Sampling a Noisy Multiple Output Channel to Maximize the Capacity

Gaston Solodky
School of Electrical Engineering
Tel Aviv University
Email: gasrp1@gmail.com

Meir Feder
School of Electrical Engineering
Tel Aviv University
Email: meir@eng.tau.ac.il

***Abstract*—This paper deals with an extension of Papoulis' generalized sampling expansion (GSE) to a case where noise is added before sampling and the total sampling rate may be higher than the Nyquist rate. We look for the best sampling scheme that maximizes the capacity of the sampled channel between the input signal and the *M* sampled outputs signals, where the channels are composed of all-pass linear time-invariant (LTI) systems with additive Gaussian white noise. For the case where the total rate is between *M-1* and *M* times the Nyquist rate, the optimal scheme samples *M-1* outputs at Nyquist rate and the last output at the remaining rate. When *M = 2* the optimal performance can also be attained by an equally sampled scheme under some condition on the LTI systems. Surprisingly, equal sampling is suboptimal in general. Nevertheless, for some total sampling rates where there is an integer relation between the number of channels and the total rate, a uniform sampling achieves the optimal performance. Finally, we discuss the relation between maximizing the capacity and minimizing the mean-square error.**

I. Introduction

In [1], [2], Papoulis introduced the generalized sampling expansion (GSE) showing that a band-limited signal $x(t)$ of finite power that passes through M linear time-invariant (LTI) systems and generating responses $\{g_k(t)\}_{k=1}^{M}$, can uniquely be reconstructed, under some conditions on the M LTI systems, from samples of the output signals $g_k(nT)$, at $1/M$ the Nyquist rate. The vector sampling expansion (VSE), introduced in [3], [4], extends the GSE to multi-input-multi-output (MIMO) LTI systems where L input signals generate M output signals, and if M/L is integer the input signals can be reconstructed from samples of the output at L/M the Nyquist rate. The GSE is a special case of the VSE where $L = 1$.

This work provides another extension to the GSE where noise is added after the LTI systems.The total sampling rate can be higher than the Nyquist rate (to combat the noise), and each output signal may be sampled at a different rate. A similar extension of the VSE is left for further work.

The criterion we suggest for choosing the best sampling scheme for a given M LTI system and a total sampling rate f is the maximal capacity representing the maximal information rate that can be achieved over the resulted sampled channel. For this, we assume that the additive noise is white and Gaussian. As will be seen, sampling the output signals will generate a non-white aliased noise; nevertheless, at high SNR, we came up with an explicit formula for the capacity of the sampled channel which is valid for any sampling scheme. This formula is utilized to establish the optimal sampling scheme.

The paper is organized as follows: In Section II, we present a time and frequency domain settings. In Section III, we develop the capacity equations. In Section IV, we provide the structure of the matrix whose determinant defines the capacity, and provide an upper-bound for that determinant. In Section V, we give a detailed analysis for the case of M outputs and show the sampling scheme that maximizes the capacity. In Section VI, we discuss the uniform sampling scheme. Finally, we provide conclusions and further work in Section VII.

II. The Model

In our model, we have a band-limited signal $x(t)$ with bandwidth B, with finite power.

$$x(t) = \frac{1}{2\pi}\int_{-B/2}^{B/2} X(\omega)\, e^{j\omega t}\, d\omega,$$
$$\mathbb{E}\left[\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} x^2(t)dt\right] \le P_x < \infty \tag{1}$$

where $\mathbb{E}[\cdot]$ is the expectation and P_x is the signal power.

A. Time Domain Model

The input signal first passes through M LTI systems with impulse response $\{h_m(t)\}_{m=1}^{M}$, which are perfectly known. Then i.i.d Additive-White-Gaussian-Noise (AWGN) is added and finally the M output signals are sampled at rate $1/T_m$, generating the output signals $y_m(nT_m)$, as can be seen in Fig. 1.

$$\boldsymbol{y}(t) = \begin{cases} y_1(t) = (x * h_1)(t) + n_1(t) \\ \vdots \\ y_M(t) = (x * h_M)(t) + n_M(t) \end{cases} \tag{2}$$

where $*$ is the convolution operator and $\{n_m(t)\}_{m=1}^{M}$ are band-limited white Gaussian with power spectrum $N_m(\omega) = N_0$, where $\omega \in [-B, B]$.

