Sampling and Distortion Tradeoffs for Indirect Source Retrieval

Elaheh Mohammadi, Alireza Fallah and Farokh Marvasti
Advanced Communications Research Institute (ACRI)
Department of Electrical Engineering
Sharif University of Technology

Abstract

In this paper, we study the problem of indirect or remote reconstruction of a continuous signal that cannot be observed directly. Instead, one has access to multiple corrupted versions of the signal. The available corrupted signals are correlated because they carry information about the common remote signal. Assuming a total budget on the number of samples that we can obtain from the corrupted signals, we consider the problem of finding the optimal number of samples and their locations for each signal to minimize the total reconstruction distortion of the remote signal. The correlation among the corrupted signals can be utilized to reduce the sampling rate. For a class of Gaussian signals, we show that in the low sampling rate regime, it is optimal to use a certain nonuniform sampling scheme on all the signals. On the other hand, in the high sampling rate regime, it is optimal to uniformly sample all the signals. We also show that both of these sampling strategies are optimal if we are interested in recovering the set of corrupted signals, rather than the remote signal.

1 Introduction

In many applications, such as monitoring or sensing systems, one may be interested in reconstructing a source $S(t)$ that is not directly observable. Instead, access is provided to $k$ correlated stochastic signals $S_1(t), S_2(t), \ldots, S_k(t)$ that are corrupted versions of $S(t)$. The goal is to reconstruct $S(t)$ from limited information that one can obtain from $S_i(t), i = 1, 2, \ldots, k$. While the source coding aspect problem is classical in information theory (see for instance [1, Sec. 3.5]), its signal processing and sampling aspect has not received much attention. In this paper, we study the sampling aspect of this problem for stochastic signals $S(t)$ and the $k$ corrupted versions $S_i(t)$. We take into account the fact that the correlation among the signals $S_i(t), i = 1, 2, \ldots, k$, can help decrease the sampling rate or improve the signal reconstruction accuracy.

In this paper, we assume that the signal $S(t)$ is defined on a finite interval $t \in [0, T]$. We cannot observe $S(t)$ directly. Instead, we have $S_1(t), S_2(t), \ldots, S_k(t)$, also defined on $t \in [0, T]$, that are corrupted versions of $S(t)$. We are allowed to take $m_i$ arbitrary noisy samples from the $i$th corrupted signal $S_i(\cdot)$ at time instances $t_{i1}, t_{i2}, \ldots, t_{imi} \in [0, T]$, for $i = 1, 2, \ldots, k$. The sampling noise of each signal $S_i(t)$ is assumed to be an independent zero-mean Gaussian random variable with variance $\sigma_i^2$. The goal is either to reconstruct the signal $S(t)$ or the collection of corrupted signals $S_1(t), S_2(t), \ldots, S_k(t)$ with minimum average distortion by optimizing over the sampling times $t_{ij}$. A further question is that if the total number of samples $\sum_i m_i$ is fixed and equal to $m$, what is the best allocation of $m_i$ samples to minimize the total distortion.

It is known that any deterministic continuous function $s(t)$ defined on the interval $[0, T]$ can be expressed in terms of sinusoids as follows:

$$s(t) = \sum_{\ell=0}^{\infty} [a_\ell \cos(\ell \omega t) + b_\ell \sin(\ell \omega t)], \quad t \in [0, T],$$

where $\omega = 2\pi/T$. If the number of non-zero coefficients $a_\ell$ and $b_\ell$ are limited, the signal $s(t)$ is sparse in the frequency domain. Herein, we are concerned with signals where their coefficients $a_\ell$ and $b_\ell$ are
zero when \( \ell > N_2 \) or \( \ell < N_1 \) for some natural numbers \( N_1 \leq N_2 \), i.e., bandpass signals. In particular, we assume that

\[
S(t) = \sum_{\ell=N_1}^{N_2} [A_{\ell} \cos(\ell \omega t) + B_{\ell} \sin(\ell \omega t)], \quad t \in [0, T],
\]

and

\[
S_i(t) = \sum_{\ell=N_1}^{N_2} [A_{i\ell} \cos(\ell \omega t) + B_{i\ell} \sin(\ell \omega t)], \quad t \in [0, T], \quad i \in \{1, 2, \cdots, k\},
\]

for some random coefficients \( A_{i\ell} = A_{\ell} + W_{i\ell} \) and \( B_{i\ell} = B_{\ell} + V_{i\ell} \), where \( W_{i\ell} \) and \( V_{i\ell} \) are independent perturbations that are added to the original signal.

We are interested in minimizing the average Minimum Mean Square Error (MMSE) distortion of reconstructing either the remote signal

\[
D_{a_{\min}} = \min_{\{t_1, t_2, \cdots, t_m\}_i^k} \frac{1}{T} \int_{t=0}^{T} \mathbb{E}\{|\hat{S}(t) - S(t)|^2\} dt,
\]

or the \( k \) corrupted signals \( S_i(t), i = 1, 2, \cdots, k \)

\[
D_{b_{\min}} = \min_{\{t_1, t_2, \cdots, t_m\}_i^k} \frac{1}{T} \int_{t=0}^{T} \sum_{i=1}^{k} \mathbb{E}\{|\hat{S}_i(t) - S_i(t)|^2\} dt.
\]

Here \( t_{ij} \) is the \( j \)th sampling time of the \( i \)th signal.

The motivation for reconstructing \( \{S_i(t), i = 1, \cdots, k\} \) is twofold: one is that these individual signals may contain some other information of interest besides \( S(t) \), e.g., the differences \( S(t) - S_i(t) \) might be correlated with some other signal of interest; and secondly, this would parallel the literature on indirect source coding, where the reconstruction distortion of intermediate signals is shown to be related to that of the original signal (the separation theorem \[2 \] \[3 \] \[4 \]). While the separation theorem does not have a complete correspondence here (and as we will see, the optimal allocation of samples can be different in the two scenarios), it is still a natural question to consider.

In general finding the optimal sampling locations to minimize the MMSE distortion depends on the joint distribution of the coefficients \( A_{\ell}, B_{\ell}, V_{i\ell}, W_{i\ell} \) and the distribution of the noises incurred at the time of sampling. Assumption of independent Gaussian distribution would correspond to the setup of the “Gaussian CEO problem” \[3 \] \[6 \]. It makes the problem more tractable analytically because MMSE and LMMSE (linear MMSE) are identical in this case (in fact, instead of any Gaussian assumptions, we can say that we are studying the performance of LMMSE estimators in this paper). Since LMMSE estimator only depends on the second moments, the problem reduces to a linear algebra optimization problem. However, this optimization problem is not easy because the variables we are optimizing over, are sampling locations \( t_{ij} \) that show up as arguments of sine and cosine functions. Sine and cosine functions are nonlinear, albeit structured, functions. The goal would be to exploit their structure to solve the optimization problem. But this is challenging even in the special case of one signal (\( k = 1 \)), as non-uniform sampling may outperform uniform sampling \[7 \]. Similar to the setup of \[6 \], herein we assume that the coefficients \( A_{\ell} \) and \( B_{\ell} \) are \( \mathcal{N}(0, 1) \), i.e., the original signal is white, and \( V_{i\ell} \) and \( W_{i\ell} \) are \( \mathcal{N}(0, \eta) \) for all \( i \) and \( \ell \). This implies that \( A_{i\ell} \) and \( B_{i\ell} \) in \[2 \] are jointly Gaussian random variables. The \( j \)th sample of the \( i \)th signal (\( 1 \leq j \leq m_i, 1 \leq i \leq k \)) is taken at time \( t_{ij} \) with an independent additive Gaussian noise of variance \( \sigma_i^2 \). These are the parameters that we use throughout the paper and are reviewed in Table 1. The goal is to find the optimal \( t_{ij} \) to minimize the reconstruction distortion.

**Related work:** Reconstruction distortion of correlated signals (lossy reconstruction) is a major theme in multi-user information theory for the class of discrete \( i.i.d. \) signals. However, the emphasis in multi-user source coding is generally on the quantization and compression rates of the sources, and not on the sampling rates. The indirect source coding problem was first introduced by Dobrushin and Tsybakov in \[3 \] in the information theory literature. This work and subsequent information theoretic ones deal with discrete sources, by assuming that we have several bandlimited signals \( S_i(t), \)
Notation | Description
--- | ---
T, ω | T is signal period and ω = 2π/T
k | Number of corrupted signals
S(t) and S_i(t) | The original signal and the ith corrupted signal, respectively.
A_ℓ, B_ℓ | Fourier series coefficients of the original signal S(t) \( \ell \in [N_1 : N_2] \), \( A_\ell, B_\ell \sim N(0, 1) \).
A_i\ell, B_i\ell | Fourier series coefficients of the ith corrupted signal \( S_i(t) \) \( \ell \in [N_1 : N_2] \), \( A_i\ell, B_i\ell \sim N(0, 1 + \eta) \).
\( \eta \) | Variance of the perturbation added to \( A_\ell \) to produce \( A_i\ell \) for \( 1 \leq i \leq k \).
N, N_1 and N_2 | The support of the input signal in frequency domain is from \( N_1 \omega \) to \( N_2 \omega \). \( N = N_2 - N_1 + 1 \).
m_i | Number of noisy samples of the ith corrupted signal
\( \sigma_i^2 \) | Variance of the sampling noise of the ith corrupted signal
\( \{t_{i1}, t_{i2}, \cdots, t_{im_i}\} \) | Sampling time instances of the ith corrupted signal

Table 1: Definition of main parameters.

that are sampled at the Nyquist rate with no distortion at the sampling phase. Then, assuming a finite quantization rate for storing the samples, the task is to minimize the total reconstruction distortion (which is only due to quantization). On the other hand, our work in this paper is on indirect source retrieval and not indirect source coding, as we do not study the quantization aspect of the problem. Rather, we focus on the distortion incurred by the sampling rate (which can be below the Nyquist rate), and the additive noise on the samples. We also allow nonuniform sampling to decrease the distortion.

In the following, we review previous works on distributed sampling that illustrate this fact in the context of compressed sensing and wireless sensor networks.

