MIMO Sampling: Necessary Density Conditions*

Raman Venkataramani and Yoram Bresler

Coordinated Science Laboratory
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
1308 W. Main Street, Urbana, IL 61801
Raman Venkataramani: E-Mail raman@research.bell-labs.com
Yoram Bresler: Ph. (217) 244 9660 E-Mail ybresler@uiuc.edu
FAX: (217) 244 1642

December 16, 2001

Abstract

We consider the problem of multiple-input multiple-output (MIMO) sampling of multiband signals. In this sampling scheme, a set of input signals are passed through a linear time-invariant continuous-time MIMO channel. The inputs are modeled as multiband signals whose spectral supports are arbitrary sets of finite measure, and the channel characteristics are assumed known. The channel outputs are sampled on different nonuniform sampling sets. The objective is to reconstruct the inputs from the output samples. This scheme is a very general and it encompasses various other schemes, including Papoulis' generalized sampling and nonuniform sampling as special cases.

We first define joint upper and lower densities for a collection of sampling sets. We then derive necessary conditions on these densities for stable sampling and consistent reconstruction of the channel inputs from the sampled outputs. These are generalizations of Landau's necessary density results for stable sampling and interpolation of single channel multiband signals. For stable MIMO sampling and consistent reconstruction we also find that the channel must satisfy additional conditions on the singular values of the submatrices of its transfer function matrix. All these necessary conditions trivially apply to the blind channel problem.

Keywords: stable sampling, frames, interpolation, reconstruction, multiband signals, multiple-input multiple-output (MIMO) systems, necessary density conditions, Landau lower bound, channel equalization.

Corresponding Author: Yoram Bresler

*This work was supported in part by a DARPA Contract F49620-98-1-0498.

I. INTRODUCTION

Multichannel deconvolution or multichannel separation of a convolutive mixture is an important problem arising in several applications and has attracted substantial interest recently. The problem, simply stated, deals with a multiple-input multiple-output (MIMO) channel whose outputs can be observed, and the primary goal is to invert or equalize the channel to recover the original input signals. In general, the channel inputs have overlapping spectra and share a common bandwidth. Some example applications where MIMO channels arise are multiuser or multiaccess wireless communications and space-time coding with antenna arrays, or telephone digital subscriber loops [1–4], multisensor biomedical signals [5, 6], multi-track magnetic recording [7], multiple speaker (or other acoustic source) separation with microphone arrays [8, 9], geophysical data processing [10], and multichannel image restoration [11, 12].

While much of the recent work on MIMO equalization has been on the so-called blind problem, we consider the non-blind problem and assume that the channel characteristics are either known or that they can be estimated accurately using known test input signals. Thus, instead of the problem of blind equalization, we focus on a simpler problem. In practice, digital processing is used to perform the channel inversion, whereas the channel inputs and outputs are continuous-time signals. Consequently, the channel outputs need to be sampled prior to processing. In other words, the objective is to reconstruct the channel inputs from the sampled output signals. Therefore, the MIMO channel inversion problem can be restated as one in sampling theory, and we call this sampling scheme *MIMO sampling*. We study this problem entirely from the perspective of sampling theory, although the problem could, equally well, be viewed as one of channel equalization. The questions that we address are to determine necessary conditions on the sampling densities and the MIMO channel to allow its inversion. These necessary conditions also trivially apply to harder problem of blind MIMO equalization in the presence of sampling.

The MIMO sampling problem is formulated as follows. Let $x_r(t)$, $r = 1, \dots, R$, be a collection of multiband signals whose spectral supports are measurable sets $\mathcal{F}_r \subseteq \mathbb{R}$ of finite measure. These R signals are the inputs to a MIMO channel consisting of linear time-invariant filters (see Figure 1) producing P output signals $y_p(t)$, $p = 1, \dots, P$. In other words

$$y_p = \sum_{r=1}^{R} g_{pr} * x_r, \quad p = 1, \dots, P,$$

where $*$ denotes convolution, and $g_{pr} \in L^2(\mathbb{R})$ are the channel filter impulse responses. The outputs $y_p(t)$ are subsequently sampled on either a uniform or a nonuniform grid $\Lambda_p = \{\lambda_{np} : n \in \mathbb{Z}\}$. We then attempt to reconstruct the channel inputs from the output samples. This sampling scheme is very general and subsumes various other sampling schemes as special cases. For instance Papoulis' generalized sampling [13] is essentially a single-input multiple-output (SIMO) sampling scheme, i.e., $R = 1$. A natural generalization of Papoulis' sampling expansion to vector valued inputs considered by Seidner and Feder [14] is also a special case where all input channels have identical low-pass spectra, i.e., $\mathcal{F}_r = [-B, B]$. See [15] for an interesting SIMO sampling scheme applicable to general signal spaces including wavelet and spline spaces. However, we restrict our attention to multiband signal spaces alone.

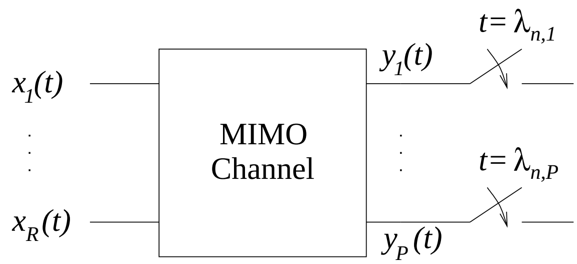

Figure 1: MIMO sampling.

Landau [16, 17] proved the following fundamental result for sampling and interpolation of multiband signals. Let $X(f)$ denote the Fourier transform of a signal $x(t)$, and

$$\mathcal{B}(\mathcal{F}) = \{x \in L^2(\mathbb{R}^d) \cap C(\mathbb{R}^d) : X(f) = 0, \ \forall f \notin \mathcal{F}\},$$

the class of continuous $L^2(\mathbb{R}^d)$ signals bandlimited to a measurable $\mathcal{F} \subseteq \mathbb{R}^d$. Suppose that a function $x \in \mathcal{B}(\mathcal{F})$ with $\mathcal{F} \subseteq \mathbb{R}^d$ is sampled at a discrete set of points $\Lambda = \{\lambda_n : n \in \mathbb{Z}\} \subseteq \mathbb{R}^d$. Then, for stable reconstruction of $x(t)$ from its samples $x(\lambda_n)$, it is necessary that the density of Λ be no less than the measure of $\mathcal{F}$, i.e., Λ must be sufficiently dense in order to stably reconstruct the input. A dual problem is that of interpolation where we seek necessary conditions on Λ that guarantee that

$$\exists x \in \mathcal{B}(\mathcal{F}) \quad \text{s.t.} \quad x(\lambda_n) = c_n \tag{1}$$

whenever $\{c_n\} \in l^2$. Equation (1) is called the *interpolation condition*, and a necessary condition for this problem is that the density of Λ be no more than the measure of $\mathcal{F}$. Roughly speaking, the samples of $x(t)$ on the set Λ can take arbitrary values only if Λ is sufficiently sparse. Alternatively, the density of Λ can be interpreted as a lower bound on the *size* of the class of multiband signals with spectral support $\mathcal{F}$, in which a solution to the interpolation problem is guaranteed to exist, where "size" refers to the Lebesgue measure of the spectral support $\mathcal{F}$.

Gröchenig and Razafinjatovo [18] recently provided a simpler proof of Landau's result. Their technique also allowed them to prove necessary density conditions for some derivative sampling schemes. However, all their results, unlike Landau's, are applicable only when the boundary of the set $\mathcal{F}$ has zero measure.

The purpose of this paper is to extend the idea of [18] to prove more general results for MIMO sampling while removing the restriction on the boundaries of the spectral supports. We consider only single variate functions in our analysis ($d = 1$), and the results easily extend to multivariate functions. The questions that we address are the following: (a) what are the necessary density conditions on the sampling densities of $\{\Lambda_p\}$ for stable reconstruction of the MIMO inputs $x_r \in \mathcal{B}(\mathcal{F}_r)$ from the MIMO output samples $\{y_p(\lambda_{np})\}$? (b) what are the necessary conditions on the sampling densities of $\{\Lambda_p\}$ such that

$$\exists x_r \in \mathcal{B}(\mathcal{F}_r) \quad \text{s.t.} \quad y_p(\lambda_{np}) = c_{np} \tag{2}$$

for any sequence $\{c_{np} : n \in \mathbb{Z},\ p = 1, \ldots, P\} \in l^2$. Problem (b) is analogous to Landau's interpolation problem for classical single input sampling, and (2) is the analogue of the interpolation condition in (1). However, we call (2) the *consistency condition*[1] rather than the *interpolation condition*. Roughly speaking, this condition implies that the channel outputs on the sets Λ_p can take arbitrary values, and this requires that Λ_p be sufficiently sparse. Equivalently, these conditions can be interpreted as minimum size requirements on the sets $\mathcal{F}_r$.

Note that although the sampling theorems for special cases considered in [13, 14] provide sufficient densities for uniform or periodic sampling, these are not shown to be necessary for arbitrary, nonuniform sampling of the channel outputs. In Section II-A, we introduce some notation and review some mathematical background. In Section III, we establish necessary conditions on $\{\Lambda_p\}$ for stable MIMO sampling and consistent MIMO reconstruction. For stable reconstruction, we prove that the sum of densities of Λ_p is lower bounded by the sum of the measures $\mathcal{F}_r$. Similarly, for the consistency problem, the sum of densities of Λ_p is upper bounded by the sum of the measures of $\mathcal{F}_r$. Apart from these natural generalizations of Landau's results, we also derive necessary conditions on the joint density for each sub-collection of sampling sets, as well as conditions on the channel transfer function. These bounds provide an outer bound on the region of achievable densities. We provide examples to illustrate the results.

II. Preliminaries

A. Signals and Matrices

The class of continuous $L^2(\mathbb{R})$ signals bandlimited to a measurable $\mathcal{F} \subseteq \mathbb{R}$ is denoted by

$$\mathcal{B}(\mathcal{F}) = \{x \in L^2(\mathbb{R}) \cap C(\mathbb{R}) : X(f) = 0,\ \forall f \notin \mathcal{F}\}, \tag{3}$$

[1] Equation (2) does not describe an interpolation problem because the multichannel samples are not samples of the input signals themselves.

where $X(f)$ is the Fourier transform of a signal $x(t)$:

$$X(f) = \int_{\mathbb{R}} x(t) e^{-j2\pi f t} dt.$$

The space $\mathcal{B}(\mathcal{F})$ is a separable Hilbert space. Let $\mu(\cdot)$ denote the Lebesgue measure, and $\chi(\cdot)$, the indicator function. For instance, $\chi(f \in \mathcal{F})$ takes the value 1 on the set $\mathcal{F}$, and 0 elsewhere. Let $\phi_{\mathcal{F}}(t)$ denote the inverse Fourier transform of $\chi(f \in \mathcal{F})$:

$$\phi_{\mathcal{F}}(t) = \int_{\mathcal{F}} e^{j2\pi f t} df.$$

Denote the time-shift operator by Θ_τ, i.e., $\Theta_\tau f(t) = f(t-\tau)$. Let $\emptyset$ denote the empty set, and $\mathcal{S}^c$, the complement of a set $\mathcal{S}$ in the appropriate universal set.

We now introduce some notation pertaining to matrices. We denote the class of complex-valued matrices of size $M \times N$ by $\mathbb{C}^{M\times N}$. Let $\boldsymbol{e}_r \in \mathbb{C}^{R\times 1}$ denote the r-th standard basis vectors, i.e., $\boldsymbol{e}_r$ has a 1 at the r-th position, and zeros elsewhere. For a given matrix $\boldsymbol{A}$, let $\boldsymbol{A}^H$ denote its conjugate-transpose, $\boldsymbol{A}_{\mathcal{R},\mathcal{C}}$, its submatrix corresponding to rows indexed by the set $\mathcal{R}$ and columns by the set $\mathcal{C}$. Also let $\boldsymbol{A}_{\bullet,\mathcal{C}}$ denote the submatrix formed by keeping all rows of $\boldsymbol{A}$, but only columns indexed by $\mathcal{C}$, and $\boldsymbol{A}_{\mathcal{R},\bullet}$, the submatrix formed by retaining rows indexed by $\mathcal{R}$ and all columns. We use a similar notation for vectors. Hence, $\boldsymbol{X}_{\mathcal{R}}$ is the subvector of $\boldsymbol{X}$ corresponding to rows indexed by $\mathcal{R}$. We always apply the subscripts before superscripts. So $\boldsymbol{A}^H_{\mathcal{R},\mathcal{C}}$ is the conjugate-transpose of $\boldsymbol{A}_{\mathcal{R},\mathcal{C}}$. When dealing with singleton index sets: $\mathcal{R} = \{r\}$ or $\mathcal{C} = \{c\}$, we omit the curly braces for readability. Therefore, $\boldsymbol{A}_{r,\bullet}$ and $\boldsymbol{A}_{\bullet,c}$ are the r-th row and the c-th column of $\boldsymbol{A}$ respectively. Let $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$ denote the largest and smallest eigenvalues of $\boldsymbol{A}$. Let $\sigma_{\max}(\boldsymbol{A})$ denote the largest singular value of matrix $\boldsymbol{A}$, and $\sigma_{\min}(\boldsymbol{A})$, the smallest *nonzero* singular value of a $\boldsymbol{A}$ if $\boldsymbol{A} \neq \boldsymbol{0}$. If $\boldsymbol{A} = \boldsymbol{0}$, we take $\sigma_{\min}(\boldsymbol{A}) = \infty$.

The following proposition, which is proved in the Appendix, is used later to characterize the region of achievable densities.

Proposition 1. *Let $\boldsymbol{G} \in \mathbb{C}^{R\times\rho}$, and let $\mathcal{A}$ and $\mathcal{B}$ be subsets of $\{1,\ldots,R\}$. Then*

$$\operatorname{rank}(\boldsymbol{G}_{\mathcal{A},\bullet}) + \operatorname{rank}(\boldsymbol{G}_{\mathcal{B},\bullet}) \geq \operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cup\mathcal{B},\bullet}) + \operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B},\bullet}). \tag{4}$$

B. Stable sampling and consistency

The following material on frames is standard (cf. [19, 20]). Let $\mathcal{H}$ is a separable Hilbert space equipped with an inner product $\langle\cdot,\cdot\rangle$. A sequence $\{\psi_n\} \subseteq \mathcal{H}$ is called a *frame* if there exist constants $A, B > 0$ such that

$$A\|x\|^2 \leq \sum_n |\langle \psi_n, x\rangle|^2 \leq B\|x\|^2, \tag{5}$$

for all $x \in \mathcal{H}$. The constants A and B are called the lower and upper *frame bounds*. If $A = B$, then the frame is a tight frame. The frame operator S, defined as

$$Sx = \sum_n \langle x, \psi_n\rangle \psi_n, \quad \forall x \in \mathcal{H},$$

is a bounded linear operator satisfying $AI \leq S \leq BI$, where I is the identity operator. Let $\tilde{\psi}_n = S^{-1}\psi_n$. Then $\{\tilde{\psi}_n\}$ is also a frame (the *dual frame*) for $\mathcal{H}$ with frame bounds B^{-1} and A^{-1}. Then any $x \in \mathcal{H}$ can be expanded as

$$x = \sum_n \langle x, \tilde{\psi}_n\rangle \psi_n = \sum_n \langle x, \psi_n\rangle \tilde{\psi}_n. \tag{6}$$

If $\{\psi_n\}$ is a frame, then for any sequence $\{c_n\} \in l^2$, we have

$$\Big\| \sum_n c_n \psi_n \Big\|^2 \leq B \sum_n |c_n|^2. \tag{7}$$

A sequence $\{\psi_n\} \subseteq \mathcal{H}$ is called a *Riesz basis* if it is fully equivalent to an orthonormal basis for $\mathcal{H}$, i.e., if there exists a bounded invertible operator T and an orthonormal basis $\{e_n\}$ such that $\psi_n = Te_n$. A Riesz basis is a frame, and hence (6) and (7) hold. In fact, for a Riesz basis, we can replace (7) by the following stronger condition:

$$A \sum_n |c_n|^2 \leq \Big\| \sum_n c_n \psi_n \Big\|^2 \leq B \sum_n |c_n|^2. \tag{8}$$

Conversely, if $\{\psi_n\}$ is a complete sequence in $\mathcal{H}$, then it is a Riesz basis for $\mathcal{H}$ whenever (8) holds for finite sequences [19]. The dual frame $\{\tilde{\psi}_n\}$ of a Riesz basis $\{\psi_n\}$ is called the *biorthogonal basis* of $\{\psi_n\}$, and it is also a Riesz basis for $\mathcal{H}$.