Rate assumptions: We assume that each of the M output signals can be sampled up to the Nyquist rate and that the total sampling rate f is between the Nyquist rate and M times the Nyquist rate:

$$f_{Nyq} \le f = \sum_{m=1}^{M} f_m \le M f_{Nyq}. \tag{3}$$

ISBN 978-90-827970-1-5 © EURASIP 2018

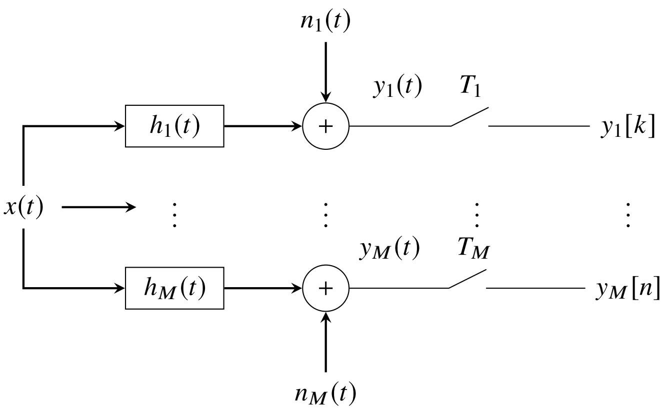

Figure 1: Sampling Scheme

If the total rate equals the Nyquist rate we have a similar model to Papoulis' GSE model, except for the noise. If the total rate is M times the Nyquist rate we sample each output signal at the Nyquist rate. In this paper, we will use normalized sampling rates:

$$r = f/f_{Nyq},\ r_m = f_m/f_{Nyq} \Rightarrow 1 \le r = \sum_{m=1}^{M} r_m \le M. \quad (4)$$

B. Frequency Domain Model

After defining the time-domain model, we consider the frequency domain model, using the notations from [5]. Let T be the smallest common denominator of the sampling times, $[T_{Nyq},\ \{T_m\}_{m=1}^M] \in \mathbb{R}$, such that:

$$LT_{Nyq} = T,\ p_m T_m = T,\ L, \{p_m\}_{m=1}^M \in \mathbb{N}. \quad (5)$$

Since $T_{Nyq} = 1$, then $p_m = r_m L$. First, we define the vector-valued function $\boldsymbol{d}(e^{j\omega})$ of length L, whose kth element is given by

$$d_k(e^{j\omega}) = \frac{1}{T} X\left(\frac{\omega}{T} + \frac{2\pi k}{T}\right), \quad 0 \le \omega \le 2\pi \quad (6)$$

where $X(\omega)$ is the discrete-time Fourier transform (DTFT) of $x[n]$. Next, we define $c_m[n]$:

$$c_m[n] = y_m(nT) = (x * h_m)[n] + n_m[n] \quad (7)$$

where $[\cdot]$ represent discrete time. $\boldsymbol{c}(e^{j\omega})$ is the output DTFT vector of length rL of the output signals $\{y_m\}_{m=1}^M$, structured as follows:

$$\boldsymbol{c}(e^{j\omega}) = \begin{bmatrix} \boldsymbol{C}_1(e^{j\omega}) \\ \boldsymbol{C}_2(e^{j\omega}) \\ \vdots \\ \boldsymbol{C}_M(e^{j\omega}) \end{bmatrix},\ \boldsymbol{C}_m(e^{j\omega}) = \begin{bmatrix} C_{m1}(e^{j\omega}) \\ C_{m2}(e^{j\omega}) \\ \vdots \\ C_{mp_m}(e^{j\omega}) \end{bmatrix}. \quad (8)$$

where the DTFT of the samples , $\{C_{mk}(e^{j\omega})\}_{k=1}^{p_m}$ is:

