Optimal Signature Sets for Transmission of Correlated Data over a Multiple Access Channel

Joydeep Acharya, Ritabrata Roy, Jasvinder Singh, Christopher Rose
Wireless Information Network Laboratory (WINLAB)
73 Brett Rd, Piscataway, NJ 08854
{joy, rito, jasingh, crose}@winlab.rutgers.edu

***Abstract*— For a single user Gaussian channel, the minimum power required to meet a specified BER at the receiver is well–known. For multiple transmitters sending independent data to a single receiver (which uses separate linear filters for decoding each transmitter's symbols), the problem of optimizing transmitter codewords to maximize capacity has been addressed in [8][9]. For the above problem, it has been found that the codeword set which maximizes capacity also minimizes TMSE (Total Mean Square Error) at the receiver. Now consider transmitters which are sending *correlated* information to a single receiver under a total power constraint. Such a scenario can typically arise in sensor networks and is usually addressed using distributed source coding [4]. In this paper, we derive optimal codeword configurations which minimize the TMSE at the receiver under a total power constraint. We also show that minimizing the TMSE is equivalent to maximizing the sum capacity.**

I. Introduction

Sensor networks are being increasingly deployed in various environments. In a typical scenario, a group of sensors observe a common phenomenon and report these observations to a central repository for processing. Since sensor nodes usually have a non–replenishable source of energy, it is highly desirable to keep individual transmission powers at their minimum levels so as to maximize the network lifetime. We consider a sensor network model where sensors use signature waveforms (codewords) to send data to the receiver and optimize the choice of signature waveforms such that TMSE is minimized under a total power constraint.

Related work [8][9] derives the optimal codewords for the case of uncorrelated symbols. Correlation among symbols however might change the structure of optimal codewords dramatically. For instance, for two transmitters sending independent symbols, the optimal codeword configuration would be along mutually orthogonal directions with equal power distribution among the codewords. However, in the extreme case when both transmitters always send the same symbols, a lower TMSE can be achieved by using equal–power identical codewords at the transmitters.

We note one can also reduce power usage in sensor networks by minimizing the number of symbols transmitted by applying distributed source coding to sensor observations [4]. However, we will show that proper choice of codewords results in an equivalent result.

The rest of this paper is arranged as follows. We present the system model in Section II and derive the relevant TMSE expression. In Section III we introduce the notion of majorization and some related results that are required for our analysis. In Section IV we derive the optimal transmitter codewords, power levels and receiver filters by minimizing the TMSE and in Section V we establish an equivalence between between TMSE minimization and sum capacity maximization. Finally, we conclude with a summary and discussion of possible future research in Section VII.

II. Problem Statement

Assuming M users transmitting symbols using unit–norm codewords of length L in an additive white Gaussian channel, the signal at the receiver is given by:

$$\mathbf{r} = \mathbf{S}\mathbf{P}^{\frac{1}{2}}\mathbf{b} + \mathbf{n} \tag{1}$$

where,

$\mathbf{P}_{M\times M}$	: diag $(p_1 p_2 \dots p_M)$
p_i	: transmit power of ith transmitter
$\mathbf{S}_{L\times M}$	: $[\mathbf{s}_1\mathbf{s}_2 \dots \mathbf{s}_M]$
$\mathbf{s}_i$	: unit norm signature of ith transmitter
$\mathbf{b}$	: symbol vector
$\mathbf{n}$	: zero–mean Gaussian noise with variance $\sigma^2\mathbf{I}_L$

We also define $\mathbf{B}_{M\times M} = \mathrm{E}\left[\mathbf{b}\mathbf{b}^\top\right]$ as the symbol correlation matrix.

