Augmented Lattice Reduction for Low-Complexity MIMO Decoding

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

Télécom ParisTech, 46 rue Barrault, 75013 Paris, France
E-mail: $\{\texttt{luzzi}, \texttt{rekaya}, \texttt{belfiore}\}$@$\texttt{telecom} - \texttt{paristech.fr}$

***Abstract*—Lattice reduction algorithms, such as the LLL algorithm, have been proposed as preprocessing tools in order to enhance the performance of suboptimal receivers in MIMO communications.**

In this paper we introduce a new kind of lattice reduction-aided decoding technique, called *augmented lattice reduction*, which recovers the transmitted vector directly from the change of basis matrix, and therefore doesn't entail the computation of the pseudo-inverse of the channel matrix or its QR decomposition.

We prove that augmented lattice reduction attains the maximum receive diversity order of the channel; simulation results evidence that it significantly outperforms LLL-SIC detection without entailing any additional complexity.

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 [8]. Right preprocessing using *lattice reduction* has been proposed in order to enhance their performance [16, 3, 15]. In particular, the classical LLL algorithm for lattice reduction, whose average complexity is polynomial in the number of antennas[1], has been proven to achieve the optimal receive diversity order in the spatial multiplexing case [14]. Very recently, it has also been shown that combined with regularization techniques such as MMSE-GDFE left preprocessing, lattice reduction-aided decoding is optimal in terms of diversity-multiplexing tradeoff [5]. However, the shift between the error probability of ML detection and LLL-ZF (respectively, LLL-SIC) detection increases greatly for a large number of antennas [10].

Recently a new lattice reduction-aided decoding technique combining the right preprocessing stage and the detection stage in a single step was proposed in [9]. 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 performance.

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[2]. Following an idea of Kannan [7], 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.

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". 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 without entailing any additional complexity.

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. Finally, the conclusions of our study are presented in Section VI.

[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 [6]. 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 [5], 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].

978-1-4244-8016-6/10/$26.00 ©2010 IEEE

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. The complex received signal is given by

$$\mathbf{y}_{\mathrm{c}} = \mathbf{H}_{\mathrm{c}}\mathbf{x}_{\mathrm{c}} + \mathbf{w}_{\mathrm{c}}, \tag{1}$$

where $\mathbf{x}_{\mathrm{c}} \in \mathbb{C}^M$, $\mathbf{y}_{\mathrm{c}}$, $\mathbf{w}_{\mathrm{c}} \in \mathbb{C}^N$, $\mathbf{H}_{\mathrm{c}} \in M_{N\times M}(\mathbb{C})$. The transmitted vector $\mathbf{x}_{\mathrm{c}}$ belongs to a finite constellation $\mathcal{S} \subset \mathbb{Z}[i]^M$; the entries of the channel matrix $\mathbf{H}_{\mathrm{c}}$ 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 $\mathbf{w}_{\mathrm{c}}$ is the Gaussian noise with i.i.d. entries of zero mean and variance N_0. We consider the coherent case where $\mathbf{H}_{\mathrm{c}}$ is known at the receiver.

Separating the real and imaginary part, the model can be rewritten as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{2}$$

in terms of the real-valued vectors

$$\mathbf{y} = \begin{pmatrix} \Re(\mathbf{y}_{\mathrm{c}}) \\ \Im(\mathbf{y}_{\mathrm{c}}) \end{pmatrix} \in \mathbb{R}^n, \quad \mathbf{x} = \begin{pmatrix} \Re(\mathbf{x}_{\mathrm{c}}) \\ \Im(\mathbf{x}_{\mathrm{c}}) \end{pmatrix} \in \mathbb{Z}^m$$

and of the equivalent real channel matrix

$$\mathbf{H} = \begin{pmatrix} \Re(\mathbf{H}_{\mathrm{c}}) & -\Im(\mathbf{H}_{\mathrm{c}}) \\ \Im(\mathbf{H}_{\mathrm{c}}) & \Re(\mathbf{H}_{\mathrm{c}}) \end{pmatrix} \in M_{n\times m}(\mathbb{R}).$$

Here $n = 2N$, $m = 2M$.

The maximum likelihood decoded vector is given by

$$\hat{\mathbf{x}}_{\mathrm{ML}} = \underset{\hat{\mathbf{x}}_{\mathrm{c}}\in\mathcal{S}}{\operatorname{argmin}} \left\|\mathbf{H}_{\mathrm{c}}\hat{\mathbf{x}}_{\mathrm{c}} - \mathbf{y}_{\mathrm{c}}\right\| = \underset{\hat{\mathbf{x}}_{\mathrm{c}}\in\mathcal{S}}{\operatorname{argmin}} \left\|\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}\right\|,$$

where $\|\cdot\|$ denotes the Euclidean norm.

