Augmented Lattice Reduction for MIMO Decoding

Laura Luzzi, Ghaya Rekaya-Ben Othman, Member, IEEE, and Jean-Claude Belfiore, Member, IEEE

Abstract—Lattice reduction algorithms, such as the Lenstra-Lenstra-Lovasz (LLL) algorithm, have been proposed as preprocessing tools in order to enhance the performance of suboptimal receivers in multiple-input multiple-output (MIMO) communications. A different approach, introduced by Kim and Park, allows to combine right preprocessing and detection in a single step by performing lattice reduction on an “augmented channel matrix”. In this paper we propose an improvement of the augmented matrix approach which guarantees a better performance. We prove that our method attains the maximum receive diversity order of the channel. Simulation results evidence that it significantly outperforms LLL reduction followed by successive interference cancellation (SIC) while requiring a moderate increase in complexity. A theoretical bound on the complexity is also derived.

Index Terms—Lattice reduction-aided decoding, LLL algorithm, right preprocessing.

I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) systems can provide high data rates and reliability over fading channels. In order to achieve optimal performance, maximum likelihood decoders such as the Sphere Decoder may be employed; however, their complexity grows prohibitively with the number of antennas and the constellation size, posing a challenge for practical implementation.

On the other hand, suboptimal receivers such as zero forcing (ZF) or successive interference cancellation (SIC) do not preserve the diversity order of the system [10]. Right preprocessing using lattice reduction has been proposed in order to enhance their performance [19, 4, 8]. In particular, the classical Lenstra-Lenstra-Lovasz (LLL) algorithm for lattice reduction, whose average complexity is polynomial in the number of antennas, has been proven to achieve the optimal receive diversity order in the spatial multiplexing case [17]. Very recently, it has also been shown that combined with regularization techniques such as minimum mean square error generalized decision-feedback equalizer (MMSE-GDFE) left preprocessing, lattice reduction-aided decoding is optimal in terms of the diversity-multiplexing tradeoff [7]. However, the shift between the error probability of maximum likelihood (ML) detection and LLL-ZF (respectively, LLL-SIC) detection increases greatly for a large number of antennas [12].

Recently a new lattice reduction-aided decoding technique combining the right preprocessing stage and the detection stage in a single step was proposed in [11]. This technique, called Improved Lattice Reduction, consists in LLL-reducing an augmented lattice which is a function of the channel matrix and of the received signal. An estimate of the transmitted message can then be recovered directly from the change of basis matrix. Improved Lattice Reduction is equivalent to LLL-SIC decoding in terms of probability of error.

In this paper we present a different kind of augmented lattice reduction decoding which significantly enhances its performance by carefully choosing the augmented lattice parameters. In the coherent case, MIMO decoding amounts to solving an instance of the closest vector problem (CVP) in a finite subset of the lattice generated by the channel matrix. Following an idea of Kannan [9], our strategy is to reduce the CVP to the shortest vector problem (SVP) by embedding the n-dimensional lattice generated by the channel matrix into an (n + 1)-dimensional lattice. We show that for a suitable choice of the embedding, the transmitted message can be recovered directly from the coordinates of the shortest vector of the augmented lattice with provably high probability.

In general, the LLL algorithm is not guaranteed to solve the SVP; however, it certainly finds the shortest vector in the lattice in the particular case where the minimum distance is exponentially smaller than the other successive minima. Equivalently, we can say that “the LLL algorithm is an SVP-oracle when the lattice gap is exponential in the lattice dimension” [16]. An appropriate choice of the embedding ensures that this condition is satisfied.

Thanks to this property, we can prove that our method also achieves the receive diversity of the channel. Numerical simulations evidence that augmented lattice reduction significantly outperforms LLL-SIC detection while requiring a moderate increase in complexity. A theoretical (albeit pessimistic) bound on the complexity is also derived.

This paper is organized as follows: in Section II we introduce the system model and basic notions concerning lattice reduction, and summarize the existing lattice reduction-aided decoding schemes. In Section III we describe augmented lattice reduction decoding, and in Sections IV and V we analyze its performance and complexity, both theoretically and through numerical simulations.

II. PRELIMINARIES

A. System model and notation

We consider a MIMO system with M transmit and N receive antennas such that $M \leq N$ using spatial multiplexing.

1 Note that the worst-case number of iterations of the LLL algorithm applied to the MIMO context is unbounded, as has been proved in [8]. However, the tail probability of the number of iterations decays exponentially, so that in many cases high complexity events can be regarded as negligible with respect to the target error rate (see [7, Theorem 3]).