In distributed compressed sensing, the structure of correlation among multiple signals is their joint-sparcity. This problem was studied in [8], where signal recovery algorithms using linear equations obtained by distributed sensors were given. Authors in [9] model the correlation of two signals by assuming that one is related to the other by an unknown sparse filtering operation. The problem of centralized reconstruction of two correlated signals based on their distributed samples is studied, and its similarities with the Slepian-Wolf theorem in information theoretic distributed compression are pointed out. Motivated by an application in array signal processing, the authors in [10] consider signal recovery for a specific type of correlated signals, assuming that the signals lie in an unknown but low dimensional linear subspace.

Spatio-temporal correlation of the distributed signals is a significant feature of wireless sensor networks and can be utilized for sampling and data collection [11]. In [12], the spatio-temporal sampling rate tradeoffs of a sensor network for minimum energy usage is studied. Authors in [13] provide a mathematical model for the spatio-temporal correlation of the signals observed by the sensor nodes. In [14], the spatio-temporal statistics of the distributed signals are used by the Principal Component Analysis (PCA) method to find transformations that sparsify the signal. Compressed sensing is then used for signal recovery. In [15], a compressive wireless sensing is given for signal retrieval at a fusion center from an ensemble of spatially distributed sensor nodes. See also [16] for a distributed algorithm based on sparse random projections for signal recovery in sensor networks.

Finally, we would like to point out that there has also been some previous works on the tradeoffs among sampling rate and distortion of single sources, e.g. see [17] [18] [19] [7]. For instance in [17], the authors consider the problem of combined source coding and sub-Nyquist reconstruction from noisy sub-Nyquist samples. They consider the special case of \( k = 1 \), i.e., when only one noisy version of S(t) is available. Here in our paper, we focus on the sampling aspect of the problem for a distributed setup. We consider the case of arbitrary \( k \) in our model (which is different from the model considered in [17]).
Our contributions:
The informal statement of our results is that whether we want to reconstruct the remote signal, or the collection of corrupted signals, the following holds:

- In the low sampling regime, i.e., \( \sum_{i=1}^{k} m_i \leq N \), it is optimal for \( t_{ij}s \) to be distinct elements of the set \( \{0,T/N,\cdots,(N-1)T/N\} \).
- In the high sampling regime, i.e., \( m_i \geq 2N \) for all \( i \), it is optimal to use uniform sampling for each signal.

Next, suppose that we have a total budget on the number of samples \( m = \sum_{i=1}^{k} m_i \) that we can take from the corrupted signals. Then, (i) for low sampling rates, \( m \leq N \): if we want to reconstruct either the remote signal or the collection of corrupted signals, it is best to take the samples from the signal with the smallest sampling noise, i.e., if \( \sigma_1 \leq \sigma_i \) for all \( i \), it is optimal to choose \( m_1 = m, m_i = 0 \) for \( i > 2 \). (ii) for high sampling rates: for reconstructing the remote signal, taking more samples from the less noisy signals is advantageous, but it is no longer true that we should take as many samples as possible from the less noisy signal if we want to reconstruct the collection of corrupted signals. The optimum number of samples that we should take from each signal is an optimization problem, with a solution depending on the parameters of the problem.

The above results are consistent with our earlier results in [7]. However, the proofs of this paper require new mathematical tools. Besides majorization inequality that has been used in our previous paper, we use the matrix version of the arithmetic and harmonic means inequality, the Löwner-Heinz theorem for operator convex functions, and more importantly a new reverse majorization inequality (Lemma [7]) that might be of independent interest. Majorization inequalities state that the diagonal entries of a Hermitian matrix \( F \) are majorized by the eigenvalues of \( F \). Therefore, if \( F_{\text{Diag}} \) is a diagonal matrix wherein we have kept the diagonal entries of \( F \) and set the off-diagonals to zero, we will have

\[
\text{Tr} \left[ F^{-1} \right] \geq \text{Tr} \left[ F_{\text{Diag}}^{-1} \right]
\]

(5)

Our reverse majorization inequality goes in the reverse direction. For certain matrices \( F \) and \( G \) of our interest, we show that

\[
\text{Tr} \left[ F^{-1} G \right] \leq \text{Tr} \left[ F_{\text{Diag}}^{-1} G_{\text{Diag}} \right].
\]

(6)

2 Preliminaries and Notation

Capital letters are used for random variables and matrices, whereas lowercase letters show (non-random) values. Bold letters are reserved for vectors (such as \( \mathbf{x} \)), and bold capital letters (such as \( \mathbf{X} \)) correspond to random vectors. The covariance of a random vector \( \mathbf{X} \) is denoted by \( C_X \). Given a matrix \( A \), we use \( A_{\text{Diag}} \) to denote the matrix formed by keeping the diagonal entries of \( A \) and changing the off-diagonal entries to zero. Given a vector \( \mathbf{x} \), \( \text{Diag}(\mathbf{x}) \) denotes the diagonal matrix where its diagonal entries are coordinates of \( \mathbf{x} \). The symbol \( \bigoplus \) is used for the direct sum and \( \otimes \) is used for the Kronecker product of matrices. We write \( A \preceq B \) if \( B - A \) is positive semi-definite. For a Hermitian matrix \( A \) with eigendecomposition \( P DP^{-1} \) and real function \( f \), \( f(A) \) is defined as \( Pf(D)P^{-1} \) in which \( f(D) \) is a diagonal matrix where function \( f \) is applied to the diagonal entries of \( D \).

2.1 Linear Minimum Mean Square Error (LMMSE)

Lemma 1 [20] Suppose that \( \mathbf{Y} = A\mathbf{X} + \mathbf{Z} \) in which \( \mathbf{Y} \) is an observation vector, \( A \) is a known matrix, \( \mathbf{X} \) is a vector to be estimated and \( \mathbf{Z} \) is an additive noise vector. In the case \( \mathbf{X} \) and \( \mathbf{Z} \) are mutually independent Gaussian vectors, LMMSE is optimal and the estimator and the mean square error, respectively, are given by

\[
\hat{x}_{\text{MMSE}}(\mathbf{y}) = \mathbb{E} \{ \mathbf{X} | \mathbf{y} \} = W \mathbf{y},
\]

\[
\mathbb{E}\| \mathbf{X} - \hat{\mathbf{X}} \|^2 = \mathbb{E}_\mathbf{Y} \{ \text{Var}[\mathbf{X} | \mathbf{Y}] \} = \text{Tr}(C_e),
\]

(7)
where the reconstruction matrix, $W$, and the error covariance matrix, $C_e$, are of the following forms:

$$W = C_X Y C_Y^{-1} = C_X A^T (A C_X A^T + C_Z)^{-1},$$

$$C_e = C_X - C_X Y C_Y^{-1} C_Y X = C_X - C_X A^T (A C_X A^T + C_Z)^{-1} A C_X. \quad (8)$$

Or alternatively [21], using the matrix identity

$$C_X A^T (A C_X A^T + C_Z)^{-1} = (A^T C_Z^{-1} A + C_X^{-1})^{-1} A^T C_Z^{-1}, \quad (9)$$

the matrices $W$ and $C_e$ are given by

$$W = (A^T C_Z^{-1} A + C_X^{-1})^{-1} A^T C_Z^{-1},$$

$$C_e = (A^T C_Z^{-1} A + C_X^{-1})^{-1}. \quad (10)$$

### 2.2 Majorization inequalities

A vector $x \in \mathbb{R}^n$ is majorized by $y \in \mathbb{R}^n$ if after sorting the two vectors in decreasing order, the following inequalities hold:

$$\sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i \quad (1 \leq k \leq n), \quad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i. \quad (11)$$

A fundamental result in majorization theory states that for any Hermitian matrix $A$ of size $n \times n$, the diagonal entries of $A$ are majorized by its eigenvalues [22]. The extension of the above result to the block Hermitian matrices is also true (e.g. see [23, Sec. 1]):

**Lemma 2 (Block majorization inequality)** If a Hermitian matrix $A$ is partitioned into block matrices

$$A = \begin{pmatrix} M_{11} & M_{12} & \ldots & M_{1k} \\ M_{21} & M_{22} & \ldots & M_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ M_{k1} & M_{k2} & \ldots & M_{kk} \end{pmatrix} \quad (12)$$

for matrices $M_{ij}, i, j = 1, 2, \ldots , k$, then the eigenvalues of $\bigoplus_{i=1}^{k} M_{ii}$ are majorized by the eigenvalues of $A$.

If a vector $x$ is majorized by $y$, then for any convex functions $f : \mathbb{R} \mapsto \mathbb{R}$, we have $\sum_i f(x_i) \leq \sum_i f(y_i)$ [22]. This implies that

**Lemma 3** Let $\Omega$ be a closed interval in $\mathbb{R}$. For any Hermitian matrix $A$ with eigenvalues in $\Omega$, and any convex function $f$ on $\Omega$,

$$\text{Tr}(f(A)) \geq \text{Tr}(f(A_{\text{Diag}})). \quad (13)$$

More generally by Lemma 2, for a Hermitian matrix $A$ partitioned into block matrices $M_{ij}, i, j = 1, 2, \ldots , k$, as in (12),

$$\text{Tr}(f(A)) \geq \text{Tr}\left( f\left( \bigoplus_{i=1}^{k} M_{ii} \right) \right). \quad (14)$$
2.3 Other useful definitions and inequalities

Lemma 4 [24] Let \( w_1, w_2, \ldots, w_k \) be non-negative weights adding up to one, and let \( B_1, B_2, \ldots, B_k \) be \( n \times n \) positive definite matrices. Consider the weighted arithmetic and harmonic means of the matrices \( B_i \)

\[
A \triangleq w_1 B_1 + w_2 B_2 + \cdots + w_k B_k, \quad (15)
\]

\[
H \triangleq (w_1 B_1^{-1} + w_2 B_2^{-1} + \cdots + w_k B_k^{-1})^{-1}. \quad (16)
\]

Then, the following inequality holds,

\[ H \leq A, \]

with equality if and only if \( B_1 = B_2 = \cdots = B_k \).

Definition 1 [25] A real-valued continuous function \( f(t) \) on a real interval \( I \) is called operator monotone if

\[ A \leq B \Rightarrow f(A) \leq f(B), \quad (17) \]

for Hermitian matrices \( A \) and \( B \) with eigenvalues in \( I \). Furthermore, \( f \) is called operator convex if

\[ f(\lambda A + (1 - \lambda)B) \leq \lambda f(A) + (1 - \lambda) f(B), \]

for any \( 0 \leq \lambda \leq 1 \) and Hermitian matrices \( A \) and \( B \) with eigenvalues that are contained in \( I \), and \( f \) is said to be operator concave if \( -f \) is operator convex.