Assuming a linear receiver filter, $\mathbf{c}_i$, corresponding to the i^{th} transmitter, the filter output is given by:

$$y_i = \mathbf{r}^\top \mathbf{c}_i \tag{2}$$

The mean square error (MSE) corresponding to the i^{th} transmitter is given by,

$$\mathrm{MSE}_i = \mathrm{E}\left[\left(\mathbf{r}^\top \mathbf{c}_i - b_i\right)^2\right] \tag{3}$$

which allows us to define total MSE as

$$
\begin{aligned}
\mathrm{TMSE} &= \sum_{i=1}^{M} \mathrm{MSE}_i \\
&= \sum_{i=1}^{M} \mathbf{c}_i^\top \left(\mathbf{SP}^{\frac{1}{2}}\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top + \sigma^2\mathbf{I}_M\right)\mathbf{c}_i + M \\
&\quad - 2\sum_{i=1}^{M} \mathbf{c}_i^\top \mathbf{SP}^{\frac{1}{2}}\mathbf{b}b_i \\
&= \mathrm{tr}\left(\mathbf{C}^\top\mathbf{SP}^{\frac{1}{2}}\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top\mathbf{C} + \sigma^2\mathbf{C}^\top\mathbf{C}\right. \\
&\quad \left. -2\mathbf{C}^\top\mathbf{SP}^{\frac{1}{2}}\mathbf{B} + \mathbf{I}_M\right)
\end{aligned} \tag{4}
$$

The optimization problem can then be stated as follows:

$$
\min_{\mathbf{S},\mathbf{P},\mathbf{C}} \mathrm{TMSE} \quad \text{subject to } \mathrm{tr}\left(\mathbf{P}\right) = P_{\text{tot}} \tag{5}
$$

III. Majorization: Definitions and Some Key Results

We will need certain mathematical relationships as outlined in this section. A detailed survey of these inequalities and their properties may be found in [2].

Definition 1: Let $x = (x_{[1]}, x_{[2]}, \ldots, x_{[n]})$ and $y = (y_{[1]}, y_{[2]}, \ldots, y_{[n]})$ be non–decreasing sequences of real numbers. Then, x is *majorized* by y (denoted by $x \prec y$) if

$$
\sum_{i=1}^{k} x_{[i]} \le \sum_{i=1}^{k} y_{[i]}, \quad k = 1, 2, \ldots, n-1
$$

$$
\text{and,} \quad \sum_{i=1}^{n} x_{[i]} = \sum_{i=1}^{n} y_{[i]}
$$

Thus, majorization of x by y suggests that the components of x are "less spread out" or "more nearly equal" than the components of y.

An important example of majorization between two vectors is the following:

Example 1: For every $a \in \Re^n$ such that $\sum_{i=1}^{n} a_i = 1$,

$$
(a_1, a_2, \ldots, a_n) \succ \left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right)
$$

Definition 2: A real–valued function ϕ, defined on a set $\mathcal{A} \subset \mathcal{R}^n$, is Schur–convex on $\mathcal{A}$ if

$$
x \prec y \text{ on } \mathcal{A} \Rightarrow \phi(x) \le \phi(y)
$$

The function ϕ is strictly Schur–convex if $x \prec y$ and $x \ne y$ implies that $\phi(x) < \phi(y)$. Also, the function ϕ is Schur–concave if $-\phi$ is Schur–convex.

An important class of Schur–convex functions is the following:

Example 2: If $g : \Re \to \Re$ is convex and increasing, then $\phi(x) = \sum_{i=1}^{n} g(x_i)$ is increasing and Schur–convex.

IV. Optimal Transmitter Codewords, Power Levels and Receiver Structure

It is well–known [7] that the structure of the optimum linear receiver that minimizes the MSE is the MMSE receiver. For this problem, the expression for the optimum receiver was obtained as:

$$
\mathbf{C}^\star = \left(\mathbf{SP}^{\frac{1}{2}}\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top + \sigma^2\mathbf{I}_L\right)^{-1}\left(\mathbf{SP}^{\frac{1}{2}}\mathbf{B}\right) \tag{6}
$$