A sequence $\{\psi_n\} \subseteq \mathcal{H}$ is called a *Riesz-Fischer* sequence if the *moment problem*:

$$\langle x, \psi_n\rangle = c_n \tag{9}$$

has a solution $x \in \mathcal{H}$ whenever $\{c_n\} \in l^2$. If $\{\psi_n\}$ is a Riesz-Fischer sequence, then there exists a solution x to (9) such that

$$\|x\|^2 \leq \frac{1}{a}\|c\|^2$$

for some $a > 0$ called the bound of the Riesz-Fischer sequence. A necessary and sufficient condition for

$\{\psi_n\}$ to be a Riesz-Fischer sequence with bound a is that

$$\Big\| \sum_n c_n \psi_n \Big\|^2 \geq a \sum_n |c_n|^2 \tag{10}$$

for every finite sequence $\{c_n\}$. Finally, note that the moment problem in (9) has a unique solution if $\{\psi_n\}$ is a *complete* Riesz-Fischer sequence in $\mathcal{H}$. Every Riesz basis is a Riesz-Fischer sequence, but the converse is not true. However, if a Riesz-Fischer sequence is also a frame, then it is a Riesz basis. The notions of frames and Riesz-Fischer sequences are used in much of our analysis in Section III.

In the context of classical multiband sampling, the class of input signals is the separable Hilbert space $\mathcal{H} = \mathcal{B}(\mathcal{F})$ having the following inner product:

$$\langle x, y \rangle = \int_{\mathbb{R}} x(t)\overline{y(t)}dt, \quad \forall x, y \in \mathcal{B}(\mathcal{F}).$$

Obviously, the norm on $\mathcal{H}$ is defined as $\|x\| = \sqrt{\langle x, x \rangle}$.

A discrete set $\Lambda = \{\lambda_n : n \in \mathbb{Z}\}$ is called a *stable set of sampling* for $\mathcal{B}(\mathcal{F})$ if there exist $A, B > 0$ such that

$$A\|x\|^2 \leq \sum_{n \in \mathbb{Z}} |x(\lambda_n)|^2 \leq B\|x\|^2, \quad \forall x \in \mathcal{B}(\mathcal{F}). \tag{11}$$

First notice that $x(\lambda_n) = \langle x, \Theta_{\lambda_n} \phi_{\mathcal{F}} \rangle$. Using this fact and (11), we see that $\{\Theta_{\lambda_n}\phi_{\mathcal{F}} : n \in \mathbb{Z}\}$ is a frame for $\mathcal{B}(\mathcal{F})$ with frame bounds A and B. Denoting its dual frame by $\{\tilde{\phi}_n\}$ and using (6), we obtain the following interpolation equation to reconstruct x:

$$x = \sum_{n \in \mathbb{Z}} \langle x, \Theta_{\lambda_n} \phi_{\mathcal{F}} \rangle \tilde{\phi}_n = \sum_{n \in \mathbb{Z}} x(\lambda_n) \tilde{\phi}_n. \tag{12}$$

The following argument shows that (12) is a stable reconstruction formula. Suppose that a perturbation $\{\delta z_n\} \in l^2$ is added to $z_n = x(\lambda_n)$. Then, owing to linearity of (12), the resulting perturbation δx in the reconstruction is given by

$$\delta x = \sum_{n \in \mathbb{Z}} \delta z_n \tilde{\phi}_n.$$

Using (7) and noting that the upper frame bound for the dual frame is $1/A$, it follows that

$$\|\delta x\|^2 = \Big\| \sum_n b_n \tilde{\phi}_n \Big\|^2 \leq \frac{1}{A} \sum_n |b_n|^2. \tag{13}$$

Finally, from (11) and (13) we conclude that

$$\frac{\|\delta x\|}{\|x\|} \leq \sqrt{\frac{B}{A}} \frac{\|\delta z\|}{\|z\|}.$$

The ratio $K = \sqrt{B/A} \geq 1$ is called the *condition number* of the sampling scheme, and K^2 is a bound on the amplification of the normalized perturbation energy. Similarly, it can be shown that a perturbation of δx in $x \in \mathcal{H}$ produces a perturbation of $\delta z \in l^2$ in $z_n = x(\lambda_n)$ such that

$$\frac{\|\delta z\|}{\|z\|} \leq \sqrt{\frac{B}{A}} \frac{\|\delta x\|}{\|x\|}.$$

Thus, the stability condition in (11) guarantees that the errors in the sampled signal or its samples cannot produce arbitrarily large errors in the reconstructed signal. Conversely, if this condition is not satisfied, then there exist bounded inputs or perturbations in the samples that produce unbounded errors in the reconstruction. Therefore, this is a natural condition to impose on any sampling system.

The set Λ is called a *set of interpolation* if there exists $x \in \mathcal{B}(\mathcal{F})$ such that $x(\lambda_n) = c_n$ whenever $\{c_n\} \in l^2$. In other words, the sampling operator from $\mathcal{B}(\mathcal{F})$ to l^2 is onto if Λ is a set of interpolation. This condition is clearly equivalent to $\{\Theta_{\lambda_n}\phi_{\mathcal{F}} : n \in \mathbb{Z}\}$ being a Riesz-Fischer sequence in $\mathcal{B}(\mathcal{F})$. There are several practical implications to Λ being a set of interpolation. First, it implies that any data sequence $\{c_n\} \in l^2$ can be interpolated to a signal $x \in \mathcal{B}(\mathcal{F})$ whose samples on Λ agree with the data sequence. Second, it implies that the samples of $x \in \mathcal{B}(\mathcal{F})$ on Λ are nonredundant because each sample is completely independent of all the others. If Λ is not a set of interpolation, then the samples of $x \in \mathcal{B}(\mathcal{F})$ can only live in a subspace of l^2, and are linearly dependent, with some samples completely determined by the others. Finally, if Λ is a set of both sampling and interpolation, then $\{\Theta_{\lambda_n}\phi_{\mathcal{F}} : n \in \mathbb{Z}\}$ is a Riesz basis for $\mathcal{B}(\mathcal{F})$. The theory of frames thus provides a convenient tool to study sampling [21].

We shall now generalize the above notions of stable sampling and interpolation for the MIMO problem. Recall that the channel input and output signals are related to each other as

$$\boldsymbol{y}(t) = \boldsymbol{g} * \boldsymbol{x}(t) = \int_{\mathbb{R}} \boldsymbol{g}(t-\tau)\boldsymbol{x}(\tau)df,$$

where $\boldsymbol{x}$ is the input vector whose components are multiband signals $x_r \in \mathcal{B}(\mathcal{F}_r)$, and $\boldsymbol{y}$ is the channel output in vector form. The class of input signals is the separable Hilbert space

$$\mathcal{H} = \mathcal{B}(\mathcal{F}_1) \times \cdots \times \mathcal{B}(\mathcal{F}_R) \tag{14}$$

equipped with the inner product

$$\langle \boldsymbol{x}, \boldsymbol{z} \rangle = \int_{\mathbb{R}} \boldsymbol{z}^H(t)\boldsymbol{x}(t)dt = \sum_{r=1}^{R} \int_{\mathbb{R}} \overline{z_r(t)} x_r(t)dt, \quad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathcal{H}. \tag{15}$$

The norm on $\mathcal{H}$ is clearly defined as $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$. In the rest of this paper let $\mathcal{R}$ and $\mathcal{P}$ denote index sets for the components of the channel inputs and outputs, i.e.,

$$\mathcal{R} = \{1, \ldots, R\} \quad \text{and} \quad \mathcal{P} = \{1, \ldots, P\}.$$

Suppose that

$$\mathcal{C}_f = \{r : f \in \mathcal{F}_r\}, \tag{16}$$

then it is clear that $\boldsymbol{X}_{\mathcal{C}_f}(f)$ captures all the nonzero elements of $\boldsymbol{X}(f)$. Hence, the channel output in the frequency domain can be expressed as

$$\boldsymbol{Y}(f) = \boldsymbol{G}(f)\boldsymbol{X}(f) = \boldsymbol{G}_{\bullet, \mathcal{C}_f}(f)\boldsymbol{X}_{\mathcal{C}_f}(f), \tag{17}$$

where $\boldsymbol{G}(f)$, the Fourier transform of $\boldsymbol{g}(t)$, is called the *channel transfer function matrix*.

Definition 1. A collection of discrete sets $\Lambda_p = \{\lambda_{np} : n \in \mathbb{Z}\}$, $p \in \mathcal{P}$ is called a *stable collection of MIMO sampling* w.r.t. $\boldsymbol{G}(f)$ for the space $\mathcal{H}$ if there exist $A, B > 0$ such that

$$A\|\boldsymbol{x}\|^2 \leq \sum_{p=1}^{P} \sum_{n \in \mathbb{Z}} |y_p(\lambda_{np})|^2 \leq B\|\boldsymbol{x}\|^2 \tag{18}$$

for every $\boldsymbol{x} \in \mathcal{H}$, where $\boldsymbol{Y}(f) = \boldsymbol{G}(f)\boldsymbol{X}(f)$.

It is clear that we can write $y_p(\lambda_{np}) = \langle \boldsymbol{x}, \Theta_{\lambda_{np}} \boldsymbol{\psi}_p \rangle$ for appropriate $\boldsymbol{\psi}_p \in \mathcal{H}$. In fact a simple calculation reveals that

$$\boldsymbol{\Psi}_p(f) = \sum_{r=1}^{R} \overline{G_{pr}(f)} \chi(f \in \mathcal{F}_r) \boldsymbol{e}_r, \tag{19}$$

where $\boldsymbol{e}_r$ is the r-th standard basis vector.

Now, (18) is equivalent to the condition that $\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : n \in \mathbb{Z}, \ p \in \mathcal{P}\}$ is a frame for $\mathcal{H}$. This observation, as in the case of classical sampling, implies that we can perform the reconstruction of the channel inputs from the output samples using the dual frame as the set of interpolating functions. Also, the stability condition in (18) guarantees that the errors in the sampled signal or its samples do not produce arbitrarily large errors in the reconstructed signals. The *condition number* for the MIMO sampling scheme is $K = \sqrt{B/A} \geq 1$.

Definition 2. A collection of discrete sets $\Lambda_p = \{\lambda_{np} : n \in \mathbb{Z}\}$, $p \in \mathcal{P}$ is called a *collection of consistent reconstruction* w.r.t. $\boldsymbol{G}(f)$ for the space $\mathcal{H}$ if there exists a solution $\boldsymbol{x} \in \mathcal{H}$ to the problem $y_p(\lambda_{np}) = c_{np}$ for every $\{c_{np}\} \in l^2$, where $\boldsymbol{Y}(f) = \boldsymbol{G}(f)\boldsymbol{X}(f)$.

Equivalently, $\{\Lambda_p : p \in \mathcal{P}\}$ is a collection of consistent reconstruction w.r.t. $\boldsymbol{G}(f)$ for the space $\mathcal{H}$ if $\{\Theta_{\lambda_{np}}\boldsymbol{\psi}_p : n \in \mathbb{Z},\ p \in \mathcal{P}\}$ is a Riesz-Fischer sequence in $\mathcal{H}$. For any finite sequence $\{c_{np}\}$, observe that

$$\begin{aligned}\Big\| \sum_{n,p} c_{np}\Theta_{\lambda_{np}}\boldsymbol{\psi}_p \Big\|^2 &= \max_{\boldsymbol{x}\in\mathcal{B}_{\mathcal{H}}} \Big| \sum_{n,p} c_{np}\langle \boldsymbol{x}, \Theta_{\lambda_{np}}\boldsymbol{\psi}_p\rangle \Big|^2 \\ &= \max_{\boldsymbol{x}\in\mathcal{B}_{\mathcal{H}}} \Big| \sum_{n,p} c_{np} y_p(\lambda_{np}) \Big|^2,\end{aligned}$$

where $B_{\mathcal{H}} = \{\boldsymbol{x} \in \mathcal{H} : \|\boldsymbol{x}\| \le 1\}$ is the unit ball in $\mathcal{H}$. In view of (10), it is clear that

$$\{\Theta_{\lambda_{np}}\boldsymbol{\psi}_p : n \in \mathbb{Z},\ p \in \mathcal{P}\}$$

is a Riesz-Fischer sequence in $\mathcal{H}$ if and only if

$$\max_{\boldsymbol{x}\in\mathcal{B}_{\mathcal{H}}} \Big| \sum_{n,p} c_{np} y_p(\lambda_{np}) \Big|^2 \ge a \sum_{n,p} |c_{np}|^2 \tag{20}$$

for every finite sequence $\{c_{np}\}$. It turns out that the above characterization of consistent MIMO reconstruction is easier to use than Definition 2. Finally, we point out that if a collection of discrete sets $\Lambda_p = \{\lambda_{np} : n \in \mathbb{Z}\}$, $p \in \mathcal{P}$ is a collection of both stable sampling and consistent reconstruction, then $\{\Theta_{\lambda_{np}}\boldsymbol{\psi}_p : n \in \mathbb{Z},\ p \in \mathcal{P}\}$ is a Riesz basis for $\mathcal{H}$.

C. Notions of sampling density

A discrete subset $\Lambda = \{\lambda_n : n \in \mathbb{Z}\} \subseteq \mathbb{R}$ is called uniformly discrete with separation δ if

$$|\lambda_m - \lambda_n| \ge 2\delta, \quad \forall m \ne n.$$

Let the maximum and minimum number of sampling points of Λ found in any interval of length 2γ be denoted by

$$\nu_\gamma^+(\Lambda) = \sup_{\tau\in\mathbb{R}} \#(\Lambda \cap B_\gamma(\tau)) \quad \text{and} \quad \nu_\gamma^-(\Lambda) = \inf_{\tau\in\mathbb{R}} \#(\Lambda \cap B_\gamma(\tau)) \tag{21}$$

respectively, where $\#(\mathcal{S})$ denotes the cardinality of a set $\mathcal{S}$, and

$$B_\gamma(\tau) = \{\sigma \in \mathbb{R} : |\sigma - \tau| \leq \gamma\}$$

is a closed interval of length 2γ centered at t. For a discrete set Λ, the upper and lower densities are defined as

$$D^+(\Lambda) = \limsup_{\gamma \to \infty} \frac{\nu_\gamma^+(\Lambda)}{2\gamma} \quad \text{and} \quad D^-(\Lambda) = \liminf_{\gamma \to \infty} \frac{\nu_\gamma^-(\Lambda)}{2\gamma} \tag{22}$$

respectively. See [21] for several other notions of density for nonuniform sampling. Although traditionally written as "$\liminf$" and "$\limsup$," the limits in (22) can be replaced by simple limits [21]. If the lower and upper densities coincide, this density is called the *uniform density* and is denoted by $D(\Lambda)$. Note that this does not mean that the sampling points in Λ are uniformly spaced. Any large interval of size l contains approximately $lD(\Lambda)$ points of Λ. If Λ is uniformly discrete, then $D^+(\Lambda)$ is finite. The converse statement is not true. However $D^+(\Lambda) < \infty$ implies that Λ can be expressed as a union K uniformly discrete sets [22].

When dealing with a collection of sampling sets, as in the MIMO setting, it is useful to define *joint* densities for the collection. In [22], we introduced the following generalizations of the densities defined earlier.

