\gamma)}{\gamma},$$

respectively. These limits exist since n^- is superadditive and n^+ is subadditive, although it is more common to replace the limits by a lim inf and lim sup, respectively.

Proposition 1.1. *[Relation between Densities]*

a. If Λ has uniform density $D_u(\Lambda) \in (0, \infty)$, then

$$D_b^-(\Lambda) = D_b^+(\Lambda) = D_u(\Lambda).$$

b. If $D_b^-(\Lambda) = D_b^+(\Lambda) = D(\Lambda)$, then the natural density $D_n(\Lambda)$ exists and

$$D_n(\Lambda) = D(\Lambda).$$

Remark 1.8. *[Beurling-Malliavin Density]*

a. By definition of $R_c(\Lambda)$, it is clear that Paley and Wiener's Theorem 1.6 is equivalent to the assertion that

$$\frac{1}{2} D_n^+(\Lambda) \leq R_c(\Lambda). \tag{1.20}$$

Pólya [Pól29] introduced the notion that is called the *Pólya maximum density* $D_p^+(\Lambda)$ of Λ, and it has the property that $D_p^+(\Lambda) \geq D_n^+(\Lambda)$. In 1935 Levinson proved that if Λ is a positive sequence, then

$$\frac{1}{2} D_p^+(\Lambda) \leq R_c(\Lambda), \tag{1.21}$$

see [Lev40] and [Red77, page 22].

Further, Paley and Wiener [PW34] (Theorem XXXIV on page 94) proved that if $\Lambda = \{\lambda_k : \lambda_{-k} = -\lambda_k, k = 0, 1, \ldots\}$ has uniform density $D_u(\Lambda) > 0$ with small enough bound in (1.19), then

$$R_c(\Lambda) \leq \frac{1}{2} D_u(\Lambda). \tag{1.22}$$

Other upper bounds on $R_c(\Lambda)$ are due to Koosis [Koo58] (1958) and Redheffer, e.g., [Red54] (1954).

b. In light of (1.20), (1.21), and (1.22), it is not unreasonable to conjecture that $\frac{1}{2} D_n^+(\Lambda) = R_c(\Lambda)$ for symmetric sequences Λ, see [Sch43, page 130] for the conjecture in terms of $D_n(\Lambda)$. Kahane [Kah58] (1958) constructed a symmetric sequence with the properties that $D_n(\Lambda) = 0$ and $R_c(\Lambda) = \infty$. Kahane's sequence Λ is not uniformly discrete, cf., the results of Koosis [Koo60] and Redheffer [Red68] from the early 1960s.

c. In one of the highlights of 20th century analysis, Beurling and Malliavin [BM62], [BM67] in 1960–1961 devised a notion of density, denoted by $D_{bm}(\Lambda)$, allowing them to prove

$$\frac{1}{2} D_{bm}(\Lambda) = R_c(\Lambda). \tag{1.23}$$

The direction, $R_c(\Lambda) \geq \frac{1}{2} D_{bm}(\Lambda)$, is the easier to prove; and the direction, $R_c(\Lambda) \leq \frac{1}{2} D_{bm}(\Lambda)$, requires a deep study of the canonical product,

$$\prod \left(1 - \frac{z^2}{\lambda_k^2}\right),$$

using potential theory. The explanation par excellence is due to Koosis [Koo96], cf., [Kah61].

Besides its influence in the study of evaluating the radius of completeness $R_c(\Lambda)$ of Λ in terms of density criteria, Paley and Wiener's book [PW34] (1934) has also inspired the study of non-harmonic Fourier series, cf., our observation in Remark 1.7 about Dini's innovations in this area. As we shall see, the following definition, which also contains some historical background, essentially includes the notion of non-harmonic Fourier series.