Substituting (6) in (4), the TMSE expression reduces to:

$$
\begin{aligned}
\mathrm{TMSE} &= M - \mathrm{tr}\left[\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top\left(\sigma^2\mathbf{I}_L + \mathbf{SP}^{\frac{1}{2}}\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top\right)^{-1}\mathbf{SP}^{\frac{1}{2}}\mathbf{B}\right] \\
&= M - \mathrm{tr}\left[\frac{\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top}{\sigma^2}\left\{\mathbf{I}_L - \frac{\mathbf{SP}^{\frac{1}{2}}\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top}{\sigma^2}\right.\right. \\
&\quad \left.\left. + \left(\frac{\mathbf{SP}^{\frac{1}{2}}\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top}{\sigma^2}\right)^2 - \cdots\right\}\mathbf{SP}^{\frac{1}{2}}\mathbf{B}\right] \\
&= M - \mathrm{tr}(B) + \sigma^2\mathrm{tr}\left[\left(\sigma^2\mathbf{B}^{-1} + \mathbf{P}^{\frac{1}{2}}\mathbf{S}^\top\mathbf{SP}^{\frac{1}{2}}\right)^{-1}\right]
\end{aligned} \tag{7}
$$

Note that $\mathbf{SP}^{\frac{1}{2}}\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top$ is positive definite, which implies that $\left(\mathbf{SP}^{\frac{1}{2}}\mathbf{BP}^{\frac{1}{2}}\mathbf{S}^\top + \sigma^2\mathbf{I}_L\right)$ is invertible. Also, it has been assumed in the above analysis that $\mathbf{B}^{-1}$ exists. However, it will be argued at the end of this section that invertibility of $\mathbf{B}$ is not necessary since it does not affect the structure of the optimum codewords.

Let $\mathbf{B} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{U}_1^\top$ and $\mathbf{A} = \mathbf{SP}^{\frac{1}{2}} = \mathbf{U}_2\mathbf{\Sigma}_2\mathbf{V}_2^\top$
where $\mathbf{\Sigma}_1 = \mathrm{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_M\right)$
such that, $\lambda_1 > \lambda_2 > \ldots > \lambda_M$
and $\mathbf{\Sigma}_2 = \left[\mathrm{diag}(\mu_1, \mu_2, \ldots, \mu_L), \mathbf{0}_{L\times(M-L)}\right]$

Note that $\mathbf{S}$ and $\mathbf{P}^{\frac{1}{2}}$ can be obtained from $\mathbf{A}$ as the normalized columns and norms of columns of $\mathbf{A}$ respectively.

Then, the optimization problem can be rewritten as:

$$
\min_{\mathbf{A}\in\mathcal{A}} \mathrm{tr}\left[\left(\sigma^2\mathbf{B}^{-1} + \mathbf{A}^\top\mathbf{A}\right)^{-1}\right] \tag{8}
$$

where, $\mathcal{A}$ is the set of all $L \times M$ matrices such that

$$
\mathrm{tr}(\mathbf{A}^\top\mathbf{A}) = \sum_{j=1}^{L} \mu_j^2 = P_{\text{tot}}
$$

Lemma 1: $\forall \mathbf{A} \in \mathcal{A}$, $\exists \widetilde{\mathbf{A}} \in \mathcal{A}$ such that $\mathrm{TMSE}(\widetilde{\mathbf{A}}) \le \mathrm{TMSE}(\mathbf{A})$ and $\widetilde{\mathbf{A}}^\top\widetilde{\mathbf{A}}$ commutes with $\mathbf{B}$.

Proof: Marshall and Olkin [2, Lemma 9.G.4] states the following:

$$
\det\left(\mathbf{G} + \mathbf{H}\right) \le \prod_{i=1}^{n}\left(\lambda_{[i]}(\mathbf{G}) + \lambda_{[n+1-i]}(\mathbf{H})\right) \tag{9}
$$

Define a function $\theta(\mathbf{A}) = \det\left(\sigma^2\mathbf{B}^{-1} + \mathbf{A}^\top\mathbf{A}\right)$.