2 Actually, LLL-ZF and LLL-SIC suboptimal decoding correspond to two classical techniques for finding approximate solutions of the CVP, due to Babai: the rounding algorithm and nearest plane algorithm respectively [1].
The complex received signal is given by
\[ y_c = H_c x_c + w_c, \] (1)
where \( x_c \in \mathbb{C}^M, y_c, w_c \in \mathbb{C}^N, H_c \in M_{N \times M}(\mathbb{C}) \). The transmitted vector \( x_c \) belongs to a finite constellation \( \mathcal{S} \subseteq \mathbb{Z}[i]^M \); the entries of the channel matrix \( H_c \) are supposed to be i.i.d. complex Gaussian random variables with zero mean and variance per real dimension equal to \( \frac{1}{2} \), and \( w_c \) is the Gaussian noise with i.i.d. entries of zero mean and variance \( N_0 \). We consider the coherent case where \( H_c \) is known at the receiver.

Separating the real and imaginary part, the model can be rewritten as
\[ y = Hx + w, \] (2)
in terms of the real-valued vectors
\[ y = \left( \Re(y_c) \Im(y_c) \right) \in \mathbb{R}^n, \quad x = \left( \Re(x_c) \Im(x_c) \right) \in \mathbb{Z}^m \]
and of the equivalent real channel matrix
\[ H = \left( \begin{array}{c} \Re(H_c) \\ \Im(H_c) \end{array} \right) \in M_{n \times m}(\mathbb{R}). \]
Here \( n = 2N, m = 2M \).

The maximum likelihood decoded vector is given by
\[ \hat{x}_{ML} = \arg \min_{x_c \in \mathcal{S}} \| Hx_c - y_c \|, \]
where \( \| \cdot \| \) denotes the Euclidean norm, and \( \hat{x} = \left( \Re(x_c) \Im(x_c) \right) \).

### B. Lattice reduction

An \( m \)-dimensional real lattice in \( \mathbb{R}^n \) is the set of points
\[ \mathcal{L}(H) = \{ Hx | x \in \mathbb{Z}^m \}, \]
where \( H \in M_{n \times m}(\mathbb{R}) \). We denote by \( d_H \) the minimum distance of the lattice, that is the smallest norm of a nonzero vector in \( \mathcal{L}(H) \). More generally, for all \( 1 \leq i \leq m \) one can define the \( i \)-th successive minimum of the lattice as follows:
\[ \lambda_i(H) = \inf \{ r > 0 \ | \ \exists v_1, \ldots, v_i \text{ linearly independent in } \mathcal{L}(H) \text{ s.t. } \| v_j \| \leq r \ \forall j \leq i \}. \] (3)

We recall that two matrices \( H, H' \) generate the same lattice if and only if \( H' = Hu \) with \( U \in M_{m \times m}(\mathbb{R}) \) unimodular, that is, \( U \) has integer entries and \( \det(U) = \pm 1 \).

Lattice reduction algorithms allow to find a new basis \( H' \) for a given lattice \( \mathcal{L}(H) \) such that the basis vectors are shorter and nearly orthogonal. Orthogonality can be measured by the absolute value of the coefficients \( \mu_{i,j} = \frac{\langle h_i, h_j^* \rangle}{\| h_j^* \|} \) in the Gram-Schmidt orthogonalization (GSO) of the basis.

We recall the following useful property of GSO: the length of the smallest of the Gram-Schmidt vectors \( h_i^* \) is always less or equal to the minimum distance \( d_{H} \) of the lattice [15]. In other words,
\[ d_H \geq a(H) = \min_{1 \leq i \leq m} \| h_i^* \|. \] (4)

A basis \( H \) is said to be \textit{LLL-reduced} [14] if its Gram-Schmidt coefficients \( \mu_{i,j} \) and Gram-Schmidt vectors satisfy the following properties:

1) **Size reduction:**
\[ |\mu_{k,l}| \leq \frac{1}{2}, \quad 1 \leq l < k \leq m, \]

2) **Lovasz condition:**
\[ \| h_k^* + \mu_{k,k-l} h_{k-l}^* \|^2 \geq \delta \| h_{k-l}^* \|^2, \quad 1 < k \leq m, \]
where \( \delta \in (\frac{1}{8}, 1) \) (a customary choice is \( \delta = \frac{3}{8} \)).

The LLL algorithm is summarized in Algorithms 1 and 2. Given a full-rank matrix \( H \in M_{n \times m}(\mathbb{R}) \), it computes an LLL-reduced version \( H_{\text{red}} = Hu \), with \( U \in M_{m \times m}(\mathbb{Z}) \) unimodular, and outputs the columns \( \{ h_i \} \) and \( \{ u_i \} \) of \( H_{\text{red}} \) and \( U \) respectively.

