On the Optimality of MMSE-GDFE Pre-Processed Sphere Decoding

Sung-Jun Hwang and Philip Schniter
The Ohio State University, Dept. of Electrical and Computer Engineering,
2015 Neil Avenue, Columbus, OH 43210.
E-mail: {hwangsu, schniter}@ece.osu.edu

Abstract

In this work, we consider maximum likelihood (ML) sequence detection in MIMO linear channels corrupted by additive white Gaussian noise. While sphere decoding (SpD) algorithms have been developed to reduce the average complexity of ML detection, the average complexity of classical SpD can itself be impractical in low-SNR settings or when the channel is ill-conditioned. In response, sequential decoding algorithms that employ a pre-processing stage based on minimum mean-squared error generalized decision-feedback equalization (MMSE-GDFE) have been proposed. They are capable of near-ML detection at a complexity that remains low over a wide SNR range and/or with ill-conditioned channels. While it has always been assumed that MMSE-GDFE pre-processing compromises the ML-optimality of the downstream minimum-distance detector, we establish, in this work, that MMSE-GDFE pre-processing preserves ML-optimality under uncoded BPSK/QPSK signaling, regardless of channel dimension and rank. The implication is that, when BPSK/QPSK signaling is used, MMSE-GDFE pre-processing can be used in conjunction with efficient SpD algorithms for true ML detection. This is particularly attractive in moderate-to-low SNR ranges or with ill-conditioned or under-determined linear channels.

1 Introduction

In this paper, we consider communication over multiple antenna wireless channels that can be described by a multiple-input multiple-output (MIMO) linear transformation corrupted by additive white Gaussian noise. Assuming a uniform sequence distribution, it is well known that the maximum likelihood (ML) sequence detector is optimal in terms of minimizing frame error rate [1]. The complexity of ML detection often makes it impractical to implement, however, especially for long sequences. These issues have motivated the development of algorithms which are capable of solving the ML detection problem with an *average* complexity that can be significantly less than that of exhaustive search, at least for well-conditioned channels at high SNR [2–6]. Since these algorithms typically search among only those sequences that map to a spherical region around the observation, we refer to them collectively as sphere decoding (SpD) algorithms.

In certain situations, however, the complexity of even these SpD algorithms remains prohibitively high. This is often the case, for example, when the channel is ill-conditioned (e.g., under-determined) or when the SNR is low. These issues have led to the development of sub-optimal sequential decoding (SqD) algorithms based on, e.g., lattice decoding [7], SpD with statistical pruning [8], the Fano algorithm [7], or other breadth-first, depth-first, and best-first strategies [9, 10]. In fact, recent work on this topic has demonstrated that near-ML performance (i.e., a fraction-of-a-dB in SNR loss relative to ML) is possible at an average search complexity that is roughly cubic in the sequence dimension over a broad range of channel/SNR conditions [7].

Most sphere and sequential decoding algorithms start with a pre-processing stage that converts the MIMO channel matrix into upper triangular form, thereby giving the detection problem a tree structure that lends itself to sequential decoding. Traditionally, coordinate rotation, based on, e.g., QR decomposition with appropriate power ordering, is used to accomplish this. Because coordinate rotation preserves the Euclidean metric, the pre-processed channel is equivalent to the original channel as far as ML detection goes. However, coordinate rotation cannot improve the channel's condition number, and ill-conditioned channels are associated with high SpD/SqD search complexity [7].

Pre-processing based on minimum mean-squared error generalized decision feedback equalization (MMSE-GDFE) has been recently proposed as an alternative to the classical coordinate-rotation approaches [11]. The potential of MMSE-GDFE pre-processing follows from the fact that it yields a well-conditioned channel matrix. Not surprisingly, it is particularly suited to the case of under-determined linear channels [12]. In fact, when combined with appropriate ordering and lattice reduction, MMSE-GDFE lies at the core of several state-of-the-art near-ML SqD schemes [7].

The fundamental disadvantage of MMSE-GDFE pre-processing is that it does not preserve the Euclidean metric. Thus, the minimum-distance search of an MMSE-GDFE pre-processed system will not, in general, produce an ML sequence estimate. For this reason, MMSE-GDFE pre-processing has been used for *near*-ML detection as opposed to *true*-ML detection. In this paper, we establish the interesting fact that MMSE-GDFE pre-processing does not compromise the optimality of minimum-distance search in the case of uncoded BPSK or QPSK signaling. Furthermore, this property holds irrespective of the channel's dimension and rank. The same claim, however, cannot be made when non constant modulus (CM) alphabets are employed. The implication is that MMSE-GDFE pre-processing can be used to dramatically reduce the average complexity of true ML sequence detection in BPSK/QPSK systems.