B. Lattice reduction

An m-dimensional lattice in $\mathbb{R}^n$ is the set of points

$$\mathcal{L}(\mathbf{H}) = \{\mathbf{H}\mathbf{x} \mid \mathbf{x} \in \mathbb{Z}^m\},$$

where $\mathbf{H} \in M_{n\times m}(\mathbb{R})$. We denote by $d_{\mathbf{H}}$ the *minimum distance* of the lattice, that is the smallest norm of a nonzero vector in $\mathcal{L}(\mathbf{H})$.

More generally, for all $1 \leq i \leq m$ one can define the *i-th successive minimum* of the lattice as follows:

$$\lambda_i(\mathbf{H}) = \inf\{r > 0 \mid \exists \mathbf{v}_1, \ldots, \mathbf{v}_i \text{ linearly independent in } \mathcal{L}(\mathbf{H}) \text{ s.t. } \|\mathbf{v}_j\| \leq r \quad \forall j \leq i\}$$

We recall that two matrices $\mathbf{H}, \mathbf{H}'$ generate the same lattice if and only if $\mathbf{H}' = \mathbf{H}\mathbf{U}$ with $\mathbf{U}$ unimodular.

Lattice reduction algorithms allow to find a new basis $\mathbf{H}'$ for a given lattice $\mathcal{L}(\mathbf{H})$ such that the basis vectors are shorter and nearly orthogonal. Orthogonality can be measured by the absolute value of the coefficients $\mu_{i,j}$ in the Gram-Schmidt orthogonalization of the basis, see the GSO Algorithm 1.

Algorithm 1: GSO (Gram-Schmidt orthogonalization)

```
$\mathbf{h}_1^* \leftarrow \mathbf{h}_1$
for $i = 2, \ldots, m$ do
    for $j = 1, \ldots, i-1$ do
        $\mu_{i,j} \leftarrow \frac{\langle \mathbf{h}_i, \mathbf{h}_j^* \rangle}{\|\mathbf{h}_j^*\|^2}$
    end
    $\mathbf{h}_i^* \leftarrow \mathbf{h}_i - \sum_{j=1}^{i-1} \mu_{i,j}\mathbf{h}_j^*$
end
```

We recall the following useful property of GSO: the length of the smallest of the Gram-Schmidt vectors $\mathbf{h}_i^*$ is always less or equal to the minimum distance $d_{\mathbf{H}}$ of the lattice [12]:

$$d_{\mathbf{H}} \geq a(\mathbf{H}) \doteq \min_{1\leq i\leq m} \|\mathbf{h}_i^*\| \tag{3}$$

A basis $\mathbf{H}$ is said to be *LLL-reduced* [11] if its Gram-Schmidt coefficients $\mu_{i,j}$ and Gram-Schmidt vectors satisfy the following properties:

1) *Size reduction:*

$$|\mu_{k,l}| < \frac{1}{2}, \quad 1 \leq l < k \leq m,$$

2) *Lovasz condition:*

$$\left\|\mathbf{h}_k^* + \mu_{k,k-1}\mathbf{h}_{k-1}^*\right\|^2 \geq \delta \left\|\mathbf{h}_{k-1}^*\right\|^2, \quad 1 < k \leq m,$$

where $\delta \in \left(\frac{1}{4}, 1\right)$ (a customary choice is $\delta = \frac{3}{4}$).

The LLL algorithm is summarized in Algorithm 2. Given a full-rank matrix $\mathbf{H} \in M_{n\times m}(\mathbb{R})$, it computes an LLL-reduced version $\mathbf{H}_{\mathrm{red}} = \mathbf{H}\mathbf{U}$, with $\mathbf{U} \in M_{m\times m}(\mathbb{Z})$ unimodular, and outputs the columns $\{\mathbf{h}_i\}$ and $\{\mathbf{u}_i\}$ of $\mathbf{H}_{\mathrm{red}}$ and $\mathbf{U}$ respectively.

Algorithm 2: The LLL algorithm

```
$\mathbf{U} = \mathbf{I}_m$
Compute the GSO of $\mathbf{H}$
$k \leftarrow 2$
while $k \leq m$ do
    RED(k,k-1)
    if $\left\|\mathbf{h}_k^* + \mu_{k,k-1}\mathbf{h}_{k-1}^*\right\|^2 < \delta \left\|\mathbf{h}_{k-1}^*\right\|^2$ then
        swap $\mathbf{h}_k$ and $\mathbf{h}_{k-1}$
        swap $\mathbf{u}_k$ and $\mathbf{u}_{k-1}$
        update GSO
        $k \leftarrow \max(k-1, 2)$
    end
    else
        for $l = k-2, \ldots, 1$ do
            RED(k,l)
        end
        $k \leftarrow k+1$
    end
end
```