Lemma 5 [26, p.260] Let \( m \) and \( n \) be given positive integers and matrices \( A \) and \( B \) be any square matrices of sizes \( m \times m \) and \( n \times n \), respectively. Then, matrix \( B \otimes A \) is permutation similar to matrix \( A \otimes B \), i.e., there is a unique matrix \( P \) such that

\[
B \otimes A = P(m,n)T (A \otimes B) P(m,n) \quad (18)
\]

where \( P(m,n) \) is the following \( mn \times mn \) permutation matrix:

\[
P(m,n) = \sum_{i=1}^{n} \sum_{j=1}^{m} E_{ij} \otimes E_{ij}^T \quad (19)
\]

in which \( E_{ij} \) is an \( m \times n \) matrix such that only the \((i,j)\)th entry is unity and the other entries are zero. Furthermore, the useful following property holds

\[
P(m,n) = P(n,m)^T = P(n,m)^{-1}. \quad (20)
\]

3 Main Results

Let \( N = N_2 - N_1 + 1 \). Each signal has \( 2N \) free variables and could be reconstructed by \( 2N \) noiseless samples. We refer to the number \( 2N \) as the Landau rate for each signal, and \( 2N_2 \) as its Nyquist rate. The number of samples that we are allowed to take from the \( i \)th signal is \( m_i \). The two main results are for the sampling rates below half the Landau rate, \( m_i \leq N \), and the rates far above the Nyquist rate, \( m_i \geq 2N_2 \).

Theorem 1 If \( m_i \)s are such that \( m = \sum_{i=1}^{k} m_i \leq N \), the minimum distortion for the reconstruction of the remote signal (as in [3]) is

\[
N - N \sum_{i=1}^{k} \frac{m_i}{2(1 + \eta)N + 2\sigma_i^2}.
\]
This minimum distortion is obtained when the sampling points, $t_i$s, are all distinct and belong to the set $\{0, T/N, \ldots, (N - 1)T/N\}$. Furthermore, the same sampling strategy is optimal for the reconstruction of the entire corrupted signals (as in (4)), and the minimum distortion is equal to

$$Nk(1 + \eta) - N ((1 + \eta)^2 + (k - 1)) \sum_{i=1}^{k} \frac{m_i}{2(1 + \eta)N + 2\sigma_i^2}.$$  

Proof of the above theorem is given in Section 5.1.

Discussion: Assume that the sampling noise variances of the corrupted signals satisfy $\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_k^2$. Let us assume that we have a fixed total budget of $m$ samples that we can distribute among the $k$ signals, i.e., $\sum_{i=1}^{k} m_i = m$. Then if $m \leq N$, it has to be the case that each $m_i \leq N$ and Theorem 1 can be used. It is not difficult to verify that the total distortion (subject to $\sum_{i=1}^{k} m_i = m$) is minimized when $m_1 = m$ and $m_i = 0$ for $i > 1$, i.e., all of the samples are taken from the first signal, $S_1(t)$, which has the minimum sampling noise variance.

**Theorem 2** For $m_i > 2N_2$ for $1 \leq i \leq k$, optimal distortion is achieved when we uniformly sample the signals for both the minimizations of the remote signal and the entire corrupted signals distortions as in (3) and (4), respectively. Let $\phi_i = \frac{m_i}{2\sigma_i^2} + \frac{1}{\eta}$, and $\Phi_p = \sum_{i=1}^{k} \phi_i^p$ for any real $p$. Then, the minimum distortion of reconstruction of the remote signal is equal to

$$D_{a_{\min}} = \frac{N}{(1 + \frac{k}{\eta}) - \frac{k^2}{\eta} (\Phi_1)^{-1}}.$$  

For reconstructing of all the noisy signals, the minimum total distortion is

$$D_{b_{\min}} = \Phi_{-1} + \frac{1}{\eta(\eta + k) + \Phi_1}.$$  

Proof of the above theorem is given in Section 5.2.

Discussion: As before, assume that $\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_k^2$ and a total budget of $m$ samples. If $m_i > 2N_2$ for all $i$, then Theorem 2 can be used. For the reconstruction of the remote signal, it can be easily shown that the total distortion (subject to $\sum_{i=1}^{k} m_i = m$, $m_i > 2N_2$) is minimized when $m_i$ for $i > 1$ reaches its minimum of $2N_2$, i.e., we take as many samples as possible from the signal with minimum sampling noise variance. On the other hand, if we wish to minimize the total distortion of the entire corrupted signals, the optimum value for $m_i$ can vary depending on the value of the parameters; it is not necessarily true that it is better to sample from the less noisy signal. For instance, let us consider the case of two signals, $k = 2$, and take the corruption noise variance to be $\eta = 0.5$. Assume that samples from the first signal are taken with variance $\sigma_1 = 1$, which is less than $\sigma_2 = 10$, the noise variance of samples from the second signal. Then for total sample budget $m = 100$, the optimum $(m_1, m_2)$ subject to $m_i > 2N_2$ is given in the following for different values of $N_2$. For $N_2 = 20$, the optimal choice is $(m_1, m_2) = (40, 60)$; surprisingly, it is better to take 60 samples from the signal with more sampling noise. For $N_2 = 10$, the optimal choice is $(m_1, m_2) = (32, 68)$. Here $m_i > 2N_2 = 20$ and the optimal choice is not one of the boundary pairs $(m_1, m_2) = (20, 80)$ or $(80, 20)$. These examples are in contrast with the case of $m \leq N$ that was discussed earlier.

### 4 Problem Formulation

Let $X$ be a column vector, consisting of the Fourier coefficients of $S(t)$

$$X = [A_{N_1}, A_{N_1+1}, \ldots, A_{N_2}, B_{N_1}, \cdots, B_{N_2}]^T,$$  

(21)
where $\top$ is used for the transpose operation. Similarly, $X_i$ is a column vector, consisting of the coefficients of $S_i(t)$

$$X_i = [A_{iN_1}, A_{i(N_1+1)}, \cdots, A_{iN_2}, B_{iN_1}, \cdots, B_{iN_2}]^T$$

$$= X + [V_{iN_1}, \cdots, V_{iN_2}, W_{iN_1}, \cdots, W_{iN_2}]^T$$

$$= X + \Delta_i,$$

where $\Delta_i = [V_{iN_1}, \cdots, V_{iN_2}, W_{iN_1}, \cdots, W_{iN_2}]^T$. We assume that $X$ consists of mutually independent Gaussian random variables $\mathcal{N}(0, 1)$, i.e., the signal $S(t)$ is white. Therefore, $C_X = I_{2N \times 2N}$. Random Variables $V_{ij}$ and $W_{ij}$ are assumed to be independent with the probability distribution $\mathcal{N}(0, \eta)$ for some $\eta > 0$.

Vectors $X_i$ and $X_j$ are correlated, because they are both corrupted versions of $X$. Their cross covariance can be computed as $C_{X_i, X_j} = (1 + \eta)I$ and $C_{X_i, X_j} = I$ for $i \neq j \in \{1, 2, \cdots, k\}$. One can verify that for any $j$, the covariance matrix for the $k$ random variables $A_{ij}$ for $i = 1, 2, \cdots, k$ is equal to

$$\Lambda = \begin{pmatrix}
1 + \eta & 1 & 1 & \cdots & 1 \\
1 & 1 + \eta & 1 & \cdots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & 1 & \cdots & 1 + \eta
\end{pmatrix}
$$

Suppose that the $i$th signal, $S_i(t)$, is sampled at time instances $t_{ij}$ for $i = 1, 2, \cdots, k$ and $j = 1, 2, \cdots, m_i$. Hence,

$$S_i(t_{ij}) = \sum_{\ell=N_1}^{N_2} [A_{i\ell} \cos(\ell \omega t_{ij}) + B_{i\ell} \sin(\ell \omega t_{ij})].$$

To represent the problem in a matrix form, we define $S_i$ to be the vector of samples as

$$S_i = [S_i(t_{i1}), S_i(t_{i2}), \cdots, S_i(t_{im_i})]^T.$$
matrix of the form

\[
Q_i = \begin{pmatrix}
\cos(N_1 \omega t_{i1}) & \cos((N_1 + 1) \omega t_{i1}) & \cdots & \cos(N_2 \omega t_{i1}) & \sin(N_1 \omega t_{i1}) & \sin((N_1 + 1) \omega t_{i1}) & \cdots & \sin(N_2 \omega t_{i1}) \\
\cos(N_1 \omega t_{i2}) & \cos((N_1 + 1) \omega t_{i2}) & \cdots & \cos(N_2 \omega t_{i2}) & \sin(N_1 \omega t_{i2}) & \sin((N_1 + 1) \omega t_{i2}) & \cdots & \sin(N_2 \omega t_{i2}) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\cos(N_1 \omega t_{im_i}) & \cos((N_1 + 1) \omega t_{im_i}) & \cdots & \cos(N_2 \omega t_{im_i}) & \sin(N_1 \omega t_{im_i}) & \sin((N_1 + 1) \omega t_{im_i}) & \cdots & \sin(N_2 \omega t_{im_i})
\end{pmatrix}.
\]

Moreover, the observation vector for the \( i \)th signal is of the following form

\[
Y_i = S_i + Z_i = Q_i X_i + Z_i = Q_i X + Q_i \Delta_i + Z_i,
\]

in which \( Z_i \) is the \( i \)th noise vector with covariance of \( C_{Z_i} = \sigma_i^2 I_{m_i \times m_i} \).

Now, we define the vector of coefficients of all the \( k \) signals, \( X \), the vector of all the samples, \( S \), the vector of all the observations, \( Y \), as follows:

\[
X = [X_1^T, X_2^T, \ldots, X_k^T]^T, \quad S = [S_1^T, S_2^T, \ldots, S_k^T]^T, \quad Y = [Y_1^T, Y_2^T, \ldots, Y_k^T]^T, \quad Z = [Z_1^T, Z_2^T, \ldots, Z_k^T]^T, \quad \Delta = [\Delta_1^T, \Delta_2^T, \ldots, \Delta_k^T]^T, \quad Q_a = [Q_1^T, Q_2^T, \ldots, Q_k^T]^T.
\]

Furthermore, let

\[
Q_b = \bigoplus_{i=1}^k Q_i
\]

be the direct sum of the individual matrices \( Q_i \). Then, we can write

\[
Y = S + Z = Q_b X + Z = Q_a X + Q_b \Delta + Z.
\]

One can verify that the covariance matrix of the noise is

\[
C_Z = \bigoplus_{i=1}^k C_{Z_i} = \bigoplus_{i=1}^k \sigma_i^2 I_{m_i \times m_i}.
\]

Moreover \( C_X = I_{2N \times 2N} \) and \( C_X = \Lambda \otimes I_{2N \times 2N} \) where \( \Lambda \) was given in (25), and \( N = N_2 - N_1 + 1 \).