The remainder of the paper is structured as follows. Section 2 formalizes the system model and reviews MMSE-GDFE pre-processing. Section 3 establishes the fact that MMSE-GDFE pre-processing preserves the optimality of the minimum-distance detector under BPSK or QPSK signaling. Section 4 demonstrates the reduction in ML search complexity that is possible using MMSE-GDFE pre-processing with greedy V-BLAST ordering [13], and shows the small, but non-zero, degradation in error performance caused by the use of MMSE-GDFE pre-processing with non-CM alphabets. Section 5 concludes.

We use $(\cdot)^T$ to denote the transpose, $\boldsymbol{I}_M$ to denote the $M \times M$ identity matrix, and $\|\cdot\|$ to denote the ℓ_2 norm. Also, $\mathbb{R}$ denotes the real field, $\mathbb{C}$ the complex field, and $\mathbb{Z}$ the integers.

2 Background

Consider a discrete-time complex-baseband communication system with transmitted signal $\underline{\boldsymbol{s}} \in \mathbb{C}^m$ and observed signal $\underline{\boldsymbol{x}} \in \mathbb{C}^n$. The k^{th} element of the column vector $\underline{\boldsymbol{s}}$, denoted by s_k, is chosen from a Q^2-ary quadrature amplitude modulation (QAM) alphabet, i.e., $s_k = a_k + jb_k$, where $a_k, b_k \in \mathbb{S}$ for

$$\mathbb{S} := \{-(Q-1)/2, -(Q-3)/2, \ldots, (Q-3)/2, (Q-1)/2\}.$$

The transmitted and observed signals are related by the MIMO channel matrix $\underline{\boldsymbol{H}} \in \mathbb{C}^{n \times m}$ and the additive white Gaussian noise (AWGN) vector $\underline{\boldsymbol{w}} \in \mathbb{C}^n$ according to

$$\underline{\boldsymbol{x}} = \underline{\boldsymbol{H}}\underline{\boldsymbol{s}} + \underline{\boldsymbol{w}}. \tag{1}$$

In the sequel, we focus on the real-valued system model

$$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{w} \tag{2}$$

where $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{H} \in \mathbb{R}^{N \times M}$, $\boldsymbol{s} \in \mathbb{R}^M$, and $\boldsymbol{w} \in \mathbb{R}^N$. More specifically, the elements of $\boldsymbol{s}$ belong to a Q-ary pulse amplitude modulation (PAM) alphabet $\mathbb{S}$, i.e., $\boldsymbol{s} \in \mathbb{S}^M$. The real-valued PAM model (2) can be related to the complex-valued QAM model (1) via

$$\begin{aligned}
\boldsymbol{x} &:= \begin{bmatrix} \mathrm{Re}\{\underline{\boldsymbol{x}}^T\} & \mathrm{Im}\{\underline{\boldsymbol{x}}^T\} \end{bmatrix}^T, \\
\boldsymbol{s} &:= \begin{bmatrix} \mathrm{Re}\{\underline{\boldsymbol{s}}^T\} & \mathrm{Im}\{\underline{\boldsymbol{s}}^T\} \end{bmatrix}^T, \\
\boldsymbol{w} &:= \begin{bmatrix} \mathrm{Re}\{\underline{\boldsymbol{w}}^T\} & \mathrm{Im}\{\underline{\boldsymbol{w}}^T\} \end{bmatrix}^T, \\
\boldsymbol{H} &:= \begin{bmatrix} \mathrm{Re}\{\underline{\boldsymbol{H}}\} & -\mathrm{Im}\{\underline{\boldsymbol{H}}\} \\ \mathrm{Im}\{\underline{\boldsymbol{H}}\} & \mathrm{Re}\{\underline{\boldsymbol{H}}\} \end{bmatrix},
\end{aligned}$$

when $M = 2m$ and $N = 2n$. Thus, the real-valued model (2) can be used to describe systems that employ either PAM or QAM alphabets.

When the channel matrix $\boldsymbol{H}$ is perfectly known to the receiver, it is well known that the ML estimate of $\boldsymbol{s}$ is obtained as the solution to the following minimum-distance detection problem [1].

$$\hat{\boldsymbol{s}}_{\mathsf{ML}} = \arg\min_{\boldsymbol{s} \in \mathbb{S}^M} \|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{s}\|^2. \tag{3}$$

Solving (3) by exhaustive search requires $\mathcal{O}(Q^M)$ operations, which is impractical unless M is very small. As discussed in the introduction, SpD can be used to solve (3) with an average complexity that is significantly lower, e.g., polynomial in M (when the channel is well behaved and the SNR is reasonably high) [2–6].