We list here some properties of LLL-reduced bases that we will need in the sequel. First of all, the LLL algorithm finds

Algorithm 3: Size reduction RED(k,l)

```
if $|\mu_{k,l}| > \frac{1}{2}$ then
    $\mathbf{h}_k \leftarrow \mathbf{h}_k - \lfloor \mu_{k,l} \rceil \mathbf{h}_l$
    $\mathbf{u}_k \leftarrow \mathbf{u}_k - \lfloor \mu_{k,l} \rceil \mathbf{u}_l$
    for $j = 1, \cdots, l-1$ do
        $\mu_{k,j} \leftarrow \mu_{k,j} - \lfloor \mu_{k,l} \rceil \mu_{l,j}$
    end
    $\mu_{k,l} \leftarrow \mu_{k,l} - \lfloor \mu_{k,l} \rceil$
end
```

at least one basis vector whose length is not too far from the minimum distance $d_{\mathbf{H}}$ of the lattice. The following inequality holds for any m-dimensional LLL-reduced basis $\mathbf{H}$ [2]:

$$\|\mathbf{h}_1\| \leq \alpha^{\frac{m-1}{2}} d_{\mathbf{H}}, \tag{4}$$

where $\alpha = \frac{1}{\delta - 1/4}$ ($\alpha = 2$ if $\delta = \frac{3}{4}$).
Moreover, the first basis vector cannot be too big compared to the Gram-Schmidt vectors $\{\mathbf{h}_i^*\}$:

$$\|\mathbf{h}_1\| \leq \alpha^{\frac{i-1}{2}} \|\mathbf{h}_i^*\|, \qquad \forall 1 \leq i \leq m.$$

In particular, if $j = \operatorname{argmin}_{1 \leq i \leq m} \|\mathbf{h}_i^*\|$,

$$d_{\mathbf{H}} \leq \|\mathbf{h}_1\| \leq \alpha^{\frac{j-1}{2}} \left\|\mathbf{h}_j^*\right\| = \alpha^{\frac{j-1}{2}} a(\mathbf{H}) \leq \alpha^{\frac{m-1}{2}} a(\mathbf{H}). \tag{5}$$

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 [16, 15, 3].
Let $\mathbf{H}_{\text{red}} = \mathbf{HU}$ be the output of the LLL algorithm on $\mathbf{H}$. We can rewrite the received vector as $\mathbf{y} = \mathbf{H}_{\text{red}}\mathbf{U}^{-1}\mathbf{x} + \mathbf{w}$.

- The *LLL-ZF decoder* outputs

$$\hat{\mathbf{x}}_{LLL-ZF} = Q_{\mathcal{S}}\left(\mathbf{U}\left(\left\lfloor \mathbf{H}_{\text{red}}^{\dagger}\mathbf{y} \right\rceil\right)\right),$$

  where $\mathbf{H}_{\text{red}}^{\dagger} = (\mathbf{H}_{\text{red}}^T \mathbf{H}_{\text{red}})^{-1}\mathbf{H}_{\text{red}}^T$ is the Moore-Penrose pseudoinverse of $\mathbf{H}_{\text{red}}$, $\lfloor \cdot \rceil$ denotes componentwise rounding to the nearest integer and $Q_{\mathcal{S}}$ is a quantization function that forces the solution to belong to the constellation $\mathcal{S}$.
- The *LLL-SIC decoder* performs the QR decomposition $\mathbf{H}_{\text{red}} = \mathbf{QR}$, computes $\widetilde{\mathbf{y}} = \mathbf{Q}^T\mathbf{y}$, finds by recursion $\widetilde{\mathbf{x}}$ defined by

$$\tilde{x}_m = \left\lfloor \frac{\tilde{y}_m}{r_{mm}} \right\rceil,$$
$$\tilde{x}_i = \left\lfloor \frac{\tilde{y}_i - \sum_{j=i+1}^m r_{ij}\tilde{x}_j}{r_{ii}} \right\rceil, \qquad i = m-1, \ldots, 1,$$

  and finally outputs $\hat{\mathbf{x}}_{LLL-SIC} = Q_{\mathcal{S}}(\mathbf{U}\widetilde{\mathbf{x}})$.

D. Improved lattice reduction

Recently, Kim and Park [9] 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 $\mathbf{y}$ be the (real) received vector in the model (2). Consider the $(n+1) \times (m+1)$ augmented matrix

$$\widetilde{\mathbf{H}} = \begin{pmatrix} \mathbf{H} & -\mathbf{y} \\ \mathbf{0}_{1 \times m} & t \end{pmatrix} = \begin{pmatrix} h_{1,1} & \cdots & h_{1,m} & -y_1 \\ \vdots & & & \vdots \\ h_{n,1} & \cdots & h_{n,m} & -y_n \\ 0 & \cdots & 0 & t \end{pmatrix} \tag{6}$$

where $t > 0$ is a parameter to be determined.
Let $\widetilde{\mathbf{H}}_{\text{red}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{U}}$ be the output of the LLL algorithm on $\widetilde{\mathbf{H}}$.