### 4.1 Reconstruction of \( S(t) \)

Here, the goal is to reconstruct \( S(t) \) with minimum distortion using the observation vector \( Y \). We use the MMSE criterion to minimize the average distortion subject to the samples. From the Parseval’s theorem, we have

\[
D_a = \frac{1}{T} \int_{t=0}^T \mathbb{E}\{|\hat{S}(t) - S(t)|^2\} dt = \frac{1}{2} \mathbb{E}\|X - \hat{X}\|^2,
\]

where \( \hat{S}(t) \) is the reconstructed signal and \( \hat{X} \) is the MMSE reconstruction of the coefficient vector \( X \) from the observation vector \( Y \). Since the random variables are jointly Gaussian, the MMSE estimator is optimal. From the equation

\[
Y = Q_a X + \tilde{Z},
\]

where \( \tilde{Z} = Q_b \Delta + Z \), the error of the linear MMSE estimator is equal to

\[
\mathbb{E}\|X - \hat{X}\|^2 = \mathbb{E}_Y \{\text{Var}[X|Y]\} = \text{Tr}(C_e^{\text{opt}}),
\]

(37)
where $C_e^a$ has the following two alternative forms
\begin{align}
C_e^a &= C_X - C_XQ_a^T(Q_aC_XQ_a^T + C_Z)^{-1}Q_aC_X, \\
&= (Q_a^TC_Z^{-1}Q_a + C_X^{-1})^{-1}. 
\end{align}

In the above formula, the covariance matrix of $\tilde{Z}$ is
\begin{equation}
C_Z = Q_bC_Q_b^T + C_Z = \eta Q_bQ_b^T + C_Z. 
\end{equation}

### 4.2 Reconstruction of $S_i(t)$ for $i = 1, 2, \cdots, k$

Here, the goal is to reconstruct all the $k$ signals with minimum distortion using the observation vector $Y$. Again, from the Parseval’s theorem, we have
\begin{equation}
D_b = \frac{1}{T}\sum_{t=0}^{T} \mathbb{E}\{|\tilde{S}_i(t) - S_i(t)|^2\} dt = \frac{1}{2} \mathbb{E}\|X - \hat{X}\|^2, 
\end{equation}
in which $\tilde{S}(t)$ and $\hat{X}$ are the reconstructed signal and the estimated coefficients, respectively. From the equation
\begin{equation}
Y = Q_bX + Z,
\end{equation}
the LMMSE error is equal to
\begin{equation}
\mathbb{E}\|X - \hat{X}\|^2 = \mathbb{E}_Y \{\text{Var}[X|Y]\} = \text{Tr}(C_e^b),
\end{equation}
where $C_e^b$ has the following two alternative forms
\begin{align}
C_e^b &= C_X - C_XQ_b^T(Q_bC_XQ_b^T + C_Z)^{-1}Q_bC_X, \\
&= (Q_b^TC_Z^{-1}Q_b + C_X^{-1})^{-1}. 
\end{align}

### 4.3 Some helpful facts

The following facts will be used in the proofs:

(i) We have
\begin{equation}
Q_bQ_b^T = \bigoplus_{i=1}^{k} Q_iQ_i^T = \begin{pmatrix}
Q_1Q_1^T & 0 & 0 & \cdots & 0 \\
0 & Q_2Q_2^T & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & Q_kQ_k^T
\end{pmatrix}_{m \times m}
\end{equation}

and
\begin{equation}
Q_aQ_a^T = \begin{pmatrix}
Q_1Q_1^T & Q_1Q_2^T & Q_1Q_3^T & \cdots & Q_1Q_k^T \\
Q_2Q_1^T & Q_2Q_2^T & Q_2Q_3^T & \cdots & Q_2Q_k^T \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
Q_kQ_1^T & Q_kQ_2^T & Q_kQ_3^T & \cdots & Q_kQ_k^T
\end{pmatrix}_{m \times m},
\end{equation}

where $m = \sum_{i=1}^{k} m_i$.

(ii) The rows of matrix $Q_i$ are vectors of norm $\sqrt{N}$. Therefore, the matrix $Q_iQ_i^T$, for $i = 1, 2, \cdots, k$, is of size $m_i \times m_i$ and with diagonal entries equal to $N$ regardless of the value of $t_{ij}$. Therefore, $(Q_aQ_a^T)_{\text{Diag}} = (Q_bQ_b^T)_{\text{Diag}} = NI_{m \times m}$. 

10
(iii) When, \( t_{ij} \in \{0,T/N, \cdots , (N-1)T/N\} \) and are distinct, the rows of matrix \( Q_i \) will be perpendicular to each other. Therefore, the matrix \( Q_i^TQ_i^T \) will be equal to \( N I_{m_i \times m_i} \). Similarly, if \( t_{ij} \) are distinct for all \( i,j \), the rows of \( Q_i \) and \( Q_j \) for \( i \neq j \) will be perpendicular to each other. Therefore, \( Q_i^TQ_j^T = 0 \) for \( i \neq j \) in this case. Hence, using the definition of \( Q_a \) and \( Q_b \) given in (32) and (33), both \( Q_a^TQ_a^T \) and \( Q_b^TQ_b^T \) will become diagonal matrices \( N I_{m \times m} \) where \( m = \sum_{i=1}^{k} m_i \).

(iv) The diagonal elements of matrices \( Q_i^TQ_i \) for \( i = 1,2, \cdots , k \) give us the norm of the column vectors of \( Q_i \). They can be calculated as follows:

\[
Q_i^TQ_i(l,l) = \begin{cases} 
\sum_{j=1}^{m_i} \cos^2\left((N_1 + l - 1)\omega t_{ij}\right); & \text{for } 1 \leq l \leq N \\
\sum_{j=1}^{m_i} \sin^2\left((N_1 + l - N - 1)\omega t_{ij}\right); & \text{for } N + 1 \leq l \leq 2N
\end{cases}
\]

(47)

When we use the uniform sampling strategy, i.e., \( \{t_{ij}\} = \{0,T/m_i,2T/m_i, \cdots , (m_i - 1)T/m_i\} \), the diagonal entries of \( Q_i^TQ_i \) will become equal to \( m_i/2 \). This is because, for instance,

\[
\sum_{j=1}^{m_i} \cos^2\left((N_1 + l - 1)\omega t_{ij}\right) = \sum_{j=0}^{m_i-1} \cos^2\left((N_1 + l - 1)\omega \frac{T}{m_i}\right) \\
= \sum_{j=0}^{m_i-1} \left( \frac{1}{2} + \frac{1}{2} \cos \left(2(N_1 + l - 1)\frac{2\pi}{m_i}\right) \right) \\
= \frac{m_i}{2}
\]

(48)

where (48) follows from the fact that \( m_i > 2N \). Moreover, the off-diagonal entries will be zero. For example, consider the entry

\[
Q_i^TQ_i(2,3) = \sum_{j=1}^{m_i} \cos\left((N_1 + 1)\omega t_{ij}\right) \cos\left((N_1 + 2)\omega t_{ij}\right) = \sum_{j=0}^{m_i-1} \frac{1}{2} \left( \cos\left(\frac{\omega T}{m_i}j\right) + \cos\left(\frac{(2N_1 + 3)\omega T}{m_i}j\right) \right) \\
= \sum_{j=0}^{m_i-1} \frac{1}{2} \left( \cos\left(\frac{2\pi}{m_i}j\right) + \cos\left(\frac{(2N_1 + 3)2\pi}{m_i}j\right) \right) \\
= 0.
\]

(49)

In fact, with uniform sampling, different columns of the matrix \( Q_i \) will be perpendicular to each other and \( Q_i^TQ_i \) will become \( (m_i/2)I_{2N \times 2N} \).

5 Proofs

5.1 Proof of Theorem 1: Rates below half the Landau Rate

Part (i): minimization of equation (3):

We start by computing the average distortion using equations (36), (37) and (38) as follows

\[
2D_a = \text{Tr}(C_a^a) = \text{Tr}\left(C_X - C_X Q_a^T(Q_a C_X Q_a^T + C_2)^{-1} Q_a C_X \right) \\
= \text{Tr}(C_X) - \text{Tr}\left((Q_a C_X Q_a^T + C_2)^{-1} Q_a C_X Q_a^T \right) \\
= 2N - \text{Tr}\left((Q_a Q_a^T + C_2)^{-1} Q_a Q_a^T \right),
\]

(50)

(51)

where (50) results from the cyclic property of the trace and (51) follows from the fact that \( C_X = I_{2N \times 2N} \).

We would like to show that the minimum distortion is achieved when we take distinct time instances, \( t_{ij} \), from the set \( \{0,T/N, \cdots , (N-1)T/N\} \). Observe that this is possible since \( \{0,T/N, \cdots , (N-1)T/N\} \) has \( N \) elements and \( \sum_{i=1}^{k} m_i \leq N \).
We begin by computing the average distortion when \( t_{ij} \) are distinct and belong to \( \{0, T/N, \ldots, (N-1)T/N\} \) for all \( i,j \). Using Fact (iii) from Section 4.3, \( C_2 \) given in (40) will be

\[ C_2 = \eta Q_a Q_b^T + C_Z = \eta N I_{m \times m} + C_Z. \]

Hence, the average distortion will be

\[ 2D_a = 2N - N \cdot \text{Tr} \left[ (1 + \eta) N I + C_Z \right]^{-1}, \]

\[ = 2N - \sum_{i=1}^{k} \frac{Nm_i}{(1 + \eta) N + \sigma_i^2}, \]

(52)

(53)

in which we have used the definition of the diagonal matrix \( C_Z \) given in (45).