$$\begin{aligned} C_{mk}(e^{j\omega}) &= C_m(e^{j\frac{\omega+2\pi k}{T}}) \\ &\overset{(a)}{=} \frac{1}{T}\sum_{l\in\mathbb{Z}} Y_m\left(\frac{\omega+2\pi k}{T} + \frac{2\pi l}{T}\right) \overset{(b)}{=} \\ &\frac{1}{T}\sum_{l\in\mathbb{Z}} X\left(\frac{\omega}{T} + \frac{2\pi(l+k)}{T}\right) H_m\left(\frac{\omega}{T} + \frac{2\pi(l+k)}{T}\right) \\ &\quad + N_m\left(\frac{\omega}{T} + \frac{2\pi(l+k)}{T}\right) = \\ &\frac{1}{T}\sum_{l\in\mathbb{Z}} X\left(\frac{\omega}{p_m T_m} + \frac{2\pi(l+k)}{p_m T_m}\right) H_m\left(\frac{\omega}{p_m T_m} + \frac{2\pi(l+k)}{p_m T_m}\right) \\ &\quad + N_m\left(\frac{\omega}{p_m T_m} + \frac{2\pi(l+k)}{p_m T_m}\right) \overset{(c)}{=} \\ &\frac{1}{T}\sum_{\frac{(l+k)}{p_m}\in\mathbb{Z}} X\left(\frac{\omega}{L} + \frac{2\pi(l+k)}{L}\right) H_m\left(\frac{\omega}{L} + \frac{2\pi(l+k)}{L}\right) \\ &\quad + N_m\left(\frac{\omega}{L} + \frac{2\pi(l+k)}{L}\right) \overset{(d)}{=} \\ &\frac{1}{T}\sum_{\substack{l=0 \\ \frac{(l+k)}{p_m}\in\mathbb{Z}}}^{L-1} X\left(\frac{\omega}{L} + \frac{2\pi(l+k)}{L}\right) H_m\left(\frac{\omega}{L} + \frac{2\pi(l+k)}{L}\right) \\ &\quad + N_m\left(\frac{\omega}{L} + \frac{2\pi(l+k)}{L}\right) \end{aligned} \quad (9)$$

where,

(a) Following Theorem 3.2 of [5]
(b) Definition of $y_m(t)$
(c) Only the relevant duplications are taken according to the ratio k/p_m
(d) $Y(\omega)$ is band-limited to $2\pi/T_{Nyq}$ and $T = LT_{Nyq}$

and where, $Y_m(\omega)$, $H_m(\omega)$, $N_m(e^{j\omega})$ are the DTFT of $y_m[n]$, $h_m[n]$, $n_m[n]$, respectively.

$\boldsymbol{H}(\omega)$ is the DTFT matrix describing the channel response of size $ML \times L$ with the following elements:

$$\begin{aligned} \boldsymbol{H}(\omega) &= \begin{bmatrix} \boldsymbol{H}_1(\omega) \\ \boldsymbol{H}_2(\omega) \\ \vdots \\ \boldsymbol{H}_M(\omega) \end{bmatrix}, \\ \boldsymbol{H}_m(\omega) &= diag\left\{H_m\left(\frac{\omega}{T} + \frac{2\pi k}{T}\right)\right\}_L, 0 \le k \le\ L-1 \end{aligned} \quad (10)$$

where $diag\{\cdot\}_L$ is a square diagonal matrix of size L. We assume that $\{\boldsymbol{H}_m(\omega)\}_{m=1}^M$ (10), are allpass filters, meaning that $\{\boldsymbol{H}_m(\omega)\}_{m=1}^M$ are unitary matrices:

$$\boldsymbol{H}_m^H(\omega)\boldsymbol{H}_m(\omega) = \boldsymbol{I}_L, \quad \forall\omega. \quad (11)$$

The choice of $\{\boldsymbol{H}_m(\omega)\}_{m=1}^M$ to be allpass filters makes sense, since in our problem we try to determine how to allocate samples between the channels, amplifying one of the channels will give that channel an unfair advantage. $\boldsymbol{A}$ is a matrix of size $rL\times ML$ where $\{\boldsymbol{A}_m\}_{m=1}^M$ are the channel aliasing matrices

ISBN 978-90-827970-1-5 © EURASIP 2018

composed of zeros and ones and depend on $\{r_m\}_{m=1}^M$, of size $r_m L \times L$:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{A}_M \end{bmatrix} \tag{12}$$

$\boldsymbol{N}(e^{j\omega})$ is the DTFT of AWGN vector of length ML distributed $N(0, N_0 \cdot \boldsymbol{I}_{ML})$ and its elements are:

$$\boldsymbol{N}(e^{j\omega}) = \begin{bmatrix} \boldsymbol{N}_1(e^{j\omega}) \\ \boldsymbol{N}_2(e^{j\omega}) \\ \vdots \\ \boldsymbol{N}_M(e^{j\omega}) \end{bmatrix}, \ \boldsymbol{N}_m(e^{j\omega}) = \begin{bmatrix} N_{m1}(e^{j\omega}) \\ N_{m2}(e^{j\omega}) \\ \vdots \\ N_{mL}(e^{j\omega}) \end{bmatrix}. \tag{13}$$

Finally, the DTFT equation model is:

$$\boldsymbol{c}(e^{j\omega}) = \boldsymbol{A}\boldsymbol{H}(\omega)\boldsymbol{d}(e^{j\omega}) + \boldsymbol{A}\boldsymbol{N}(e^{j\omega}) = \boldsymbol{G}_\omega \boldsymbol{d}(e^{j\omega}) + \boldsymbol{N}_c(e^{j\omega}) \tag{14}$$

where, $\boldsymbol{G}_\omega = \boldsymbol{G}(\omega)$ is the channel matrix.