Definition 1.4. *[Frames]* Let H be a separable Hilbert space with inner product $\langle x, y\rangle$ and norm $\|x\| = \langle x, x\rangle^{1/2}$.

a. A sequence $\{x_n : n \in \mathbb{Z}^d\} \subseteq H$ is a *Schauder basis* or *basis* for H if each $y \in H$ has a unique decomposition $y = \sum c_n(y)x_n$ in H, where $\{c_n(y)\} \subseteq \mathbb{C}$.

b. A basis $\{x_n\}$ for H is an *unconditional basis* for H if

$$\exists C > 0 \text{ such that } \forall F \subseteq \mathbb{Z}^d, \text{ finite, and } \forall\{b_j, c_j : j \in F\} \subseteq \mathbb{C},$$

$$\text{where } |b_j| \leq |c_j| \text{ for each } j \in F, \text{ we have}$$

$$\|\sum_{n\in F} b_n x_n\| \leq C\|\sum_{n\in F} c_n x_n\|.$$

An unconditional basis $\{x_n\} \subseteq H$ is *bounded* if

$$\exists A, B > 0 \text{ such that } \forall n \in \mathbb{Z}^d,\ A \leq \|x_n\| \leq B.$$

c. It is well known that separable Hilbert spaces have orthonormal bases (ONBs); and it is elementary to see that ONBs are bounded unconditional bases.

d. A sequence $\{x_n : n \in \mathbb{Z}^d\} \subseteq H$ is a *frame* for H if there exist $A, B > 0$ such that

$$\forall y \in H, \quad A\|y\|^2 \leq \sum |\langle y, x_n\rangle|^2 \leq B\|y\|^2. \tag{1.24}$$

A and B are *frame bounds*, and a frame is *tight* if $A = B$. A frame is *exact* if it is no longer a frame whenever any one of its elements is removed.

The *frame operator* of the frame $\{x_n\}$ is the function $S : H \longrightarrow H$ defined as $Sy = \sum\langle y, x_n\rangle x_n$ for all $y \in H$.

The theory of frames is due to Duffin and Schaeffer [DS52], cf., [You80], [DGM86], [Dau92], [BW94].

e. An exact frame is a bounded unconditional basis and vice-versa, e.g., [You80]. In particular, ONBs are exact frames and there are frames which are not exact frames.

An essential feature of frames $\{x_n\} \subseteq H$ is that they provide the harmonics for signal reconstruction formulas, see (1.25). They may not be ONBs, but ONBs are not necessarily an advantage when it comes to noise reduction and stable decompositions.

Theorem 1.7. *[Frame Decomposition Theorem] Let $\{x_n : n \in \mathbb{Z}^d\} \subseteq H$ be a frame for H with frame bounds A and B.*

a. The frame operator S is a topological isomorphism with inverse $S^{-1} : H \longrightarrow H$. $\{S^{-1}x_n\}$ is a frame with frame bounds B^{-1} and A^{-1}, and

$$\forall y \in H, \; y = \sum \langle y, S^{-1}x_n \rangle x_n = \sum \langle y, x_n \rangle S^{-1}x_n \text{ in } H. \tag{1.25}$$

b. If $\{x_n\}$ is a tight frame for H, if $\|x_n\| = 1$ for all n, and if $A = B = 1$, then $\{x_n\}$ is an orthonormal basis for H.

c. If $\{x_n\}$ is an exact frame for H, then $\{x_n\}$ and $\{S^{-1}x_n\}$ are biorthonormal, i.e.,

$$\forall m, n, \quad \langle x_m, \; S^{-1}x_n \rangle = \delta(m, n) = \begin{cases} 0 & \text{if } m \neq n, \\ 1 & \text{if } m = n. \end{cases}$$

$\{S^{-1}x_n\}$ is the unique sequence in H which is biorthonormal to $\{x_n\}$.

d. If $\{x_n\}$ is an exact frame for H, then the sequence resulting from the removal of any one element is not complete in H, i.e., the linear span of the resulting sequence is not dense in H.

We saw in Theorem 1.3 that the Classical Sampling Theorem for uniform sampling can be considered a special case of a signal decomposition theorem in terms of Gabor systems; and there are comparable results for wavelet systems, e.g., [Mey90], [Dau92], [Wal94c], [BL98]. The following example, stated in terms of general frames, can be considered a quantitative step to implement Theorem 1.7 in the context of sampling.