We would like to show that this is the minimum possible average distortion for all choices of \( t_{ij} \), i.e., for any arbitrary choice of sampling time instances, the average distortion will be bounded from below as follows:

\[ 2D_a \geq 2N - \sum_{i=1}^{k} \frac{Nm_i}{(1 + \eta) N + \sigma_i^2}. \]

(54)

In other words, from (51), we wish to prove that

\[ \text{Tr} \left( (Q_a Q_b^T + C_Z)^{-1} Q_a Q_a^T \right) \leq \sum_{i=1}^{k} \frac{Nm_i}{(1 + \eta) N + \sigma_i^2}. \]

(55)

Equivalently, we would like to show that

\[ \text{Tr} \left( (Q_a Q_a^T + \eta Q_b Q_b^T + C_Z)^{-1} Q_a Q_a^T \right) \leq \sum_{i=1}^{k} \frac{Nm_i}{(1 + \eta) N + \sigma_i^2}. \]

(56)

This can be derived using Lemma 7 (given in Appendix A) with matrices \( F = Q_a Q_a^T + \eta Q_b Q_b^T \), \( G = Q_a Q_a^T \) and \( C = C_Z \). Using Fact 4 from Section 4.3 observe that the matrices \( F \) and \( G \) are of the forms

\[ F = \begin{pmatrix} (1 + \eta) Q_1 Q_1^\dagger & Q_1 Q_2^\dagger & \cdots & Q_1 Q_k^\dagger \\ Q_2 Q_1^\dagger & (1 + \eta) Q_2 Q_2^\dagger & \cdots & Q_2 Q_k^\dagger \\ \vdots & \vdots & \ddots & \vdots \\ Q_k Q_1^\dagger & Q_k Q_2^\dagger & \cdots & (1 + \eta) Q_k Q_k^\dagger \end{pmatrix}_{m \times m} \]

(57)

and

\[ G = Q_a Q_a^T = \begin{pmatrix} Q_1 Q_1^\dagger & Q_1 Q_2^\dagger & \cdots & Q_1 Q_k^\dagger \\ Q_2 Q_1^\dagger & Q_2 Q_2^\dagger & \cdots & Q_2 Q_k^\dagger \\ \vdots & \vdots & \ddots & \vdots \\ Q_k Q_1^\dagger & Q_k Q_2^\dagger & \cdots & Q_k Q_k^\dagger \end{pmatrix}_{m \times m} \]

(58)

and they satisfy the required properties of Lemma 7. Hence, the desired inequality in (55) follows from Fact (ii) from Section 4.3 which states that \( F_{\text{Diag}} = (Q_a Q_a^T)_{\text{Diag}} + \eta (Q_b Q_b^T)_{\text{Diag}} = N(1 + \eta) I_{m \times m} \) and \( G_{\text{Diag}} = (Q_a Q_a^T)_{\text{Diag}} = N I_{m \times m} \).

**Part (ii): minimization of equation (41)**

Here we are interested in reconstructing the signals \( S_i(t) \). We show that the optimal sampling points are the same as the ones we had in the previous part.
To compute the average distortion, here we use equations (11), (12) and (41). Hence,

\[ 2D_b = \text{Tr} \left( C_X - C_XQ_b^T (Q_bC_XQ_b^T + C_Z)^{-1}Q_bC_X \right) \]
\[ = \text{Tr}(C_X) - \text{Tr} \left( (Q_bC_XQ_b^T + C_Z)^{-1}Q_bC_X^2Q_b^T \right) \]
\[ = k(2N)(1 + \eta) - \text{Tr} \left( (Q_bC_XQ_b^T + C_Z)^{-1}Q_bC_X^2Q_b^T \right), \]  

(59)

(60)

where (59) and (60) are achieved, respectively, by the trace cyclic property and the fact that \( C_X = \Lambda \otimes I_{2N \times 2N} \) (The matrix \( \Lambda \) has been defined in (25)).

Following the similar steps from the previous part, we use Lemma 7 with the choice of \( F = Q_bC_XQ_b^T \), \( G = Q_bC_X^2Q_b^T \) and \( C = C_Z \). To demonstrate that these matrices have the required properties of Lemma 7, one can verify (by explicit evaluation) that \( F \) has the same expression as in (57), i.e., \( F = Q_bC_XQ_b^T = Q_b(\Lambda \otimes I_{2N \times 2N})Q_b^T \) is also equal to \( Q_bQ_b^T + \eta \sigma_b Q_b^T \). Furthermore to compute \( G \), observe that \( C_X^2 = \Lambda^2 \otimes I_{2N \times 2N} \) in which \( \Lambda^2 \) is a matrix of the following form

\[ \Lambda^2(i, j) = \begin{cases} (1 + \eta)^2 + (k - 1) \equiv \alpha & ; \ i = j, \\ 2(1 + \eta) + (k - 2) \equiv \beta & ; \ i \neq j, \end{cases} \]

(61)

for \( i, j = 1, 2, \ldots, k \). Then, one can verify that

\[ G = \begin{bmatrix} \alpha Q_1Q_1^1 & \beta Q_1Q_2^1 & \beta Q_1Q_3^1 & \cdots & \beta Q_1Q_k^1 \\ \beta Q_2Q_1^1 & \alpha Q_2Q_2^1 & \beta Q_2Q_3^1 & \cdots & \beta Q_2Q_k^1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta Q_kQ_1^1 & \beta Q_kQ_2^1 & \beta Q_kQ_3^1 & \cdots & \alpha Q_kQ_k^1 \end{bmatrix}_{m \times m}. \]

(62)

Therefore, applying Lemma 7 for the matrices \( F, G \) and \( C = C_Z \), we have

\[ \text{Tr} \left( (Q_bC_XQ_b^T + C_Z)^{-1}Q_bC_X^2Q_b^T \right) = \text{Tr}[(F + C_Z)^{-1}G] \]
\[ \leq \text{Tr}[(F_{\text{diag}} + C_Z)^{-1}G_{\text{diag}}] \]
\[ = \text{Tr} \left[ ((1 + \eta)NI + C_Z)^{-1} \alpha NI \right] \]
\[ = \sum_{i=1}^{k} \frac{\alpha Nm_i}{(1 + \eta)N + \sigma_i^2}, \]

(63)

(64)

where (63) results from the fact that the diagonal entries of the matrices \( F \) and \( G \) are \((1 + \eta)N\) and \( \alpha N \), respectively (Fact 11 from Section 4.3 and (62)). Consequently, the average distortion, for any arbitrary choice of sampling times, can be bounded as

\[ 2D_b = 2Nk(1 + \eta) - \text{Tr} \left( (Q_bC_XQ_b^T + C_Z)^{-1}Q_bC_X^2Q_b^T \right) \]
\[ \geq 2Nk(1 + \eta) - \sum_{i=1}^{k} \frac{\alpha Nm_i}{(1 + \eta)N + \sigma_i^2}, \]

in which \( \alpha \) is the one defined in (61).

\[ \square \]

5.2 Proof of Theorem 2: Rates above the Nyquist Rate

Part (i): minimization of equation (33): In this section, we assume that each \( m_i > 2N \) for \( i = 1, 2, \ldots, k \). Here, we prove that for fixed values of \( m_i \)'s, one can uniformly sample each of the signals to get the minimum mean square reconstruction distortion. To compute the minimum average distortion, we use the alternative form of LMMSE, \( D_a = 1/2\text{Tr}(C_a^\alpha) \), in which \( C_a^\alpha \) is of the form

\[ C_a^\alpha = (Q_bC_XQ_b^T + C_Z)^{-1}Q_a + C_X^{-1})^{-1}. \]

(65)
Hence, the average distortion will be
\[ 2D_a = \text{Tr}(C_{e^i}^a) = \text{Tr} \left[ \left( \sum_{i=1}^{k} Q_i^T (\eta Q_i Q_i^T + \sigma_i^2 I)^{-1} Q_i + I \right)^{-1} \right], \]  
(66)

which results from the facts that \( C_2 = \eta Q_i Q_i^T + C_2 \) (Section 4.1) and \( C_X = I_{2N \times 2N} \).

Using the matrix identity given in (9) for matrix \( A \) to be \( \sqrt{\eta Q_i} \), we have
\[ Q_i^T \left( \eta Q_i Q_i^T + C_2 \right)^{-1} Q_i = \left( (\eta Q_i C_{Z_i}^{-1} Q_i + I)^{-1} Q_i C_{Z_i}^{-1} Q_i \right), \]

Therefore, the matrix in the left-hand side of (66) will be
\[
\sum_{i=1}^{k} Q_i^T \left( \eta Q_i Q_i^T + C_2 \right)^{-1} Q_i = \sum_{i=1}^{k} \left( Q_i^T \left( C_{Z_i}^{-1} Q_i + I \right)^{-1} Q_i \right) \]

\[
= \frac{1}{\eta} \sum_{i=1}^{k} \left( \eta Q_i^T Q_i + \sigma_i^2 I \right)^{-1} \left( \eta Q_i^T Q_i + \sigma_i^2 I - \sigma_i^2 I \right)
\]
\[
= \frac{1}{\eta} \sum_{i=1}^{k} \left( \eta Q_i^T Q_i + \sigma_i^2 I \right)^{-1} \left( I - \left( \eta Q_i^T Q_i + \sigma_i^2 I \right)^{-1} \right)
\]
\[
= \frac{1}{\eta} \left( I - \sum_{i=1}^{k} B_i \right)
\]
\[
= \frac{1}{\eta} \left( kI - \sum_{i=1}^{k} B_i \right)
\]
\[
\leq \frac{1}{\eta} \left( kI - k^2 \left( \sum_{i=1}^{k} B_i^{-1} \right)^{-1} \right),
\]  
(69)

where (67) results from the fact that \( C_{Z_i} = \sigma_i^2 I \), and in (68) matrix \( B_i \) stands for \( \left( \eta Q_i^T Q_i + \sigma_i^2 I \right)^{-1} \). Moreover, (69) is derived using Lemma 4 for positive definite matrices \( B_i \) with the equality if and only if \( B_1 = B_2 = \cdots = B_k \). Notice that \( 0 < B_i < I \).