III. Capacity Calculation - Water-filling Formula

Next, our goal is to find the sampling scheme that achieves maximum capacity. The capacity of a general vector channel is given by Telatar [6], using the water-filling equations [7]. In order to use the water-filling formula in [6], the model noise should be AWGN. Thus we need to whiten the noise in (14).

A. Noise Whitening

First let's calculate the $rL \times rL$ covariance matrix of $\boldsymbol{N}_c(\omega)$:

$$\boldsymbol{\sigma}_c^2 = \mathbb{E}[\boldsymbol{N}_c(\omega)\boldsymbol{N}_c^H(\omega)] = \boldsymbol{A} \cdot \mathbb{E}[\boldsymbol{N}_\omega \boldsymbol{N}_\omega^H] \cdot \boldsymbol{A}^H = N_0 \boldsymbol{A}\boldsymbol{A}^H \tag{15}$$

where $(\cdot)^H$ is the conjugate transpose operator. In order to whiten the noise, we will multiply the noise with the whitening matrix $(N_0 \boldsymbol{A}\boldsymbol{A}^H)^{-0.5}$.

$$\begin{aligned} &\mathbb{E}[(N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5}\boldsymbol{N}_c(\omega)] = (N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5}\mathbb{E}[\boldsymbol{N}_c(\omega)] = 0, \\ &\mathbb{E}[(N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5}\boldsymbol{N}_c(\omega)\boldsymbol{N}_c(\omega)^H(N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5H}] \\ &= (N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5}N_0\boldsymbol{A}\boldsymbol{A}^H(N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5} \\ &= (\boldsymbol{A}\boldsymbol{A}^H)^{-0.5}(\boldsymbol{A}\boldsymbol{A}^H)^{0.5}(\boldsymbol{A}\boldsymbol{A}^H)^{0.5}(\boldsymbol{A}\boldsymbol{A}^H)^{-0.5} = \boldsymbol{I}_{rL} \end{aligned} \tag{16}$$

Next, the model (14) is multiplied with the whitening matrix to get a new model with white noise:

$$\tilde{\boldsymbol{c}}(e^{j\omega}) = \tilde{\boldsymbol{G}}_\omega \boldsymbol{d}(e^{j\omega}) + \tilde{\boldsymbol{N}}(e^{j\omega}) \tag{17}$$

where, $\tilde{\boldsymbol{c}}(e^{j\omega}) = (N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5}\boldsymbol{c}(e^{j\omega})$,
$\tilde{\boldsymbol{G}}_\omega = (N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5}\boldsymbol{A}\boldsymbol{H}(\omega)$, $\tilde{\boldsymbol{N}}(e^{j\omega}) = (N_0\boldsymbol{A}\boldsymbol{A}^H)^{-0.5}\boldsymbol{A}\boldsymbol{N}(e^{j\omega})$.

B. Water-filling formula

The water-filling equations are:

$$\begin{aligned} P(\mu) &= \frac{1}{2\pi}\int_0^{2\pi}\sum_l \left(\mu - \lambda_{l\omega}^{-1}\right)^+ d\omega, \\ C(\mu) &= \frac{1}{2\pi}\int_0^{2\pi}\sum_l \left(\ln\left(\mu\lambda_{l\omega}\right)\right)^+ d\omega \end{aligned} \tag{18}$$

where μ is the water-filling level such that $P(\mu) = P_x$, $(g)^+ = \max(g, 0)$, and $\lambda_l^{-0.5}$ are the eigenvalues of matrix $\tilde{\boldsymbol{G}}_\omega$. In our problem we normalize the water-filling equations with L (5):

$$\begin{aligned} P(\mu) &= \frac{1}{2\pi L}\int_0^{2\pi}\sum_l \left(\mu - \lambda_{l\omega}^{-1}\right)^+ d\omega, \\ C(\mu) &= \frac{1}{2\pi L}\int_0^{2\pi}\sum_l \left(\ln\left(\mu\lambda_{l\omega}\right)\right)^+ d\omega. \end{aligned} \tag{19}$$