In SpD, pre-processing is first applied to convert (2) into upper triangular form. Traditionally this is accomplished using a QR decomposition of the form $\boldsymbol{H} = \boldsymbol{Q}\boldsymbol{R}$, where $\boldsymbol{Q}$ is orthogonal and $\boldsymbol{R}$ is upper triangular. The transformed observation $\boldsymbol{x}' := \boldsymbol{Q}^T\boldsymbol{x}$ obeys

$$\boldsymbol{x}' = \boldsymbol{R}\boldsymbol{s} + \boldsymbol{w}', \tag{4}$$

where the transformed noise $\boldsymbol{w}' := \boldsymbol{Q}^T\boldsymbol{w}$ is statistically equivalent to $\boldsymbol{w}$. Due to the noise equivalence, the detection problem (3) can be equivalently restated as

$$\hat{\boldsymbol{s}}_{\mathsf{ML}} = \arg\min_{\boldsymbol{s} \in \mathbb{S}^M} \|\boldsymbol{x}' - \boldsymbol{R}\boldsymbol{s}\|^2. \tag{5}$$

It is usually advantageous to incorporate power-ordering with QR decomposition. For example, [11] suggested that the columns of $\boldsymbol{H}$ might be ordered in non-decreasing Euclidean norm prior to QR decomposition. The same paper also proposed to employ the greedy V-BLAST ordering from [13], whereby the columns of $\boldsymbol{H}$ are ordered so that the minimum diagonal element of $\boldsymbol{R}$ is maximized. In our numerical simulations, we focus on the latter ordering scheme.

In MIMO channels, it is not unusual for the QR pre-processed channel matrix $\boldsymbol{R}$ to be ill-conditioned, or, when $N < M$, singular. In these cases, the complexity of SpD has been observed to grow significantly [7]. In response, MMSE-GDFE pre-processing [11, 12] was proposed as an alternative to QR pre-processing. It was motivated by the fact that, under perfect decision feedback, the MMSE-GDFE [14] maximizes signal to interference-plus-noise ratio (SINR) at the decision point.

We now outline the MMSE-GDFE pre-processing algorithm of [12]. Consider the "augmented" channel matrix $\tilde{\boldsymbol{H}}$ defined according to a signal-to-noise ratio parameter[1] $\gamma > 0$.

$$\tilde{\boldsymbol{H}} \;:=\; \begin{bmatrix} \boldsymbol{H} \\ \frac{1}{\sqrt{\gamma}}\boldsymbol{I}_M \end{bmatrix}. \tag{6}$$

Say that QR decomposition yields $\tilde{\boldsymbol{H}} = \tilde{\boldsymbol{Q}}\tilde{\boldsymbol{R}}$, where $\tilde{\boldsymbol{Q}} \in \mathbb{R}^{(N+M)\times M}$ has orthonormal columns and where $\tilde{\boldsymbol{R}} \in \mathbb{R}^{M\times M}$ is upper triangular with positive diagonal entries. If we partition $\tilde{\boldsymbol{Q}} = \begin{bmatrix} \boldsymbol{Q}_1 \\ \boldsymbol{Q}_2 \end{bmatrix}$ so that $\boldsymbol{Q}_1 \in \mathbb{R}^{N\times M}$ and $\boldsymbol{Q}_2 \in \mathbb{R}^{M\times M}$, then

$$\boldsymbol{H} \;=\; \boldsymbol{Q}_1\tilde{\boldsymbol{R}}. \tag{7}$$

MMSE-GDFE pre-processing (PP) transforms the observation according to $\tilde{\boldsymbol{x}} := \boldsymbol{Q}_1^T\boldsymbol{x}$ prior to minimum-distance (MD) decision making:

$$\hat{\boldsymbol{s}}_{\mathsf{PPMD}} \;=\; \arg\min_{\boldsymbol{s}\in\mathbb{S}^M} \|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{R}}\boldsymbol{s}\|^2. \tag{8}$$

Since $\boldsymbol{Q}_1$ is not, in general, orthogonal, we cannot claim that the PPMD rule (8) is equivalent to the ML rule (3). In other words, solutions to (8) are not, in general, ML. However, in the next section, we show that solutions to (8) are indeed ML when BPSK is employed in the PAM model (2) or when QPSK is employed in the QAM model (1).

3 Optimality of MMSE-GDFE Pre-Processed Estimates

In this section we establish that, when BPSK is employed in (2), MMSE-GDFE pre-processing does not compromise the ML-optimality of minimum-distance decision making. Since MMSE-GDFE pre-processing can yield a significant reduction in average SpD search complexity relative to standard QR-based pre-processing, especially when $N < M$, our finding implies that the complexity savings of MMSE-GDFE pre-processing can be leveraged for *true* ML (rather than only *near*-ML) detection.