For any two symmetric positive definite matrices \( A \) and \( B \), the relation \( A \leq B \) implies that \( B^{-1} \leq A^{-1} \). This is because the function \( f(t) = -t^{-1} \) is operator monotone [27]. Hence, (69) implies that
\[
\left( I + \sum_{i=1}^{k} Q_i^T (\eta Q_i Q_i^T + \sigma_i^2 I)^{-1} Q_i \right)^{-1} \geq \left[ I + \frac{1}{\eta} \left( kI - k^2 \left( \sum_{i=1}^{k} B_i^{-1} \right)^{-1} \right) \right]^{-1}.
\]

Let \( A = \sum_{i=1}^{k} B_i^{-1} = \sum_{i=1}^{k} \left( \eta Q_i^T Q_i + \sigma_i^2 I \right) \). Then, the relation between the traces of the above matrices is
\[
\text{Tr} \left[ I + \sum_{i=1}^{k} Q_i^T (\eta Q_i Q_i^T + \sigma_i^2 I)^{-1} Q_i \right] \geq \text{Tr} \left[ I + \frac{1}{\eta} \left( kI - k^2 A^{-1} \right) \right]^{-1}
\]
\[
\geq \text{Tr} \left[ I + \frac{1}{\eta} \left( kI - k^2 A_{\text{Diag}}^{-1} \right) \right]^{-1},
\]  
(70)

where (70) results from Lemma 3 for the convex function \( f(t) = (1 + k/\eta - k^2 t^{-1}/\eta)^{-1} \) when \( t \geq k \) and the Hermitian matrix \( A \geq kI \). To find \( A_{\text{Diag}} \), we need to calculate the diagonal entries of matrix \( A \). They are
\[
A(l, l) = \begin{cases} 
\sum_{i=1}^{k} \left( 1 + \sum_{j=1}^{m_l} \frac{\eta}{\sigma_i^2} \cos^2 \left( (N_1 + l - 1) \omega t_{ij} \right) \right) \triangleq a_l; & \text{for } 1 \leq l \leq N \\
\sum_{i=1}^{k} \left( 1 + \sum_{j=1}^{m_l} \frac{\eta}{\sigma_i^2} \sin^2 \left( (N_1 + l - N - 1) \omega t_{ij} \right) \right) \triangleq b_l; & \text{for } N + 1 \leq l \leq 2N
\end{cases},
\]  
(71)
since the diagonal elements of matrices \( Q_i^TQ_i \) for \( i = 1, 2, \ldots, k \) are of the following forms (Fact [iv] from Section 4.3)

\[
Q_i^TQ_i(l, l) = \begin{cases} 
\sum_{j=1}^{m_i} \cos^2((N_i + l - 1)\omega t_{ij}) & \text{for } 1 \leq l \leq N \\
\sum_{j=1}^{m_i} \sin^2((N_i + l - N - 1)\omega t_{ij}) & \text{for } N + 1 \leq l \leq 2N
\end{cases}
\]  

(72)

Substituting the diagonal entries of the matrix \( A \) in (70), we obtain

\[
\text{Tr} \left( I + \sum_{i=1}^{k} Q_i^T(\eta_i Q_i^T + \sigma_i^2 I)^{-1}Q_i \right)^{-1} \geq \text{Tr} \left[ I + \frac{1}{\eta}(kI - k^2A_\text{Diag})^{-1} \right]^{-1} 
\]

\[= \sum_{\ell=1}^{N} \frac{1}{1 + \frac{k}{\eta} - \frac{k^2}{\eta} a_\ell^{-1}} + \frac{1}{(1 + \frac{k}{\eta} - \frac{k^2}{\eta} b_\ell^{-1})} \]

\[\geq \sum_{\ell=1}^{N} \frac{2}{1 + \frac{k}{\eta} - \frac{k^2}{\eta} (k + \eta \sum_{i=1}^{k} \frac{m_i}{2\sigma_i^2})^{-1}} \]

\[= \frac{2N}{(1 + \frac{k}{\eta} - \frac{k^2}{\eta} (k + \eta \sum_{i=1}^{k} \frac{m_i}{2\sigma_i^2})^{-1})} 
\]

(73)

(74)

where (73) results from convexity of the function \( f(t) = (1 - at^{-1})^{-1} \) for \( t \geq a \).

Using Fact [iv] from Section 4.3, equality in the above equations holds when we uniformly sample the signals, i.e., sample \( S_i(t) \) at time instances \( \{0, T/m_i, 2T/m_i, \ldots, (m_i - 1)T/m_i\} \). \( \square \)

Part (ii): minimization of equation (41): In this section, suppose that each \( m_i > 2N_2 \) for \( i = 1, 2, \ldots, k \). To get the minimum mean squared error, for fixed values of \( m_i \), one can uniformly sample the signals. To compute the minimum average distortion, we use the alternative form of LMMSE, \( D_b = 1/2\text{Tr}(C_e^b) \), in which \( C_e^b \) is of the form

\[C_e^b = (Q_i^T C_i^{-1}Q_i + C_x^{-1})^{-1}.
\]

(75)

In the above formula,

\[C_z^{-1} = \bigoplus_{i=1}^{k} \frac{1}{\sigma_i^2} I_{2N \times 2N}, \quad C_x^{-1} = \Gamma_{k \times k} \otimes I_{2N \times 2N}.
\]

(76)

where the matrix \( \Gamma \) is the inverse of the matrix \( \Lambda \), given in (25). Therefore, the average distortion will be

\[2D_b = \text{Tr}(C_e^b) = \text{Tr} \left[ \left( \bigoplus_{i=1}^{k} \frac{Q_i^T Q_i}{\sigma_i^2} \right) + \Gamma \otimes I \right]^{-1}.
\]

(77)

We divide the proof into two parts. In the first part, we show that the optimal sampling strategy is uniform sampling and the minimum distortion is equal to

\[D_{b_{\text{min}}} = N \cdot \text{Tr} \left( \text{Diag}(\frac{m_1}{2\sigma_1^2}, \frac{m_2}{2\sigma_2^2}, \ldots, \frac{m_k}{2\sigma_k^2}) + \Gamma \right)^{-1}.
\]

In the second part, we simplify the above equation to obtain the expression given in the statement of the theorem.

Part ii-1: If we employ uniform strategy, the matrices \( Q_i^T Q_i (i = 1, 2, \ldots, k) \) become diagonal matrices with diagonal entries equal to \( m_i/2 \) (Fact [iv] from Section 4.3). Then, the matrix in the right hand side of (77) will be

\[\left[ \left( \bigoplus_{i=1}^{k} \frac{m_i}{2\sigma_i^2} I_{2N \times 2N} \right) + \Gamma \right]^{-1} = \left( \text{Diag}(\frac{m_1}{2\sigma_1^2}, \frac{m_2}{2\sigma_2^2}, \ldots, \frac{m_k}{2\sigma_k^2}) \otimes I + \Gamma \right)^{-1} \]

\[= \left( \text{Diag}(\frac{m_1}{2\sigma_1^2}, \frac{m_2}{2\sigma_2^2}, \ldots, \frac{m_k}{2\sigma_k^2}) + \Gamma \right) \otimes I_{2N \times 2N}.
\]

(78)

(79)
Therefore, from (77), we get
\[ D_{b_{\min}} \leq N \cdot \mathbf{Tr} \left( \text{Diag}(\left[ \frac{m_1}{2\sigma_1^2}, \frac{m_2}{2\sigma_2^2}, \ldots, \frac{m_k}{2\sigma_k^2} \right]) + \Gamma \right)^{-1}. \]

To complete the proof, it remains to show that
\[ D_{b_{\min}} \geq N \cdot \mathbf{Tr} \left( \text{Diag}(\left[ \frac{m_1}{2\sigma_1^2}, \frac{m_2}{2\sigma_2^2}, \ldots, \frac{m_k}{2\sigma_k^2} \right]) + \Gamma \right)^{-1}. \]

First notice that due to Lemma 4, matrices \( \Gamma \otimes I \) and \( I \otimes \Gamma \) are permutation similar, i.e., there exists a unique permutation matrix \( P(k, 2N) \) of size \( 2Nk \times 2Nk \) such that
\[ \Gamma \otimes I = (P(k, 2N))^T(I \otimes \Gamma)P(k, 2N), \]
and moreover \( P(k, 2N) \) has the property
\[ P(k, 2N) = (P(k, 2N))^T = P(k, 2N)^{-1}. \]

Using the permutation matrix \( P(k, 2N) \) and the cyclic property of the trace, we have
\[
2D_b = \mathbf{Tr}(C_{cb}) = \mathbf{Tr} \left( (P(k, 2N))^TC_{cb}P(k, 2N) \right) \\
= \mathbf{Tr} \left[ (P(k, 2N))^T \left( \bigoplus_{i=1}^{k} \frac{Q_i^T Q_i}{\sigma_i^2} \right) P(k, 2N) + (P(k, 2N))^T (\Gamma \otimes I) P(k, 2N) \right]^{-1} \\
= \mathbf{Tr} (H + I \otimes \Gamma)^{-1},
\]
where matrix \( H \) denotes the permuted matrix \( (\bigoplus_{i=1}^{k} \frac{Q_i^T Q_i}{\sigma_i^2})P(k, 2N) \). Notice that the matrix
\[ \bigoplus_{i=1}^{k} Q_i^T Q_i = \sum_{i=1}^{k} G_i \otimes Q_i^T Q_i, \]

where \( G_i \) is a \( k \times k \) matrix defined as follows: \( G_i(i', j') = 0 \) if \( (i', j') \neq (i, i) \), and \( G_i(i', j') = 1 \) if \( (i', j') = (i, i) \). Therefore, the matrix \( H \) can be written as
\[ H = (P(k, 2N))^T \left( \bigoplus_{i=1}^{k} \frac{Q_i^T Q_i}{\sigma_i^2} \right) P(k, 2N) \]
\[ = \sum_{i=1}^{k} Q_i^T Q_i \otimes G_i. \]