We use the normalized equations since for different sampling scheme, L is different. Then the spectrum is divided into different parts number without increasing the spectrum support. While the total capacity does not increase, it is divided through different spectrum parts number. To simplify the calculations, we will find the eigenvalues of matrix $\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega = N_0^{-1}\boldsymbol{H}^H(\omega)\boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{A}^H)^{-1}\boldsymbol{A}\boldsymbol{H}(\omega)$. Our solution is for high Signal to Noise Ratio (SNR) cases. In that case, μ is large enough (larger than $\lambda_{l\omega}^{-1}$), so the water-filling formulas can be written as:

$$P(\mu) = \int_0^{2\pi}\frac{\sum_l(\mu - \lambda_{l\omega}^{-1})^+}{2\pi L}d\omega \approx \int_0^{2\pi}\frac{\sum_l(\mu - \lambda_{l\omega}^{-1})}{2\pi L}d\omega =>$$

$$\mu = P + \frac{1}{2\pi L}\int_0^{2\pi} tr\left((\tilde{\boldsymbol{G}}_\omega^H\tilde{\boldsymbol{G}}_\omega)^{-1}\right)d\omega \tag{20}$$

$$C(\mu) = \int_0^{2\pi}\frac{\sum_l\left(\ln\left(\mu\lambda_{l\omega}\right)\right)^+}{2\pi L}d\omega \approx \int_0^{2\pi}\frac{\sum_l\left(\ln\left(\mu\lambda_{l\omega}\right)\right)}{2\pi L}d\omega =>$$

$$C(\mu) \approx K + \frac{1}{2\pi L}\int_0^{2\pi}\ln\left(\det\left(\tilde{\boldsymbol{G}}_\omega^H\tilde{\boldsymbol{G}}_\omega\right)\right)d\omega \tag{21}$$

where $K = \ln(\mu)$ is a constant, $tr(\cdot)$ is the trace function, and the approximation is due to high SNR. High SNR causes the water-level to be high, then all the eigenvalues are considered and the water-level is constant. Thus, to maximize the capacity at high SNR, we need the maximize $\tilde{\boldsymbol{G}}_\omega^H\tilde{\boldsymbol{G}}_\omega$ determinant.

IV. $\tilde{\boldsymbol{G}}_\omega^H\tilde{\boldsymbol{G}}_\omega$ Properties

Equations (10) and (12) provide the components of $\tilde{\boldsymbol{G}}_\omega^H\tilde{\boldsymbol{G}}_\omega$ matrix. With direct calculation we get that:

$$\tilde{\boldsymbol{G}}_\omega^H\tilde{\boldsymbol{G}}_\omega = N_0^{-1}\sum_{m=1}^M \boldsymbol{H}_m^H(\omega)\boldsymbol{A}_m^H(\boldsymbol{A}_m\boldsymbol{A}_m^H)^{-1}\boldsymbol{A}_m\boldsymbol{H}_m(\omega) \tag{22}$$

It is easy to see that $\{\boldsymbol{H}_m^H(\omega)\boldsymbol{A}_m^H(\boldsymbol{A}_m\boldsymbol{A}_m^H)^{-1}\boldsymbol{A}_m\boldsymbol{H}_m(\omega)\}_{m=1}^M$ are projection matrices, using (11) assumption. Thus $\tilde{\boldsymbol{G}}_\omega^H\tilde{\boldsymbol{G}}_\omega$ is a summation of projection matrices, and so its normalized trace is constant:

$$\begin{aligned} \frac{tr(\tilde{\boldsymbol{G}}_\omega^H\tilde{\boldsymbol{G}}_\omega)}{L} &= \sum_{m=1}^M \frac{tr(\boldsymbol{H}_{m\omega}^H\boldsymbol{A}_m^H(\boldsymbol{A}_m\boldsymbol{A}_m^H)^{-1}\boldsymbol{A}_m\boldsymbol{H}_{m\omega})}{LN_0} \\ &= \frac{1}{LN_0}\sum_{m=1}^M \text{rank}(\boldsymbol{A}_m) = \frac{1}{LN_0}\sum_{m=1}^M r_m L = N_0^{-1}r \end{aligned} \tag{23}$$