Example 1.2. [Frame Sampling Formulas and Implementation] Let $\{x_n : n \in \mathbb{Z}^d\} \subseteq H$ be a frame for H with frame bounds A and B.

a. The *Bessel mapping* $L : H \longrightarrow \ell^2(\mathbb{Z}^d)$ is defined by $Ly = \{\langle y, x_n \rangle\}$, and $L^* : \ell^2(\mathbb{Z}^d) \longrightarrow H$ denotes the adjoint of L. Clearly, $S = L^*L$. L can be viewed as a *sampling operator* in the following sense. Suppose we are given a sequence $\{a_n(y)\}$ of sampled values of some unknown signal y. The problem is to design a frame $\{x_n\}$ so that for all such signals y, $a_n(y) = \langle y, x_n \rangle$ for each n. Once the design problem is solved, the basic idea is to use (1.25) to reconstruct y from the sampled values $\{a_n(y)\}$. An example of the effectiveness of this procedure is in *wavelet auditory modeling* for dealing with compression problems and signal reconstruction in noisy environments, see [BT93], [BT95].

b. With regard to part a, it is easy to see that

$$\forall y \in H, \quad y = (S^{-1}L^*)Ly,$$

i.e., the unknown signal y can be reconstructed by means of its sequence of sampled values.

In the same spirit, but perhaps more in the spirit of digital signal processing, we define the *Gram operator* $R = LL^* : \ell^2(\mathbb{Z}^d) \longrightarrow \ell^2(\mathbb{Z}^d)$, and we obtain the signal reconstruction formula

$$\forall y \in H, \quad y = (L^*R^{-1})Ly. \tag{1.26}$$

In fact, an explanation of (1.26) requires the notion of the pseudo-inverse of R and leads to the following iterative procedure for the reconstruction of y from its sequence Ly of sampled values, see [Ben98] for a proof and [TB95], [Har98], and [Strb] for insights about implementation in special cases. The iteration proceeds as follows. Let $y \in H$ and let $c_{(0)} = Ly \in \ell^2(\mathbb{Z}^d)$. Set $y_0 = 0$ and $\lambda = 2/(A+B)$, and assume $\alpha = \|I - \lambda R\|_{L(H)} < 1$, where I is the identity mapping. Define u_m, $y_m \in H$ and $c_{(m)} \in L(H)$, $m = 0, 1, \ldots$, recursively as

$$u_m = \lambda L^* c_{(m)}, \quad c_{(m+1)} = c_{(m)} - Lu_m,$$

and

$$y_{m+1} = y_m + u_m.$$

Then

$$\forall m = 1, 2, \ldots, \quad \|y - y_m\| < \alpha^m \frac{B}{A}\|y\|,$$

and, in particular, $\lim_{m\to\infty} y_m = y$ in H.

c. It is easy to see that

$$\|I - \frac{2}{A+B}S\| \leq \frac{B-A}{A+B} < 1. \tag{1.27}$$

The inequality (1.27) allows us to prove that

$$S^{-1} = \frac{2}{A+B}\sum_{k=0}^{\infty}\left(I - \frac{2}{A+B}S\right)^k,$$

which, in turn, can be used to prove Theorem 1.7a.

We also mention (1.27) because of the notion of *visibility* V, which is defined as

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}},$$

where $I_{\max}$ and $I_{\min}$ are maximum and minimum light intensities, e.g., [Kla57]. In the case of full interference of light waves (for the classical two-slit experiment), one has $I_{\min} = 0$; and, hence, the visibility is 1, a value associated with *coherent light*. In the context of frames, the analogy is that $A = 0$, so that a frame is not obtained.

d. We shall verify Theorem 1.7b because of a stronger result proved by Vitali which deserves comment.

Since $\{x_n\}$ is tight and $A = 1$ we can write

$$\|x_m\|^2 = \|x_m\|^4 + \sum_{n \neq m} |\langle x_m, x_n \rangle|^2,$$

thereby obtaining the orthonormality of $\{x_n\}$ since each $\|x_n\| = 1$. Theorem 1.7b then follows by the following well-known result: if $\{x_n\} \subseteq H$ is orthonormal, then it is an orthonormal basis for H if and only if

$$\forall y \in H, \quad \|y\|^2 = \sum |\langle y, x_n \rangle|^2.$$

In 1921, Vitali proved that an orthonormal sequence $\{x_n\} \subseteq L^2[a,b]$ is complete, and so $\{x_n\}$ is an orthonormal basis if and only if

$$\forall t \in [a,b], \quad \sum \left| \int_a^t x_n(u)du \right|^2 = t - a. \tag{1.28}$$

For the case $H = L^2[a,b]$, Vitali's result is stronger than Theorem 1.7b since (1.28) is tightness with $A = 1$ for functions $y = \mathbf{1}_{[a,t]}$. Other remarkable contributions by Vitali are highlighted in [Ben76].