- The *Improved lattice reduction decoder* outputs

$$\hat{\mathbf{x}}_{ILR} = Q_{\mathcal{S}}(\widetilde{u}_{1,m+1}, \ldots, \widetilde{u}_{m,m+1})^T \tag{7}$$

In [9], the parameter t is chosen such that $t > r_{m,m}$, where $\mathbf{R} = (r_{i,j})$ is the upper triangular matrix in the the QR decomposition $\mathbf{H}_{\text{red}} = \mathbf{QR}$. This ensures that the Lovasz condition on the last column of the augmented matrix $\widetilde{\mathbf{H}}$ is always verified. Therefore, LLL-reducing $\widetilde{\mathbf{H}}$ amounts to LLL-reducing the submatrix $\mathbf{H}$ and then performing a final round of size reduction without swaps on the last column.
Although not explicitly stated in [9], the performance of improved lattice reduction is exactly the same as LLL-SIC. In fact, it is not hard to prove by induction that

$$Q_{\mathcal{S}}(\widetilde{\mathbf{u}}_{m+1}) = Q_{\mathcal{S}}\begin{pmatrix} \sum_{i=1}^n \widetilde{x}_i \mathbf{u}_i \\ 1 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{x}}_{LLL-SIC} \\ 1 \end{pmatrix}$$

where $\mathbf{u}_i$ and $\widetilde{\mathbf{u}}_i$ are the columns of $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ respectively.

III. Augmented Lattice Reduction

We introduce here a different decoder based on the augmented matrix (6) 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 $\mathcal{L}(\widetilde{\mathbf{H}})$ are of the form $\begin{pmatrix} \mathbf{Hx}' - q\mathbf{y} \\ qt \end{pmatrix}$, $\mathbf{x}' \in \mathbb{Z}^m$, $q \in \mathbb{Z}$. In particular, the vector $\mathbf{v} = \begin{pmatrix} \mathbf{Hx} - \mathbf{y} \\ t \end{pmatrix} = \begin{pmatrix} \mathbf{w} \\ t \end{pmatrix}$ belongs to the augmented lattice. We will show that for a suitable choice of the parameter t, and supposing that the noise $\mathbf{w}$ is exponentially smaller than the minimum distance in the original lattice $\mathcal{L}(\mathbf{H})$, $\mathbf{v}$ is the shortest vector in the lattice and the LLL algorithm finds this vector. That is, $\pm\mathbf{v}$ is the first column of $\widetilde{\mathbf{H}}_{\text{red}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{U}}$, the output of LLL algorithm on $\widetilde{\mathbf{H}}$. Clearly, since $\widetilde{\mathbf{H}}$ is full-rank with probability 1, in this case the first column of the change of basis matrix $\widetilde{\mathbf{U}}$ is $\begin{pmatrix} \pm\mathbf{x} \\ \pm 1 \end{pmatrix}$. Thus we can "read" the transmitted message directly from the change of basis matrix $\widetilde{\mathbf{U}}$.
To summarize, in order to decode we can perform the LLL

algorithm on $\widetilde{\mathbf{H}}$, and given the output $\widetilde{\mathbf{H}}_{\text{red}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{U}}$, we can choose

$$\hat{\mathbf{x}} = Q_{\mathcal{S}}\left(\left\lfloor \frac{1}{\widetilde{u}_{m+1,1}}(\widetilde{u}_{1,1},\ldots,\widetilde{u}_{m,1})^T \right\rceil\right), \tag{8}$$

where $\widetilde{\mathbf{U}} = (\widetilde{u}_{i,j})$. The previous decoder can be improved by including all the columns of $\mathbf{H}_{\text{red}}$ in the search: let

$$\mathbf{u}_k = \frac{1}{\widetilde{u}_{m+1,k}}(\widetilde{u}_{1,k},\ldots,\widetilde{u}_{m,k})^T, \quad k = 1,\ldots,m.$$

If $|\widetilde{u}_{m+1,k}| = 1$ for some $k \in \{1,\ldots,m\}$, we define

$$k_{\min} = \underset{k \text{ s.t. } |\widetilde{u}_{m+1,k}|=1}{\operatorname{argmin}} \|\mathbf{H}\mathbf{u}_k - \mathbf{y}\|,$$

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

$$\hat{\mathbf{x}}_{\text{ALR}} = Q_{\mathcal{S}}\left(\lfloor \mathbf{u}_{k_{\min}} \rceil\right), \tag{9}$$

IV. Performance

A. Diversity

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 (6). The diversity gain d of a decoding scheme is defined as follows:

$$d = -\lim_{\rho\to\infty} \frac{\log(P_e)}{\log(\rho)},$$

where P_e denotes the error probability as a function of the signal to noise ratio ρ.