Let us define $\mathcal{X}(\boldsymbol{s})$ as the set of (non-pre-processed) observations for which the ML estimate of the PAM sequence $\boldsymbol{s} \in \mathbb{S}^M$ will be error free. In other words,

$$\mathcal{X}(\boldsymbol{s}) \;:=\; \{\boldsymbol{x} : \|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{s}\| \le \|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{s}'\|,\ \forall \boldsymbol{s}' \in \mathbb{S}^M\}. \tag{9}$$

Since $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{w}$, we find that

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{s}\| &\le \|\boldsymbol{x} - \boldsymbol{H}\boldsymbol{s}'\| \\ \Leftrightarrow \|\boldsymbol{w}\| &\le \|\boldsymbol{w} - \boldsymbol{H}(\boldsymbol{s}' - \boldsymbol{s})\| \\ \Leftrightarrow 2\boldsymbol{w}^T\boldsymbol{H}(\boldsymbol{s}' - \boldsymbol{s}) &\le \|\boldsymbol{H}(\boldsymbol{s}' - \boldsymbol{s})\|^2. \end{aligned} \tag{10}$$

[1]Usually, $\gamma = \sigma_s^2/\sigma_w^2$ where σ_w^2 denotes the noise variance and σ_s^2 the symbol variance under the assumption that $\mathrm{E}\{\boldsymbol{s}\boldsymbol{s}^T\} = \sigma_s^2\boldsymbol{I}_M$ [12, 14]. As we will see, however, any $\gamma > 0$ suffices for the main result of this paper.

We now define $\mathcal{A}(\boldsymbol{s}) := \{\boldsymbol{s}' - \boldsymbol{s} : \boldsymbol{s}' \in \mathbb{S}^M\}$ as the set of error sequences relative to the true sequence $\boldsymbol{s}$. Putting (9) and (10) together, we find the following equivalence.

$$\boldsymbol{x} \in \mathcal{X}(\boldsymbol{s}) \;\;\Leftrightarrow\;\; 2\boldsymbol{w}^T \boldsymbol{H}\boldsymbol{\alpha} \;\leq\; \|\boldsymbol{H}\boldsymbol{\alpha}\|^2,\ \forall \boldsymbol{\alpha} \in \mathcal{A}(\boldsymbol{s}). \tag{11}$$

Next let us define $\tilde{\mathcal{X}}(\boldsymbol{s})$ as the set of MMSE-GDFE pre-processed observations for which the MD estimate of the PAM sequence $\boldsymbol{s}$ will be error free. In other words,

$$\tilde{\mathcal{X}}(\boldsymbol{s}) \;:=\; \{\tilde{\boldsymbol{x}} : \|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{R}}\boldsymbol{s}\| \leq \|\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{R}}\boldsymbol{s}'\|,\ \forall \boldsymbol{s}' \in \mathbb{S}^M\} \tag{12}$$

where

$$\begin{aligned} \tilde{\boldsymbol{x}} &= \boldsymbol{Q}_1^T \boldsymbol{x} \\ &= \boldsymbol{Q}_1^T(\boldsymbol{H}\boldsymbol{s} + \boldsymbol{w}) \\ &= \tilde{\boldsymbol{R}}\boldsymbol{s} + \underbrace{(\boldsymbol{Q}_1^T \boldsymbol{H} - \tilde{\boldsymbol{R}})\boldsymbol{s} + \boldsymbol{Q}_1^T \boldsymbol{w}}_{:=\,\boldsymbol{n}}. \end{aligned} \tag{13}$$

Recall that, since $\boldsymbol{Q}_1$ is typically non-orthogonal, we cannot claim $\boldsymbol{Q}_1^T \boldsymbol{H} = \tilde{\boldsymbol{R}}$. Repeating the arguments in (10), we obtain the following equivalence.

$$\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}(\boldsymbol{s}) \;\;\Leftrightarrow\;\; 2\boldsymbol{n}^T \tilde{\boldsymbol{R}}\boldsymbol{\alpha} \;\leq\; \|\tilde{\boldsymbol{R}}\boldsymbol{\alpha}\|^2,\ \forall \boldsymbol{\alpha} \in \mathcal{A}(\boldsymbol{s}). \tag{14}$$

Finally, let us define the $\boldsymbol{Q}_1^T$-transformation of the region $\mathcal{X}(\boldsymbol{s})$:

$$\mathcal{X}_{\boldsymbol{Q}_1^T}(\boldsymbol{s}) \;:=\; \{\boldsymbol{Q}_1^T \boldsymbol{x} : \boldsymbol{x} \in \mathcal{X}(\boldsymbol{s})\}. \tag{15}$$

Note that $\boldsymbol{x} \in \mathcal{X}(\boldsymbol{s}) \;\Leftrightarrow\; \tilde{\boldsymbol{x}} \in \mathcal{X}_{\boldsymbol{Q}_1^T}(\boldsymbol{s})$ since $\tilde{\boldsymbol{x}} := \boldsymbol{Q}_1^T \boldsymbol{x}$. Thus (11) implies

$$\tilde{\boldsymbol{x}} \in \mathcal{X}_{\boldsymbol{Q}_1^T}(\boldsymbol{s}) \;\;\Leftrightarrow\;\; 2\boldsymbol{w}^T \boldsymbol{H}\boldsymbol{\alpha} \;\leq\; \|\boldsymbol{H}\boldsymbol{\alpha}\|^2,\ \forall \boldsymbol{\alpha} \in \mathcal{A}(\boldsymbol{s}). \tag{16}$$

