Parallel QR Decomposition in LTE-A Systems

Sébastien Aubert*, Manar Mohaisen†, Fabienne Nouvel‡, and KyungHi Chang†

* ST-ERICSSON, 505, route des lucioles, CP 06560 Sophia-Antipolis, France
† Inha University, Graduate school of IT & T, 402-751 Incheon, Korea
‡ Research unit IETR, 20, avenue des buttes de Coësmes, CP 35043 Rennes, France
Email: sebastien.aubert@stericsson.com, lemanar@hotmail.com, fabienne.nouvel@insa-rennes.fr, khchang@inha.ac.kr

Abstract—The QR Decomposition (QRD) of communication channel matrices is a fundamental prerequisite to several detection schemes in Multiple-Input Multiple-Output (MIMO) communication systems. Herein, the main feature of the QRD is to transform the non-causal system into a causal system, where consequently efficient detection algorithms based on the Successive Interference Cancellation (SIC) or Sphere Decoder (SD) become possible. Also, QRD can be used as a light but efficient antenna selection scheme. In this paper, we address the study of the QRD methods and compare their efficiency in terms of computational complexity and error rate performance. Moreover, a particular attention is paid to the parallelism of the QRD algorithms since it reduces the latency of the matrix factorization.

Index Terms—MIMO detection, QR decomposition, parallelism.

I. INTRODUCTION

Multiple-Input Multiple-Output (MIMO) technologies have gained an increasing attention in the recent standardization bodies due to their capabilities to improve the reliability of the communication link and boost the channel capacity without requiring additional spectral resources [1]. In 3GPP Long-Term Evolution (3GPP LTE) and 3GPP LTE-Advanced, MIMO techniques with up to 8 transmit antennas are combined with the Orthogonal Frequency Division Multiplexing (OFDM) to offer both low complexity equipments and robustness to the frequency selective fading [2].

The main challenge on the receiver side is to introduce low-complexity but efficient detection algorithms able to recover quasi-optimally the transmitted signals. QR Decomposition (QRD)-based Successive Interference Cancellation (SIC) schemes [3], Sphere Decoder (SD) [4] and QRD with M-algorithm (QRD-M) [5] detection algorithms use the QRD to transfer the MIMO system from a non-causal, where interference from future symbols exists, into a causal system where the detection of the current symbol only depends on the already-detected symbols. Hence, SIC becomes possible. Lenstra-Lenstra-Lovasz (LLL) [6] lattice basis reduction algorithm also requires the QRD to decrease computational complexity in the calculation of its orthogonality deficit measure.

The main idea of the QRD is to factorize the channel matrix $H$ as the product of an unitary matrix $Q$ and an upper triangular matrix $R$. Since the matrix $Q$ is unitary, the Euclidean norm, the singular values, and the determinant of $H$ and $R$ are equal. Thanks to the QRD, the MIMO system is transferred into a causal system by simply filtering the received vector using the Hermitian transpose of the $Q$ matrix.

It turns out that the QRD, although it plays an important role, has not been sufficiently studied in the literature of the wireless communication systems.

Contributions. The contributions of this paper can be summarized as follows:

• First, several computation methods of the QRD are introduced. The parallelism capabilities in calculation of the QRD are emphasized;
• A reduced complexity solution is proposed and discussed;
• The complexity of the introduced algorithms is calculated in terms of real multiplications (MUL), with realistic implementation considerations.

The rest of this paper is organized as follows. In Section II, we introduce the system model and review several detection schemes that are based on the QRD. QRD techniques are introduced in details in Section III and simulation results are exposed in Section IV. Section V provides a conclusion and a discussion.