ISBN 978-90-827970-1-5 © EURASIP 2018

As can be seen from (23), the normalized trace is constant and does not depend on the sampling scheme. Since, $\{\boldsymbol{H}_m(\omega)^H \boldsymbol{A}_m^H (\boldsymbol{A}_m \boldsymbol{A}_m^H)^{-1} \boldsymbol{A}_m \boldsymbol{H}_m(\omega)\}_{m=1}^M$ are projection matrices, each one has $r_m L$ eigenvalues that are equal to 1 and $(1-r_m)L$ that are equal to 0. Therefore, their rank is $r_m L$. The arithmetic-geometric mean inequality and the normalized trace property (23) provide an upper-bound for the determinant and therefore, an upper-bound to the capacity as well:

$$\det\left(\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega\right) \le N_0^{-1} r^L. \tag{24}$$

V. Best Sampling Scheme

Here we present the optimal sampling scheme that provides the maximal capacity for the case when the input signal passes through M LTI systems and the total sampling rate satisfies:

$$M-1 \le r \le M. \tag{25}$$

First, we find the sampling scheme that maximize the upper-bound on the capacity and then we show that they are equal:

$$\begin{aligned} \max_{\sum_{m=1}^M r_m = r} C(\mu) &= \max_{\sum_{m=1}^M r_m = r} \int_0^{2\pi} \frac{\ln\left(\det\left(\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega\right)\right)}{2\pi L} d\omega \\ &\le \frac{1}{2\pi L} \int_0^{2\pi} \max_{\sum_{m=1}^M r_m = r} \ln\left(\det\left(\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega\right)\right) d\omega \end{aligned} \tag{26}$$

If $\max_{\sum_{m=1}^M r_m = r} \det\left(\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega\right) \neq f(\omega)$, then the inequality becomes equality. As shown in the following theorem and the discussion that follows, for the case of (25), the best sampling scheme, for all-pass $\boldsymbol{H}_m$'s, will be:

$$\{r_m\}_{m=1}^{M-1} = 1, \; r_M = r-(M-1). \tag{27}$$

Theorem 1. *Let $\{\boldsymbol{P}_m\}_{m=1}^M$ be $L\times L$ projection matrices with ranks $r_m L$, $1 \le m \le M$, respectively and where*

$$\begin{gathered} \{r_m\}_{m=1}^M \le 1, \; r = \sum_{m=1}^M r_m, \\ M-1 \le r \le M, \; r_m L \in \mathbb{N} \quad 1 \le m \le M \end{gathered} \tag{28}$$

Let $\boldsymbol{H}$ be a $L\times L$ full rank matrix summation of $\{\boldsymbol{P}_m\}_{m=1}^M$:

$$\boldsymbol{H} = \sum_{m=1}^M \boldsymbol{P}_m. \tag{29}$$

Then $\boldsymbol{H}$ has a least $\left(r-(M-1)\right)L$ eigenvalues with value M.

Proof. Let $\{r_m\}_{m=1}^M$ have restrictions as in (28). Let V be a vector space with dimension L and $\{V_m\}_{m=1}^M$ be subspaces of V. Let $\{\boldsymbol{P}_m\}_{m=1}^M$ be $L\times L$ projection matrices to the subspaces $\{V_m\}_{m=1}^M$ with dimensions $r_m L$ $1 \le m \le M$, respectively.

Each $\boldsymbol{P}_m$ has $r_m L$ eigenvectors that spans the subspace V_m with eigenvalue 1, per projection matrix properties. First we show that the dimension of the subspace $V_1 \cap V_2 \cap \cdots \cap V_M$ is at least $\left(r-(M-1)\right)L$ when $M-1 \le r \le M$. This follows by induction. Proof for the $M=2$ case:

$$\dim(V_1 \cap V_2) = \dim(V_1) + \dim(V_2) - \dim(V_1 \cup V_2) \ge (r-1)L \tag{30}$$

Next assume that statement is correct for $M-1$:

$$\dim(V_1 \cap V_2 \cap \cdots \cap V_{M-1}) \ge \left(r-(M-2)\right)N \tag{31}$$

It follows then:

$$\begin{aligned} \dim(V_1 \cap V_2 \cap \cdots \cap V_M) &= \dim(V_M) \\ &\quad + \dim(V_1 \cap V_2 \cap \cdots \cap V_{M-1}) \\ &\quad - \dim(\{V_1 \cap V_2 \cap \cdots \cap V_{M-1}\} \cup V_M) \\ &\ge r_M L + \left(r - r_M - (M-2)\right)L \\ &\quad - \dim(\{V_1 \cap V_2 \cap \cdots \cap V_{M-1}\} \cup V_M) \\ &\ge \left(r-(M-2)\right)L - L = \left(r-(M-1)\right)L \end{aligned} \tag{32}$$

where the first inequality is due to the induction assumption. That is, the statement is true for M.