### Algorithm 1: The LLL algorithm

1. \( U = I_m \)
2. Compute the GSO of \( H \)
3. \( k \leftarrow 2 \)
4. While \( k \leq m \) do
5. RED(k,k-1)
6. if \( \| h_k^* + \mu_{k,k-l} h_{k-l}^* \|^2 < \delta \| h_{k-l}^* \|^2 \) then
7. swap \( h_k \) and \( h_{k-1} \)
8. swap \( u_k \) and \( u_{k-1} \)
9. update GSO
10. \( k \leftarrow \max(k-1, 2) \)
11. end
12. else
13. for \( l = k-2, \ldots, 1 \) do
14. RED(k,l)
15. end
16. \( k \leftarrow k + 1 \)
17. end
18. end

### Algorithm 2: Size reduction RED(k,l)

1. if \( |\mu_{k,l}| > \frac{1}{2} \) then
2. \( h_k \leftarrow h_k - \| \mu_{k,j} \| h_j \)
3. \( u_k \leftarrow u_k - \| \mu_{k,j} \| u_j \)
4. for \( j = 1, \ldots, l-1 \) do
5. \( \mu_{k,j} \leftarrow \mu_{k,j} - \| \mu_{k,l} \| \mu_{k,j} \)
6. end
7. \( \mu_{k,l} \leftarrow \mu_{k,l} - \| \mu_{k,l} \| \)
8. end

We list here some properties of LLL-reduced bases that we will need in the sequel. First of all, the LLL algorithm finds at least one basis vector whose length is not too far from the minimum distance \( d_H \) of the lattice. The following inequality holds for any \( m \)-dimensional LLL-reduced basis \( H \) [3]:
\[ \| h_1 \| \leq \alpha \frac{1}{\sqrt{m}} d_H, \] (5)
where \( \alpha = \frac{1}{1-\delta} \) (\( \alpha = 2 \) if \( \delta = \frac{3}{8} \)). Moreover, the first basis vector cannot be too big compared to the Gram-Schmidt vectors \( \{ h_i^* \} \):
\[ \| h_1 \| \leq \alpha \frac{1}{\sqrt{m}} \| h_i^* \|, \quad \forall 1 \leq i \leq m. \]
In particular, if \( j = \arg \min_{1 \leq i \leq m} \| h_i \| \),

\[
d_H \leq \| h_i \| \leq \alpha \frac{\pi}{2} \| h_j \| = \alpha \frac{\pi}{2} a(H) \leq \alpha \frac{\pi}{2} a(H).
\]  

(6)

C. Lattice reduction-aided decoding

In this section we briefly review existing detection schemes which use the LLL algorithm to preprocess the channel matrix, in order to improve the performance of suboptimal decoders such as ZF or SIC [19, 18, 4].

Let \( H_{\text{red}} = HU \) be the output of the LLL algorithm on \( H \).

We can rewrite the received vector as \( y = H_{\text{red}}U^{-1}x + w \).

- The **LLL-ZF decoder** outputs

\[
\hat{x}_{\text{LLL-ZF}} = Q_S \left( U \left( H_{\text{red}}^\dagger y \right) \right),
\]

where \( H_{\text{red}}^\dagger = (H_{\text{red}}^T H_{\text{red}})^{-1} H_{\text{red}}^T \) is the Moore-Penrose pseudoinverse of \( H_{\text{red}} \). \( \lfloor \cdot \rfloor \) denotes component-wise rounding to the nearest integer and \( Q_S \) is a quantization function that forces the solution to belong to the constellation \( S \).

- The **LLL-SIC decoder** performs the QR decomposition \( H_{\text{red}} = QR \), computes \( \tilde{y} = Q^T y \), finds by recursion \( \tilde{x} \) defined by

\[
\tilde{x}_m = \begin{bmatrix} \tilde{y}_m \\ r_{m1} \\ \vdots \\ r_{m1} \\ \vdots \\ r_{m1} \\ \vdots \\ r_{m1} \\ \vdots \\ r_{m1} \end{bmatrix},
\]

\[
\tilde{x}_i = \tilde{y}_i - \sum_{j=1}^{m} r_{ij} \tilde{x}_j, \quad i = m-1, \ldots, 1,
\]

and finally outputs \( \hat{x}_{\text{LLL-SIC}} = Q_S \left( U \tilde{x} \right) \).

D. Improved lattice reduction

Recently, Kim and Park [11] have proposed a new decoding technique based on the LLL algorithm, called **Improved lattice reduction**, which allows to estimate the transmitted message directly from the unimodular reduction matrix.

Let \( y \) be the (real) received vector in the model (2). Consider the \( (n + 1) \times (m + 1) \) augmented matrix

\[
\tilde{H} = \begin{pmatrix} H & -y \\ 0_{1 \times m} & t \end{pmatrix} = \begin{pmatrix} h_{1,1} & \cdots & h_{1,m} & -y_1 \\ \vdots & \cdots & \vdots \\ h_{n,1} & \cdots & h_{n,m} & -y_n \\ 0 & \cdots & 0 & t \end{pmatrix}
\]  

(7)

where \( t > 0 \) is a parameter to be determined.