Choose $\mathbf{G} = \sigma^2\mathbf{B}^{-1}$ and $\mathbf{H} = \mathbf{A}^\top\mathbf{A}$. Define $\widetilde{\mathbf{A}} = \mathbf{AQ}$, where $\mathbf{Q}$ is an orthogonal matrix chosen so that $\sigma^2\mathbf{B}^{-1}$ and $\mathbf{Q}^\top\mathbf{A}^\top\mathbf{AQ}$ commute and the eigenvector corresponding to the ith largest eigenvalue of $\sigma^2\mathbf{B}^{-1}$ is the same as that corresponding to the $(n+1-i)$th largest eigenvalue of $\widetilde{\mathbf{A}}^\top\widetilde{\mathbf{A}}$.

Note that $\widetilde{\mathbf{A}} \in \mathcal{A}$ since $\mathrm{tr}(\widetilde{\mathbf{A}}^\top\widetilde{\mathbf{A}}) = \mathrm{tr}\left(\mathbf{Q}^\top\mathbf{A}^\top\mathbf{A}\mathbf{Q}\right) = P_{\text{tot}}$

Using (9), $\theta(\widetilde{\mathbf{A}}) \geq \theta(\mathbf{A})$. Since $\theta(\mathbf{A})$ is Schur–concave and TMSE is Schur–convex in eigenvalues of $\left(\sigma^2\mathbf{B}^{-1} + \mathbf{A}^\top\mathbf{A}\right)$, it follows that $\mathrm{TMSE}(\widetilde{\mathbf{A}}) \leq \mathrm{TMSE}(\mathbf{A})$. ■

Lemma 1, combined with the fact that two matrices commute if and only if they share the same eigenvectors [6], restricts the optimization space to that subset of $\mathcal{A}$ for which the condition $\mathbf{V}_2 = \mathbf{U}_1$ holds. Note that this condition is sufficient but not necessary.

Substituting $\mathbf{V}_2 = \mathbf{U}_1$ in (7), the following two cases arise.

1) $M \geq L$:

$$\begin{aligned}\mathrm{TMSE} = M &- \mathrm{tr}\left[\mathbf{U}_1\boldsymbol{\Sigma}_1\mathbf{U}_1^\top + \right.\\ &\left.\sigma^2\mathrm{tr}\left\{\left(\sigma^2\mathbf{U}_1\boldsymbol{\Sigma}_1^{-1}\mathbf{U}_1^\top + \mathbf{U}_1\boldsymbol{\Sigma}_2^\top\boldsymbol{\Sigma}_2\mathbf{U}_1^\top\right)^{-1}\right\}\right]\\ = M &- \sum_{i=1}^{M}\lambda_i + \sigma^2\sum_{i=1}^{L}\frac{1}{\frac{\sigma^2}{\lambda_i}+\mu_i^2} + \sum_{i=L+1}^{M}\frac{\lambda_i}{\sigma^2}\end{aligned} \tag{10}$$

The Lagrangian corresponding to the optimization problem at hand can be written as follows:

$$\mathcal{L}\left(\mu_1^2,\ldots,\mu_L^2,\beta\right) = \mathrm{TMSE} + \beta\left(\sum_{i=1}^{L}\mu_i^2 - P_{\text{tot}}\right)$$

It is required that $\dfrac{\partial\mathcal{L}}{\partial\mu_i} = 0$ and $\dfrac{\partial\mathcal{L}}{\partial\beta} = 0$.

Using Kuhn-Tucker conditions [1], this leads to the following optimal solution:

$$\mu_i = \sqrt{\max\left(0, \frac{P_{\text{tot}}}{L} + \frac{\sigma^2}{L}\sum_{i=1}^{L}\frac{1}{\lambda_i} - \frac{\sigma^2}{\lambda_i}\right)} \tag{11}$$

Note that the optimal solution depends only on the first L eigenvalues of B, *i.e.*, $\{\lambda_i\}_{i=1}^{L}$. Also, the optimal solution has the property that if $\lambda_i \geq \lambda_j$, then $\mu_i \leq \mu_j$ as described in the proof for Lemma 1. It will now be shown that the ordering $O_1 : \lambda_1 > \lambda_2 > \ldots > \lambda_M$ achieves the optimal solution.