If we partition \( H \) into \( k \times k \) submatrices \( H_{ij} \) as follows
\[ H = \begin{pmatrix}
H_{11} & H_{12} & \cdots & H_{1(2N)} \\
H_{21} & H_{22} & \cdots & H_{2(2N)} \\
& \vdots & \ddots & \vdots \\
H_{(2N)1} & H_{(2N)2} & \cdots & H_{(2N)(2N)}
\end{pmatrix}, \]
all the \( H_{ij} \) submatrices will be diagonal matrices because they are weighted sums of diagonal matrices \( G_i \). More precisely, the submatrices \( H_{ll} \) for \( l \leq N \) can be computed as follows, using Fact (iv) from Section 1.3 that gives us the diagonal entries of \( Q_i^T Q_i \) (the entries of matrix \( H_{ll} \) for \( 1 \leq l \leq N \) and \( N \leq l \leq 2N \) are the \( l \)th and the \( (l + N) \)th diagonal entries of matrices \( Q_i^T Q_i/\sigma_i^2 \), given in (72), respectively):
\[ H_{ll} = \begin{pmatrix}
\frac{1}{\sigma_1^2} \sum_{j=1}^{m_1} \cos^2((N_1 + l - 1)\omega t_{1j}) & 0 & 0 & \cdots & 0 \\
0 & \frac{1}{\sigma_2^2} \sum_{j=1}^{m_2} \cos^2((N_1 + l - 1)\omega t_{2j}) & 0 & \cdots & 0 \\
0 & 0 & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \ddots & \frac{1}{\sigma_k^2} \sum_{j=1}^{m_k} \cos^2((N_1 + l - 1)\omega t_{kj})
\end{pmatrix}_{k \times k}. \]
and for \( l = N + 1, N + 2, \ldots, 2N \),

\[
H_{ll} = \begin{pmatrix}
\frac{1}{\sigma^2} \sum_{j=1}^{m_1} \sin^2((N_l + l - N_1)\omega_{t_j}) & 0 & \cdots & 0 \\
0 & \frac{1}{\sigma^2} \sum_{j=1}^{m_2} \sin^2((N_l + l - N_1)\omega_{t_j}) & 0 & \cdots \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{\sigma^2} \sum_{j=1}^{m_k} \sin^2((N_l + l - N_1)\omega_{t_j})
\end{pmatrix}_{k \times k}
\]

Applying Lemma 3 for the Hermitian matrix \( A \) and the convex function \( f(x) = x^{-1} \) for \( x > 0 \), we attain a lower bound on the average distortion as:

\[
2D_b = \text{Tr}(C_b^h) = \text{Tr}(H + I \otimes \Gamma)^{-1},
\]

in which the matrix \( H_{\text{BDiag}} \) is the block diagonal form of matrix \( H \), where all the submatrices other than the \( 2N \) block diagonal submatrices, \( H_{ll} \), are zero. Consequently, the matrix \( (H_{\text{BDiag}} + I \otimes \Gamma) \) is of the form

\[
(H_{\text{BDiag}} + I \otimes \Gamma) = \begin{pmatrix}
H_{11} + \Gamma & 0 & \cdots & 0 \\
0 & H_{22} + \Gamma & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & H_{(2N)(2N)} + \Gamma
\end{pmatrix}_{2Nk \times 2Nk}
\]

Therefore,

\[
\text{Tr}(H_{\text{BDiag}} + I \otimes \Gamma)^{-1} = \sum_{l=1}^{2N} \text{Tr}(H_{ll} + \Gamma)^{-1}.
\]

Using Lemma 6 (given in Appendix A), for any \( 1 \leq l \leq N \), we obtain

\[
\frac{1}{2}(\text{Tr}(H_{ll} + \Gamma)^{-1} + \text{Tr}(H_{(l+N)(l+N)} + \Gamma)^{-1}) \geq \text{Tr}\left(\frac{1}{2}(H_{ll} + H_{(l+N)(l+N)} + \Gamma)^{-1}\right)
\]

\[
= \text{Tr}\left(\text{Diag}\left([\frac{m_1}{2\sigma^2}, \frac{m_2}{2\sigma^2}, \ldots, \frac{m_k}{2\sigma^2}] + \Gamma\right)^{-1}\right).
\]

Consequently,

\[
\text{Tr}(H_{\text{BDiag}} + I \otimes \Gamma)^{-1} = \sum_{l=1}^{2N} \text{Tr}(H_{ll} + \Gamma)^{-1} \geq 2N \cdot \text{Tr}\left(\text{Diag}\left([\frac{m_1}{2\sigma^2}, \frac{m_2}{2\sigma^2}, \ldots, \frac{m_k}{2\sigma^2}] + \Gamma\right)^{-1}\right).
\]

Therefore, from (83), we obtain

\[
D_b \geq N \cdot \text{Tr}\left(\text{Diag}\left([\frac{m_1}{2\sigma^2}, \frac{m_2}{2\sigma^2}, \ldots, \frac{m_k}{2\sigma^2}] + \Gamma\right)^{-1}\right).
\]

**Part ii-2:** So far we have shown that

\[
D_{b_{\text{min}}} = N \cdot \text{Tr}\left(\text{Diag}\left([\frac{m_1}{2\sigma^2}, \frac{m_2}{2\sigma^2}, \ldots, \frac{m_k}{2\sigma^2}] + \Gamma\right)^{-1}\right).
\]

Observe that \( \Gamma = \Lambda^{-1} \) can be computed from (27) as follows:

\[
\Gamma = \begin{pmatrix}
\alpha & \beta & \beta & \cdots & \beta \\
\beta & \alpha & \beta & \cdots & \beta \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\beta & \beta & \beta & \cdots & \alpha
\end{pmatrix}_{k \times k},
\]

(88)
in which
\[ \alpha = \frac{\eta + k - 1}{\eta(\eta + k)} \quad \text{and} \quad \beta = \frac{-1}{\eta(\eta + k)}. \] (89)

For simplicity define:
\[ A := \text{Diag}(\frac{m_1}{2\sigma_1^2}, \frac{m_2}{2\sigma_2^2}, \ldots, \frac{m_k}{2\sigma_k^2}) + \Gamma \] (90)

Therefore, we need to find \( \text{Tr}(A^{-1}) \) which is equal to \( \sum_{i=1}^{k}(A^{-1})_{i,i} \). To calculate the diagonal entries of \( A^{-1} \), we use the Cramer’s rule as follows:
\[ (A^{-1})_{i,i} = \frac{\det(A_i)}{\det(A)} \] (91)

where \( A_i \) is the remaining matrix after removing the \( i \)th row and column of \( A \). According to Lemma 8 (given in Appendix A), we deduce:
\[ (A^{-1})_{i,i} = \frac{\prod_{j=1}^{k} \phi_j + \beta \sum_{j=1}^{k} \prod_{l \neq i,j} \phi_l}{\prod_{j=1}^{k} \phi_j + \beta \sum_{j=1}^{k} \prod_{l \neq j} \phi_l} \] (92)

in which \( \phi_i = \frac{m_i}{2\sigma_i^2} + \frac{1}{\eta} \). Hence,
\[ \text{Tr}(A^{-1}) = \sum_{i=1}^{k}(A^{-1})_{i,i} \] (93)
\[ = \frac{\sum_{i=1}^{k} \frac{1}{\phi_i} + \beta \sum_{i,j=1}^{k} \frac{1}{\phi_i \phi_j}}{1 + \beta(\sum_{i=1}^{k} \frac{1}{\phi_i})}, \] (94)

in which (94) is derived by replacing (92) in (93) and dividing both numerator and denominator by \( \prod_{i=1}^{k} \phi_i \). Observe that by our definition \( \Phi_{-1} := \sum_{i=1}^{k} \frac{1}{\phi_i} \), we get
\[ \text{Tr}(A^{-1}) = \frac{\Phi_{-1} + \beta(\Phi_{-1}^2 - \sum_{i=1}^{k} \frac{1}{\phi_i})}{1 + \beta \Phi_{-1}} = \Phi_{-1} - \frac{\beta}{1 + \beta \Phi_{-1}} \Phi_{-2}. \] (95)

We get the desired result by replacing \( \beta = -1/(\eta(\eta + k)) \). \( \square \)

A Lemmas

**Lemma 6** Assume that \( B_1, B_2 \) are positive definite matrices. Let \( B = (B_1 + B_2)/2 \), then
\[ 2\text{Tr}(B^{-1}) \leq \text{Tr}(B_1^{-1}) + \text{Tr}(B_2^{-1}). \]

**Proof:** The Löwner-Heinz theorem implies that the function \( f(t) = t^{-1} \) for \( t > 0 \) is operator convex [27]. From the fact that \( B = (B_1 + B_2)/2 \), we conclude the desired inequality. \( \square \)
Lemma 7  Take two positive semidefinite matrices $F$ and $G$ of sizes $m \times m$ satisfying $G = F \circ L$ where $\circ$ is the Hadamard product and $L$ is a matrix of the following form:

$$L = \begin{pmatrix}
a_{1_{m_1 \times m_1}} & b_{1_{m_1 \times m_2}} & \cdots & b_{1_{m_1 \times m_k}} \\
b_{1_{m_2 \times m_1}} & a_{1_{m_2 \times m_2}} & \cdots & b_{1_{m_2 \times m_k}} \\
\vdots & \vdots & \ddots & \vdots \\
b_{1_{m_k \times m_1}} & b_{1_{m_k \times m_2}} & \cdots & a_{1_{m_k \times m_k}}
\end{pmatrix},$$

(96)

where $1$ is a matrix with all one coordinates and $a$ and $b$ are two positive real numbers where $0 \leq a \leq b$. Then, for any positive definite diagonal matrix $C$, we have

$$\text{Tr} \left[ (F + C)^{-1} G \right] \leq \text{Tr} \left[ (F_{\text{Diag}} + C)^{-1} G_{\text{Diag}} \right],$$

(97)

where $F_{\text{Diag}}$ is a diagonal matrix formed by taking the diagonal entries of $F$, and the matrix $G_{\text{Diag}}$ is defined similarly.

**Proof:** Since if the statement of theorem holds for $F$, it will also hold for $kF$ for any positive constant $k$, without loss of generality, we assume that $a = 1$. From the Hadamard product relation, this implies that $F$ and $G$ are equal on block matrices on the diagonal. Let $A$ denote this common part, i.e.,

$$A = \begin{pmatrix}
F_{1_{m_1 \times 1_{m_1}}} & 0 & 0 & \cdots & 0 \\
0 & F_{(m_1+1_{m_1+1_{m_2}})(m_1+1_{m_1+1_{m_2}})} & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & F_{(m_1+\cdots+m_k-1+1_{m_1+\cdots+m_k-1+1_{m_k})}}
\end{pmatrix}_{m \times m}$$

One can find matrix $B$ such that $F = A + B$ and $G = A + bB$. Observe that wherever $A$ is non-zero, $B$ is zero and vice versa. Substituting $F$ and $G$ in the left hand side of (97) attains

$$\text{Tr} \left[ (F + C)^{-1} G \right] = \text{Tr} \left[ (A + B + C)^{-1} (A + bB) \right]$$

$$= \text{Tr} \left[ (A + B + C)^{-1} (A + B) \right] + \text{Tr} \left[ (A + B + C)^{-1} (b - 1)B \right].$$

(98)

We will show that the first term in the above formula is less than or equal to the right hand side of (97) and the second term is non-positive.