Let \( H_{\text{red}} = \tilde{H} \hat{U} \) be the output of the LLL algorithm on \( \tilde{H} \).

- The **Improved lattice reduction decoder** outputs

\[
\hat{x}_{\text{LRR}} = Q_S (\tilde{u}_{m+1,1}, \ldots, \tilde{u}_{m,m+1})^T
\]  

(8)

In [11], the parameter \( t \) is chosen such that \( t > r_{m,m} \), where \( R = (r_{i,j}) \) is the upper triangular matrix in the the QR decomposition \( H_{\text{red}} = QR \). This ensures that the Lovasz condition on the last column of the augmented matrix \( \tilde{H} \) is always verified. Therefore, LLL-reducing \( H \) amounts to LLL-reducing the submatrix \( \tilde{H} \) and then performing a final round of size reduction without swaps on the last column.

Although not explicitly stated in [11], the performance of improved lattice reduction is exactly the same as LLL-SIC. In fact, it is not hard to prove by induction that

\[
Q_S (\tilde{u}_{m+1}) = Q_S \left( \sum_{i=1}^{n} \tilde{u}_i \right) = \left( \hat{x}_{\text{LLL-SIC}} \right)
\]

where \( u_i \) and \( \tilde{u}_i \) are the columns of \( U \) and \( \tilde{U} \) respectively.

III. AUGMENTED LATTICE REDUCTION

We introduce here a different decoder based on the augmented matrix (7) which, by carefully choosing the parameter \( t \), greatly enhances the performance with respect to Improved lattice reduction.

Observe that the points of the augmented lattice \( \mathcal{L}(\tilde{H}) \) are of the form \( (Hx' - qy, q^T t) \), \( x' \in \mathbb{Z}^n \), \( q \in \mathbb{Z} \). In particular, the vector \( v = (Hx - y, t) \) belongs to the augmented lattice. We will show that for a suitable choice of \( t \), and supposing that the noise is small enough, \( v \) is the shortest vector in the lattice and the LLL algorithm finds this vector. That is, \( \pm v \) is the first column of \( H_{\text{red}} = \tilde{H} \hat{U} \), the output of LLL algorithm on \( \tilde{H} \). Clearly, since \( \tilde{H} \) is full-rank with probability 1, in this case the first column of the change of basis matrix \( \hat{U} \) is \( \pm \hat{u}_1 \). Thus we can “read” the transmitted message directly from the change of basis matrix \( \hat{U} \).

To summarize, in order to decode we can perform the LLL algorithm on \( \tilde{H} \), and given the output \( \tilde{H}_{\text{red}} = \tilde{U} \hat{U} \), we can choose

\[
\hat{x} = Q_S \left( \frac{1}{\tilde{u}_{m+1,1}} (\tilde{u}_{1,1}, \ldots, \tilde{u}_{m,1})^T \right)
\]  

(9)

where \( \hat{U} = (\tilde{u}_{i,j}) \). (Division by \( \tilde{u}_{m+1,1} \) is needed because under the previous assumptions, \( \tilde{u}_{m+1,1} = \pm 1 \) and \( (\tilde{u}_{1,1}, \ldots, \tilde{u}_{m,1})^T = \pm \hat{x} \).

The previous decoder can be improved by including all the columns of \( H_{\text{red}} \) in the search for the vector \( v \). Let

\[
s_k = \frac{1}{\tilde{u}_{m+1,k}} (\tilde{u}_{1,k}, \ldots, \tilde{u}_{m,k})^T, \quad k = 1, \ldots, m.
\]

If \( \exists k \in \{1, \ldots, m\} \) such that \( |\tilde{u}_{m+1,k}| = 1 \), we define

\[
k_{\min} = \arg \min_{k \text{ s.t. } |\tilde{u}_{m+1,k}| = 1} \| Hs_k - y \|
\]

otherwise \( k_{\min} = 1 \). Then the **Augmented Lattice Reduction decoder** outputs

\[
\hat{x}_{\text{ALR}} = Q_S \left( [s_{k_{\min}}] \right)
\]  

(10)

IV. PERFORMANCE ANALYSIS AND SIMULATION RESULTS

A. Diversity of augmented lattice reduction

In this paragraph we will investigate the performance of augmented lattice reduction. We begin by proving that our method, like LLL-ZF and LLL-SIC, attains the maximum receive diversity gain of \( N \), for an appropriate choice of the parameter \( t \) in (7). The diversity gain \( d \) of a decoding scheme is defined as follows:

\[
d = - \lim_{\rho \to \infty} \frac{\log(P_c)}{\log(\rho)},
\]
where $\rho$ denotes the signal to noise ratio and $P_e$ is the error probability.