For ordering O_1, the eigenvalues $\{\gamma_i\}_{i=1}^{M}$ of $\sigma^2\mathbf{B}^{-1} + \mathbf{A}^\top\mathbf{A}$ are given by:

$$\gamma_i\left(O_1\right) = \begin{cases}\frac{\sigma^2}{L}\sum_{j=1}^{L}\frac{1}{\lambda_j} + \frac{P_{\text{tot}}}{L}, & i \leq L_1 \leq L\\ \frac{\sigma^2}{\lambda_i}, & i > L_1\end{cases}$$

It can be verified that for any other ordering O_2,

$$\{\gamma_i\}_{i=1}^{M}\left(O_2\right) \succ \{\gamma_i\}_{i=1}^{M}\left(O_1\right) \tag{12}$$

Now consider the function $f(x) = \frac{a}{x}$. It can be shown that $f(x)$ is convex if $a, x \in \Re^+$. Using Example 2, it follows that TMSE is a Schur–convex function in the eigenvalues $\{\gamma_i\}_{i=1}^{M}$ of $\sigma^2\mathbf{B}^{-1} + \mathbf{A}^\top\mathbf{A}$, which in conjunction with (12) implies that O_1 achieves the optimal solution.

Fig. 1. Waterfilling is achieved by distributing the sum of the eigenvalues of $\mathbf{A}$ over the eigenvalues of $\mathbf{B}^{-1}$.

2) $M<L$:

It can be verified that only the first M μ_is need to be optimized, and the remaining $(L-M)$ eigenvalues may be set to zero for obtaining the optimal solution.

In other words, for any M, L, the optimal solution corresponds to waterfilling (Fig. IV) the smallest $K(=\min(L,M))$ eigenvalues of $\mathbf{B}^{-1}$ with those of $\mathbf{A}^\top\mathbf{A}$, and aligning the eigenvectors of $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{B}$ as described in the proof of Lemma 1.

The above analysis assumed that $\mathbf{B}$ is invertible. However, the result holds even for a non–invertible $\mathbf{B}$ since it can be made invertible by adding an infinitesimally small perturbation matrix (while ensuring that $\mathbf{B}$ is still a correlation matrix). As a result, previously non–zero eigenvalues of $\mathbf{B}^{-1}$ will suffer very little change, while the other eigenvalues (previously zero) will now attain large finite values, but the corresponding dimensions will be avoided by the waterfilling solution [3].

V. Relationship Between TMSE and Sum Capacity

Verdu [8] derives the information theoretic capacity region for a white Gaussian synchronous CDMA system. Proceeding in a similar manner, the sum capacity for the system under consideration can be expressed as:

$$C_{\text{sum}} = \frac{1}{2}\log\left[\det\left(\sigma^2\mathbf{I}_L + \mathbf{S}\mathbf{P}^{\frac{1}{2}}\mathbf{B}\mathbf{P}^{\frac{1}{2}}\mathbf{S}^\top\right)\right] - \frac{L}{2}\log\sigma^2 \tag{13}$$

when we assume that the symbols b_i are jointly Gaussian known covariance $\mathbf{B}$.

We will now show that TMSE minimization and sum capacity maximization are equivalent problems. Using the notation defined previously,

$$C_{\text{sum}} = \frac{1}{2}\log\left[\det\left(\sigma^2\mathbf{I}_L + \mathbf{A}\mathbf{B}\mathbf{A}^\top\right)\right] - \frac{L}{2}\log\sigma^2 \tag{14}$$

Lemma 2: $\forall\mathbf{A} \in \mathcal{A}$, $\exists\widetilde{\mathbf{A}} \in \mathcal{A}$ such that $C_{\text{sum}}(\widetilde{\mathbf{A}}) \geq C_{\text{sum}}(\mathbf{A})$ and $\widetilde{\mathbf{A}}^\top\widetilde{\mathbf{A}}$ commutes with $\mathbf{B}$.