Definition 3. Given a finite collection of discrete sets Λ_p, $p = 1, \ldots, P$, their *joint upper and lower densities* are defined as

$$D^+(\Lambda_1, \ldots, \Lambda_P) = \limsup_{\gamma \to \infty} \frac{\nu_\gamma^+(\Lambda_1, \ldots, \Lambda_P)}{2\gamma}, \tag{23}$$

$$D^-(\Lambda_1, \ldots, \Lambda_P) = \liminf_{\gamma \to \infty} \frac{\nu_\gamma^-(\Lambda_1, \ldots, \Lambda_P)}{2\gamma} \tag{24}$$

respectively, where

$$\nu_\gamma^+(\Lambda_1, \ldots, \Lambda_P) = \sup_{\tau \in \mathbb{R}} \sum_{p=1}^{P} \#(\Lambda_p \cap B_\gamma(\tau)),$$

$$\nu_\gamma^-(\Lambda_1, \ldots, \Lambda_P) = \inf_{\tau \in \mathbb{R}} \sum_{p=1}^{P} \#(\Lambda_p \cap B_\gamma(\tau))$$

are the maximum and minimum number of sampling points of the collection $\{\Lambda_p : p = 1, \ldots, P\}$ found in any interval of length 2γ.

If these densities coincide, then $\{\Lambda_1, \dots, \Lambda_P\}$ has uniform joint density of

$$D(\Lambda_1, \dots, \Lambda_P) = D^{\pm}(\Lambda_1, \dots, \Lambda_P).$$

If each Λ_p has uniform density, then so does the collection $\{\Lambda_1, \dots, \Lambda_P\}$. However, the converse is not true. From these definitions it is clear that

$$D^+(\Lambda_1, \dots, \Lambda_P) \le \sum_{p=1}^{P} D^+(\Lambda_p),$$
$$D^-(\Lambda_1, \dots, \Lambda_P) \ge \sum_{p=1}^{P} D^-(\Lambda_p).$$

Moreover, if each Λ_p has uniform density, the collection $\{\Lambda_p\}$ also has uniform joint densities given by

$$D(\Lambda_1, \dots, \Lambda_P) = \sum_{p=1}^{P} D(\Lambda_p).$$

We use the above properties later without explicitly stating them. finally, we restate the following proposition, which is proved in [22].

Proposition 2. *The* $\limsup$ *in (23) and the* $\liminf$ *in (24) can be replaced by simple limits. In fact,*

$$\nu_\gamma^+(\Lambda_1, \dots, \Lambda_P) \ge 2\gamma D^+(\Lambda_1, \dots, \Lambda_P), \tag{25}$$
$$\nu_\gamma^-(\Lambda_1, \dots, \Lambda_P) \le 2\gamma D^-(\Lambda_1, \dots, \Lambda_P) \tag{26}$$

for all $\gamma > 0$*.*

III. Necessary Density Conditions

A. Previous Results and a New Comparison Theorem

Our aim in this section is to prove necessary density conditions for MIMO sampling of multiband signals. These results are analogous to Landau's density result for nonuniform sampling of multiband signals [16, 17]. Gröchenig and Razafinjatovo [18] provided a simpler proof of Landau's result. Their idea allowed them to prove some results for derivative sampling. With some modifications, the results in [18] can also be extended to SIMO sampling and interpolation. However, these results apply only to signals in the class of multiband signals $\mathcal{B}(\mathcal{F})$ for which $\mu(\partial\mathcal{F}) = 0$, i.e., the boundary of $\mathcal{F}$ has measure zero. Most sets of practical interest satisfy this condition, while several pathological sets such as nowhere dense sets are excluded.

Unfortunately, the condition also excludes some reasonable sets. For example, let $\mathcal{F} = [0,1] \cap Q^c$, where Q is the set of rationals. Then $\partial\mathcal{F} = [0,1]$, implying that $\mu(\partial\mathcal{F}) = 1$. But $\mathcal{B}(\mathcal{F}) = \mathcal{B}([0,1])$ since $\mathcal{F}$ differs from $[0,1]$ by a set of measure zero. In other words their results do not apply to some elementary classes of signals under a simple disguise.

The following theorem, which is proved in [22], is a stronger version of the main result in [18]. This theorem allows us to compute necessary density conditions for the stable MIMO sampling and consistent reconstruction. However, the theorem is very general, involving arbitrary signal spaces, and can potentially be used for proving necessary density conditions for sampling problems in wavelet or spline spaces.

Theorem 1 (Comparison Theorem). *Let $\mathcal{H}_S$ and $\mathcal{H}_L$ be closed subspaces of $\mathcal{H}_\infty$, and let $\Sigma_1, \ldots, \Sigma_Q$, and $\Lambda_1, \ldots, \Lambda_P$ be discrete subsets of $\mathbb{R}$ such that all $D^+(\Lambda_p) < \infty$. Suppose that $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_Q$ and $\boldsymbol{l}_1, \ldots, \boldsymbol{l}_P$ are such that*

$$\{\Theta_\sigma \boldsymbol{s}_q : \sigma \in \Sigma_q,\ q = 1, \ldots, Q\} \subseteq \mathcal{H}_S$$

is a Riesz-Fischer sequence in $\mathcal{H}_S$ with bound $a > 0$, and that

$$\{\Theta_\lambda \boldsymbol{l}_p : \lambda \in \Lambda_p,\ p = 1, \ldots, P\} \subseteq \mathcal{H}_L$$

is a frame for $\mathcal{H}_L$. Then

$$D^\pm(\Lambda_1, \ldots, \Lambda_P) \geq D^\pm(\Sigma_1, \ldots, \Sigma_Q) - \sum_{q=1}^{Q} \alpha_q D^+(\Sigma_q), \tag{27}$$

if all $D^+(\Sigma_q) < \infty$, where

$$\alpha_q = \frac{1}{\sqrt{a}} \sup_{\sigma \in \Sigma_q} \|\Theta_\sigma \boldsymbol{s}_q - P_{\mathcal{H}_L} \Theta_\sigma \boldsymbol{s}_q\|.$$

In particular, $D^+(\Sigma_q) < \infty$ is guaranteed whenever all $\alpha_q < 1$.

Note that $\mathcal{H}_L$ and $\mathcal{H}_S$ are arbitrary subspaces in $\mathcal{H}_\infty$. However, the comparison theorem is most powerful when we let the spaces be nearly the same. In this case, the coefficients α_q would be small, thereby yielding the following density bound:

$$D^\pm(\Lambda_1, \ldots, \Lambda_P) \geq D^\pm(\Sigma_1, \ldots, \Sigma_Q) - \epsilon,$$

where $\epsilon > 0$ is a small quantity representing the summation in (27) involving the terms α_q. The import of this statement is roughly that a frame, being an overcomplete sequence in a Hilbert space $\mathcal{H}$, is "denser" (contains more vectors) than a Riesz-Fischer sequence $\mathcal{H}$. By using an appropriate limiting argument, we can then show that $\epsilon > 0$ can be made arbitrarily small, yielding

$$D^{\pm}(\Lambda_1, \ldots, \Lambda_P) \geq D^{\pm}(\Sigma_1, \ldots, \Sigma_Q).$$

We illustrate the use of this theorem in the next section, where we derive necessary density conditions for the MIMO sampling problem.

B. Density conditions for stable sampling

We begin with the following lemma, whose proof can be found in [22]:

Lemma 1. *Let $h \in \mathcal{B}([\nu_1, \nu_2])$. then*

$$h^{\#}(t) \stackrel{\text{def}}{=} \sup_{|\tau - t| \leq 1} |h(\tau)|,$$

satisfies $h^{\#} \in L^2(\mathbb{R})$ and $\|h^{\#}\|^2 \leq C\|h\|^2$ for some $C = C(\nu_2 - \nu_1) > 0$ that depends only on the difference $(\nu_2 - \nu_1)$. Moreover, if $\Lambda = \{\lambda_m : n \in \mathbb{Z}\} \subseteq \mathbb{R}$ is a discrete set with $D^+(\Lambda) < \infty$, then

$$\sum_{|\lambda_n - \sigma| \geq \Gamma} |h(\lambda_n)|^2 \leq C' \int_{|t-\sigma| \geq \Gamma - 1} |h^{\#}(t)|^2 dt \tag{28}$$

for all $\sigma \in \mathbb{R}, \Gamma \geq 0$, and some $C' = C'(\Lambda) > 0$. In particular

$$\sum_{n \in \mathbb{Z}} |h(\lambda_n)|^2 \leq C'C\|h\|^2.$$

Lemma 1 says that the samples of a bandlimited signal on a sampling set of finite upper density cannot be arbitrarily large. As we shall see later, it is a simple but powerful result.

We now introduce a few quantities relevant to the main result that follows shortly. Define the following separable Hilbert spaces:

$$\begin{aligned} \mathcal{H}_\beta &= \left(\mathcal{B}([-\beta, \beta])\right)^R, \quad \beta > 0, \\ \mathcal{H}_\infty &= \left(L^2(\mathbb{R})\right)^R, \end{aligned}$$

and let the inner product on both spaces be defined as in (15). Note that $\mathcal{H}_\beta$ is the space of vector functions whose R components are bandlimited to the frequencies $[\beta, \beta]$. Let $P_{\mathcal{S}} : \mathcal{H}_\infty \to \mathcal{S}$ denote the orthogonal projection operator onto a closed subspace $\mathcal{S} \subseteq \mathcal{H}_\infty$.

Definition 4. A subspace $\mathcal{S} \subseteq \mathcal{H}_\infty$ is called *shift-invariant* if $\Theta_\sigma \boldsymbol{x} \in \mathcal{S}$ for all $\sigma \in \mathbb{R}$ whenever $\boldsymbol{x} \in \mathcal{S}$.

Evidently $\mathcal{H}_\infty$ and $\mathcal{H}_\beta$ are shift-invariant spaces. We write $\boldsymbol{x} \perp \mathcal{S}$ whenever $\langle \boldsymbol{x}, \boldsymbol{z} \rangle = 0$ for all $\boldsymbol{z} \in \mathcal{H}_\beta$. The following properties of a closed shift-invariant subspace $\mathcal{S} \subseteq \mathcal{H}_\infty$ can be verified easily.

Proposition 3. *Suppose that $\boldsymbol{x} \in \mathcal{H}_\beta$, $\sigma \in \mathbb{R}$, and $\mathcal{S} \subseteq \mathcal{H}_\infty$ is a closed shift-invariant subspace. Then (a) $\boldsymbol{x} \perp \mathcal{S} \implies \Theta_\sigma \boldsymbol{x} \perp \mathcal{S}$, and (b) $P_\mathcal{S} \Theta_\sigma \boldsymbol{x} = \Theta_\sigma P_\mathcal{S} \boldsymbol{x}$, i.e., translation commutes with orthogonal projection onto $\mathcal{S}$.*

Proof. (a) Suppose that $\boldsymbol{x} \perp \mathcal{S}$. Then $\langle \Theta_\sigma \boldsymbol{x}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \Theta_{-\sigma} \boldsymbol{z} \rangle = 0$ whenever $\boldsymbol{z} \in \mathcal{S}$ and $\sigma \in \mathbb{R}$ because $\Theta_{-\sigma} \boldsymbol{z} \in \mathcal{S}$. Hence $\Theta_\sigma \boldsymbol{x} \perp \mathcal{S}$. To prove (b), note $P_\mathcal{S} \Theta_\sigma \boldsymbol{x} - \Theta_\sigma P_\mathcal{S} \boldsymbol{x} \in \mathcal{S}$. For arbitrary $\boldsymbol{z} \in \mathcal{S}$, we have

$$\begin{aligned}
\langle P_\mathcal{S} \Theta_\sigma \boldsymbol{x} - \Theta_\sigma P_\mathcal{S} \boldsymbol{x}, \boldsymbol{z} \rangle &= \langle P_\mathcal{S} \Theta_\sigma \boldsymbol{x}, \boldsymbol{z} \rangle - \langle \Theta_\sigma P_\mathcal{S} \boldsymbol{x}, \boldsymbol{z} \rangle \\
&= \langle \Theta_\sigma \boldsymbol{x}, P_\mathcal{S} \boldsymbol{z} \rangle - \langle \boldsymbol{x}, P_\mathcal{S} \Theta_{-\sigma} \boldsymbol{z} \rangle \\
&= \langle \Theta_\sigma \boldsymbol{x}, \boldsymbol{z} \rangle - \langle \boldsymbol{x}, \Theta_{-\sigma} \boldsymbol{z} \rangle = 0,
\end{aligned}$$

proving that $P_\mathcal{S} \Theta_\sigma \boldsymbol{x} = \Theta_\sigma P_\mathcal{S} \boldsymbol{x}$. ☐

We are now ready to compute necessary density results for stable sampling in the MIMO setting using Theorem 1.

Theorem 2. *Suppose that $\mathcal{F}_r$, $r \in \mathcal{R}$ are real sets of finite measure, and Λ_p, $p \in \mathcal{P}$ are discrete sets with $D^+(\Lambda_p) < \infty$ that constitute a stable collection of MIMO sampling w.r.t. $\boldsymbol{G}(f)$ for $\mathcal{H} = \mathcal{B}(\mathcal{F}_1) \times \cdots \times \mathcal{B}(\mathcal{F}_R)$. Then for every $\Pi \subseteq \mathcal{P}$,*

$$D^-(\{\Lambda_p : p \in \Pi\}) \geq \sum_{r=1}^{R} \mu(\mathcal{F}_r) - \int_\mathbb{R} \operatorname{rank}\big(\boldsymbol{G}_{\Pi^c, \mathcal{C}_f}(f)\big) df, \tag{29}$$

where $\mathcal{C}_f = \{r : f \in \mathcal{F}_r\}$ and Π^c denotes the complement of Π in $\mathcal{P}$. Furthermore, if

$$\operatorname*{ess\,inf}_{f \in \mathcal{F}} \sigma_{\min}\big(\boldsymbol{G}_{\Pi^c, \mathcal{C}_f}(f)\big) = 0, \quad \mathcal{F} = \bigcup_{r \in \mathcal{R}} \mathcal{F}_r, \tag{30}$$

for some $\Pi \neq \mathcal{P}$, then the inequality in (29) is strict.

Proof. Note that $\mathcal{F}$ is the set where $\mathcal{C}_f$ is not empty. Let $\Pi \subseteq \mathcal{P}$ be a fixed subset. We consider two cases: first suppose that either $\Pi = \mathcal{P}$ or (30) does not hold. In this case take $\mathcal{D}_0 = \emptyset$. Otherwise $\Pi \neq \mathcal{P}$, so we can define

$$K = \max_{p \in \Pi^c} C'(\Lambda_p) C(1), \tag{31}$$

where C' and C are quantities defined in Lemma 1. Let $\epsilon_0 > 0$ be such that $K\epsilon_0^2 \leq A/2$, where A is the lower stability bound in (18). Since (30) is satisfied in the second case, there exists a set $\mathcal{D}_0$ such that $\mu(\mathcal{D}_0) > 0$ and

$$\sigma_{\min}\big(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\big) \leq \epsilon_0, \quad \forall f \in \mathcal{D}_0. \tag{32}$$

Without loss of generality, assume that $\mathcal{D}_0 \subseteq [\nu, \nu+1]$ for some $\nu \in \mathbb{R}$. In fact, (32) is satisfied in both cases. Let the cardinality of the set $\mathcal{C}_f$ be denoted by $|\mathcal{C}_f|$, i.e.,

$$|\mathcal{C}_f| = \sum_{r=1}^{R} \chi(f \in \mathcal{F}_r). \tag{33}$$

Let the dimension of the null space of $\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)$ be denoted by

$$\rho(f) = |\mathcal{C}_f| - \operatorname{rank}\big(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\big), \tag{34}$$

and let the columns of $\boldsymbol{U}'(f) \in \mathbb{C}^{|\mathcal{C}_f| \times \rho(f)}$ form an orthonormal basis for the null space of $\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)$. For $f \in \mathcal{D}_0$, let $\boldsymbol{U}''(f) \in \mathbb{C}^{|\mathcal{C}_f| \times 1}$ be a unit-norm right singular vector of $\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)$ corresponding to its smallest nonzero singular value. We can always choose $\boldsymbol{U}'(f)$ and $\boldsymbol{U}''(f)$ to be measurable functions. Clearly, $\boldsymbol{U}''(f)$ is orthogonal to the columns of $\boldsymbol{U}'(f)$ for $f \in \mathcal{D}_0$. Therefore,

$$\boldsymbol{U}(f) = \begin{cases} \big[\boldsymbol{U}'(f) \;\; \boldsymbol{U}''(f)\big] & \text{if } f \in \mathcal{D}_0, \\ \boldsymbol{U}'(f) & \text{otherwise,} \end{cases}$$

has orthonormal columns for all f. Let $\mathcal{G}_r$ be the set where $\boldsymbol{U}(f)$ contains r columns, i.e.,

$$\mathcal{G}_r = \{f : \rho(f) + \chi(f \in \mathcal{D}_0) = r\}, \quad r \in \mathcal{R}. \tag{35}$$