Definition 1.5. *[Fourier Frames]*

a. Let $R > 0$, and assume that the sequence $\{e_\lambda : \lambda \in \Lambda\}$ is a frame for $H = L^2[-R,R]$. This is clearly equivalent to the assertion that there exist $A, B > 0$ such that

$$\forall F \in PW_R, \quad A\|F\|^2_{L^2(\widehat{\mathbb{R}})} \leq \sum_{\lambda \in \Lambda} |F(\lambda)|^2 \leq B\|F\|^2_{L^2(\widehat{\mathbb{R}})}, \tag{1.29}$$

see [Ben97, page 69] for the proper application of the Fourier inversion formula. As such we say that $\{e_\lambda : \lambda \in \Lambda\}$ is a *Fourier frame* for $L^2[-R,R]$, and by (1.25) we have

$$\forall f \in L^2[-R,R], \quad f = \sum_{\lambda \in \Lambda} a_\lambda(f) e_\lambda \quad \text{in} \quad L^2[-R,R]. \tag{1.30}$$

Equation (1.30) is a *non-harmonic Fourier series*, see Chapter VII of Paley and Wiener [PW34]; and in fact the Fourier frame condition (1.29) is precisely the inequality (30.56) of Paley and Wiener [PW34, page 115], cf., [Kah62] dating from results in 1953, where Kahane uses a condition similar to (1.29) in order to deal with completeness/density problems described above.

b. The *frame radius* $R_f(\Lambda)$ of Λ is

$$R_f(\Lambda) = \sup\{R \geq 0 : \{e_\lambda\} \text{ is a Fourier frame for } L^2[-R, R]\}.$$

The following theorem is a characterization of Fourier frames in terms of density. Part *a* is due to Duffin and Schaeffer [DS52] (Theorem 1); part *b* is due to Landau [Lan67a], although not using the term "frame"; and part *c* is due to Jaffard [Jaf91].

Theorem 1.8. *[Fundamental Theorem of Fourier Frames]*

a. If Λ has uniform density $D_u(\Lambda) > 2R$, then $\{e_\lambda : \lambda \in \Lambda\}$ is a Fourier frame for $L^2[-R, R]$.

b. If $\{e_\lambda : \lambda \in \Lambda\}$ is a Fourier frame for $L^2[-R, R]$, then $D_b^-(\Lambda) \geq 2R$.

c. If $R_f(\Lambda) \in (0, \infty)$, then

$$R_f(\Lambda) = \frac{1}{2} \sup\{D_u(\Lambda')\},$$

where $\Lambda' \subseteq \Lambda$ has finite uniform density.

The following elementary result proved in [BH90], [Ben92] illustrates the role of Fourier frames in formulating non-uniform sampling formulas. Such formulas can not generally have sampled values $f(t_n)$ as coefficients in the case of non-uniformly spaced translates $\tau_{t_n} s$ of a given sampling function s. (This is a consequence of the ubiquitous "no free lunch" metatheorem.) However, the coefficients do contain the sampled data in a quantitatively estimable way, see [Ben92], pages 481–482, and Remark 1.9b.

Theorem 1.9. *[Fourier Frame Non-Uniform Sampling Theorem] Suppose $\Omega > 0$ and $\Omega_1 > \Omega$, and let the sequence $\{t_n : n \in \mathbb{Z}\} \subseteq \mathbb{R}$ have the property that $\{e_{-t_n}\}$ is a Fourier frame for $L^2[-\Omega_1, \Omega_1]$ with frame bounds A and B and frame operator S. Further, let $s \in L^2(\mathbb{R})$ have the properties that $\widehat{s} \in L^\infty(\widehat{\mathbb{R}})$, $\operatorname{supp} \widehat{s} \subseteq [-\Omega_1, \Omega_1]$, and $\widehat{s} = 1$ on $[-\Omega, \Omega]$. Then*

$$\forall f \in PW_\Omega, \quad f = \sum a_n(f) \tau_{t_n} s \quad \text{in} \quad L^2(\mathbb{R}), \tag{1.31}$$

where

$$a_n(f) = \langle S^{-1}(\widehat{f} \mathbf{1}_{[-\Omega_1, \Omega_1)}), e_{t_n} \rangle_{L^2[-\Omega_1, \Omega_1]}. \tag{1.32}$$

Remark 1.9. *[On Theorem 1.24 and Theorem 1.25]*

a. After almost 50 years, Theorem 1.8a is still difficult to prove; and parts b and c of Theorem 1.8 are not only deep results, but are special cases of the authors' original theorems. Part a, resp., part b, is used to prove the inequality $R_f(\Lambda) \geq \frac{1}{2}\sup\{D_u(\Lambda')\}$, resp., $R_f(\Lambda) \leq \frac{1}{2}\sup\{D_u(\Lambda')\}$, in part c. Part a should be compared with the earlier completeness theorem (1.22) of Paley and Wiener.