**Proposition 1.** If the augmented lattice reduction is performed using $t = \varepsilon a(H_{\text{red}})$, where $a(H_{\text{red}})$ is the length of the smallest vector in the GSO of $H_{\text{red}}$, and $\varepsilon \leq \frac{1}{\sqrt{2^m d_H}}$, then it achieves the maximum receive diversity $N$.

**Remark.** It is essential to use $a(H_{\text{red}})$ in place of $a(H)$. In fact, for general bases $H$ that are not LLL-reduced, there is no lower bound of the type (6) limiting how small the smallest Gram-Schmidt vector can be. For $a(H_{\text{red}})$, putting together the bounds (4) and (6), we obtain

$$
\frac{d_H}{a(H_{\text{red}})} \leq a(H_{\text{red}}) \leq d_H
$$

(11)

Note that the LLL reduction of $H$ does not entail any additional complexity, since it is the same as the LLL reduction on the first $m$ columns of $H$. In fact the parameter $t$ can be chosen during the LLL reduction of $H$, after carrying out the LLL algorithm on the first $m$ columns.

In order to prove the previous Proposition, we will show that in the $(n+1)$-dimensional lattice $L(H)$ there is an exponential gap between the first two successive minima. Then, using the estimate (5) on the norm of the first vector in an LLL-reduced basis, one can conclude that in this particular case the LLL algorithm finds the shortest vector in the lattice $L(H)$ with high probability. This, in turn, allows to recover the closest lattice vector $Hx$ to $y$ in $L(H)$ supposing that the noise $w$ is small enough.

The following definition makes the notion of “gap” more precise:

**Definition.** Let $v$ be a shortest nonzero vector in the lattice $L(H)$, and let $\gamma > 1$. $v$ is called $\gamma$-unique if $\forall u \in L(H)$,

$$
\|u\| \leq \gamma \|v\| \Rightarrow u \text{ is an integer multiple of } v.
$$

Note that if $v$ is a $\gamma$-unique shortest vector, there is a gap between the first two successive minima of the lattice $L(H)$ defined in equation (3): $\lambda_2(H) > \gamma \lambda_1(H) = \gamma d_H$. We now prove the existence of such a gap under suitable conditions:

**Lemma 1.** Let $H$ be the matrix defined in (7), and let $t = \varepsilon a(H_{\text{red}})$, with $\varepsilon \leq \frac{1}{\sqrt{2^m d_H}}$

Suppose $v = \left(\begin{array}{c} Hx - y \end{array}\right)$ is an $\alpha\frac{\gamma}{2}$-unique shortest vector of $L(H)$.

**Remark.** Observe that the hypothesis on $\|w\|$ implies in particular that $\|w\| < \frac{d_H}{2}$ and $Hx$ is indeed the closest lattice point to $y$.

**Proof:** We need to show that any vector $u \in L(H)$ that is not a multiple of $v$ must have length greater than $\alpha\frac{\gamma}{2} \|v\|$.

By contradiction, suppose that $\exists u = \left(\begin{array}{c} Hx' - qy \end{array}\right) \in L(H)$ linearly independent from $v$ such that $\|u\| \leq \alpha\frac{\gamma}{2} \|v\|$. Since $\|u\| \geq \|q\| t$,

$$
\|q\| \leq \frac{\|u\|}{t} \leq \frac{\alpha\frac{\gamma}{2} \|v\|}{t}.
$$

On the other side, $\|u\| \leq \alpha\frac{\gamma}{2} \|v\|$ implies that also $\|Hx' - qy\| \leq \alpha\frac{\gamma}{2} \|v\|$. Consider

$$
\|Hx' - qHx\| \leq \|Hx' - qy\| + \|qy - qHx\| \leq
$$

$$
\leq \alpha\frac{\gamma}{2} \|v\| + |q| \|y - Hx\| \leq \alpha\frac{\gamma}{2} \|v\| + \frac{\alpha\frac{\gamma}{2} \|v\| \|w\|}{t} \leq
$$

$$
\leq \alpha\frac{\gamma}{2} t \left(1 + \frac{w^2}{t^2}\right) \left(1 + \frac{\|w\|}{t}\right)
$$

(12)

The bound (11) on $a(H_{\text{red}})$ implies

$$
\frac{\varepsilon}{\alpha\frac{\gamma}{2} d_H} \leq t \leq \varepsilon d_H, \quad \|w\| < t
$$

Using these inequalities, we can bound the expression (12) with

$$
\alpha\frac{\gamma}{2} \varepsilon d_H 2\sqrt{2} < d_H.
$$

Thus $\|Hx' - qHx\| < d_H$. But this is a contradiction because $Hx' - qHx \in L(H)$ and is nonzero since $v$ and $u$ are linearly independent. Therefore $v$ is an $\alpha\frac{\gamma}{2}$-unique. (Since the last coordinate of $v$ in the basis $H$ is 1, $v$ cannot be a nontrivial multiple of another lattice vector.)