The sets $\{\mathcal{G}_r\}$ are clearly disjoint sets of finite measure. Therefore, for any $\delta > 0$, there exist finite collection of disjoint intervals $\{\mathcal{I}_{rk} : r \in \mathcal{R},\; k = 1, \ldots, K_r\}$ such that the sets

$$\mathcal{G}'_r \stackrel{\text{def}}{=} \bigcup_{k=1}^{K_r} \mathcal{I}_{rk}, \quad r \in \mathcal{R} \tag{36}$$

approximate $\mathcal{G}_r$ in the sense that $\mu(\mathcal{G}'_r \cap \mathcal{G}_r^c) \leq \delta/R^2$ and $\mu(\mathcal{G}'^{\,c}_r \cap \mathcal{G}_r) \leq \delta/R^2$. It follows that

$$\sum_{r=1}^{R} r\mu(\mathcal{G}'_r \cap \mathcal{G}_r^c) \leq \sum_{r=1}^{R} \frac{r\delta}{R^2} \leq \delta. \tag{37}$$

It is also clear that $|\mu(\mathcal{G}'_r) - \mu(\mathcal{G}_r)| \le \delta/R^2$. Consequently, we have

$$\left| \sum_{r=1}^{R} r\mu(\mathcal{G}_r) - \sum_{r=1}^{R} r\mu(\mathcal{G}'_r) \right| \le \delta. \tag{38}$$

Now, define $\boldsymbol{W}^r(f) \in \mathbb{C}^{R\times r}$ on $\mathcal{G}'_r$ for each r as follows:

$$\boldsymbol{W}^r(f) = \begin{cases} \boldsymbol{I}_{\bullet,\mathcal{C}_f}\boldsymbol{U}(f) & \text{if } f \in \mathcal{G}'_r \cap \mathcal{G}_r, \\ \big(\boldsymbol{e}_1, \cdots, \boldsymbol{e}_r\big) & \text{if } f \in \mathcal{G}'_r \cap \mathcal{G}_r^c, \end{cases} \tag{39}$$

where $\boldsymbol{I}$ is the $R \times R$ identity matrix. Note that the columns of $\boldsymbol{W}^r(f)$ form an orthonormal set of vectors for each $f \in \mathcal{G}'_r$. For each $r \in \mathcal{R}$, let k and m be indices such that $1 \le k \le K_r$ and $1 \le m \le r$. For convenience let $q(r, k, m)$ denote an invertible mapping from the triplet (r, k, m) to a single index q:

$$q(r, k, m) : \{(r, k, m) : r \in \mathcal{R},\ k = 1, \ldots, K_r,\ m = 1, \ldots, r\} \to \mathcal{Q},$$

where $\mathcal{Q} = \{1, \ldots, Q\}$ and

$$Q = \sum_{r=1}^{R} rK_r.$$

In the rest of the proof, assume that q, r, k, and m are related to each other by $q = q(r, k, m)$. We shall now define several quantities with the intention of eventually using Theorem 1 to derive the necessary density conditions. Let $\{\boldsymbol{s}_q\} \subseteq \mathcal{H}_\infty$ be defined as follows in terms of their Fourier transforms:

$$\boldsymbol{S}_q(f) = \begin{cases} \boldsymbol{W}^r_{\bullet,m}(f)/\sqrt{\mu(\mathcal{I}_{rk})} & \text{if } f \in \mathcal{I}_{rk}, \\ \boldsymbol{0} & \text{otherwise}, \end{cases} \tag{40}$$

where $\boldsymbol{W}^r_{\bullet,m}(f)$ is the m-th column of $\boldsymbol{W}^r(f)$. The sampling set

$$\Sigma_q \overset{\text{def}}{=} \left\{ \frac{n}{\mu(\mathcal{I}_{rk})} : n \in \mathbb{Z} \right\}, \quad \forall m \in \{1, \ldots, r\} \tag{41}$$

has uniform density of $\mu(\mathcal{I}_{rk})$. Since the intervals $\mathcal{I}_{rk}$ are disjoint, and $\{\boldsymbol{W}^r_{\bullet,m}(f) : m = 1, \ldots, r\}$ is a set of orthonormal vectors for each r and f, it follows that $\{\Theta_{\sigma_{nq}}\boldsymbol{s}_q : q \in \mathcal{Q},\ n \in \mathbb{Z}\}$ is an orthonormal sequence. Let $\mathcal{H}_S$ be the closure of the span of this orthonormal sequence, i.e.,

$$\mathcal{H}_S = \overline{\text{span}\{\Theta_{\sigma_{nq}}\boldsymbol{s}_q : q \in \mathcal{Q},\ n \in \mathbb{Z}\}} \subseteq \mathcal{H}_\infty. \tag{42}$$

Then clearly $\{\Theta_{\sigma_{nq}} \boldsymbol{s}_q : q \in \mathcal{Q},\ n \in \mathbb{Z}\}$ is an orthonormal Riesz basis for $\mathcal{H}_S$ with lower frame bound $a = 1$. In particular, it is a Riesz-Fischer sequence with bound $a = 1$.

Now define

$$\mathcal{H}_L = \{\boldsymbol{x} \in \mathcal{H} : \boldsymbol{X}_{\mathcal{C}_f}(f) = \boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{X}_{\mathcal{C}_f}(f) \text{ a.e.}\}. \tag{43}$$

It is obvious that $\mathcal{H}_L$ is a shift-invariant subspace. To see that $\mathcal{H}_L$ is closed, consider the following argument. Let $\{\boldsymbol{x}^i\} \in \mathcal{H}_L$ be a sequence converging to $\boldsymbol{x}^\infty \in \mathcal{H}_\infty$. Then we have

$$\boldsymbol{X}^i_{\mathcal{C}_f}(f) = \boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{X}^i_{\mathcal{C}_f}(f) \text{ a.e.}$$

Also, $\boldsymbol{X}^i(f)$ converges to $\boldsymbol{X}^\infty(f)$ in the L^2 sense. Hence there exists a subsequence $\{i_j\}$ such that as $j \to \infty$, we have $\boldsymbol{X}^{i_j}(f) \to \boldsymbol{X}^\infty(f)$ a.e. Therefore

$$\boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{X}^\infty_{\mathcal{C}_f}(f) = \lim_{j\to\infty} \boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{X}^{i_j}_{\mathcal{C}_f}(f) = \boldsymbol{X}^\infty_{\mathcal{C}_f}(f) \text{ a.e.,}$$

or equivalently $\boldsymbol{x}_\infty \in \mathcal{H}_L$, proving that $\mathcal{H}_L$ is closed.

Suppose that $\boldsymbol{x} \in \mathcal{H}_L$. Then using (17) and (43) we see that

$$\boldsymbol{Y}_{\Pi^c}(f) = \boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\boldsymbol{X}_{\mathcal{C}_f}(f) = \boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{X}_{\mathcal{C}_f}(f).$$

Using the definitions of $\boldsymbol{U}(f)$ and $\boldsymbol{U}''(f)$, we conclude that

$$\boldsymbol{Y}_{\Pi^c}(f) = \boldsymbol{0}, \quad f \notin \mathcal{D}_0, \tag{44}$$

$$\begin{aligned} \|\boldsymbol{Y}_{\Pi^c}(f)\| &= \|\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\boldsymbol{U}''(f)\boldsymbol{U}''^H(f)\boldsymbol{X}_{\mathcal{C}_f}(f)\| \\ &\leq \epsilon_0 \|\boldsymbol{X}(f)\|, \quad f \in \mathcal{D}_0. \end{aligned} \tag{45}$$

Equations (44) and (45) imply that $\|\boldsymbol{Y}_{\Pi^c}(f)\| \leq \epsilon_0 \|\boldsymbol{X}(f)\|$. Hence

$$\int_{\mathbb{R}} \|\boldsymbol{Y}_{\Pi^c}(f)\|^2 df \leq \epsilon_0^2 \|\boldsymbol{x}\|^2. \tag{46}$$

By (44) for each $p \in \Pi^c$, $Y_p(f)$ is supported on $\mathcal{D}_0 \subseteq [\nu, \nu + 1]$. Applying Lemma 1 to y_p, $p \in \Pi^c$ and using (46) yields

$$\sum_{p\in\Pi^c} \sum_{n\in\mathbb{Z}} |y_p(\lambda_{np})|^2 \leq K \int_{\mathbb{R}} \|\boldsymbol{Y}_{\Pi^c}(f)\|^2 df \leq K\epsilon_0^2 \|\boldsymbol{x}\|^2, \tag{47}$$

where K is the constant defined in (31). Combining (47) with the first inequality in the sampling stability condition in (18), we obtain

$$\sum_{p\in\Pi}\sum_{n\in\mathbb{Z}}|y_p(\lambda_{np})|^2 \geq (A-K\epsilon_0^2)\|\boldsymbol{x}\|^2 \geq \frac{A}{2}\|\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x}\in\mathcal{H}_L, \tag{48}$$

where the second inequality above follows from the choice of ϵ_0. From (18), we obviously also have

$$\sum_{p\in\Pi}\sum_{n\in\mathbb{Z}}|y_p(\lambda_{np})|^2 \leq B\|\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x}\in\mathcal{H}_L \tag{49}$$

because $\mathcal{H}_L \subseteq \mathcal{H}$. Combining (48) and (49), we obtain

$$\frac{A}{2}\|\boldsymbol{x}\|^2 \leq \sum_{p\in\Pi}\sum_{n\in\mathbb{Z}}|y_p(\lambda_{np})|^2 \leq B\|\boldsymbol{x}\|^2, \quad \forall \boldsymbol{x}\in\mathcal{H}_L. \tag{50}$$

Let $\boldsymbol{l}_p = P_{\mathcal{H}_L}\boldsymbol{\psi}_p$, where $\boldsymbol{\psi}_p$ is defined in (19). Recall that $\mathcal{H}_L$ is shift-invariant. Hence, using Proposition 3, we obtain

$$\Theta_{\lambda_{np}}\boldsymbol{l}_p = \Theta_{\lambda_{np}}P_{\mathcal{H}_L}\boldsymbol{\psi}_p = P_{\mathcal{H}_L}\Theta_{\lambda_{np}}\boldsymbol{\psi}_p \in \mathcal{H}_L. \tag{51}$$

Also,

$$\langle \boldsymbol{x}, \Theta_{\lambda_{np}}\boldsymbol{l}_p\rangle = \langle \boldsymbol{x}, P_{\mathcal{H}_L}\Theta_{\lambda_{np}}\boldsymbol{\psi}_p\rangle = \langle \boldsymbol{x}, \Theta_{\lambda_{np}}\boldsymbol{\psi}_p\rangle = y_p(\lambda_{np}), \quad \forall \boldsymbol{x}\in\mathcal{H}_L. \tag{52}$$

It follows from (50), (51), and (52) that $\{\Theta_{\lambda_{np}}\boldsymbol{l}_p : p\in\Pi,\ n\in\mathbb{Z}\}$ is a frame for $\mathcal{H}_L$. Having verified all the required hypotheses, we can now apply Theorem 1 to obtain the following inequality relating the densities of $\{\Lambda_p : p\in\Pi\}$ and $\{\Sigma_q : q\in\mathcal{Q}\}$:

$$D^-(\{\Lambda_p : p\in\Pi\}) \geq D^-(\Sigma_1,\ldots,\Sigma_Q) - \sum_{q\in\mathcal{Q}}\alpha_q D^+(\Sigma_q), \tag{53}$$

where

$$\alpha_q = \frac{1}{\sqrt{a}}\sup_{\sigma\in\Sigma_q}\|\Theta_\sigma\boldsymbol{s}_q - P_{\mathcal{H}_L}\Theta_\sigma\boldsymbol{s}_q\|.$$

Since $\mathcal{H}_L$ is shift-invariant, we can use Proposition 3 to obtain

$$\alpha_q = \frac{1}{\sqrt{a}}\sup_{\sigma\in\Sigma_q}\|\Theta_\sigma\boldsymbol{s}_q - \Theta_\sigma P_{\mathcal{H}_L}\boldsymbol{s}_q\| = \frac{1}{\sqrt{a}}\|\boldsymbol{s}_q - P_{\mathcal{H}_L}\boldsymbol{s}_q\| = \|\boldsymbol{s}_q - P_{\mathcal{H}_L}\boldsymbol{s}_q\|$$

where the last equality holds because $a=1$. We shall estimate α_q in a moment, but first, define $\boldsymbol{v}_q\in\mathcal{H}$ as follows:

$$\boldsymbol{V}_q(f) = \begin{cases} \boldsymbol{S}_q(f) & \text{if } f \in \mathcal{I}_{rk} \cap \mathcal{G}_r, \\ \boldsymbol{0} & \text{otherwise.} \end{cases} \tag{54}$$

For all $f \in \mathcal{I}_{rk} \cap \mathcal{G}_r$, we use (39), (40) and (54) to conclude that

$$\boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{V}_q(f) = \frac{\boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{U}_{\bullet,m}(f)}{\sqrt{\mu(\mathcal{I}_{rk})}} = \frac{\boldsymbol{U}_{\bullet,m}(f)}{\sqrt{\mu(\mathcal{I}_{rk})}} = \boldsymbol{V}_q(f).$$

This proves that $\boldsymbol{v}_q \in \mathcal{H}_L$. Therefore, $\alpha_q \leq \|\boldsymbol{s}_q - \boldsymbol{v}_q\|$. Using Parseval's Theorem and (40) and (54) we obtain

$$\alpha_q \leq \left(\int \|\boldsymbol{S}_q(f) - \boldsymbol{V}_q(f)\|^2 df \right)^{\frac{1}{2}} = \left(\int_{\mathcal{I}_{rk} \cap \mathcal{G}_r^c} \left\| \frac{\boldsymbol{W}^r_{\bullet,m}(f)}{\sqrt{\mu(\mathcal{I}_{rk})}} \right\|^2 df \right)^{\frac{1}{2}}.$$

Since $\boldsymbol{W}^r_{\bullet,m}(f)$ is a normal vector, we arrive at the following estimate for α_q:

$$\alpha_q \leq \sqrt{\frac{\mu(\mathcal{I}_{rk} \cap \mathcal{G}_r^c)}{\mu(\mathcal{I}_{rk})}}. \tag{55}$$

Combining (53) and (55) and using the fact that Σ_q has a uniform density of $\mu(\mathcal{I}_{rk})$, we obtain

$$\begin{aligned} D^-(\{\Lambda_p : p \in \Pi\}) &\geq \sum_{r,m,k} \mu(\mathcal{I}_{rk}) - \sum_{r,m,k} \alpha_q \mu(\mathcal{I}_{rk}) \\ &\geq \sum_r r\,\mu(\mathcal{G}'_r) - \sum_{r,k} r\sqrt{\mu(\mathcal{I}_{rk} \cap \mathcal{G}_r^c)\mu(\mathcal{I}_{rk})}. \end{aligned}$$

Using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} D^-(\{\Lambda_p : p \in \Pi\}) &\geq \sum_r r\mu(\mathcal{G}'_r) - \Big(\sum_{rk} r\mu(\mathcal{I}_{rk} \cap \mathcal{G}_r^c)\Big)^{\frac{1}{2}} \Big(\sum_{rk} r\mu(\mathcal{I}_{rk})\Big)^{\frac{1}{2}} \\ &= \sum_r r\mu(\mathcal{G}'_r) - \Big(\sum_r r\mu(\mathcal{G}'_r \cap \mathcal{G}_r^c)\Big)^{\frac{1}{2}} \Big(\sum_r r\mu(\mathcal{G}'_r)\Big)^{\frac{1}{2}}. \end{aligned} \tag{56}$$

Now, (37), (38), and (56) imply that

$$D^-(\{\Lambda_p : p \in \Pi\}) \geq \Big[\sum_r r\mu(\mathcal{G}_r)\Big] - \delta - \Big[\delta\Big(\delta + \sum_r r\mu(\mathcal{G}_r)\Big)\Big]^{\frac{1}{2}}. \tag{57}$$

Meanwhile, using (33), (34) and (35), and the definition of the Lebesgue integral, we obtain

$$\begin{aligned}\sum_{r=1}^{R} r\mu(\mathcal{G}_r) &= \int_{\mathbb{R}} \left[\rho(f) + \chi(f \in \mathcal{D}_0)\right] df \\ &= \mu(\mathcal{D}_0) + \int_{\mathbb{R}} \left[|\mathcal{C}_f| - \operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right)\right] df \\ &= \mu(\mathcal{D}_0) + \sum_{r=1}^{R} \mu(\mathcal{F}_r) - \int_{\mathbb{R}} \operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right) df. \end{aligned} \tag{58}$$

Putting together (57) and (58), and letting $\delta \to 0$ yields

$$D^-(\{\Lambda_p : p \in \Pi\}) \geq \mu(\mathcal{D}_0) + \sum_{r=1}^{R} \mu(\mathcal{F}_r) - \int_{\mathbb{R}} \operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right) df.$$

This proves (29). Finally, recall that if (30) is satisfied for some $\Pi \neq \mathcal{P}$, then $\mu(\mathcal{D}_0) > 0$, proving that the strict inequality in (29) holds in this case. □

Theorem 2 is a generalization of Landau's classical result to the MIMO sampling problem. Letting $\Pi = \mathcal{P}$ in (29), we obtain

$$D^-(\Lambda_1, \ldots, \Lambda_P) \geq \sum_{r=1}^{R} \mu(\mathcal{F}_r). \tag{59}$$

In other words, the combined sampling density on all the output channels must be no less than the combined bandwidth of all the input signals. Theorem 2 also provides lower bounds on the joint densities of sub-collections of $\{\Lambda_p\}$, and some of them may even be strict inequalities.