Proposition 1. *If the augmented lattice reduction is performed using $t = \varepsilon a(\mathbf{H}_{\text{red}})$, where $a(\mathbf{H}_{\text{red}})$ is the length of the smallest vector in the Gram-Schmidt orthogonalization of $\mathbf{H}_{\text{red}}$, and $\varepsilon \le \frac{1}{2\sqrt{2}\alpha^{\frac{m}{2}}}$, then it achieves the maximum receive diversity N.*

Remark. It is essential to use $a(\mathbf{H}_{\text{red}})$ in place of $a(\mathbf{H})$. In fact, for general bases $\mathbf{H}$ that are not LLL-reduced, there is no lower bound of the type (5) limiting how small the smallest Gram-Schmidt vector can be. For $a(\mathbf{H}_{\text{red}})$, putting together the bounds (3) and (5), we obtain

$$\frac{d_{\mathbf{H}}}{\alpha^{\frac{m-1}{2}}} \le a(\mathbf{H}_{\text{red}}) \le d_{\mathbf{H}} \tag{10}$$

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

In order to prove the previous Proposition, we will show that in the $(m+1)$-dimensional lattice $\mathcal{L}(\widetilde{\mathbf{H}})$ there is an exponential gap between the first two successive minima. Then, using the estimate (4) 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 $\mathcal{L}(\widetilde{\mathbf{H}})$ with high probability. This, in turn, allows to recover the closest lattice vector $\mathbf{Hx}$ to $\mathbf{y}$ in $\mathcal{L}(\mathbf{H})$ supposing that the noise $\mathbf{w}$ is small enough.

The following definition makes the notion of "gap" more precise:

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

$$\|\mathbf{u}\| \le \gamma \|\mathbf{v}\| \quad \Rightarrow \quad \mathbf{u}, \mathbf{v} \quad \text{are linearly dependent.}$$

We now prove the existence of such a gap under suitable conditions:

Lemma 1. *Let $\widetilde{\mathbf{H}}$ be the matrix defined in (6), and let $t = \varepsilon a(\mathbf{H}_{\text{red}})$, with $\varepsilon \le \frac{1}{2\sqrt{2}\alpha^{\frac{m}{2}}}$.*
Suppose that $\|\mathbf{w}\| = \|\mathbf{y} - \mathbf{Hx}\| \le \frac{\varepsilon}{\alpha^{\frac{m-1}{2}}} d_{\mathbf{H}}$.
Then $\mathbf{v} = \begin{pmatrix} \mathbf{Hx} - \mathbf{y} \\ t \end{pmatrix}$ is an $\alpha^{\frac{m}{2}}$-unique shortest vector of $\mathcal{L}(\widetilde{\mathbf{H}})$.

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

Proof: We need to show that any vector $\mathbf{u} \in \mathcal{L}(\widetilde{\mathbf{H}})$ that is not a multiple of $\mathbf{v}$ must have length greater than $\alpha^{\frac{m}{2}} \|\mathbf{v}\|$. By contradiction, suppose that $\exists \mathbf{u} = \begin{pmatrix} \mathbf{Hx}' - q\mathbf{y} \\ qt \end{pmatrix} \in \mathcal{L}(\widetilde{\mathbf{H}})$ linearly independent from $\mathbf{v}$ such that $\|\mathbf{u}\| \le \alpha^{\frac{m}{2}} \|\mathbf{v}\|$. Since $\|\mathbf{u}\| \ge |q|\, t$,

$$|q| \le \frac{\|\mathbf{u}\|}{t} \le \frac{\alpha^{\frac{m}{2}} \|\mathbf{v}\|}{t}.$$

On the other side, $\|\mathbf{u}\| \le \alpha^{\frac{m}{2}} \|\mathbf{v}\|$ implies that also $\|\mathbf{Hx}' - q\mathbf{y}\| \le \alpha^{\frac{m}{2}} \|\mathbf{v}\|$. Consider

$$\begin{aligned} &\|\mathbf{Hx}' - q\mathbf{Hx}\| \le \|\mathbf{Hx}' - q\mathbf{y}\| + \|q\mathbf{y} - q\mathbf{Hx}\| \le \\ &\le \alpha^{\frac{m}{2}} \|\mathbf{v}\| + |q| \|\mathbf{y} - \mathbf{Hx}\| \le \alpha^{\frac{m}{2}} \|\mathbf{v}\| + \frac{\alpha^{\frac{m}{2}} \|\mathbf{v}\|}{t} \|\mathbf{w}\| \le \\ &\le \alpha^{\frac{m}{2}} t \sqrt{\frac{1 + \|\mathbf{w}\|^2}{t^2}} \left(1 + \frac{\|\mathbf{w}\|}{t}\right) \end{aligned} \tag{11}$$