**Lemma 2.** Under the hypotheses of Lemma 1, the augmented lattice reduction methods (9) and (10) correctly decode the transmitted signal $x$.

**Proof:** Let $\tilde{H}_{\text{red}} = \tilde{H} \tilde{U}$ denote the output of the LLL reduction of $\tilde{H}$, and let $\tilde{h}_1 = \tilde{H} \left(\begin{array}{c} x' \\ q \end{array}\right) = \left(\begin{array}{c} (Hx' - qy) \\ qt \end{array}\right)$ be its first column. The property (5) of LLL reduction in dimension $m+1$ entails that $\|\tilde{h}_1\| \leq \alpha\frac{\gamma}{2} d_H$. But since $v = \left(\begin{array}{c} Hx - y \\ t \end{array}\right)$ has been shown to be an $\alpha\frac{\gamma}{2}$-unique in the previous Lemma, it means that $\tilde{h}_1$ is an integer multiple of $v$: $\exists a \in \mathbb{Z} \setminus \{0\}$ such that $\tilde{h}_1 = av$. In particular $a = qt$, so $a = q$ and $\tilde{h}_1 = qv$. Then by definition of $\tilde{H}$, $\tilde{h}_1 = \tilde{H} \left(\begin{array}{c} qx \\ q \end{array}\right)$. This means that the first column of the reduction matrix $\tilde{U}$ is $\left(\begin{array}{c} qx \\ q \end{array}\right)$, and so $x_{\text{ALR}} = Q_S(|u_1|) = Q_S(qx/q) = x$ and the augmented lattice reduction methods (9) and (10) correctly decode the transmitted message.

(Observable that this is possible only if $|q| = 1$, since $a$ is also a multiple of $q$ and $\tilde{U}$ is unimodular.)

Thus for any channel realization $H$, we have the following bound on the error probability for the augmented lattice reduction method:

$$
P_{e,\text{ALR}}(H) \leq P\{\|w\| > \varepsilon' d_H\},
$$

with $\varepsilon' = \alpha\frac{\gamma}{2} \varepsilon$. To conclude the proof of Proposition 1, we need to show that

$$
\lim_{\rho \to \infty} \frac{-\log P\{\|w\| > \varepsilon' d_H\}}{\log \rho} \geq N
$$

This turns out to be true. In fact, it has been shown in [17] (Proof of Theorem 2), that for any constant $\varepsilon_M$ depending only on the number of transmit antennas, there exist two constants
Fig. 1. Performance comparison of augmented lattice reduction with LLL-ZF, LLL-SIC and Improved Lattice Reduction with MMSE-GDFE preprocessing for a 6 × 6 uncoded MIMO system using 16-QAM. The LLL algorithm is performed using δ = 4/6.

\[ C_1, C_2 > 0 \text{ depending on } c_M, M \text{ and } N \text{ such that} \]
\[ P\{|\|w\| > c_M \|d_H\|\} \leq \frac{C_1 \ln(\rho)}{\rho N} \text{ for } N = M, \]
\[ P\{|\|w\| > c_M \|d_H\|\} \leq \frac{C_2}{\rho^2} \text{ for } N > M. \]

Thus augmented lattice reduction achieves the maximum receive diversity \( N \) with the choice \( t = \varepsilon a(H_{\text{red}}) \).

B. Simulation results

Figure 1 shows the comparison of augmented lattice reduction with LLL-SIC detection, both using MMSE-GDFE preprocessing. The simulations refer to an uncoded 6 × 6 MIMO system using 16-QAM (quadrature amplitude modulation) constellations. The improved decoder defined in (10) is used for augmented lattice reduction.

Two versions of augmented lattice reduction with different values of the parameter \( \varepsilon \) are compared. Clearly it is preferable to choose \( \varepsilon \) as big as possible in order to minimize the probability \( P\{|\|w\| > c_M \|d_H\|\} \). Version 1 corresponds to the choice \( \varepsilon = \frac{1}{\alpha} d_H \), the highest value of \( \varepsilon \) that verifies the hypothesis of Proposition 1. At the symbol error rate (SER) of \( 1 \cdot 10^{-4} \), its performance is within 0.8 dB from ML decoding and gains 1.9 dB with respect to LLL-SIC decoding.

Version 2 corresponds to a value of \( \varepsilon \) optimized by computer search (experimentally, this is around \( 2^{-12} \)), whose performance is within only 0.4 dB of ML decoding at the SER of \( 1 \cdot 10^{-4} \). From now on, we will always consider this optimized version. For higher values of \( \varepsilon \), we are not able to prove that the LLL algorithm finds the shortest lattice vector in \( L(H) \).

However, it is well-known that the LLL algorithm performs much better on average than the theoretical bounds predict.