Proof: Similar to Lemma 1. ■

As in the case of TMSE, Lemma 2 when combined with the fact that two matrices commute if and only if they share the same eigenvectors [6], restricts the optimization space to that subset of $\mathcal{A}$ for which the condition $\mathbf{V}_2 = \mathbf{U}_1$ holds. Again, this condition is sufficient but not necessary.

A similar analysis reveals that sum capacity is Schur–concave under the total power constraint, and hence minimizing TMSE is equivalent to maximizing C_{sum}.

VI. Simulation

WHAT IS YOUR CONCLUSION FROM THIS FIGURE

Fig. 2. Plot of TMSE and sum capacity for 10000 independent trials (random choices of $\mathbf{S}$ and $\mathbf{P}^{\frac{1}{2}}$). The results (TMSE and sum capacity) of the trials were sorted before plotting for convenience. The number of users was $M = 6$ and the number of available dimensions was $L = 4$. $\mathbf{B}$ was a randomly chosen symbol correlation matrix. The dotted lines indicate the values of TMSE and sum capacity corresponding to the optimally chosen $\mathbf{S}$ and $\mathbf{P}^{\frac{1}{2}}$.

VII. Conclusion And Future Work

We have considered a sensor network model where sensors transmit correlated information to a receiver using a set of signature waveforms. We found the optimal signature set covariance for minimizing total mean square correlation (TMSE) at the receiver under a total power constraint. Then, any of a number of methods could be used to find actual codewords. **IS THIS TRUE?????? You did nto actually find codewords, you just showed waterfilling is the answer** The expressions for optimal receiver filter and transmit power levels were also formulated.

There remain several open issues such as when the number of users exceeds the number of available dimensions $M > L$. **Show why this is not considered here – it'll be missed by the casual reader since it's embedded in the assumptions in section IV** Another important area of work is to more carefully compare the efficiency of correlated data transmission using the scheme presented in the paper to that using Distributed Source Coding and define suitable metrics for comparing and contrasting the two. **I would think they would be EXACTLY equivalent. The only thing you need to do is compare power budgets and the capacity. Of course, that's probably not right since since the coding method implies short range links to do the combining and you're transmitting everything back to the receiver, right?** Also, throughout the paper we have considered that the different transmitters operate under a total power constraint. In a sensor network scenario, it might be more reasonable to expect each sensor to operate under an individual power constraint. Search for optimal codewords under individual power constraints is therefore another problem, and we expect ideas from [9], [5] to prove especially useful.

References

[1] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. John Wiley, New York, NY, 1991.

[2] A.W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications*. Academic Press, New York, NY, 1979.

[3] D. C. Popescu and C. Rose. Codeword Optimization for Uplink CDMA Dispersive Channels. *IEEE Transactions on Wireless Systems*. to appear.

[4] S.S. Pradhan and K. Ramchandran. Distributed Source Coding Using Syndromes (DISCUS). *IEEE Transactions on Information Theory*, 49(3):626–643, March 2003.

[5] C. Rose. Cdma codeword optimization: interference avoidance and convergence via class warfare. *IEEE Transactions on Information Theory*, 47(7):2368–2382, September 2001.

[6] Gilbert Strang. *Linear Algebra and its Applications*. International Thompson Publishing, Singapore, 1988.

[7] Sennur Ulukus and Aylin Yener. Iterative Transmitter and Receiver Optimization for CDMA Networks. *IEEE Transactions on Wireless Communications*, to appear.

[8] Sergio Verdu. Capacity Region of Gaussian CDMA Channels: The Symbol–Synchronous Case. *Proceedings of the 24th Allerton Conference on Communication, Control and Computing*, 1989.

[9] Pramod Viswanath and Venkat Anantharam. Optimal Sequences and Sum Capacity of Synchronous CDMA Systems. *IEEE Transactions on Information Theory*, 45(6):1984–1991, September 1999.