In (29), the joint density $D^-(\{\Lambda_p : p \in \Pi\})$ is interpreted as an upper bound on the "signal information" captured by the samples of y_p on those sampling sets. The bound on the joint density of $\{\Lambda_p : p \in \Pi\}$ depends only on the $\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)$, i.e., the submatrix of $\boldsymbol{G}(f)$ whose rows are indexed by the complement of Π and columns by $\mathcal{C}_f$. Note that $G_{pr}(f)$ is irrelevant for $f \notin \mathcal{F}_r$ because $X_r(f)$ vanishes outside $\mathcal{F}_r$. This explains restriction of the columns to the set $\mathcal{C}_f = \{r : f \in \mathcal{F}_r\}$. Suppose that the outputs $y_p(t)$, $p \in \Pi^c$ are completely known for all $t \in \mathbb{R}$. Then $\boldsymbol{Y}_{\Pi^c}(f) = \boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\boldsymbol{X}_{\mathcal{C}_f}(f)$ is also known for all f. Then $\operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right)$ is the number of independent components of $\boldsymbol{X}(f)$ that can be determined from $\boldsymbol{Y}_{\Pi^c}(f)$ alone. Consequently,

$$\int_{\mathbb{R}} \operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right)$$

is a measure of input signal information that can be captured by knowing the outputs $y_p(t)$, $p \in \Pi^c$ completely (for all t). The additional signal information required from the samples of $\{y_p(\lambda_{np}) : p \in \Pi\}$ is indicated by the difference of the two terms in (29). This also explains why (29) depends only on $G_{pr}(f)$

for $p \in \Pi^c$.

Next, if some singular value of $\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)$ takes arbitrarily small nonzero values, then there is not enough information in $\boldsymbol{Y}_{\Pi^c}(f) = \boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\boldsymbol{X}_{\mathcal{C}_f}(f)$ to stably recover $\operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right)$ independent components of $\boldsymbol{X}(f)$. Therefore, the information contained in the samples of $y_p(t)$ must be a little bit more than right-hand side of (29) for stable reconstruction, and this explains the strictness of the bound.

The following corollary shows that stability of sampling requires an additional condition on the channel transfer function matrix $\boldsymbol{G}(f)$.

Corollary 1. *Suppose that the hypotheses of Theorem 2 are satisfied, and let $\mathcal{F} = \bigcup_{r\in\mathcal{R}} \mathcal{F}_r$. Then*

$$\operatorname*{ess\,inf}_{f\in\mathcal{F}} \lambda_{\min}\big(\boldsymbol{G}^H_{\Pi^c,\mathcal{C}_f}(f)\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\big) > 0, \tag{60}$$

for every $\Pi \subset \mathcal{P}$, $\Pi \neq \mathcal{P}$ such that $D^-(\{\Lambda_p : p \in \Pi\}) = 0$. In particular,

$$\operatorname*{ess\,inf}_{f\in\mathcal{F}} \lambda_{\min}\big(\boldsymbol{G}^H_{\bullet,\mathcal{C}_f}(f)\boldsymbol{G}_{\bullet,\mathcal{C}_f}(f)\big) > 0. \tag{61}$$

Proof. From Theorem 2 and (33) and (34) we have

$$D^-(\{\Lambda_p : p \in \Pi\}) \geq \sum_{r=1}^{R} \mu(\mathcal{F}_r) - \int_{\mathbb{R}} \operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right) df \geq \sum_{r=1}^{R} \mu(\mathcal{F}_r) - \int_{\mathbb{R}} |\mathcal{C}_f| df \geq 0.$$

If $D^-(\{\Lambda_p : p \in \Pi\}) = 0$, all the inequalities above must, in fact, be equalities. In particular, (29) holds with an equality, which implies that

$$\operatorname*{ess\,inf}_{f\in\mathcal{F}} \sigma_{\min}\big(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\big) > 0,$$

by Theorem 2. We also have $\operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right) = |\mathcal{C}_f|$ a.e., implying that

$$\lambda_{\min}\big(\boldsymbol{G}^H_{\Pi^c,\mathcal{C}_f}(f)\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\big) = \big[\sigma_{\min}\big(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\big)\big]^2.$$

Now (60) follows by combining the last two observation. Applying this result to $\Pi = \emptyset$, we obtain (61). ☐

Equation (61), which states that the singular values of $\boldsymbol{G}_{\bullet,\mathcal{C}_f}(f)$ are uniformly bounded away from the origin, must always hold for stable MIMO sampling. In fact, even if all outputs $y_p(t)$ are known for all t, we cannot stable recover the channel inputs unless (61) holds because this condition is necessary to satisfy the lower stability bound in (18). In particular, a more elementary necessary condition that emerges from (61) is that $P \geq |\mathcal{C}_f|$ a.e., i.e., the number of channels cannot be less than the number of overlapping input spectral supports at any frequency. Next $D^-(\{\Lambda_p : p \in \Pi\}) = 0$ implies that the output samples on the sampling sets $\{\Lambda_p : p \in \Pi\}$ are too sparse to contain any signal information. Therefore, we must rely entirely on the outputs samples taken on $\{\Lambda_p : p \in \Pi^c\}$ to achieve stable reconstruction, and an argument as before justifies the validity of (60).

The following theorem provides another necessary condition for stable sampling.

Theorem 3. *Under the hypotheses of Theorem 2,*

$$\operatorname*{ess\,sup}_{f \in \mathcal{F}} \sigma_{\max}\big(\boldsymbol{G}_{\Pi^+,\mathcal{C}_f}(f)\big) < \infty, \tag{62}$$

where $\Pi^+ = \{p \in \mathcal{P} : D^+(\Lambda_p) > 0\}$ *and* $\mathcal{F} = \bigcup_{r \in \mathcal{R}} \mathcal{F}_r$.

Proof. Suppose that (62) fails to hold, then some entries of $\boldsymbol{G}_{\Pi^+,\mathcal{C}_f}(f)$ are necessarily unbounded on $\mathcal{F}$. So let $p_\circ \in \Pi^+$ and $r_\circ \in \mathcal{R}$ be indices such that for every $\epsilon > 0$, there exists

$$\mathcal{G} \subseteq \mathcal{F}_{r_\circ} \cap \{f : |G_{p_\circ,r_\circ}(f)|^2 \geq 1/\epsilon\},$$

satisfying $\mu(\mathcal{G}) > 0$. Without loss of generality assume that $\mu(\mathcal{G}) < \infty$. Let $\{\mathcal{I}_k : k = 1, \ldots, K\}$ be a finite collection of disjoint intervals such that

$$\mathcal{G}' = \bigcup_{k=1}^{K} \mathcal{I}_k,$$

satisfies $\mu(\mathcal{G}' \cap \mathcal{G}^c) \leq \delta$ and $\mu(\mathcal{G}'^c \cap \mathcal{G}) \leq \delta$, where $\delta = \epsilon\mu(\mathcal{G})/(1+\epsilon)$. It follows easily that $\mu(\mathcal{G}') \geq \mu(\mathcal{G}) - \delta$. Now at least one interval $\mathcal{I}_k$ satisfies

$$\frac{\mu(\mathcal{I}_k \cap \mathcal{G}^c)}{\mu(\mathcal{I}_k)} \leq \epsilon. \tag{63}$$

Otherwise, we would have

$$\mu(\mathcal{G}' \cap \mathcal{G}^c) = \sum_{k=1}^{K} \mu(\mathcal{I}_k \cap \mathcal{G}^c) > \sum_{k=1}^{K} \epsilon\mu(\mathcal{I}_k) = \epsilon\mu(\mathcal{G}') \geq \epsilon(\mu(\mathcal{G}) - \delta) = \delta,$$

violating our assumption that $\mu(\mathcal{G}' \cap \mathcal{G}^c) \leq \delta$. So, let $\mathcal{I}_k$ denote an interval that satisfies (63). Define $\gamma = 1/(2\mu(\mathcal{I}_k))$. Since $D^+(\Lambda_{p_\circ}) > 0$, Proposition 2 and (21) imply that there exists $\tau \in \mathbb{R}$ such that

$$\#(\Lambda_{p_\circ} \cap B_\gamma(\tau)) \geq 2\gamma(D^+(\Lambda_{p_\circ})/2). \tag{64}$$

Define

$$X_{r_\circ}(f) = \begin{cases} e^{-j2\pi f\tau}/G_{p_\circ,r_\circ}(f) & \text{if } f \in \mathcal{I}_k \cap \mathcal{G}, \\ 0 & \text{otherwise,} \end{cases}$$

and $X_r(f) = 0$ for all $r \neq r_\circ$. Then, we clearly have

$$\|\boldsymbol{x}\|^2 \leq \epsilon\mu(\mathcal{I}_k). \tag{65}$$

Using (17), we conclude that

$$Y_{p_\circ}(f) = e^{-j2\pi f\tau}\chi(f \in \mathcal{I}_k \cap \mathcal{G}) = e^{-j2\pi f\tau}\big[\chi(f \in \mathcal{I}_k) - \chi(f \in \mathcal{I}_k \cap \mathcal{G}^c)\big],$$

whose inverse Fourier transform is

$$y_{p_\circ}(t) = \mu(\mathcal{I}_k)\,\mathrm{sinc}\,\big(\mu(\mathcal{I}_k)(t-\tau)\big)e^{-j2\pi f_0(t-\tau)} - \phi_{\mathcal{I}_k\cap\mathcal{G}^c}(t-\tau), \quad t \in \mathbb{R}, \tag{66}$$

where $\mathrm{sinc}(t) = \sin(\pi t)/(\pi t)$ and f_0 is the midpoint of $\mathcal{I}_k$. Note that

$$\sup_t |\phi_{\mathcal{I}_k\cap\mathcal{G}^c}(t)| \leq \mu(\mathcal{I}_k \cap \mathcal{G}^c) \leq \epsilon\mu(\mathcal{I}_k). \tag{67}$$

Then, for all t such that $|t-\tau| \leq \gamma = 1/(2\mu(\mathcal{I}_k))$, it follows from (66) and (67) that

$$|y_{p_\circ}(t)| \geq \mu(\mathcal{I}_k)(\mathrm{sinc}(1/2) - \epsilon). \tag{68}$$

Using (64) and (68), we obtain

$$\begin{aligned} \sum_{p=1}^{P}\sum_{n\in\mathbb{Z}} |y_p(\lambda_{np})|^2 \geq \sum_{n\in\mathbb{Z}} |y_{p_\circ}(\lambda_{np_\circ})|^2 &\geq \sum_{\lambda\in\Lambda_{p_\circ}\cap B_\gamma(\tau)} |y_{p_\circ}(\lambda)|^2 \\ &\geq 2\gamma(D^+(\Lambda_{p_\circ})/2)\big[\mu(\mathcal{I}_k)(\mathrm{sinc}(1/2)-\epsilon)\big]^2 \\ &= \frac{1}{2}D^+(\Lambda_{p_\circ})\mu(\mathcal{I}_k)\big[\,\mathrm{sinc}(1/2)-\epsilon\big]^2. \end{aligned}$$

Combining this result with (65), we conclude that

$$\sum_{p=1}^{P}\sum_{n\in\mathbb{Z}} |y_p(\lambda_{np})|^2 \geq \frac{1}{2\epsilon}D^+(\Lambda_{p_\circ})\big[\,\mathrm{sinc}(1/2)-\epsilon\big]^2\|\boldsymbol{x}\|^2.$$

Since $\epsilon > 0$ is arbitrary, and $\boldsymbol{x} \in \mathcal{H}$ is nonzero, the above conclusion violates the second inequality of the stability condition (18), proving the necessity of (62). □

In Theorem 3, suppose that $D^+(\Lambda_p) = 0$ for some $p \in \mathcal{P}$. Then the samples of y_p on Λ_p are too sparse to provide any useful information. Consequently, the p-th row of the $\boldsymbol{G}(f)$ is irrelevant. Thus, Π^+ is the set of outputs whose samples are sufficiently dense to provide information about the inputs. Now, (62) can be interpreted as being equivalent to the upper stability bound in (18) for stable sampling.

To explore the implications of Theorem 2) observe that we can view the set of densities

$$\left\{D^-(\{\Lambda_p : p \in \Pi\}) : \Pi \subseteq \mathcal{P},\ \Pi \neq \emptyset\right\}$$

as a point in $\mathbb{R}^{2^P-1}$. However, if the sampling sets have uniform densities[2], then

$$D^-(\{\Lambda_p : p \in \Pi\}) = \sum_{p\in\Pi} D(\Lambda_p),$$

i.e., the joint densities become linearly related to each other, and they can all be described in terms of the individual densities. The resulting density space is now $\mathbb{R}^P$, which has a much smaller dimension. In general, we are interested in constructing stable sampling sets whose lower and upper densities are the smallest possible. Obviously, it becomes desirable to achieve minimum density sampling with sampling sets of uniform densities.

Definition 5. The *density region for stable sampling* is defined as the collection of all $(d_1, \ldots, d_P)$ such that stable MIMO sampling is realizable using sampling sets $\{\Lambda_p\}$ of uniform densities $D(\Lambda_p) = d_p$, $p \in \mathcal{P}$.

Suppose that every Λ_p has uniform density $D(\Lambda_p) = d_p$, then the sampling density conditions in (29) reduce to

$$\sum_{p\in\Pi} d_p \geq \theta_S(\Pi), \quad \forall \Pi \subseteq \mathcal{P}, \tag{69}$$
$$\theta_S(\Pi) = \sum_{r=1}^{R} \mu(\mathcal{F}_r) - \int_{\mathbb{R}} \operatorname{rank}\left(\boldsymbol{G}_{\Pi^c,\mathcal{C}_f}(f)\right) df.$$

Since (69) is a set of necessary conditions whose sufficiency is unknown, the region specified by (69) is generally an outer bound on the density region. It is clear that $\theta_S(\emptyset) = 0$ and $\theta_S(\Pi_1) \leq \theta_S(\Pi_2)$ whenever $\Pi_1 \subseteq \Pi_2$. Using Proposition 1, it is an easy exercise to show that

$$\theta_S(\Pi_1) + \theta_S(\Pi_2) \leq \theta_S(\Pi_1 \cup \Pi_2) + \theta_S(\Pi_1 \cap \Pi_2), \quad \forall \Pi_1, \Pi_2 \subseteq \mathcal{P}.$$

[2] Recall that a sampling set need not have uniform sample spacing in order to have uniform density.

These properties of $\theta_S(\Pi)$ imply that the outer bound on the density region specified by the system of inequalities in (69) forms a contra-polymatroid [23, 24]. Consequently, every constraint in (69) is active, i.e., the equality in each constraint in (69) holds for some point in the region.