Let the set $\{\boldsymbol{v}_i\}_{i=1}^{\left(r-(M-1)\right)L}$ be joint eigenvectors of matrices $\{\boldsymbol{P}_m\}_{m=1}^M$, that spans the subspace $V_1 \cap V_2 \cap \cdots \cap V_M$. Let $\boldsymbol{H}$ be as in (29). Now we can show that $\{\boldsymbol{v}_i\}_{i=1}^{\left(r-(M-1)\right)L}$ are eigenvectors of $\boldsymbol{H}$ as well with the eigenvalue M:

$$\boldsymbol{H}\boldsymbol{v}_i = \sum_{m=1}^M \boldsymbol{P}_m \boldsymbol{v}_i = \sum_{m=1}^M \lambda_m \boldsymbol{v}_i = M\boldsymbol{v}_i = \lambda \boldsymbol{v}_i \tag{33}$$

The eigenvalues of the set $\{\boldsymbol{v}_i\}_{i=1}^{\left(r-(M-1)\right)L}$ are $\{\lambda_m\}_{m=1}^M = 1$. That conclude the proof, that $\boldsymbol{H}$ has at least $\left(r-(M-1)\right)L$ eigenvalues with value M when $M-1 \le r \le M$. □

Using Theorem 1, we know that $\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega$ has at least $\big(r-(M-1)\big)L$ eigenvalues with the value M. The other $(M-r)L$ eigenvalues have a constant summation by the normalized trace property of $\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega$ (23):

$$\frac{tr\left(\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega\right)}{L} = \frac{r}{N_0} = N_0^{-1}\left(\left(r-(M-1)\right)M + (M-r)(M-1)\right) \tag{34}$$

Then, if we can maximize the product of the other $(M-r)L$ eigenvalues, we are maximizing the determinant. The arithmetic-geometric mean inequality,

$$\begin{aligned} M-1 &= \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_{(M-r)L}}{(M-r)L} \\ &\ge \sqrt[(M-r)L]{\lambda_1 \cdot \lambda_2 \cdots \lambda_{(M-r)L}} \end{aligned} \tag{35}$$

with an equality, i.e. maximizes the determinant, when $\lambda_1 = \lambda_2 = \cdots = \lambda_{(M-r)L} = M-1$. This case occurs when we sample as in (27), regardless of the LTI systems. The determinant in this case will be:

$$\begin{aligned} \det\left(\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega\right) &= N_0^{-1} M^{r_M L} (M-1)^{(1-r_M)L} \\ &= N_0^{-1} M^{\left(r-(M-1)\right)L} (M-1)^{(M-r)L} \end{aligned} \tag{36}$$

Since the determinant is not dependent on ω (26) becomes an equality, and this sampling scheme maximizes the capacity. The best sampling scheme reaches the upper-bound (24), when the total normalized sampling rate is $r = \{M-1, \; M\}$.

Unfortunately, we do not know the optimal scheme for the case where the total normalized sampling rate is lower than $M-1$. It turns out that the techniques used to find the optimal

ISBN 978-90-827970-1-5

sampling when the rate is (25), do not apply and it requires to calculate the determinant of a summation of matrices $\boldsymbol{H}_m^H(\omega)\boldsymbol{A}_m^H(\boldsymbol{A}_m\boldsymbol{A}_m^H)^{-1}\boldsymbol{A}_m\boldsymbol{H}_m(\omega)$. Each of those matrices has non-zero elements on some of the matrix diagonals, where the location of those elements and the number of diagonals are determined by the sampling rates. For example, when one of the sampling rates is higher than half of the Nyquist rate, $r_m \geq 1/2$, the matrix has three diagonals with non-zero elements - the main diagonal and two secondary diagonals. The location of the secondary diagonals is very sensitive to a variation in r_m value. Then for every change of r_m different techniques are required to calculate the determinant.