The gain with respect to LLL-SIC decoding increases with the number of antennas: it is 3.5 dB for an 8 × 8 MIMO system, at the SER of \( 10^{-4} \). On the other side, augmented lattice reduction is still within 0.8 dB from ML performance (see Fig. 2).

V. COMPLEXITY ANALYSIS AND SIMULATION RESULTS

In this section we propose to estimate the additional complexity required by augmented lattice reduction with respect to LLL-ZF and LLL-SIC decoding. We are interested in the complexity order as a function of the number of transmit and receive antennas. For the sake of simplicity, we consider the case \( n = m \).

A. Theoretical bounds

The complexity of LLL reduction of a Gaussian matrix \( H \) has been studied in [8] and [13]. As we have seen in Section II, every instance of the LLL-SIC (respectively LLL-ZF) decoder consists of three main phases: GSO, the main body of the LLL algorithm and detection.

The full GSO which is performed at the beginning of the LLL algorithm requires \( O(m^3) \) elementary operations. The number \( K(H) \) of iterations (Lovasz tests) in the main body of the LLL algorithm for a fixed realization \( H \) of the channel is bounded by [5, 13, 8]

\[ K(H) \leq m^2 \log_2 \left( \frac{A(H)}{a(H)} \right) + m, \quad (13) \]

where \( A(H) \) and \( a(H) \) denote respectively the maximum and minimum norm of the Gram-Schmidt vectors of \( H \). For general \( H \), \( K(H) \) can be arbitrarily large. However, it was shown in [13] and [8] that \( E(K(H)) \sim O(m^2 \ln m) \).

Each iteration of the classical LLL algorithm requires \( O(m^2) \) elementary operations. However, this estimate can be reduced to \( O(m) \) by considering a weaker version of the algorithm, called effective LLL reduction, which has been proposed in [13]. It consists in removing the size reduction operations RED(k,l) for \( l = k - 2, \ldots, 1 \) (lines 13-15 in Algorithm 1). As shown in [13], effective LLL reduction outputs the same Gram-Schmidt vectors as the standard LLL reduction. Therefore, SIC decoding using effective LLL reduction has the
same performance as LLL-SIC. In the case of ZF decoding, a final cycle of size-reduction, requiring \(O(m^3)\) operations, has to be performed on the effectively LLL-reduced basis.

Since LLL reduction outputs the GSO of \(H_{\text{red}}\), its QR decomposition is easily obtained and so the complexity of SIC detection is negligible. The QR decomposition can be also used to perform ZF detection, by back substitution using \(R\).

In conclusion, the average complexity of LLL-SIC and LLL-ZF detection is of the order \(O(m^4 \ln m)\) if using the standard LLL algorithm and \(O(m^3 \ln m)\) for effective LLL.

For fixed \(H\), we can use the estimate (13) to obtain a bound of the number of iterations of the LLL reduction of \(H\). The GSO of \(H\) yields

\[
\begin{pmatrix}
h_1^* & \cdots & h_m^* & 0_{m \times 1}
\end{pmatrix}.
\]

Therefore \(a(H) \geq \min(t, a(H)) = \min(\varepsilon a(H_{\text{red}}), a(H))\). LLL reduction increases the minimum of the Gram-Schmidt vectors [5], so \(a(H_{\text{red}}) \geq a(H)\), and \(a(H) \geq \varepsilon a(H)\). On the other side \(t < a(H_{\text{red}}) \leq A(H_{\text{red}}) \leq A(H)\) and so \(A(H) = \max(t, A(H)) = A(H)\). Then

\[
K(\tilde{H}) \leq (m+1)^2 \log_2 \left( \frac{A(H)}{a(H)} \right) + m + 1 \leq (m+1)^2 \ln \left( \frac{A(H)}{\varepsilon a(H)} \right) + m + 1
\]

where \(c = \log \frac{1}{\sqrt{\varepsilon}}\). Following [8], we can estimate the average \(E[K(\tilde{H})]\), recalling that \(\frac{A(H)}{\varepsilon a(H)} \leq k(H)\), the condition number of \(H\), and that [2]

\[
E[\ln k(H)] \leq \ln m + 2.24.
\]

We thus obtain

\[
E[K(\tilde{H})] \leq \frac{(m+1)^2}{c} (-\ln \varepsilon + E[\ln k(H)]) + m + 1 \leq \frac{(m+1)^2}{c} (-\ln \varepsilon + \ln m + 2.24) + m + 1. \tag{14}
\]

For the choice \(\varepsilon = \frac{1}{2\sqrt{2\pi eN}}\), the average number of iterations of the main loop of the LLL algorithm using augmented lattice reduction is at most of the order of \(O(m^3)\). This yields a complexity bound of \(O(m^3)\) operations using standard LLL reduction and \(O(m^2)\) operations using effective LLL followed by a final cycle of size reduction. (The latter is necessary to obtain the best performance for the improved decoder (10)).