The bound (10) on $a(\mathbf{H}_{\text{red}})$ implies

$$\frac{\varepsilon}{\alpha^{\frac{m-1}{2}}} d_{\mathbf{H}} \le t \le \varepsilon d_{\mathbf{H}}, \quad \|\mathbf{w}\| < t$$

Using these inequalities, we can bound the expression (11) with

$$\alpha^{\frac{m}{2}} \varepsilon d_{\mathbf{H}} 2\sqrt{2} < d_{\mathbf{H}}.$$

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

Lemma 2. *Under the hypotheses of Lemma 1, the augmented lattice reduction methods (8) and (9) correctly decode the transmitted signal $\mathbf{x}$.*

Proof: Let $\widetilde{\mathbf{H}}_{\text{red}} = \widetilde{\mathbf{H}}\widetilde{\mathbf{U}}$ denote the output of the LLL reduction of $\widetilde{\mathbf{H}}$, and let $\hat{\mathbf{h}}_1 = \widetilde{\mathbf{H}}\begin{pmatrix}\mathbf{x}' \\ q\end{pmatrix}$ be its first column. The property (4) of LLL reduction in dimension $m+1$ entails that $\left\|\hat{\mathbf{h}}_1\right\| \leq \alpha^{\frac{m}{2}} d_{\widetilde{\mathbf{H}}}$. But since $\mathbf{v} = \begin{pmatrix}\mathbf{Hx}-\mathbf{y} \\ t\end{pmatrix}$ has been shown to be $\alpha^{\frac{m}{2}}$-unique in the previous Lemma, it means that $\hat{\mathbf{h}}_1$ and $\mathbf{v}$ are linearly dependent; equivalently, $\exists a, b \in \mathbb{Z} \setminus \{0\}$ such that $a\mathbf{v} + b\hat{\mathbf{h}}_1 = 0$. In particular $at + bqt = 0$, that is $a = -bq$ and $\hat{\mathbf{h}}_1 = q\mathbf{v}$. Then by definition of $\widetilde{\mathbf{H}}$, $\hat{\mathbf{h}}_1 = \widetilde{\mathbf{H}}\begin{pmatrix}q\mathbf{x} \\ q\end{pmatrix}$. This means that the first column of the reduction matrix $\widetilde{\mathbf{U}}$ is $\begin{pmatrix}q\mathbf{x} \\ q\end{pmatrix}$, and so $\hat{\mathbf{x}}_{\text{ALR}} = Q_{\mathcal{S}}(\lfloor \mathbf{u}_1 \rceil) = Q_{\mathcal{S}}(q\mathbf{x}/q) = \mathbf{x}$ and the augmented lattice reduction methods (8) and (9) correctly decode the transmitted message.
(Observe that this is possible only if $|q| = 1$, since $\det(\widetilde{\mathbf{U}})$ is also a multiple of q and $\widetilde{\mathbf{U}}$ is unimodular.) □

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

$$P_{e,\text{ALR}}(\mathbf{H}) \leq P\left\{\|\mathbf{w}\| > \frac{\varepsilon}{\alpha^{\frac{m-1}{2}}} d_{\mathbf{H}}\right\} = P\{\|\mathbf{w}\| > \varepsilon' d_{\mathbf{H}}\}.$$

To conclude the proof of Proposition 1, we need to show that

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

This turns out to be true. In fact, it has been shown in [14] (Proof of Theorem 2), that for any constant c_M depending only on the number of transmit antennas[3],

$$P\{\|\mathbf{w}\| > c_M d_{\mathbf{H}}\} \leq \frac{C(\ln(\rho))^{N+1}}{\rho^N} \qquad \text{for } N = M,$$

$$P\{\|\mathbf{w}\| > c_M d_{\mathbf{H}}\} \leq \frac{C}{\rho^N} \qquad \text{for } N > M.$$

Thus we have shown that augmented lattice reduction achieves the maximum receive diversity N with the choice $t = \varepsilon a(\mathbf{H}_{\text{red}})$.

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 (9) is used for augmented lattice reduction.
Two versions of augmented lattice reduction with different values of the parameter ε are compared. Clearly it is preferable to choose ε as big as possible in order to minimize the probability $P\left\{\|\mathbf{w}\| > \frac{\varepsilon}{\alpha^{\frac{m-1}{2}}} d_{\mathbf{H}}\right\}$. Version 1 corresponds to the choice $\varepsilon = \frac{1}{2\sqrt{2}\alpha^{\frac{m}{2}}}$, the highest value of ε that verifies the hypothesis of Proposition 1. At the symbol error rate (SER) of $1 \cdot 10^{-4}$, its performance is within $0.8\,\text{dB}$ from ML decoding and gains $1.9\,\text{dB}$ with respect to LLL-SIC decoding.