VI. Uniform Sampling Scheme

An interesting sampling scheme that was expected to be the best sampling scheme is the uniform scheme, similar to the GSE and VSE, i.e.:

$$\{r_m\}_{m=1}^{M} = \frac{r}{M}. \tag{37}$$

The maximal determinant for this case was shown in [8]:

$$\max_{r_m=r/M} \det(\tilde{\boldsymbol{G}}_\omega^H \tilde{\boldsymbol{G}}_\omega) = N_0^{-1}\left(\frac{M}{d_f}\right)^L \left(\frac{d_f}{d_f+1}\right)^{(d_f+1)d_p}. \tag{38}$$

where $d_f = \lfloor M/r \rfloor$ and $d_p = (1 - rd_f/M)L$. The maximum is achieved when $\sum_{m=1}^{M} \boldsymbol{H}_m(\omega) = 0$. When the $\boldsymbol{H}_m$'s are pure delay systems, this condition makes the entire sampling uniform, in standard sampling - uniform sampling is known to be the best. When $M = 2$ (36) and (38) are equal, while this scheme provides another sampling scheme that achieves the maximal determinant (for $M = 2$ case), it is very sensitive to changes of $\boldsymbol{H}_m$'s, unlike the best sampling scheme. For the cases where $M > 2$, the maximal capacity that can be achieved in this sampling scheme is suboptimal to (27). When the relation between the number of channels and rate is:

$$\frac{M}{r} \in \mathbb{N} \tag{39}$$

The uniform sampling scheme achieves the upper-bound (24). For those rates, it is the best sampling scheme. Notice that this relation also valid when $r \leq M - 1$. In the previous results, the best sampling scheme was the scheme that maximized the capacity for a given rate and number of channels. This result provides the optimal sampling scheme without the channels number constrain. This is an important result since it allows to construct a sampling scheme that achieves the maximal performance.

VII. Conclusion

In this paper we analyzed the best sampling scheme, in the sense of maximizing the capacity, for the scenario where a band-limited signal passes through M LTI systems and sampled at a constant total rate. In the case where $M - 1 \leq r \leq M$ we found the best sampling scheme in which $M - 1$ outputs are sampled at the Nyquist rate and the last output is sampled at the remaining rate. In addition, we showed that for $M = 2$ and for systems that fulfill $\sum_{m=1}^{2} \boldsymbol{H}_m(\omega) = 0$ condition, there is another optimal solution where the sampling rates are equal. Surprisingly, when $M > 2$ with systems that satisfies similar condition, an equally sampled scheme turned out to be suboptimal. Nevertheless, a sampling scheme in which the output signals are sampled equally, is shown to be optimal for some numbers of output signals and total sampling rates.

The criterion we used in this paper is maximum capacity, with high SNR assumption. Another optional criterion is to minimize the mean-square-error (MSE) of a Least-Square (LS) estimator. We were able to show [8] that for the case where $M - 1 \leq r \leq M$, minimizing this MSE is equivalent to maximizing the capacity. The proof is omitted here for lack of space.

Interestingly, although this problem seems simple, it turned out to be quite challenging. There are still quite a few open challenges in this problem. To mention a few:

- What is the optimal scheme when the total sampling rate is $1 \leq r \leq M - 1$ and $M > 2$?
- Extending the results to the Vector Sampling Expansion case (VSE), i.e., when there are $L > 1$ input band-limited signals.
- Relating the capacity criterion to the MMSE criterion in general. Also, finding solution for any SNR.
- What happens when the noise is not Gaussian?

These questions and some more are left for further research.

VIII. Acknowledgment

A special acknowledgment is due to Prof. Yonina Eldar. A discussion with her raised the problem. In addition, she referred us to the frequency domain model and provided insights and comments.

We also with to thank Lital Yodla, whose thesis preceded this work and Prof. Arie Yeredor for helpful discussions.

References

[1] A. Papoulis, "Generalized sampling expansion," *IEEE transactions on circuits and systems*, vol. 24, no. 11, pp. 652–654, 1977.

[2] ——, *Signal analysis*. McGraw-Hill New York, 1977, vol. 191.

[3] D. Seidner, M. Feder, D. Cubanski, and S. Blackstock, "Introduction to vector sampling expansion," *IEEE Signal Processing Letters*, vol. 5, no. 5, pp. 115–117, 1998.

[4] D. Seidner and M. Feder, "Vector sampling expansion," *IEEE Transactions on Signal Processing*, vol. 48, no. 5, pp. 1401–1416, 2000.

[5] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge University Press, 2015.

[6] E. Telatar, "Capacity of multi-antenna gaussian channels," *Transactions on Emerging Telecommunications Technologies*, vol. 10, no. 6, pp. 585–595, 1999.

[7] R. G. Gallager, *Information theory and reliable communication*. Springer, 1968, vol. 2.

[8] G. Solodky and M. Feder, "Sampling a noisy multiple output channel to maximize the capacity," Master's thesis, Tel Aviv University, 2018.

ISBN 978-90-827970-1-5 © EURASIP 2018