Start with the first term of (98)

$$\text{Tr} \left[ (A + B + C)^{-1} (A + B) \right] = \text{Tr}(I_{m \times m}) - \text{Tr} \left( (A + B + C)^{-1} C \right),$$

(99)

where the second term in (99) can be bounded as follows:

$$\text{Tr} \left( (A + B + C)^{-1} C \right) = \text{Tr} \left[ (\frac{1}{2} A C^{-\frac{1}{2}} + C^{-\frac{1}{2}} BC^{-\frac{1}{2}} + I)^{-1} \right]$$

$$\geq \text{Tr} \left[ (\frac{1}{2} A_{\text{Diag}} C^{-\frac{1}{2}} + I)^{-1} \right].$$

(100)

The last inequality comes from Lemma 3.

Hence, (99) will be bounded as

$$\text{Tr} \left[ (A + B + C)^{-1} (A + B) \right] \leq \text{Tr}(I_{m \times m}) - \text{Tr} \left[ (\frac{1}{2} A_{\text{Diag}} C^{-\frac{1}{2}} + I)^{-1} \right],$$

(101)

$$= \text{Tr}[(A_{\text{Diag}} + C)^{-1} (A_{\text{Diag}} + C)] - \text{Tr}[C^{-\frac{1}{2}} (A_{\text{Diag}} + C)^{-1} C^{-\frac{1}{2}}]$$

(102)

$$= \text{Tr}[(A_{\text{Diag}} + C)^{-1} A_{\text{Diag}}] - \text{Tr}[C^{-\frac{1}{2}} A_{\text{Diag}} C^{-\frac{1}{2}}]$$

(103)

$$= \text{Tr}[(F_{\text{Diag}} + C)^{-1} G_{\text{Diag}}],$$

(104)

in which (103) comes from the trace interchange property and (104) is derived since $A$ is defined to be the common part of the two matrices $F$ and $G$. 

19
To complete the proof, it remains to show that the right hand side of (98) is non-positive, i.e.,

\[(b - 1) \cdot \text{Tr} \left[ (A + B + C)^{-1} B \right] \leq 0. \tag{105} \]

Since we have assumed that \( b \geq 1 \), we need to show that the trace function is non-positive. For the positive definite matrix \((A + C)\), we have

\[
\text{Tr} \left( (A + B + C)^{-1} B \right) = \text{Tr}(I) - \text{Tr} \left( (A + C + B)^{-1} (A + C) \right)
\]

\[
= \text{Tr}(I) - \text{Tr} \left( (A + C + B)^{-1} (A + C + B)^{-\frac{1}{2}} (A + C + B)^{\frac{1}{2}} \right)
\]

\[
= \text{Tr}(I) - \text{Tr} \left( (I + (A + C)^{\frac{1}{2}} B (A + C)^{\frac{1}{2}})^{-1} \right)
\]

\[
\leq \text{Tr}(I) - \text{Tr}(I) = 0.
\tag{107} \]

in which the inequality \((106)\) follows from the trace interchange property and \((107)\) is derived using Lemma 3 for the Hermitian matrix \((I + (A + C)^{\frac{1}{2}} B (A + C)^{\frac{1}{2}})^{-1}\). Note that the matrix \((A + C + B)^{-\frac{1}{2}} (A + C + B)^{\frac{1}{2}}\) is a block off-diagonal matrix, since matrices \(A\) and \(B\) has been defined to be, respectively, block diagonal and off-diagonal matrices such that wherever \(A_{ij}\) is zero, \(B_{ij}\) is non-zero and vice versa. This completes the proof of Lemma 4. \(\square\)

**Lemma 8** Given real non-negative \(a_1, a_2, \ldots, a_k\) and positive \(b\), let

\[
M(a_1, \ldots, a_k, b) := \begin{pmatrix}
a_1 & -b & -b & \cdots & -b \\
-b & a_2 & -b & \cdots & -b \\
& \vdots & \vdots & \ddots & \vdots \\
& & -b & -b & \cdots & a_k
\end{pmatrix}_{k \times k}.
\tag{108} \]

Then,

\[
\det(M(a_1, \ldots, a_k, b)) = \prod_{i=1}^{k} (a_i + b) - b \sum_{i=1}^{k} \prod_{j=1}^{k} (a_j + b). \tag{109} \]

**Proof:** The elementary row operations do not change the determinant. If we first subtract the first row from all the other rows, and then multiply the \(i\)th row of the matrix by \(b/(a_i + b)\) for \(i = 2, 3, \ldots, k\), and add it to the first row, we end up with an upper triangular matrix with diagonal entries \(\{a', a_2 + b, a_3 + b, \ldots, a_k + b\}\) where \(a' = a_1 - \sum_{i=2}^{k} \frac{(a_i + b)}{(a_i + b)}\). Since the determinant of an upper triangular matrix is equal to product of the diagonal elements, we have

\[
\det(M(a_1, \ldots, a_k, b)) = a'(a_2 + b) \cdots (a_k + b)
\]

\[
= a_1(a_2 + b) \cdots (a_k + b) - b \sum_{i=2}^{k} \prod_{j=1}^{k} (a_j + b)
\]

\[
= \prod_{i=1}^{k} (a_i + b) - b \sum_{i=1}^{k} \prod_{j=1}^{k} (a_j + b).
\tag{110} \]

\(\square\)

**References**

[1] T. Berger, *Rate-Distortion Theory*, Wiley Online Library, 1971.
[2] H. Witsenhausen, “Indirect rate distortion problems,” IEEE Trans. Inf. Theory, vol. 26, no. 5, pp. 518-521, Sep. 1980.

[3] J. Wolf and J. Ziv, “Transmission of noisy information to a noisy receiver with minimum distortion,” IEEE Trans. Inf. Theory, vol.16, no. 4, pp. 406-411, Apr. 1970.

[4] R. Dobrushin and B. Tsybakov, “Information transmission with additional noise,” IRE Trans. Inform. Theory, vol. 8, no. 5, pp. 293304, Jun. 1962.

[5] T. Berger, Z. Zhang and and H. Viswanathan, “The CEO problem [multiterminal source coding],” IEEE Trans. Inf. Theory, vol. 42, no. 3, pp. 887-902, 1996.

[6] H. Viswanathan, and T. Berger, “The quadratic Gaussian CEO problem,” IEEE Trans. Inf. Theory, vol. 43, no. 5, pp. 1549-1559, 1997.

[7] E. Mohammadi and F. Marvasti, “Sampling and Distortion Tradeoffs for Bandlimited Periodic Signals,” Sampling Theory and Applications (SampTA), 2015 International Conference on, pp. 468 - 472, arXiv 1405.3980.

[8] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, “Distributed Compressed Sensing,” Elect. Comput. Eng. Dep., Rice University, Houston, TX, Tech. Rep. ECE-0612, Dec. 2006.

[9] A. Hormati, O. Roy, Y.M. Lu and M. Vetterli, “Distributed sampling of signals linked by sparse filtering: Theory and applications,” IEEE Trans. Signal Process., 58(3), pp.1095-1109, 2010.

[10] A. Ahmed and J. Romberg, “Compressive Sampling of Ensembles of Correlated Signals,” arXiv:1501.06654 2015.

[11] D. Ganesan, D. Estrin and J. Heidemann, “DIMENSIONS: Why do we need a new data handling architecture for sensor networks?,” ACM SIGCOMM Computer Communication Review, 33(1), pp.143-148, 2003.

[12] S. Bandyopadhyay, Q. Tian and E.J. Coyle, “Spatio-temporal sampling rates and energy efficiency in wireless sensor networks,” IEEE/ACM Transactions on Networking (TON) 13, no. 6, pp. 1339-1352, 2005.

[13] D. Zordan, G. Quer, M. Zorzi and M. Rossi, “Modeling and generation of space-time correlated signals for sensor network fields,” IEEE Global Telecommunications Conference (GLOBECOM 2011), (pp. 1-6).

[14] R. Masiero, G. Quer, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, “Data Acquisition through joint Compressive Sensing and Principal Component Analysis,” in IEEE Globecom 2009, Honolulu, Hawaii, US, Nov.-Dec. 2009.

[15] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, “Compressive wireless sensing,” in Proc. Int. Conf. Inf. Process. Sens. Netw., Apr. 2006, pp. 134142.

[16] W. Wang, M. Garofalakis, and K. Ramchandran, “Distributed sparse random projections for refinable approximation,” in Proc. Int. Conf. Inf. Process. Sens. Netw., Apr. 2007, pp. 331339.

[17] A. Kipnis, A. Goldsmith, Y. C. Eldar and T. Weissman, “Distortion Rate Function of sub-Nyquist Sampled Gaussian Sources,” IEEE Trans. Inf. Theory, vol. 62, no. 1, pp. 401-429, Jan. 2016.

[18] S. Feizi, V. K. Goyal and M. Medard, “Time-stampless adaptive nonuniform sampling for stochastic signals”, IEEE Trans. Signal Process., vol. 60, no. 10, pp. 5440-5450, Oct. 2012.

[19] V. P. Boda and P. Narayan, “Sampling rate distortion,” in Proc. IEEE Int. Symp. on Inf. Theory (ISIT), 2014, pp. 3057 - 3061.
[20] S. M. Kay, *Fundamentals of statistical signal processing: Estimation theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[21] D. G. Luenberger, *Optimization by vector space methods*, John Wiley & Sons, 1990.

[22] R. Bhatia, *Matrix analysis*, Springer Science & Business Media, 1997.

[23] M. Lin and H. Wołkowicz, “An eigenvalue majorization inequality for positive semidefinite block matrices,” Linear and Multilinear Algebra, no. 60.11-12, pp. 1365-1368, 2012.

[24] M. Sagae and K. Tanabe, “Upper and lower bounds for the arithmetic-geometric-harmonic means of positive definite matrices,” Linear and Multilinear Algebra, pp. 279-282, 1994.

[25] X. Zhan, *Matrix inequalities*, Springer Science & Business Media, 2002.

[26] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*, Vol. 2, Camb. Univ. Press, England, 1991.

[27] E. Carlen, “Trace inequalities and quantum entropy: an introductory course,” Entropy and the quantum 529, pp. 73-140, 2010.