   Landau's work in [Lan67a], [Lan67b] was influenced by Beurling's ideas, and some of Beurling's results can be expressed in terms of Fourier frames. For example, Beurling proved that *if* $\Lambda \subseteq \widehat{\mathbb{R}}$ *is uniformly discrete and*

$$\rho = \sup_{\gamma \in \widehat{\mathbb{R}}} \operatorname{dist}(\gamma, \Lambda)$$

   ($\operatorname{dist}(\gamma, \Lambda)$ *is the Euclidean distance between* γ *and* Λ), *then the condition* $R\rho < \frac{1}{4}$ *implies* $\{e_\lambda : \lambda \in \Lambda\}$ *is a Fourier frame for* $L^2[-R, R]$, see [Beu60], [Beu66a], [Beu89]. Beurling's d-dimensional version of this result has been reformulated as a covering theorem in [BW99], and then used by the authors as a constructive non-uniform sampling theorem in the spirit of Theorem 1.9; it has applications in topics such as fast MRI, cf., with the approach in Chapter 16.

b. Equation (1.31) in Theorem 1.9, with coefficients given by (1.32), should be compared with the uniform sampling formula (1.11) of Theorem 1.3. In fact, (1.32) gives rise to the expansion

$$a_n(f) = \sum_{k=0}^{\infty} \frac{2}{A+B} \left\langle \left(I - \frac{2}{A+B} S\right)^k \left(\widehat{f}\mathbf{1}_{[-\Omega_1, \Omega_1)}\right), e_{-t_n} \right\rangle_{L^2[-\Omega_1, \Omega_1]} . \tag{1.33}$$

   Thus, if we consider the $k = 0$ term of (1.33) as an approximation of $a_n(f)$, we have

$$a_n(f) \approx \frac{2}{A+B} f(t_n).$$

   In this sense, Equation (1.31) can be considered a non-uniform sampling formula. An error analysis of (1.33) in terms of sampled values and truncations is not difficult.

c. Constructive versions and extensions of Theorem 1.8 are found in [BH90], [Ben92], and [GO95].

   Different "Nyquist hypotheses" and methods for non-uniform sampling signal reconstruction have also been developed by Feichtinger and Gröchenig, see [FG94], [Grö92], [Grö93], and [Grö99].

d. Duffin and Schaeffer's work [DS52] is not only seminal in defining and developing the theory of frames, but it also provides a smooth transition from the point of view in Paley and Wiener [PW34] of characterizing completeness to the current interest, greatly motivated by signal processing problems, of obtaining signal reconstruction formulas.

Fifteen pages of [DS52] are devoted to the proof of Theorem 1.8a, and the authors also have a version of the iterative procedure in Example 1.2b.

Being masters of [PW34], they understood Paley and Wiener's inequality (30.56), referenced in Definition 1.5a, and modestly asserted that "the theory of Paley and Wiener and the theory of exact frames are equivalent" [DS52, page 362]. It should be noted that Duffin and Schaeffer introduced the notion of *exact frames* in analogy with Paley and Wiener's notion of *exact complete sequences* [PW34, page 92]; and Duffin and Schaeffer clearly understood the relevance of non-exact frames, which they designate *overcomplete frame*.