B. Simulation results

Our complexity simulations evidence the fact that the upper bounds (13) and (14) on the average number of iterations of the LLL algorithm for LLL-aided linear decoding and the augmented lattice reduction method are both quite pessimistic. The number of iterations for both methods appears in fact to be quadratic in practice, see Fig. 3.

Augmented lattice reduction and LLL-SIC appear to have the same complexity order but augmented lattice reduction requires roughly 1.5 times the number of flops\(^4\) as LLL-SIC, at least for small values of \(m\) (Fig. 4).

Our simulations are performed for \(\beta = \frac{3}{2}\). Two versions of augmented lattice reduction are considered, one using standard LLL, the other using effective LLL followed by a final cycle of size reduction. As was shown also in [13], for small dimensions (\(N < 10\)) effective LLL reduction doesn’t offer a real complexity advantage with respect to standard LLL.

C. Case of slow fading channels

If the channel matrix \(H\) varies so slowly that it can be considered constant during the transmission of several frames, the complexity of LLL-SIC and LLL-ZF decoding is greatly reduced since lattice reduction needs to be performed only once. In the case of augmented lattice reduction, the matrix \(H\) depends also on the received vector \(y\), so an update of the

\(^4\)Here we define a “flop” as any floating-point operation (addition, multiplication, division or square root).
LLL algorithm will still be required for each frame. However, a significant complexity gain will be achieved, since in many cases of interest, LLL-reducing the first \( m \) columns roughly accounts for \( 1/4 \) of the total complexity, as suggested by Fig. 4.

\[ D. \text{Complex LLL reduction} \]

A generalization of the LLL algorithm to complex lattices has been applied to MIMO decoding in [6]. It has been show experimentally in [6] that the complex versions of LLL-ZF and LLL-SIC decoding have essentially the same performance of their real counterparts but with substantially reduced complexity.

A complex version of the augmented lattice reduction can be implemented by LLL-reducing the \(( N + 1 ) \times ( M + 1 )\)-dimensional matrix

\[
\mathbf{H}_c = \begin{pmatrix} \mathbf{H} & -\mathbf{y}_c \\ 0_{1 \times N} & t \end{pmatrix},
\]

and allows to save about 40\% of computational costs without any change in performance.

\[ \text{VI. Conclusions} \]

In this paper, we have presented a low-complexity decoding method for MIMO systems based on augmented lattice reduction, which improves a previous technique introduced in [11]. Simulation results evidence that our augmented lattice reduction technique has a substantial performance gain with respect to the classical LLL-ZF and LLL-SIC decoders, while having approximately the same complexity order.

\[ \text{ACKNOWLEDGEMENTS} \]

We wish to thank Cong Ling for helpful discussions. We also would like to thank the three anonymous reviewers for their valuable suggestions and comments that helped to improve the quality of this paper, in particular concerning the complexity analysis.

\[ \text{REFERENCES} \]


Laura Luzzi was born in Pavia, Italy, in 1980. She received the degree (Laurea) in Mathematics from the University of Pisa, Italy, in 2003 and the Ph.D. degree in Mathematics for Technology and Industrial Applications from Scuola Normale Superiore, Pisa, Italy, in 2007. During the completion of this work, she was a postdoctoral fellow at the Department of Communications and Electronics at TELECOM ParisTech, Paris, France. Her research interests include algebraic space-time coding and decoding, and continued fractions.

Ghaya Rekaya-Ben Othman was born in Tunis, Tunisia, in 1977. She received the degree in electrical engineering from ENIT, Tunisia, in 2000 and the Ph.D. degree from the Ecole Nationale Supérieure des Télécommunications (ENST) Paris, France, in 2004. Since 2005, she has been with the Department of Communications and Electronics, TELECOM ParisTech, Paris, France, as Assistant Professor. Her research interests are in space-time coding and lattice reduction and decoding.

Jean-Claude Belfiore received the Diplôme d’ingénieur (Eng. degree) from Supelec in 1985, the doctorat (PhD) from Ecole Nationale Supérieure des Télécommunications (ENST) Paris in 1989 and the “Habilitation à diriger des Recherches” (HdR) from Université Pierre et Marie Curie (UPMC) in 2001. He was enrolled in Alcatel in 1985 and then in 1989 in ENST, Paris where he is now a full professor. He is carrying out research at the “Laboratoire de Traitement et Communication de l’Information” (LTCI), joint research laboratory between ENST and the “Centre National de la Recherche Scientifique” (CNRS), UMR 5141, where he is in charge of research activities in the fields of wireless communications, coding for wireless networks and space-time coding. Jean-Claude Belfiore has made pioneering contributions on signal design for wireless communication systems, space-time coding, cooperative and multi-user communications.