We now present a simple example to illustrate the necessary conditions for stable MIMO sampling.

Example 1. Consider a MIMO channel with $R = 2$ inputs and $P = 2$ outputs having the following transfer function matrix:

$$\boldsymbol{G}(f) = \begin{bmatrix} 1 & K(f) \\ 0 & 1 \end{bmatrix},$$

where $K(f) = (1 - 2f/3)\chi(f \in [0, 1.5])$ is shown in Figure 2. Let $\mathcal{F}_1 = [-1, 1)$ and $\mathcal{F}_2 = [0, 2)$ be the input spectral supports. Figure 3 illustrates the input and output spectra for a typical set of channel inputs.

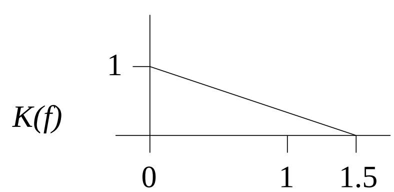

Figure 2: $K(f)$.

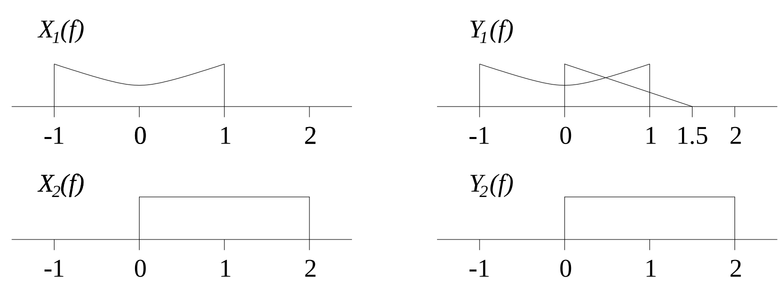

Figure 3: Typical spectra of the channel inputs and outputs.

We seek conditions on the sampling sets Λ_1 and Λ_2 for stable MIMO sampling w.r.t. $\boldsymbol{G}(f)$. We have $\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2 = [-1, 2]$ and

$$\boldsymbol{C}_f = \{r : f \in \mathcal{F}_r\} = \begin{cases} \{1\} & \text{if } f \in [-1, 0), \\ \{1, 2\} & \text{if } f \in [0, 1), \\ \{2\} & \text{if } f \in [1, 2), \\ \emptyset & \text{otherwise.} \end{cases}$$

It is easy to check that (62) is satisfied regardless of Π^+. Also

$$\sigma_{\min}\big(\boldsymbol{G}_{\bullet,\mathcal{C}_f}(f)\big) = \frac{[2+K^2(f)] - \sqrt{[2+K^2(f)] - 4}}{2}.$$

This quantity is uniformly bounded from below because $K(f)$ is a bounded function. Hence, the necessary condition in (61) is satisfied. Applying Theorem 2, we obtain the following density conditions:

$$\begin{aligned}
D^-(\Lambda_1,\Lambda_2) &\geq \mu(\mathcal{F}_1) + \mu(\mathcal{F}_2) = 4\\
D^-(\Lambda_1) &\geq \mu(\mathcal{F}_1) + \mu(\mathcal{F}_2) - \int_{\mathbb{R}} \operatorname{rank}\big(\boldsymbol{G}_{2,\mathcal{C}_f}(f)\big)df\\
&= 4 - \int_{[-1,0]} 0df - \int_{[0,1]} 1df - \int_{[1,2]} 1df = 2\\
D^-(\Lambda_2) &\geq \mu(\mathcal{F}_1) + \mu(\mathcal{F}_2) - \int_{\mathbb{R}} \operatorname{rank}\big(\boldsymbol{G}_{1,\mathcal{C}_f}(f)\big)df\\
&= 4 - \int_{[-1,0]} 1df - \int_{[0,1]} 1df - \int_{[1,2]} \chi(K(f)\neq 0)df\\
&= 1.5.
\end{aligned}$$

Now, a simple calculation reveals that

$$\sigma_{\min}\big(\boldsymbol{G}_{1,\mathcal{C}_f}(f)\big) = \begin{cases} 1 & \text{if } f\in[-1,0),\\ \sqrt{1+K^2(f)} & \text{if } f\in[0,1),\\ |K(f)| & \text{if } f\in[1,2). \end{cases}$$

Clearly, this quantity takes arbitrarily small values in the vicinity of $f=1.5$, where $K(f)$ vanishes. Hence the bound on $D^-(\Lambda_2)$ is indeed a strict inequality. Another calculation reveals that[3]

$$\sigma_{\min}\big(\boldsymbol{G}_{2,\mathcal{C}_f}(f)\big) = \begin{cases} \infty & \text{if } f\in[-1,0),\\ 1 & \text{if } f\in[0,1),\\ 1 & \text{if } f\in[1,2). \end{cases}$$

Hence, we cannot guarantee that the bound on $D^-(\Lambda_1)$ is a strict inequality. Thus, we obtain the following conditions on the joint densities:

$$D^-(\Lambda_1,\Lambda_2)\geq 4,\quad D^-(\Lambda_1)\geq 2,\quad \text{and}\quad D^-(\Lambda_2) > 1.5.$$

[3]Recall that we take $\sigma_{\min}(\boldsymbol{A}) = \infty$ if $\boldsymbol{A}=\boldsymbol{0}$.

We shall now explain the above bounds intuitively. First of all, we can interpret Landau's classical sampling density result as follows: the sampling density must be no less than the number of *units of signal information*, where a unit of signal information, roughly speaking, equals the information contained in a unit bandwidth of a signal spectrum. In other words, sampling at unit density provides no more than one unit of signal information. In the MIMO problem, the condition $D^{-}(\Lambda_1, \Lambda_2) \geq 4$ is a natural generalization of Landau's result because the combined bandwidth of the input signals in 4 units.

Next, from at Figure 3, we see that $Y_2(f)$ is unaffected by $X_1(f)$. Therefore $X_1(f)$ on the set $[-1, 0)$ must be reconstructed from the samples of y_2 alone. In order to reconstruct $X_1(f)$ and $X_2(f)$ on $[0, 1)$, it is clearly necessary that $Y_1(f)$ and $Y_2(f)$ on $[0, 1)$ be reconstructible from their samples. Now, $X_2(f)$ on $[1.5, 2)$ must be reconstructed based entirely on $Y_2(f)$. Combining all these observations, we see that the samples of y_1 contain enough information to reconstruct $Y_1(f)$ on $[-1, 1)$, while the samples of y_2 contain enough information to reconstruct $Y_2(f)$ on $[0, 1) \cup [1.5, 2)$. This explains that $D^{-}(\Lambda_1) \geq 2$ and that $D^{-}(\Lambda_2) \geq 1.5$. Suppose that $D^{-}(\Lambda_2) = 1.5$, then roughly speaking, there is just enough information in the samples of y_2 to reconstruct $Y_2(f)$ on $[0, 1) \cup [1.5, 2)$. Consequently, we must reconstruct $Y_2(f) = X_2(f)$ for all $f \in [1, 1.5)$ based entirely on the samples of y_1. Now, $Y_1(f) = K(f)X_2(f)$ for $f \in [1, 1.5)$, where $K(f)$ takes arbitrarily small values on $[1, 1.5)$. Hence, this inversion cannot be accomplished stably, justifying the need for $D^{-}(\Lambda_2) > 1.5$.

Finally, we point out that we can have under-sampling at each output and yet, be able to reconstruct all the inputs jointly from the available information. For instance, we do not need $D^{-}(\Lambda_1) \geq 2.5$, even though y_1 has a bandwidth of 2.5. To see this, we construct a sampling scheme for which the densities $(d_1, d_2) = (2, 2)$ are *achievable*, where $d_p = D(\Lambda_p)$, i.e., Λ_p has a uniform density of d_p. Let Λ_1 and Λ_2 be uniform sampling lattices:

$$\Lambda_1 = \Lambda_2 = \{n/2 : n \in \mathbb{Z}\}.$$

Clearly, $y_2 = x_2$ can be reconstructed stably. Now, the samples of x_1 on Λ_1 can be computed as follows:

$$x_1(\lambda_{n1}) = y_1(\lambda_{n1}) - k * x_2(\lambda_{n1}),$$

because $x_2(t)$ is known for all t. Thus, x_2 can also be reconstructed stably. However, it is not immediately clear whether all densities satisfying the necessary conditions are achievable, or how to achieve them.

C. Density conditions for consistent reconstruction

We now present the necessary condition for consistent MIMO reconstruction, which is a dual problem to the one of stable sampling.

Theorem 4. *Suppose that $\mathcal{F}_r$, $r \in \mathcal{R}$ are real sets of finite measure, and Λ_p, $p \in \mathcal{P}$ are discrete sets that constitute a collection of consistent reconstruction w.r.t. $\boldsymbol{G}(f)$ for $\mathcal{H} = \mathcal{B}(\mathcal{F}_1) \times \cdots \times \mathcal{B}(\mathcal{F}_R)$. Then for every $\Pi \subseteq \mathcal{P}$,*

$$D^+(\{\Lambda_p : p \in \Pi\}) \leq \int_{\mathbb{R}} \operatorname{rank}\big(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\big) df, \tag{70}$$

where $\mathcal{C}_f = \{r : f \in \mathcal{F}_r\}$. Furthermore, if

$$\operatorname*{ess\,inf}_{f \in \mathcal{F}} \sigma_{\min}\big(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\big) = 0, \quad \mathcal{F} = \bigcup_{r \in \mathcal{R}} \mathcal{F}_r, \tag{71}$$

for some $\Pi \neq \emptyset$, then the inequality in (70) is strict.

Proof. First note that the consistency condition implies that

$$\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : p \in \mathcal{P},\ n \in \mathbb{Z}\}, \tag{72}$$

is a Riesz-Fischer sequence in $\mathcal{H}$, where $\boldsymbol{\psi}_p$ is defined in (19). Let $\Pi \subseteq \mathcal{P}$ be a fixed subset. Consider the following two cases. First suppose that either $\Pi = \emptyset$, or $D^+(\Lambda_p) = \infty$ for some $p \in \Pi$, or (71) does not hold: in this case take $\mathcal{D}_0 = \emptyset$. Otherwise, $D^+(\Lambda_p) < \infty$ for all $p \in \Pi \neq \emptyset$. So, we can define

$$K = \max_{p \in \Pi} C'(\Lambda_p) C(1) < \infty, \tag{73}$$

where C' and C are quantities defined in Lemma 1. Since (71) is satisfied in the second case, we can find a set $\mathcal{D}_0$ such that $\mu(\mathcal{D}_0) > 0$, and

$$\sigma_{\min}\big(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\big) \leq \epsilon_0, \quad \forall f \in \mathcal{D}_0. \tag{74}$$

where $\epsilon_0 > 0$ is such that $K\epsilon_0^2 \leq a/4$, and a is the bound for the Riesz-Fischer sequence in (72). Assume without loss of generality that $\mathcal{D}_0 \subseteq [\nu, \nu+1]$ for some $\nu \in \mathbb{R}$. Thus, (74) is satisfied in both cases. Let the dimension of the range space of $\boldsymbol{G}^H_{\Pi,\mathcal{C}_f}(f)$ be denoted by

$$\rho(f) = \operatorname{rank}\big(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\big), \tag{75}$$

and let the columns of $\boldsymbol{U}'(f) \in \mathbb{C}^{|\mathcal{C}_f| \times \rho(f)}$ form an orthonormal basis for the range space of $\boldsymbol{G}^H_{\Pi,\mathcal{C}_f}(f)$. Note that $\rho(f)$ is defined differently from $\rho(f)$ in the proof of Theorem 2. For $f \in \mathcal{D}_0$, let $\boldsymbol{U}''(f) \in \mathbb{C}^{|\mathcal{C}_f| \times 1}$ be a unit-norm right singular vector of $\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)$ corresponding to its smallest nonzero singular value. There is no loss of generality is assuming that the first column of $\boldsymbol{U}'(f)$ equals $\boldsymbol{U}''(f)$ for all $f \in \mathcal{D}_0$. Hence for

$f \in \mathcal{D}_0$, we can write

$$\boldsymbol{U}'(f) = \left[\boldsymbol{U}''(f) \;\; \boldsymbol{U}(f)\right],$$

for some $\boldsymbol{U}(f)$. For $f \notin \mathcal{D}_0$, let $\boldsymbol{U}(f) = \boldsymbol{U}'(f)$. The columns of $\boldsymbol{U}(f)$ are clearly orthonormal. Note that $\boldsymbol{U}'(f)$ and $\boldsymbol{U}''(f)$ can be assumed to be measurable. The matrix $\boldsymbol{U}(f)$ has r columns for $f \in \mathcal{G}_r$, where

$$\mathcal{G}_r = \{f : \rho(f) - \chi(f \in \mathcal{D}_0) = r\}, \quad r \in \mathcal{R}. \tag{76}$$

Each $\mathcal{G}_r$ has finite measure. Since the sets $\{\mathcal{G}_r\}$ are disjoint, we can find, as in the proof of Theorem 2, a collection of disjoint intervals $\{\mathcal{I}_{rk} : r \in \mathcal{R},\; k = 1, \ldots, K_r\}$ and sets $\mathcal{G}'_r$ as in (36) such that (37) and (38) hold. In the rest of the proof, assume as in the proof of Theorem 2 that $q = q(r, k, m)$ for some invertible index-mapping.

Define $\boldsymbol{W}^r(f) \in \mathbb{C}^{R \times r}$, $\boldsymbol{S}_q(f)$, and Σ_q exactly as in (39), (40), and (41) respectively. Also let $\mathcal{H}_S$ be the closed subspace of $\mathcal{H}_\infty$ defined as in (42). Using arguments similar to those in Theorem 2, we see that

$$\{\Theta_{\sigma_{nq}} \boldsymbol{s}_q : q \in \mathcal{Q},\; n \in \mathbb{Z}\}$$

is an orthonormal Riesz basis for $\mathcal{H}_S$. In particular, it is a frame for $\mathcal{H}_S$. It is also easily verified that $\mathcal{H}_S$ is a shift-invariant subspace of $\mathcal{H}_\infty$.