Finally, in the context of Fourier frames, it is natural to ask the extent to which Duffin and Schaeffer understood their ideas in terms of non-uniform sampling. In their last brief section, called *pointwise convergence*, they provide a decomposition formula for Fourier frames, in the form (1.30), with an emphasis on analogy with Fourier series, but never with an analysis of the coefficients in terms of sampling.

e. There are non-uniform sampling theorems in the engineering and scientific literature, and there are applications of non-uniform sampling formulas. An early and celebrated result in the first category is due to Yao and Thomas [YT67] (1967): *if* $\{e_{-t_n}\}$ *is an exact frame for* $L^2[-\Omega,\Omega]$*, then*

$$\forall f \in PW_\Omega, \quad f = \sum f(t_n) s_n \ \text{ in } \ L^2(\mathbb{R}),$$

where

$$\overline{s_n(-t)} = \int_{-\Omega}^{\Omega} \overline{h_n(\gamma)} e^{2\pi i t\gamma}\, d\gamma,$$

and where $\{h_n\} \subseteq L^2[-\Omega,\Omega]$ *is the unique sequence for which* $\{h_n\}$ *and* $\{e_{-t_n}\}$ *are biorthonormal.* A straightforward proof in terms of frames is given in [Ben92, page 465]. (Of course, Yao and Thomas did not use the term "exact frame", but this formulation is equivalent to theirs.) In the second category, the book by Marvasti [Mar87] presents a comprehensive list of references, and commentary on them, up to 1987. Since then the applications and references have expanded manyfold. We shall resist presenting the litany we know, to save space as well as to save the embarrassment of showing what we don't know, see [Mar00], [BWa].

1.3 Outline of the Book

Chapter 2 is Victor E. Katsnelson's translation for this book of the classical paper by Kotel'nikov. Kotel'nikov's paper was originally written in Russian.

We have organized the remaining chapters under three general headings:

1. Sampling, Wavelets, and the Uncertainty Principle
2. Sampling Topics from Mathematical Analysis
3. Sampling Tools and Applications

In fact, the chapters cannot be so easily compartmentalized, but these headings are not unreasonable.

1.3.1 Sampling, Wavelets, and the Uncertainty Principle

Chapter 3 by Walter is "Wavelets and Sampling". It explores the theme of obtaining sampling theorems in the context of Sobolev spaces and other distribution spaces. Time-limited and other classes of non-bandlimited functions can be treated using the methods developed in this chapter.

Walter also shows that, beginning with any scaling function satisfying certain properties, there is a sampling function ξ in the space V_0 such that, for any $f \in V_0$,

$$f(t) = \sum_n f(n)\xi(t-n).$$

When the samples of both f and its derivative f' are known, one expects to be able to sample at a less dense set of points. Exploring the connections with the subject of multiwavelets, and finite element multiwavelets, Walter considers the problem of constructing multisampling functions in a given multiresolution analysis. This leads to Hermite interpolation and classes of interpolating multiwavelets.

In Chapter 4, "Embeddings and Uncertainty Principles for Generalized Modulation Spaces", Hogan and Lakey deal with two classical themes: the interplay between uncertainty principle inequalities, weighted norm inequalities for the Fourier transform, and embedding theorems for modulation spaces, on one hand; and their interpretations in terms of localization of energy in the time-frequency plane, on the other.

Modulation norms measure joint time-frequency localization of a function f by replacing the L_2 norm of the short-time Fourier transform by a mixed L_p norm. The embedding theorems are key to a firm understanding of how these norms measure smoothness versus decay. The chapter points out several ways of doing this, and gives the underlying connections with sampling. The generalizations by Hogan and Lakey of an embedding theorem due to Gröchenig are related to the Poisson summation formula and hence with sampling, and the metaplectic frames are related to the problem of recovering a signal from samples in phase space.

"Sampling Theory for Certain Hilbert Spaces of Bandlimited Functions" by Gabardo is Chapter 5. It relates sampling theory to the problem of extending a positive-definite continuous function on an interval $(-R, R)$ to one defined and positive-definite on the whole real line. The problem is fundamental in Fourier analysis, e.g., see the important early contribution by Krein [Kre40].

Gabardo deals with the extension of *distributions* positive-definite on $(-R, R)$ in the so-called *indeterminate* case, where more than one extension exists, and with the problem of parametrizing all such extensions. Because of the Paley–Wiener–Schwartz theorem, the distributional Fourier transform of an extension leads to a tempered measure, which, if discrete, leads to a sampling formula.

Zayed's Chapter 6, "Shannon-Type Wavelets and the Convergence of their Associated Wavelet Series", studies a class of wavelets that contains the Shannon wavelet as a special case and that shares its general properties. The generalization is in the well-known context of obtaining the Shannon wavelet from the multiresolution analysis constructed with the sinc sampling function being used as the scaling function.

Features of this chapter are closed form expressions of the author's Shannon-type wavelets and his study of the pointwise convergence properties of the corresponding wavelet series.