Lemma 1. *$\mathcal{X}_{\boldsymbol{Q}_1^T}(\boldsymbol{s}) \subset \tilde{\mathcal{X}}(\boldsymbol{s})$ for any $\boldsymbol{s} \in \{-\frac{1}{2}, \frac{1}{2}\}^M$.*

Proof. Examining the left side of the inequality in (14), we see that

$$\begin{aligned} \boldsymbol{n}^T \tilde{\boldsymbol{R}}\boldsymbol{\alpha} &= \left(\boldsymbol{w}^T \boldsymbol{Q}_1 + \boldsymbol{s}^T \left(\boldsymbol{H}^T \boldsymbol{Q}_1 - \tilde{\boldsymbol{R}}^T\right)\right) \tilde{\boldsymbol{R}}\boldsymbol{\alpha} \\ &= \boldsymbol{w}^T \boldsymbol{H}\boldsymbol{\alpha} + \boldsymbol{s}^T \left(\boldsymbol{H}^T \boldsymbol{H} - \tilde{\boldsymbol{R}}^T \tilde{\boldsymbol{R}}\right) \boldsymbol{\alpha} \\ &= \boldsymbol{w}^T \boldsymbol{H}\boldsymbol{\alpha} - \gamma^{-1} \boldsymbol{s}^T \boldsymbol{\alpha}, \end{aligned} \tag{17}$$

where we have used the facts that $\boldsymbol{Q}_1 \tilde{\boldsymbol{R}} = \boldsymbol{H}$ and $\tilde{\boldsymbol{R}}^T \tilde{\boldsymbol{R}} = \boldsymbol{H}^T \boldsymbol{H} + \gamma^{-1} \boldsymbol{I}_M$. The latter fact also implies

$$\|\tilde{\boldsymbol{R}}\boldsymbol{\alpha}\|^2 \;=\; \|\boldsymbol{H}\boldsymbol{\alpha}\|^2 + \gamma^{-1}\|\boldsymbol{\alpha}\|^2. \tag{18}$$

Equations (17)-(18) can be used to rewrite (14) as

$$\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}(\boldsymbol{s}) \;\;\Leftrightarrow\;\; 2\boldsymbol{w}^T \boldsymbol{H}\boldsymbol{\alpha} \;\leq\; \|\boldsymbol{H}\boldsymbol{\alpha}\|^2 + \underbrace{\gamma^{-1}\left(\|\boldsymbol{\alpha}\|^2 + 2\boldsymbol{s}^T \boldsymbol{\alpha}\right)}_{:=\,D},\ \forall \boldsymbol{\alpha} \in \mathcal{A}(\boldsymbol{s}). \tag{19}$$

It is readily verified that $D \geq 0$ for any $\boldsymbol{s} \in \{-\frac{1}{2}, \frac{1}{2}\}^M$. Thus, comparing (19) to (16), we see that $\mathcal{X}_{\boldsymbol{Q}_1^T}(\boldsymbol{s}) \subset \tilde{\mathcal{X}}(\boldsymbol{s})$ for arbitrary $\boldsymbol{s} \in \{-\frac{1}{2}, \frac{1}{2}\}^M$. Note that it is not possible to guarantee $D \geq 0$ for larger alphabets, e.g., $\boldsymbol{s} \in \{-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}\}^M$. □

Corollary 2. *Consider model (2) with arbitrary $\boldsymbol{H}$, $\boldsymbol{w}$, and $\boldsymbol{s} \in \{-\frac{1}{2}, \frac{1}{2}\}^M$. If $\hat{\boldsymbol{s}}_{\text{ML}} = \boldsymbol{s}$, then $\hat{\boldsymbol{s}}_{\text{PPMD}} = \boldsymbol{s}$.*

Proof. Recall that $\hat{\boldsymbol{s}}_{\text{ML}} = \boldsymbol{s}$ iff $\boldsymbol{x} \in \mathcal{X}(\boldsymbol{s})$, and that $\hat{\boldsymbol{s}}_{\text{PPMD}} = \boldsymbol{s}$ iff $\tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}(\boldsymbol{s})$, where $\tilde{\boldsymbol{x}} := \boldsymbol{Q}_1^T \boldsymbol{x}$. Then, from (15) and Lemma 1, we see that