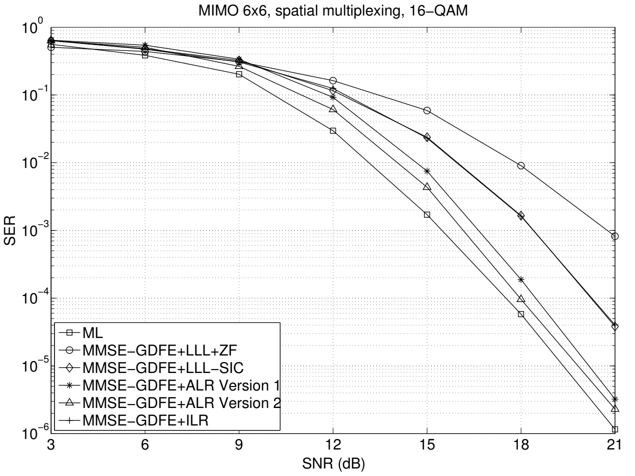

Figure 1. Performance comparison of augmented lattice reduction with LLL-ZF, LLL-SIC and Improved Lattice Reduction with MMSE-GDFE preprocessing for a 6×6 MIMO system using 16-QAM.

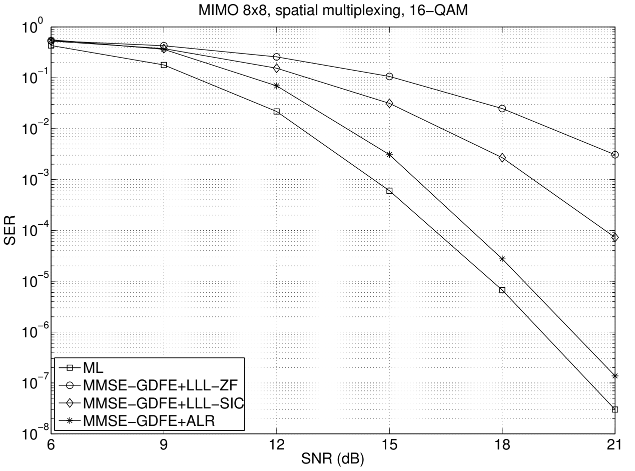

Figure 2. Performance comparison of augmented lattice reduction with LLL-ZF and LLL-SIC detection with MMSE-GDFE preprocessing for a 8×8 MIMO system using 16-QAM.

Version 2 corresponds to a value of ε optimized by computer search (experimentally, this is around $2^{-\frac{m}{4}}$), whose performance is within only $0.4\,\text{dB}$ 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 ε, we are not able to prove that the LLL algorithm finds the shortest lattice vector in $\mathcal{L}(\widetilde{\mathbf{H}})$. 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 $3.5\,\text{dB}$ for an 8×8 MIMO system, at the SER of 10^{-4}. On the other side, augmented lattice reduction is still within $0.8\,\text{dB}$ from ML performance (see Figure 2).

[3] This result was used in [14] in order to prove that the LLL-ZF decoder achieves the receive diversity order. The proof in [14] actually refers to the complex model (1), but the statement also holds for the real model since $d_{\mathbf{H}} = d_{\mathbf{H}_c}$, $\|\mathbf{w}\| = \|\mathbf{w}_c\|$.

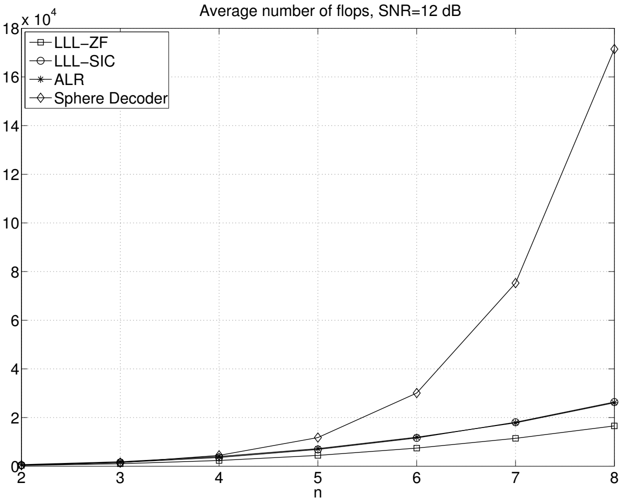

Figure 3. Complexity comparison (in flops) of augmented lattice reduction with LLL-ZF, LLL-SIC and sphere decoding as a function of the number n of transmit and receive antennas, at $\mathrm{SNR} = 12\,\mathrm{dB}$, using 16-QAM constellations.