1.3.2 Sampling Topics from Mathematical Analysis

Gröchenig's Chapter 7, "Non-Uniform Sampling in Higher Dimensions: From Trigonometric Polynomials to Bandlimited Functions", employs interpolation and approximation by trigonometric polynomials for the correct finite-dimensional discretization of the sampling problem for bandlimited functions in higher dimensions.

The main result reduces the infinite dimensional problem to a matrix problem in a finite dimensional space, and it can be also regarded as a new approximation result for entire functions of exponential type (several complex variables) from finitely many of its samples.

Torres' title, "The Analysis of Oscillatory Behavior in Signals Through Their Samples", for Chapter 8 aptly describes the material. In practice, it is desirable to measure oscillations directly from signal samples. This explains the usefulness of sampling theory for this topic.

Researchers seeking to understand the progress made recently in the field should be able to deal with mean oscillation spaces and Besov spaces, and their properties vis a vis sampling. Torres goes from the Plancherel–Pólya inequality and sampling in Paley–Wiener spaces to certain other function spaces, whose sampled versions can be viewed as discrete Besov spaces. The chapter shows how to quantify oscillations at large scale and deviations of a signal from its average value at different scales.

Chapter 9 by Casey and Walnut is "Residue and Sampling Techniques in Deconvolution". A guiding principle is that many problems in harmonic analysis can be translated into interpolation problems in spaces of functions subject to growth conditions. The authors study the following deconvolution problem with this point of view.

Consider a collection of m compactly supported distributions $\{\mu_i\}_{i=1}^m$, a function f, and the m data $\{s_i\}_{i=1}^m$, obtained from f and the μ_i by convolution:

$$s_i = f * \mu_i, \quad (i = 1, 2, \ldots, m).$$

It is natural to ask under what conditions it is possible to recover f from the finite set $\{s_i\}_{i=1}^m$. If it is possible, how can it be accomplished?

Translated to a language more familiar to engineers, consider the outputs of m linear and time-invariant systems, subject to the same input signal f. Can f be found from the m outputs? One has a bank of finitely many filters, and the problem is to analyze the extent to which the outputs of the m filters determine f, and to determine if f can be recovered from the data.

Some of the ingredients required for the authors' formulation and solution are Bezout's equation, the strongly coprime condition, and properties concerning rational approximation of irrational numbers. The results include conditions under which f is determined by finitely many averages, and, more importantly from the viewpoint of applications, *practical deconvolution procedures* are obtained using residue and sampling techniques.

In Chapter 10, "Sampling Theorems from the Iteration of Low Order Differential Operators", Higgins suggests a new direction in the procedure by which a sampling series is derived from differential operators.

Kramer's lemma is the basic result in this context. The kernel required by the lemma is usually obtained as a general solution of an eigenvalue problem. Evaluation of the kernel at each of the eigenvalues leads to a basis for L^2, which in turn leads to a sampling series. Higgins' chapter shows that *functions* of the operator defined by the eigenvalue problem not only give rise to new sampling expansions but also shed light on already existing ones.

The method leads to results related to both self-adjoint and non-self-adjoint operators (including fourth order cases), and also to some new problems. For example, one would like to know the extent to which one can determine sampling series associated with an operator $\mathcal{T}^p$ from knowledge of the series associated with operators $\mathcal{T}^q$, for $q < p$.

Kivinukk's Chapter 11, "Approximation of Continuous Functions by Rogosinsky-Type Sampling Series", studies the approximation properties of sampling operators of the form

$$(S_W f)(t) := \sum_{k=-\infty}^{+\infty} f(k/W)\, s(Wt - k).$$

The operator S_W, in which the function s replaces the sinc function found in the simplest form of the Classical Sampling Theorem, was studied by Butzer and his school, e.g., see [BSS87]. Kivinukk explores the parallel between such series and the summation methods for Fourier series, emphasizing the Rogosinsky means.

1.3.3 Sampling Tools and Applications

Chapter 12 by Potts, Steidl, and Tasche is "Fast Fourier Transforms for Nonequispaced Data: A Tutorial". It discusses fast and robust algorithms for computing discrete Fourier expansions similar to

$$f(v_j) = \sum_{k \in I_N} f_k e^{-i2\pi x_k v_j} \quad (j \in I_M) \tag{1.34}$$

where

$$I_K = \left\{ n \in \mathbb{Z}^d : -\frac{K}{2} \leq n \leq \frac{K}{2} \right\}.$$

The well-known and large family of FFT algorithms cover one particular case of this problem, corresponding to regular time-frequency grids. The list of applications for the more general case is rich and varied, and ranges from geophysics to antenna theory and scattered data approximation.