$$\hat{\boldsymbol{s}}_{\text{ML}} = \boldsymbol{s} \;\; \Leftrightarrow \;\; \boldsymbol{x} \in \mathcal{X}(\boldsymbol{s}) \;\; \Leftrightarrow \;\; \tilde{\boldsymbol{x}} \in \mathcal{X}_{\boldsymbol{Q}_1^T}(\boldsymbol{s}) \;\; \Rightarrow \;\; \tilde{\boldsymbol{x}} \in \tilde{\mathcal{X}}(\boldsymbol{s}) \;\; \Leftrightarrow \;\; \hat{\boldsymbol{s}}_{\text{PPMD}} = \boldsymbol{s} \tag{20}$$

when $\mathbb{S} = \{-\frac{1}{2}, \frac{1}{2}\}$. □

Corollary 2 establishes the fact that MMSE-GDFE pre-processing does not affect the optimality of ML detection when BPSK signaling is used. Since the use of QPSK symbols in model (1) gives a special case of BPSK symbols in model (2), Corollary 2 also applies to QPSK systems in complex-valued channels. From the proof of Lemma 1, it is clear that the property $\mathcal{X}_{\boldsymbol{Q}_1^T}(\boldsymbol{s}) \subset \tilde{\mathcal{X}}(\boldsymbol{s})$ will not hold for larger PAM/QAM alphabets, though, implying that MMSE-GDFE pre-processing can render MD decision making sub-optimal when non-BPSK/QPSK alphabets are used. This sub-optimality will be investigated further in Section 4.

It is interesting to note that the MMSE-GDFE property holds for arbitrary positive γ. (This fact will be confirmed numerically in Section 4.) Thus, ML estimates can be obtained via MMSE-GDFE pre-processing without knowledge of SNR. However, the search complexity remains a function of γ, and numerical experiments suggest choosing γ as specified in [12] (see also Footnote 1).

4 Numerical Experiments

In this section we present the results of three numerical experiments. In all experiments, the MIMO channel matrix $\boldsymbol{H} \in \mathbb{R}^{N \times M}$ was generated from i.i.d. zero-mean Gaussian elements with variance M^{-1}, and the noise vector from i.i.d. zero-mean Gaussian elements with variance $(4\mathsf{SNR})^{-1}$ for BPSK and $(4\mathsf{SNR}/5)^{-1}$ for 4-PAM, recalling that $\mathbb{S} = \{-\frac{1}{2}, \frac{1}{2}\}$ for BPSK and $\mathbb{S} = \{-\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}\}$ for 4-PAM. In other words, SNR is the signal-energy to noise-energy ratio at each receive antenna. Unless noted otherwise, we set $\gamma = \mathsf{SNR}$, as specified in [12].

The first experiment shows the typical reduction of search complexity that comes from the use of MMSE-GDFE pre-processing and ordering in place of the traditional QR pre-processing and ordering. We employed the greedy ordering scheme suggested by [11] which was originally proposed for V-BLAST in [13]. Figure 1 shows average Schnorr-Euchner (SE) SpD search complexity for a system with $M = N = 32$ under BPSK signaling (or, equivalently, $m = n = 16$ under QPSK signaling). By "search complexity" we mean the number of real multiplications plus additions per frame consumed by the SpD search stage[2] of the SE-SpD algorithm from [11]. Note that the complexity is reported on a $\log_N$ scale. Here the SE-SpD sphere radius was initialized at 1.5 times the average distance between the observation and the closest lattice point. Figure 1 demonstrates that MMSE-GDFE pre-processing can lead to significant complexity savings over a moderate-to-low SNR range. For example, a factor of about 10 in complexity savings can be observed for SNR at 8 dB. More detailed investigations of MMSE-GDFE complexity savings can be found in [7, 11, 12].

The second experiment compares the frame error rate (FER) achieved by the ML detector (3) to that achieved by the MMSE-GDFE pre-processed MD detector (8) under BPSK signaling for several combinations of M and N and for several choices of MMSE-GDFE parameter γ.

[2]This definition assumes slow-fading, where the matrix computations associated with pre-processing could be amortized over many frames, thereby making the SpD search complexity dominant.

Specifically, Fig. 2 examines $(M, N) \in \{(6, 8), (8, 8), (8, 6)\}$, which includes over- and under-determined linear channels, and $\gamma \in \{\frac{1}{10}\mathsf{SNR}, \mathsf{SNR}, 10\mathsf{SNR}\}$. Consistent with Corollary 2, the FER of MMSE-GDFE sphere decoding is identical to that of ML detection in all cases. In this experiment, the ML detector was implemented using the QR pre-processed SE-SpD algorithm from [11].

The third experiment verifies the sub-optimality of MMSE-GDFE pre-processed MD estimates when non-BPSK/QPSK alphabets are used. In Fig. 3, the FER of the ML detector is compared to that of the MMSE-GDFE pre-processed SpD for a 4-PAM system with $M = N = 8$. Observe that the FER degradation caused by MMSE-GDFE pre-processing is small but measurable. As before, the ML detector was implemented using the QR pre-processed SE-SpD algorithm from [11].