V. Complexity

In this section we compare the computational complexity of augmented lattice reduction, LLL-ZF and LLL-SIC decoding. We are interested in the complexity order as a function of the number of transmit and receive antennas.

A. Simulation results

Figure 4 shows the average number of iterations of the LLL algorithm for LLL-aided linear decoding and the augmented lattice reduction method. We have chosen $\delta = \frac{3}{4}$ in all the numerical simulations.

While the number of iterations of LLL is indeed higher, approximately by a factor 2, for the augmented lattice reduction, the total complexity expressed in flops[4] is about the same for LLL-SIC and the augmented lattice method (see Figure 3). The additional complexity of the LLL algorithm is balanced out by the complexity savings due to the fact that QR decomposition is not needed.

B. Complex LLL reduction

A generalization of the LLL algorithm to complex lattices has been studied in [13] and applied to MIMO decoding in [4]. It has been show experimentally in [4] 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

$$\widetilde{\mathbf{H}}_{\mathrm{c}} = \begin{pmatrix} \mathbf{H}_{\mathrm{c}} & -\mathbf{y}_{\mathrm{c}} \\ \mathbf{0}_{1\times N} & t \end{pmatrix},$$

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

VI. Conclusions

In this paper, we introduced a new kind of lattice-reduction aided decoding which does not require a linear or decision-feedback receiver at the last stage. We proved that this method attains the maximum receive diversity order. Simulation results evidence that the new technique has a substantial performance gain with respect to the classical LLL-ZF and LLL-SIC decoders, while having approximately the same complexity order as LLL-SIC.

References

[1] L. Babai, “On Lovasz’ lattice reduction and the nearest lattice point problem”, *Combinatorica*, vol. 6, n.1, pp 1–13 (1986)

[2] H. Cohen, “A course in computational algebraic number theory”, Graduate Texts in Mathematics, Springer, 2000

[3] M. O. Damen, H. El Gamal, G. Caire, “On maximum-likelihood detection and the search for the closest lattice point”, *IEEE Trans. Inform. Theory*. vol. 49, 2389–2402, 2003

[4] Y. H. Gan, C. Ling, W. H. Mow, “Complex Lattice Reduction Algorithm for Low-Complexity MIMO Detection”, *IEEE Trans. Signal Process.*, vol 57 n.7 (2009)

[5] J. Jaldén, P. Elia, “DMT optimality of LR-aided linear decoders for a general class of channels, lattice designs, and system models”, submitted to *IEEE Trans. Inform. Theory*

[6] J. Jaldén, D. Seethaler, G. Matz, “Worst- and average-case complexity of LLL lattice reduction in MIMO wireless systems”, *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (2008), 2685 – 2688

[7] R. Kannan, “Minkowski’s convex body theorem and integer programming”, *Math. Oper. Res.* 12, 415–440 (1987)

[8] K. Raj Kumar, G. Caire, A. L. Moustakas, “Asymptotic performance of linear receivers in MIMO fading channels”, submitted.

[9] N. Kim, H. Park, “Improved lattice reduction aided detections for MIMO systems”, *Vehicular Technology Conference* 2006

[10] C. Ling, “On the proximity factors of lattice reduction-aided decoding”, submitted.

[11] A. K. Lenstra, J. H. W. Lenstra, L. Lovasz, “Factoring polynomials with rational coefficients”, *Math. Ann.*, vol. 261, pp. 515-534, 1982

[12] J. C. Lagarias, H. W. Lenstra Jr., C. P. Schnorr, “Korkin-Zolotarev bases and successive minima of a lattice and its reciprocal lattice”, *Combinatorica*, vol. 10 n.4 (1990), 333–348

[13] H. Napias, “A generalization of the LLL-algorithm over Euclidean rings or orders”, *Journal de Théorie des Nombres de Bordeaux* 8 (1996), 387-396

[14] M. Taherzadeh, A. Mobasher, A. K. Khandani, “LLL reduction achieves the receive diversity in MIMO decoding”, *IEEE Trans. Inform. Theory*, vol 53 n. 12, 2007, pp 4801–4805

[15] C. Windpassinger, R. Fischer, “Low-complexity near-maximum likelihood detection and precoding for MIMO systems using lattice reduction”, *Proc IEEE Information Theory Workshop*, 2003, 345–348

[16] H. Yao, G. W. Wornell, “Lattice-reduction-aided detectors for MIMO communication systems”, *Proc. Global Telecommunications Conference* 2002, vol 1, 424–428

[4] Here we define a “flop” as any floating-point operation (addition, multiplication, division or square root).