Now, $\{\Theta_{\lambda_{np}} \boldsymbol{\psi}_p : p \in \Pi,\; n \in \mathbb{Z}\}$, being a sub-collection of the set in (72), is also a Riesz-Fischer sequence in $\mathcal{H}$. Let $\{c_{np}\}$, $p \in \Pi$ be some finite sequences. Then (20) implies that

$$\max_{\boldsymbol{x} \in \mathcal{B}_\mathcal{H}} \Big| \sum_{n,p} c_{np} y_p(\lambda_{np}) \Big|^2 \geq a \sum_{n,p} |c_{np}|^2. \tag{77}$$

Let $\boldsymbol{x}^\circ \in \mathcal{B}_\mathcal{H}$ be the maximizer of the left-hand side of (77), and $\boldsymbol{Y}^\circ(f) = \boldsymbol{G}(f)\boldsymbol{X}^\circ(f)$, its corresponding MIMO channel output. Then,

$$\Big| \sum_{n,p} c_{np} y^\circ_p(\lambda_{np}) \Big| \geq \sqrt{a} \|c\|. \tag{78}$$

Next, the subspace

$$\mathcal{H}_L = \{\boldsymbol{x} \in \mathcal{H} : \boldsymbol{X}_{\mathcal{C}_f}(f) = \boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{X}_{\mathcal{C}_f}(f) \text{ a.e.}\} \tag{79}$$

is closed and shift-invariant by the same argument as in the proof of Theorem 2. Note that

$$\boldsymbol{X}(f) = \boldsymbol{I}_{\bullet,\mathcal{C}_f} \boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{I}_{\mathcal{C}_f,\bullet}\boldsymbol{X}(f) \tag{80}$$

is an equivalent way of stating $\boldsymbol{x} \in \mathcal{H}_L$ because $\boldsymbol{X}_{\mathcal{C}_f}(f)$ captures all the nonzero elements of $\boldsymbol{X}(f)$ for every $\boldsymbol{x} \in \mathcal{H}$. Let $\boldsymbol{x}' \in \mathcal{H}$ be defined as follows:

$$\boldsymbol{X}'_{\mathcal{C}_f}(f) = \boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{X}^\circ_{\mathcal{C}_f}(f).$$

Evidently $\boldsymbol{x}' \in \mathcal{H}_L$ and $\|\boldsymbol{x}'\| \leq 1$ because $\boldsymbol{U}(f)$ has orthonormal columns. Let $\boldsymbol{Y}'(f) = \boldsymbol{G}(f)\boldsymbol{X}'(f)$ and $\boldsymbol{Y}''(f) = \boldsymbol{Y}^\circ(f) - \boldsymbol{Y}'(f)$, and recall that for all $f \notin \mathcal{D}_0$, we have $\boldsymbol{U}(f) = \boldsymbol{U}'(f)$. Hence, using the definition of $\boldsymbol{U}'(f)$, we conclude that

$$\boldsymbol{Y}''_\Pi(f) = \boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\big[\boldsymbol{X}^\circ_{\mathcal{C}_f}(f) - \boldsymbol{U}'(f){\boldsymbol{U}'}^H(f)\boldsymbol{X}^\circ_{\mathcal{C}_f}(f)\big] = \boldsymbol{0}, \quad \forall f \in \mathcal{D}_0 \tag{81}$$

For $f \in \mathcal{D}_0$, we have $\boldsymbol{U}(f)\boldsymbol{U}^H(f) = \boldsymbol{U}'(f){\boldsymbol{U}'}^H(f) - \boldsymbol{U}''(f){\boldsymbol{U}''}^H(f)$. Hence

$$\begin{aligned}\|\boldsymbol{Y}''_\Pi(f)\| &= \|\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\big[\boldsymbol{X}^\circ_{\mathcal{C}_f}(f) - \boldsymbol{U}(f)\boldsymbol{U}^H(f)\boldsymbol{X}^\circ_{\mathcal{C}_f}(f)\big]\| \\ &= \|\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\boldsymbol{U}''(f){\boldsymbol{U}''}^H(f)\boldsymbol{X}^\circ_{\mathcal{C}_f}(f)\| \\ &\leq \epsilon_0\|\boldsymbol{X}^\circ_{\mathcal{C}_f}(f)\|, \quad \forall f \in \mathcal{D}_0 \end{aligned} \tag{82}$$

Combining (81) and (82), and using $\|\boldsymbol{x}^\circ\| \leq 1$, we have

$$\int_{\mathbb{R}} \|\boldsymbol{Y}''_\Pi(f)\|^2 df \leq \begin{cases} \epsilon_0^2 & \text{if } \mu(\mathcal{D}_0) > 0, \\ 0 & \text{if } \mu(\mathcal{D}_0) = 0. \end{cases} \tag{83}$$

Notice that $Y''_p(f)$ is supported on $\mathcal{D}_0 \subseteq [\nu, \nu+1]$ for all $p \in \Pi$. Recall that $\mu(\mathcal{D}_0) > 0$ implies that $D^+(\Lambda_p) < \infty$. In this case, we can invoke Lemma 1 to conclude that

$$\sum_{p\in\Pi}\sum_{n\in\mathbb{Z}} |y''_p(\lambda_{np})|^2 \leq K\epsilon_0^2, \tag{84}$$

where K is the constant defined in (73). However, if $\mu(\mathcal{D}_0) = 0$, then $\boldsymbol{Y}''_\Pi(f) = \boldsymbol{0}$ from (83), and hence (84) holds trivially. In other words (84) always holds, and using the Cauchy-Schwarz inequality, we conclude that

$$\Big|\sum_{p\in\Pi}\sum_{n\in\mathbb{Z}} c_{np} y''_p(\lambda_{np})\Big| \leq \sqrt{K}\epsilon_0\|c\|, \tag{85}$$

Recall that ϵ_0 is chosen so that $K\epsilon_0^2 \leq a/4$. So, combining (78) and (85) and noting that $y'_p = y^\circ_p - y''_p$, we obtain

$$\Big|\sum_{p\in\Pi}\sum_{n\in\mathbb{Z}} c_{np} y'_p(\lambda_{np})\Big| \geq (\sqrt{a} - \sqrt{K}\epsilon_0)\|c\| \geq \frac{\sqrt{a}}{2}\|c\|.$$

Since the quantities $\{y'_p\}$ are the channel outputs corresponding to an input $\boldsymbol{x}' \in \mathcal{H}_L$ satisfying $\|\boldsymbol{x}'\| \le 1$, we have

$$\max_{\boldsymbol{x} \in B_{\mathcal{H}_L}} \Big| \sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} c_{np} y_p(\lambda_{np}) \Big|^2 \ge \frac{a}{4} \|c\|^2. \tag{86}$$

Define $\boldsymbol{l}_p = P_{\mathcal{H}_L} \boldsymbol{\psi}_p$. Since $\mathcal{H}_L$ is shift-invariant, we obtain $\Theta_{\lambda_{np}} \boldsymbol{l}_p \in \mathcal{H}_L$ and $y_p(\lambda_{np}) = \langle \boldsymbol{x}, \Theta_{\lambda_{np}} \boldsymbol{l}_p \rangle$ using the same argument as in (51) and (52). Therefore, (86) implies that

$$\Big\| \sum_{p \in \Pi} \sum_{n \in \mathbb{Z}} c_{np} \Theta_{\lambda_{np}} \boldsymbol{l}_p \Big\|^2 \ge \frac{a}{4} \|c\|^2$$

for all finite sequences $\{c_{np}\}$, $p \in \Pi$. Then we conclude, using (10), that $\{\Theta_{\lambda_{np}} \boldsymbol{l}_p : p \in \Pi,\ n \in \mathbb{Z}\}$ is a Riesz-Fischer sequence in $\mathcal{H}_L$ with bound $a/4$.

To avoid confusion, we point out that the quantities associated with the frame in this proof are $\mathcal{H}_S$, Σ_q, $\boldsymbol{s}_q$ etc., while those associated with the Riesz-Fischer sequence are $\mathcal{H}_L$, Λ_p, $\boldsymbol{l}_p$ etc. This is opposite from the convention adopted in Theorem 1 and the proof of Theorem 2. We shall estimate the quantities

$$\alpha_p = \sqrt{\frac{4}{a}} \sup_{\lambda \in \Lambda_p} \|\Theta_\lambda \boldsymbol{l}_p - P_{\mathcal{H}_S} \Theta_\lambda \boldsymbol{l}_p\|$$

defined in Theorem 1 shortly, but first, define $\boldsymbol{v}_p \in \mathcal{H}_\infty$ for $p \in \Pi$ as follows:

$$\boldsymbol{V}_p(f) = \begin{cases} \boldsymbol{L}_p(f) & \text{if } f \in \mathcal{G}'_r \cap \mathcal{G}_r, \text{ for some } r, \\ \boldsymbol{0} & \text{otherwise.} \end{cases} \tag{87}$$

Since $\boldsymbol{l}_p \in \mathcal{H}_L$, (80) implies that

$$\boldsymbol{L}_p(f) = \boldsymbol{I}_{\bullet, \mathcal{C}_f} \boldsymbol{U}(f) \boldsymbol{U}^H(f) \boldsymbol{I}_{\mathcal{C}_f, \bullet} \boldsymbol{L}_p(f).$$

Therefore, whenever $f \in \mathcal{G}'_r \cap \mathcal{G}_r$, we have

$$\begin{aligned} \boldsymbol{V}_p(f) &= \boldsymbol{I}_{\bullet, \mathcal{C}_f} \boldsymbol{U}(f) \boldsymbol{U}^H(f) \boldsymbol{I}_{\mathcal{C}_f, \bullet} \boldsymbol{L}_p(f) \\ &= \sum_{m=1}^{r} \boldsymbol{I}_{\bullet, \mathcal{C}_f} \boldsymbol{U}_{\bullet, m}(f) \big[\boldsymbol{U}^H_{\bullet, m}(f) \boldsymbol{I}_{\mathcal{C}_f, \bullet} \boldsymbol{L}_p(f) \big]. \end{aligned}$$

Using (36), (40), and the above observation, we can express $\boldsymbol{V}_p(f)$ as a linear combination involving $\boldsymbol{S}_q(f)$:

$$\boldsymbol{V}_p(f) = \sum_{r=1}^{R} \sum_{k=1}^{K_r} \sum_{m=1}^{r} \sqrt{\mu(\mathcal{I}_{rk})} \boldsymbol{S}_q(f) \big[\boldsymbol{U}^H_{\bullet, m}(f) \boldsymbol{I}_{\mathcal{C}_f, \bullet} \boldsymbol{L}_p(f) \big] \chi(f \in \mathcal{G}_r), \quad f \in \mathbb{R}.$$

Recall that $\mathcal{H}_S$ is shift-invariant. Now, the quantity $[\boldsymbol{U}^H_{\bullet,m}(f)\boldsymbol{I}_{\mathcal{C}_f,\bullet}\boldsymbol{L}_p(f)]$ is clearly square-integrable. Hence, $\boldsymbol{v}_p$ in the time-domain is obtained by first convolving each $\boldsymbol{s}_q$ with a square-integrable function and then adding them together. Hence, we conclude that $\boldsymbol{v}_p \in \mathcal{H}_S$. Thus, using Proposition 3, we obtain

$$\begin{aligned}\alpha_p &= \sqrt{\frac{4}{a}} \sup_{\lambda\in\Lambda_p} \|\Theta_\lambda \boldsymbol{l}_p - P_{\mathcal{H}_S}\Theta_\lambda \boldsymbol{l}_p\| \\ &= \frac{2}{\sqrt{a}}\|\boldsymbol{l}_p - P_{\mathcal{H}_S}\boldsymbol{l}_p\| \le \frac{2}{\sqrt{a}}\|\boldsymbol{l}_p - \boldsymbol{v}_p\|.\end{aligned}$$

Using Parseval's theorem and (87), we obtain the following estimate for α_p:

$$\begin{aligned}\alpha_p &\le \frac{2}{\sqrt{a}}\left(\int_{\mathbb{R}} \|\boldsymbol{L}_p(f) - \boldsymbol{V}_p(f)\|^2 df\right)^{\frac{1}{2}} \\ &= \frac{2}{\sqrt{a}}\left(\sum_{r=1}^{R}\int_{\mathcal{G}'_r\cap\mathcal{G}^c_r} \|\boldsymbol{L}_p(f)\|^2 df\right)^{\frac{1}{2}}.\end{aligned}$$

The last expression can be made arbitrarily small for sufficiently small δ, because each $\boldsymbol{L}_p(f)$ is square-integrable, and $\mu(\mathcal{G}'_r \cap \mathcal{G}^c_r) \le \delta/R^2$. Hence for any $\epsilon > 0$ and sufficiently small $\delta > 0$, we can guarantee that $\alpha_p \le \epsilon$. Applying Theorem 1, we obtain $D^+(\Lambda_p) < \infty$ for $p \in \Pi$ and

$$D^+(\Sigma_1, \ldots, \Sigma_Q) \ge D^+(\{\Lambda_p : p \in \Pi\}) - \sum_{p\in\Pi} \alpha_p D^+(\Lambda_p). \tag{88}$$

Using the estimate for α_p and the fact that Σ_q has a uniform density of $\mu(\mathcal{I}_{rk})$ in (88), we obtain

$$\begin{aligned}D^+(\{\Lambda_p : p \in \Pi\}) &\le \sum_{r,n,k} \mu(\mathcal{I}_{rk}) + \epsilon \sum_{p\in\Pi} D^+(\Lambda_p) \\ &= \sum_{r=1}^{R} r\mu(\mathcal{G}'_r) + \epsilon \sum_{p\in\Pi} D^+(\Lambda_p) \\ &\le \sum_{r=1}^{R} r\mu(\mathcal{G}_r) + \delta + \epsilon \sum_{p\in\Pi} D^+(\Lambda_p).\end{aligned} \tag{89}$$

Using (75), (76), and the definition of the Lebesgue integral, we have

$$\sum_{r=1}^{R} r\mu(\mathcal{G}_r) = \int_{\mathbb{R}} \left[\rho(f) - \chi(f \in \mathcal{D}_0)\right] df = \int_{\mathbb{R}} \operatorname{rank}\left(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) df - \mu(\mathcal{D}_0). \tag{90}$$

Combining (89) and (90), and letting $\delta, \epsilon \to 0$, we obtain

$$D^+(\{\Lambda_p : p \in \Pi\}) \le \int_{\mathbb{R}} \text{rank}\left(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) df - \mu(\mathcal{D}_0).$$

This proves (70). We have already demonstrated that $D^+(\Lambda_p) < \infty$. Therefore if (71) is satisfied for some $\Pi \neq \emptyset$, we have $\mu(\mathcal{D}_0) > 0$, implying that the inequality in (70) is strict. □

Theorem 4 is the generalization of Landau's necessary density condition for interpolation. In particular, if $\boldsymbol{G}(f)$ satisfies (61), then

$$\text{rank}\left(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) \le |\mathcal{C}_f|.$$

Under this condition, it follows from (70) for $\Pi = \mathcal{P}$ that

$$D^+(\Lambda_1, \ldots, \Lambda_P) \le \sum_{r=1}^{R} \mu(\mathcal{F}_r), \tag{91}$$

which states that the joint density of Λ_p cannot exceed the combined bandwidth of the input signals. Note that (61) need not hold for consistent reconstruction, and we would get a stronger upper bound on the joint density than (91) when (61) does not hold.

Theorem 4 also provides conditions on the joint densities of all sub-collections of $\{\Lambda_p\}$. We have already seen that the quantity

$$\int_{\mathbb{R}} \text{rank}\left(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) df$$

is a measure of "signal information" contained in the samples of $\{y_p(\lambda_{np}) : p \in \Pi\}$. We can think of $y_p(\lambda_{np}) = c_{np}$ as constraints that restrict the freedom of the input $\boldsymbol{x}$. In the consistent reconstruction problem, we interpret the *constraint density* $D^+(\{\Lambda_p : p \in \Pi\})$ as a lower bound of this freedom lost, measured in units of signal information. For the existence of a solution to the consistency problem, we require that the constraint density be no more than the amount of signal information present in the inputs, thereby justifying (70).

The following corollary is an exact dual of Corollary 1.

Corollary 2. *Suppose that the hypotheses of Theorem 4 are satisfied and $\mathcal{F} = \bigcup_{r \in \mathcal{R}} \mathcal{F}_r$, then*

$$\operatorname*{ess\,inf}_{f \in \mathcal{F}} \lambda_{\min}\left(\boldsymbol{G}^H_{\Pi,\mathcal{C}_f}(f) \boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) > 0, \tag{92}$$

for every $\Pi \subseteq \mathcal{P}$, $\Pi \neq \emptyset$ such that

$$D^+(\{\Lambda_p : p \in \Pi\}) = \sum_{r=1}^{R} \mu(\mathcal{F}_r). \tag{93}$$

Proof. From Theorem 4 and (33), we have

$$D^{+}(\{\Lambda_p : p \in \Pi\}) \leq \int_{\mathbb{R}} \operatorname{rank}\left(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) df \leq \int_{\mathbb{R}} |\mathcal{C}_f| = \sum_{r=1}^{R} \mu(\mathcal{F}_r).$$

If (93) holds, then both the above inequalities are, in fact, equalities. Then (70) is satisfied with an equality, implying that (71) fails to hold. Also, $\operatorname{rank}\left(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) = |\mathcal{C}_f|$, implying that

$$\lambda_{\min}\left(\boldsymbol{G}^{H}_{\Pi,\mathcal{C}_f}(f)\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) = \left[\sigma_{\min}\left(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right)\right]^2$$

Now (92) follows by combining the last two observations. □

Corollary 2 can be interpreted as follows. Suppose that the smallest singular value of $\boldsymbol{G}_{\Pi,\mathcal{C}_f}$ takes arbitrarily small values, then there are inputs with unit energy that can produce outputs y_p, $p \in \Pi$ of arbitrarily small energies. Roughly speaking, this means that we can find an l^2 sequences $\{c_{np}\}$ for which the consistency problem does not have a finite-energy solution $\boldsymbol{x} \in \mathcal{H}$ if we are operating at the critical sampling density, i.e., with an equality in (70).