5 Conclusions

In this paper, we established that MMSE-GDFE pre-processing does not compromise the ML-optimality of minimum-distance decisions for MIMO systems with uncoded BPSK or QPSK signaling. This property holds for systems of arbitrary size (i.e., over- or under-determined linear channels), though not for larger PAM/QAM alphabets. The result is attractive because MMSE-GDFE pre-processing is known to yield significant reductions in the average search complexity of sphere decoding algorithms, especially in moderate-to-low SNR ranges and/or with ill-conditioned/under-determined linear channels.

References

[1] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 4th ed., 2001.

[2] U. Fincke and M. Pohst, “Improved methods for calculating vectors of short length in a lattice,” *Math. of Comput.*, vol. 44, pp. 463–471, Apr. 1985.

[3] C. P. Schnorr and M. Euchner, “Lattice basis reduction: Improved practical algorithms and solving subnet sum problems,” *Math. of Programming*, vol. 66, pp. 181–191, 1994.

[4] E. Viterbo and J. Boutros, “A universal lattice code decoder for fading channels,” *IEEE Trans. on Information Theory*, vol. 45, pp. 1639–1642, July 1999.

[5] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, “Closest point search in lattices,” *IEEE Trans. on Information Theory*, vol. 48, pp. 2201–2214, Aug. 2002.

[6] W. H. Mow, “Universal lattice decoding:: Principles and recent advances,” *Wireless Communications and Mobile Computing*, vol. 3, pp. 553–569, Aug. 2003.

[7] A. Murugan, H. El Gamal, M. O. Damen, and G. Caire, “A unified framework for tree search decoding: Rediscovering the sequential decoder,” *IEEE Trans. on Information Theory*, submitted May 2005. available at http://www.ece.osu.edu/ helgamal/publications.html.

[8] R. Gowaikar and B. Hassibi, “Efficient statistical pruning algorithms for maximum likelihood decoding,” in *Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 49–52, 2003.

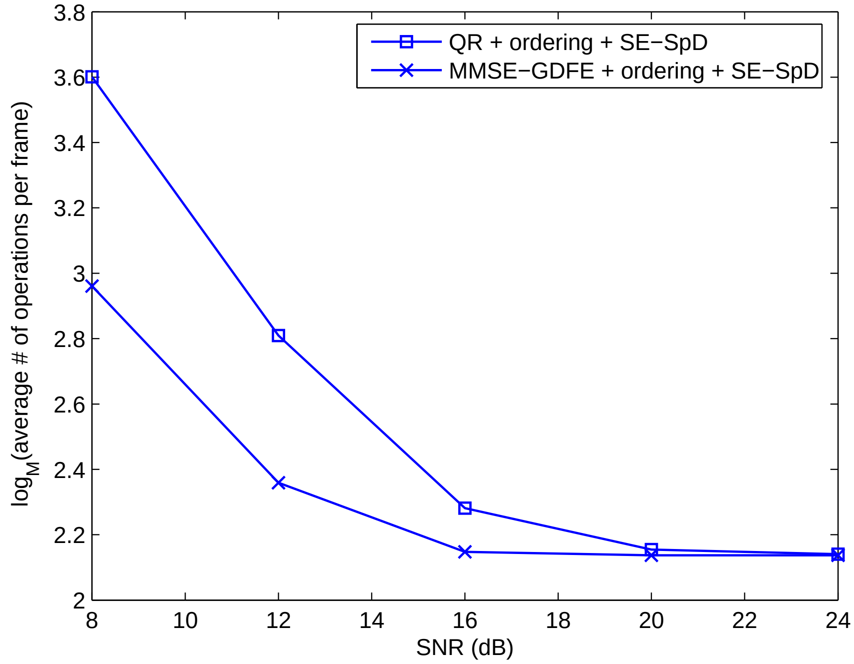

Figure 1: A comparison of the per-frame average search complexity of Schnorr-Euchner SpD with two forms of pre-processing: QR versus MMSE-GDFE, both with greedy ordering. A BPSK system of dimension $M = N = 32$ was employed and $\gamma = \text{SNR}$ was used in MMSE-GDFE.

[9] J. B. Anderson and S. Mohan, "Sequential decoding algorithms: A survey and cost analysis," *IEEE Trans. on Communications*, vol. 32, pp. 169–172, 1984.

[10] S. J. Simmons, "Breadth-first trellis decoding with adaptive effort," *IEEE Trans. on Communications*, vol. 38, pp. 3–12, Jan. 1990.

[11] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. on Information Theory*, vol. 49, pp. 2389–2402, Oct. 2003.

[12] M. O. Damen, H. El Gamal, and G. Caire, "MMSE-GDFE lattice decoding for underdetermined linear channels," in *Proc. Conf. on Information Science and Systems*, 2004.