The chapter provides the right tool to anyone facing nontrivial computations of the form (1.34). The approach taken is connected with interpolation methods by translates of a single function, known from approximation theory, and yields appropriate error estimates.

"Efficient Minimum Rate Sampling of Signals with Frequency Support over Non-Commensurable Sets" by Herley and Wong is Chapter 13. The chapter addresses the fact that sampling of narrow band bandpass or multiband signals using the Classical Sampling Theorem, at twice the highest frequency, entails considerable loss of efficiency. The Nyquist-Landau critical sampling density for a multiband signal can be arbitrarily smaller than twice the highest frequency, and consequently there is a need for *practical, stable* sampling schemes that allow reconstruction at or close to the minimum sampling rate.

Herley and Wong characterize the signals that can be reconstructed at the minimum rate, for a given fixed non-uniform sampling pattern, obtained by periodically discarding samples from an otherwise uniform distribution. They also show that the class of signals that can be reconstructed in this way is much larger than previously considered. The chapter draws on the theory of multichannel filter banks and on methods such as POCS (projection onto convex sets).

In Chapter 14, "Finite and Infinite-Dimensional Models for Oversampled Filter Banks", Strohmer discusses the relation between certain finite-dimensional models used for numerical procedures and infinite-dimensional filter bank theory. He establishes the convergence of the synthesis filter bank, obtained by solving a finite dimensional problem, to the synthesis filter bank of the original infinite dimensional problem. He is also able to estimate the rate of approximation. This is a fundamental component in actually implementing various theoretical results.

Chapter 15, "Statistical Aspects of Sampling for Noisy and Grouped Data", is due to Pawlak and Stadtmüller, and it deals with the the problem of recovering a bandlimited signal from noisy and grouped samples. The motivation for this study is easy to understand. Not only are real-world signals always subject to noise, but the standard data acquisition and quantization methods round the data to a amplitude *versus* time grid, introducing a perturbation that, although deterministic in nature, is usually called "quantization noise". It is of course important to assess their combined effects on signal reconstruction methods.

Data grouping, on the other hand, can be regarded either as a form of data compression or a way of converting a non-uniform (clustered) sampling problem into a uniform sampling. Understanding its advantages and disadvantages is also important from the practical viewpoint.

Pawlak and Stadtmüller give an account of the statistical aspects of reconstruction algorithms derived from the Classical Sampling Theorem, examine the problem of recovery from grouped data, and discuss the statistical accuracy of the proposed algorithms.

In Chapter 16, "Reconstruction of MRI Images from Non-Uniform Sampling, Application to Intrascan Motion Correction in Functional MRI", Bourgeois, Wajer, van Ormondt, and Graveron-Demilly deal with an important tomography modality: Magnetic Resonance Imaging (MRI).

The chapter gives a brief survey of the basic physical principles underlying MRI. In MRI, there are several ways of encoding spatial information including the number of trajectories and their shapes. Both of these can vary, subject to the crucial constraint that the space must be covered at or above the Nyquist-Landau sampling density.

Image reconstruction based on regularly sampled data requires a standard inverse FFT and is straightforward. However, other types of sampling require more complex reconstruction procedures. The chapter deals with the reconstruction from non-uniform samples and Bayesian image estimation, cf., the deterministic frame theoretic approach in [BW99]. It also considers the influence of motion and the correction of intrascan artifacts, as well as analyzing simulation results.

Chapter 17, "Efficient Sampling of the Rotation Invariant Radon Transform", by Desbat and Mennessier, is also connected with tomography, and it extends results on multidimensional sampling to the rotation invariant Radon transform.

The work was motivated by the problem of Doppler imaging in astronomy, and the authors explain the Doppler imaging technique and establish the relation with tomography. Then standard results on tomographic reconstruction are extended to the rotation invariant Radon transform with polynomial weights. A main conclusion is that interlaced sampling in Doppler imaging of rotating stars is efficient when the star rotation axis is perpendicular to the line of sight, in which case the number of measurements can be reduced by roughly one half.