To explore the implications of Theorem 4 further we consider, once again, the case where the sampling sets have uniform densities, as it enables us to reduce the dimension of the space of densities from $\mathbb{R}^{2^P-1}$ to $\mathbb{R}^P$.

Definition 6. The *density region for consistency* is defined as the collection of all $(d_1, \ldots, d_P)$ such that consistent MIMO reconstruction is realizable using sampling sets $\{\Lambda_p\}$ of uniform densities $D(\Lambda_p) = d_p$, $p \in \mathcal{P}$.

Suppose that every Λ_p has uniform density with $D(\Lambda_p) = d_p$, then the consistency conditions in (70) reduce to

$$\sum_{p\in\Pi} d_p \geq \theta_C(\Pi), \quad \forall \Pi \subseteq \mathcal{P}, \tag{94}$$
$$\theta_C(\Pi) = \int_{\mathbb{R}} \operatorname{rank}\left(\boldsymbol{G}_{\Pi,\mathcal{C}_f}(f)\right) df.$$

As in the case of stable MIMO sampling, the above region is an outer bound on the density region for consistency. Next, observe that $\theta_C(\emptyset) = 0$ and $\theta_C(\Pi_1) \leq \theta_C(\Pi_2)$ whenever $\Pi_1 \subseteq \Pi_2$. Consequently, we can use Proposition 1 to show that

$$\theta_C(\Pi_1) + \theta_C(\Pi_2) \geq \theta_C(\Pi_1 \cup \Pi_2) + \theta_C(\Pi_1 \cap \Pi_2), \quad \forall \Pi_1, \Pi_2 \subseteq \mathcal{P}.$$

These properties of $\theta_C(\Pi)$ imply that the system of inequalities in (94) forms a polymatroid [23, 24], implying that every constraint in (94) is active for some point in the region.

We now present at an example to illustrate the results for consistent reconstruction.

Example 2. Let the MIMO channel and the input spectral supports be as defined in Example 1. We seek necessary conditions on the sampling sets Λ_1 and Λ_2 for consistent reconstruction. Fortunately, we have already performed all the necessary calculations in Example 1. Applying Theorem 4, we obtain the following bounds on the joint densities:

$$\begin{aligned} D^+(\Lambda_1, \Lambda_2) &\leq \int_{\mathbb{R}} \operatorname{rank}\big(\boldsymbol{G}_{\bullet,\mathcal{C}_f}(f)\big) df = 4, \\ D^+(\Lambda_1) &< \int_{\mathbb{R}} \operatorname{rank}\big(\boldsymbol{G}_{1,\mathcal{C}_f}(f)\big) df = 2.5, \\ D^+(\Lambda_2) &\leq \int_{\mathbb{R}} \operatorname{rank}\big(\boldsymbol{G}_{2,\mathcal{C}_f}(f)\big) df = 2. \end{aligned}$$

These inequalities can be justified as follows. The combined bandwidth of the inputs is 4, requiring $D^+(\Lambda_1, \Lambda_2) \leq 4$ for consistency. Now, if consistent reconstruction of x_1 and x_2 is possible, then in particular, we must also have consistent reconstruction (or interpolation) of y_1 and y_2 from their respective samples. Looking at Figure 3, we see that $Y_2(f) = X_2$ is bandlimited to $[0, 2)$. Therefore, by Landau's interpolation density result, we require $D^+(\Lambda_2) \leq 2$ for consistent reconstruction of y_2. Finally, $Y_1(f)$ is bandlimited to $[-1, 2.5)$, thereby requiring $D^+(\Lambda_1) \leq 2.5$. However, $D^+(\Lambda_1) = 2.5$ is not allowed because $K(f)$ is arbitrarily small in the vicinity of $f = 1.5$. We now show that the point $(d_1, d_2) = (2, 2)$ is achievable, where $d_p = D(\Lambda_p)$ for sampling sets $\{\Lambda_p\}$ of uniform density. Let Λ_1 and Λ_2 be defined as in Example 1. Let $\{c_{n1}\}$ and $\{c_{n2}\}$ be l^2 sequences. Then the problem $y_2(\lambda_{n2}) = c_{n2}$ clearly has a solution $y_2 \in \mathcal{B}([0, 2])$. Now, the sequence $d_n = k * x_2(\lambda_{n1})$ is square-summable because $K(f)$ is a bounded function, implying that

$$x_1(\lambda_{n1}) = y_1(\lambda_{n1}) - k * x_2(\lambda_{n1}) = c_{n1} - d_n,$$

also has a solution $x_1 \in \mathcal{B}([-1, 1])$. This proves that $(d_1, d_2) = (2, 2)$ is achievable.

If Λ_1 and Λ_2 have uniform densities of d_1 and d_2 respectively, the resulting outer bounds on the density regions for stable sampling (Example 1) and consistent reconstruction (Example 2) can be viewed as sets in $\mathbb{R}^2$. These regions are illustrated in Figure 4.

Remark. Recall that a collection of sets that is both a collection of stable sampling and consistent reconstruction provides a Riesz basis for the space $\mathcal{H}$. For instance, such collections would have to lie on the line segment that adjoins the regions in Figure 4 for the MIMO channel considered in Examples 1 and 2, provided that the sampling sets have uniform densities. Of course, each Λ_p need not be uniformly dense,

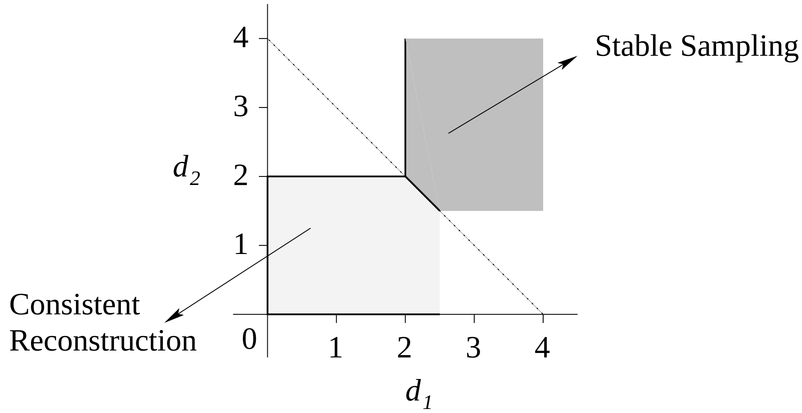

Figure 4: Density regions for stable sampling and consistent reconstruction.

however, we evidently require that

$$D^{+}(\Lambda_1, \ldots, \Lambda_P) = D^{-}(\Lambda_1, \ldots, \Lambda_P).$$

Unfortunately, the existence of collections is difficult to prove (or disprove) for a given $\mathcal{H}$. In fact, the simpler problem of finding exponential Riesz bases for $\mathcal{B}(\mathcal{F})$ is still unsolved except for special sets. For instance, the simplest case $\mathcal{F} = [a, b]$ is very well studied, and Riesz bases are easy to construct for $\mathcal{B}(\mathcal{F})$. The problem is also solved when $\mathcal{F}$ is a finite union of intervals of whose lengths are commensurate, or an arbitrary union of two intervals [25–27]. However, even the case where $\mathcal{F}$ is an arbitrary finite union of intervals is still an open problem.

IV. Summary

In this paper, we formulated the MIMO sampling scheme, and defined the notions of stable MIMO sampling and consistent MIMO reconstruction. These notions are generalizations of the definitions of stable sampling and interpolation for classical sampling. We also generalized the definitions of upper and lower sampling densities applicable to *collections* of sampling sets.

We derived necessary density conditions for stable sampling and consistent reconstruction in the MIMO setting. For the stability of MIMO sampling, we find that a family of $2^P - 1$ bounds hold—a lower bound on the joint lower density of each nonempty set of output sampling sets. Similarly we find that a family of $2^P - 1$ bounds hold for the consistency problem. This time, they are upper bounds on the joint upper densities of the sampling sets. These bounds generalize density results that Landau derived for the classical sampling problem, and like Landau's results, they are fundamental bounds. Since the MIMO sampling scheme is extremely general, and encompasses various sampling schemes such as Papoulis' generalized sampling, and multicoset or periodic nonuniform sampling as special cases, we have automatically generated necessary conditions for all these sampling schemes. Finally, we point out that our results, being necessary conditions, are also applicable to the harder problem of blind channel equalization.

APPENDIX

Proof of Proposition 1

Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be the null spaces of $\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B},\bullet}$ and $\boldsymbol{G}_{\mathcal{B},\bullet}$ respectively. Then obviously $\mathcal{S}_2 \subseteq \mathcal{S}_1$ implying that $\boldsymbol{P}_{\mathcal{S}_2} = \boldsymbol{P}_{\mathcal{S}_1}\boldsymbol{P}_{\mathcal{S}_2}$, where $\boldsymbol{P}_{\mathcal{S}_2}$ and $\boldsymbol{P}_{\mathcal{S}_2}$ are the orthogonal projection operator onto the spaces $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively. Then we have

$$\begin{aligned}\operatorname{rank}(\boldsymbol{G}_{\mathcal{A},\bullet}) &= \operatorname{rank}\begin{pmatrix}\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\\ \boldsymbol{G}_{\mathcal{A}\cap\mathcal{B},\bullet}\end{pmatrix} = \operatorname{rank}\begin{pmatrix}\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\boldsymbol{P}_{\mathcal{S}_1}\\ \boldsymbol{G}_{\mathcal{A}\cap\mathcal{B},\bullet}\end{pmatrix}\\ &= \operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\boldsymbol{P}_{\mathcal{S}_1}) + \operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B},\bullet}),\end{aligned}$$

where the last step follows because every row of $\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\boldsymbol{P}_{\mathcal{S}_1}$ is orthogonal to every row of $\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B},\bullet}$. Hence, we have

$$\operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\boldsymbol{P}_{\mathcal{S}_1}) = \operatorname{rank}(\boldsymbol{G}_{\mathcal{A},\bullet}) - \operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B},\bullet}). \tag{.1}$$

Similarly, we can show that

$$\operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\boldsymbol{P}_{\mathcal{S}_2}) = \operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cup\mathcal{B},\bullet}) - \operatorname{rank}(\boldsymbol{G}_{\mathcal{B},\bullet}). \tag{.2}$$

Finally, using $\boldsymbol{P}_{\mathcal{S}_2} = \boldsymbol{P}_{\mathcal{S}_1}\boldsymbol{P}_{\mathcal{S}_2}$, we obtain

$$\begin{aligned}\operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\boldsymbol{P}_{\mathcal{S}_2}) &= \operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\boldsymbol{P}_{\mathcal{S}_1}\boldsymbol{P}_{\mathcal{S}_2})\\ &\le \operatorname{rank}(\boldsymbol{G}_{\mathcal{A}\cap\mathcal{B}^c,\bullet}\boldsymbol{P}_{\mathcal{S}_1}). \end{aligned} \tag{.3}$$

Combining (.1), (.2), and (.3), we obtain (4). □

REFERENCES

[1] B. R. Petersen and D. D. Falconer, "Supression of adjacent-channel, co-channel, and intersimbol interference by equalizers and linear combiners," *IEEE Trans. Comm.*, vol. 42, pp. 3109–3118, December 1994.

[2] J. Yang and S. Roy, “On joint receiver and transmitted optimization for multiple-input-multiple-output (MIMO) transmission systems,” *IEEE Trans. Comm.*, vol. 42, pp. 3221–3231, December 1994.

[3] G. G. Raleigh and J. M. Cioffi, “Spatio-temporal coding for wireless communication,” *IEEE Trans. Comm.*, vol. 46, no. 3, pp. 357–366, March 1998.

[4] L. Ye and K. J. R. Liu, “Adaptive blind source separation and equalization for multiple-input/multiple-output systems,” *IEEE Trans. Info. Theory*, vol. 44, no. 7, pp. 2864–2876, November 1998.

[5] W. Zishun and J. D. Z. chen, “Blind separation of slow waves and spikes from gastrointestinal myo-electrical recordings,” *IEEE Trans. Info. Tech. Biomed.*, vol. 5, no. 2, pp. 133–137, June 2001.

[6] V. Zarzoso and A. K. Nandi, “Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation,” *IEEE Trans. Biomed. Eng.*, vol. 48, no. 1, pp. 12–18, January 2001.

[7] P. a. Voois and J. M. Cioffi, “Multichannel signal processing for multiple-head digital magnetic recording,” *IEEE Trans. Magnetics*, vol. 30, no. 6, pp. 5100–5114, November 1994.

[8] K.-C. Yen and Y. Zhao, “Adaptive co-channel speech separation and recognition,” *IEEE Trans. Speech Audio Process.*, vol. 7, no. 2, pp. 138–151, March 1999.

[9] A. González and J. J. Lopéz, “Fast transversal filters for deconvolution in multichannel sound reproduction,” *IEEE Trans. Speech Audio Process.*, vol. 9, no. 4, pp. 429–440, May 2001.

[10] J. Idier and Y. Goussard, “Multichannel seismic deconvolution,” *IEEE Trans. Geoscience Remote Sensing*, vol. 31, no. 5, pp. 961–979, September 1993.

[11] G. Harikumar and Y. Bresler, “Exact image deconvolution from multiple FIR blurs,” *IEEE Trans. Image Process.*, vol. 8, no. 6, pp. 846–862, June 1999.

[12] G. B. Giannakis and R. W. Heath Jr., “Blind identification of multichannel FIR blurs and perfect image restoration,” *IEEE Trans. Image Process.*, vol. 9, no. 11, pp. 1877–1896, November 2000.

[13] R. J. Papoulis, “Generalized sampling expansions,” *IEEE Trans. Circuits Syst.*, vol. CAS-24, pp. 652–654, November 1977.

[14] D. Seidner and M. Feder, “Vector sampling expansions,” *IEEE Trans. Sig. Process.*, vol. 48, no. 5, pp. 1401–1416, May 2000.

[15] M. Unser and J. Zerubia, “Generalized sampling: Stability and analysis,” *IEEE Trans. Sig. Process.*, vol. 45, no. 12, pp. 2941–2950, December 1997.

[16] H. Landau, "Necessary density conditions for sampling and interpolation of certain entire functions," *Acta Math.*, vol. 117, pp. 37–52, 1967.

[17] H. J. Landau, "Sampling, data transmission, and the Nyquist rate," *Proc. IEEE*, vol. 55, pp. 1701–1706, October 1967.

[18] K. Gröchenig and H. Razafinjatovo, "On Landau's necessary density conditions for sampling and interpolation of bandlimited functions," *J. London Math. Soc.*, vol. 2, no. 54, pp. 557–565, 1996.

[19] R. M. Young, *An Introduction to Nonharmonic Fourier Analysis*, Academic Press, New York, 2001.

[20] J. R. Higgins, *Sampling Theory in Fourier and Signals Analysis Foundations*, Oxford Science Pub., New York, 1996.

[21] J. Benedetto and P. J. S. G. Ferreira, Eds., *Modern Sampling Theory: Mathematics and Applications*, Birkhäuser, Boston, 2001.

[22] R. Venkataramani and Y. Bresler, "Necessary density conditions for generalized sampling in certain signal subspaces of $L^2(\mathbf{R})$," *J. Fourier Anal.*, submitted.

[23] J. Edmonds, "Submodular functions, matroids and certain polyhedra," in *Proc. Calgary Int. Conf. Combinatorial Structures and Applications*, Calgary, Alberta, June 1969, pp. 69–87.

[24] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels–Part I: Ploymatroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Info. Theory*, vol. 44, no. 7, pp. 2796–2815, November 1998.

[25] Y. I. Lyubarskii and K. Seip, "Sampling and interpolating sequences for multiband-limited functions and exponential bases on disconnected sets," *J. Fourier Anal.*, vol. 3, no. 5, pp. 597–615, September 1997.

[26] S. Avdonin and W. Moran, "Sampling and interpolation of functions with multiband spectra and controllability problems," in *Optimal Control of Partial Differential Equations (Chemnitz, 1998)*, vol. 133 of *International Series of Numerical Mathematics*, pp. 43–51. Birkhäuser, Basel, 1999.

[27] V. E. Katnelson, "Sampling and interpolation for functions with multi-band spectrum: the mean periodic continuation method," in *Signal and Image Reproduction in Combined Spaces*, vol. 7 of *Wavelet analysis and Applications*, pp. 525–553. Academic Press, San Diego, CA, 1998.