[13] G. J. Foschini, G. Golden, R. Valanzuela, and P. Wolniansky, "Simplified processing for high spectral efficiency wireless communication employing multi-element arrays," *IEEE Journal on Selected Areas In Communications*, vol. 17, pp. 1841–1852, Nov. 1999.

[14] N. Al-Dhahir and A. H. Sayed, "The finite-length MIMO MMSE-DFE," *IEEE Trans. on Signal Processing*, vol. 48, pp. 2921–2936, Oct. 2000.

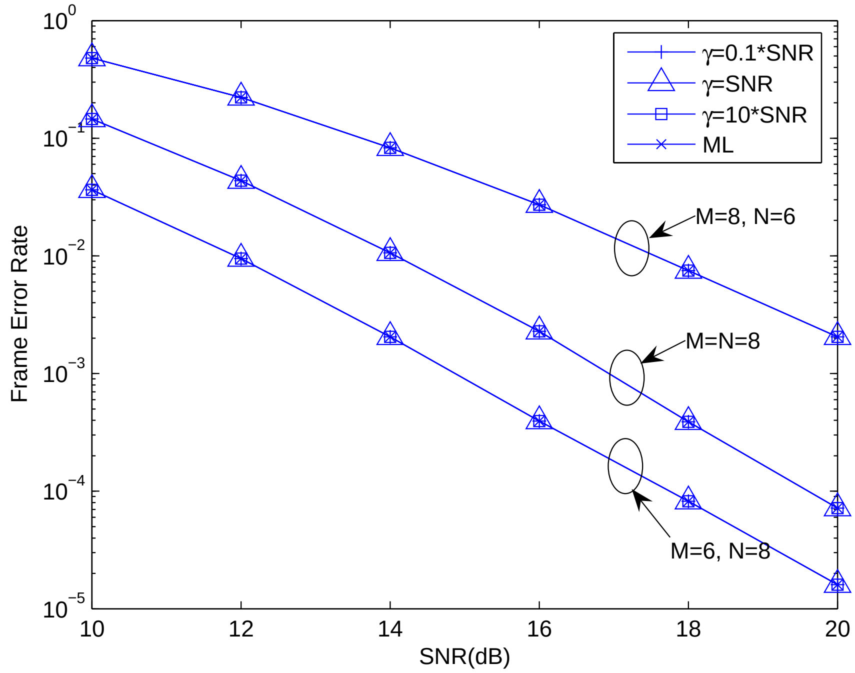

Figure 2: Frame error rate of the ML detector (via QR pre-processed SpD) versus the MMSE-GDFE pre-processed SpD for a BPSK system under several channel dimensions (M, N) and several values of MMSE-GDFE parameter γ.

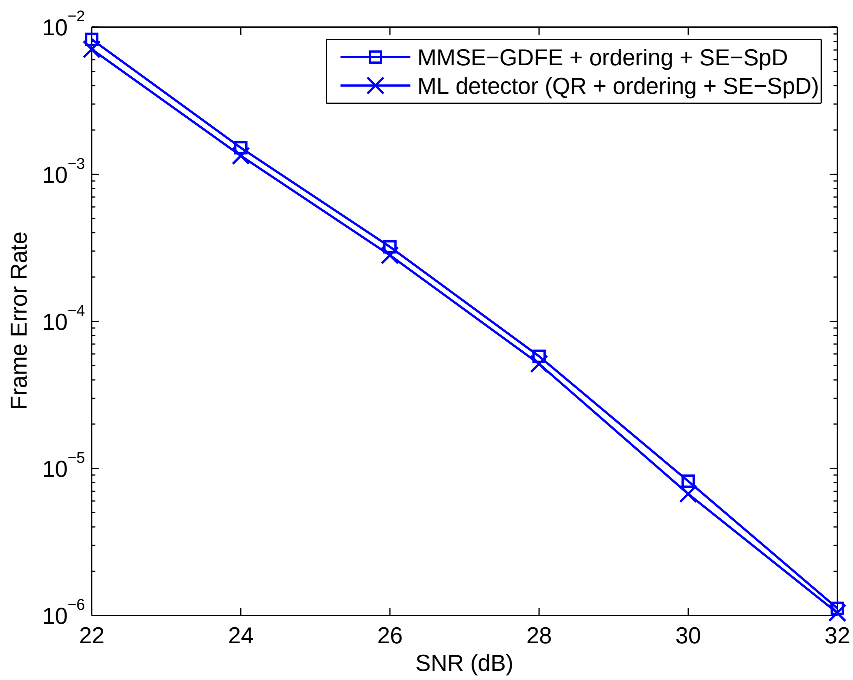

Figure 3: Frame error rate of the ML detector (via QR pre-processed SpD) versus the MMSE-GDFE pre-processed SpD for a 4-PAM system with $M = N = 8$.