II. SYSTEM MODEL AND MIMO DETECTION

A. System Model

In this paper, we consider a MIMO-Spatial Multiplexing (MIMO-SM) system, where independent data symbols are transmitted via sufficiently-separated antennas. By considering $n_T$ transmit antennas and $n_R$ receive antennas - with $n_R \geq n_T$ to avoid rank-deficiency effects [7] - the received vector $y \in \mathbb{C}^{n_R}$ is given by:

$$y = Hx + n,$$

(1)

where $x \in \xi^{n_T}$ is the transmitted vector with $E[xx^H] = I_{n_T}$ ($\xi$ is the modulation set and $I_{n_T}$ is the $n_T \times n_T$ identity matrix) and $H$ is the channel matrix with complex element $h_{i,j}$ whose real and imaginary parts are independent and follow $\mathcal{N}(0,0.5)$. Also, $n$ is the additive white Gaussian vector with $E[mm^H] = \sigma_n^2 I_{n_R}$.

B. MIMO Detection

In MIMO-SM, the optimum Maximum-Likelihood Detector (MLD) employs a brute-force detection to estimate the transmitted vector:

$$\hat{x} = \arg\min_{x \in \xi^{n_T}} \|y - Hx\|^2.$$  

(2)
By defining the QRD $H = QR$ of the $n_R \times n_T$ complex channel matrix, with $Q$ an unitary matrix of dimensions $n_R \times n_T$ ($Q^HQ = I_{n_T}$) and $R$ an upper triangular matrix of dimensions $n_T \times n_T$. By considering $\hat{y} = Q^Hy$, the search problem in (2) can be re-written as:

$$\hat{x} = \arg\min_{x \in \mathbb{C}^{n_T}} \|y - Rx\|^2,$$

$$= \arg\min_{x \in \mathbb{C}^{n_T}} \sum_{i=1}^{n_T} \|y_i - R_{i,i}x_i - \sum_{j=i+1}^{n_T} R_{i,j}\hat{x}_j\|^2,$$  

where $y_i$ is the $i$-th entry of $\hat{y}$ and $x_j$ is the hypothetical value for the $j$-th entry of $\hat{x}$ and $R_{i,j}$ is the $(i,j)$ entry of $R$. Due to the QRD, the $n_T$-dimensional lattice search in (3) is transformed into $n_T$ parallel one-dimensional search problems. Therefore, the search is done in parallel over lines, which reduces the computational complexity.

Since embedded communication systems are limited in terms of computational complexity, several detection schemes have been proposed in the literature to solve the search problem in (3) with lower computational costs.

The QRD-based SIC technique was originally introduced in [8] and further studied by Wuebben et al. [3], [9]. The decision process becomes, for all $i$ from $n_T$ to 1:

$$\hat{x}_i = Q_{\xi} \left\{ \frac{y_i - \sum_{j=i+1}^{n_T} R_{i,j}\hat{x}_j}{R_{i,i}} \right\},$$

where $Q_{\xi}$ denotes the quantification to the modulation set $\xi$ and $\hat{x}_1$ and $\hat{x}_j$ the estimates of the $i$-th and $j$-th transmit symbols respectively, for $j > i$.

Although the SIC schemes outperform the linear schemes, they are still far from achieving the optimum performance of the MLD [10] and in particular the same diversity.

SD and in particular QRD-M detection algorithms are considered as prominent schemes that achieve quasi-ML performance while requiring lower complexity compared to the MLD. The main idea of the SD is to restrict the search problem in (3) to the lattice points inside a hyper-sphere of predefined radius $d$. The decision process takes into account the sphere constraint and becomes, for all $i$ from $n_T$ to 1:

$$\hat{x} = \arg\min_{x \in \xi^{n_T}} \sum_{i=1}^{n_T} \|y_i - R_{i,i}x_i - \sum_{j=i+1}^{n_T} R_{i,j}\hat{x}_j\|^2 \leq d^2,$$

Although SD has a low average computational complexity, its worst-case complexity is still high. The conventional QRD-M algorithm has a fixed complexity which is a desired property for latency-limited communication systems. The idea of the QRD-M is to retain a fixed number of symbol candidates at each search level.

III. QR-DECOMPOSITION TECHNIQUES

A. Gram-Schmidt Orthogonalization

Gram-Schmidt (GS) orthogonalization is frequently used in communication systems. The GS QRD algorithm consists of two steps, namely orthogonalization and normalization steps. In the orthogonalization step, a normal vector of the matrix $Q$ that is already normalized in the normalization step is obtained and the remaining columns of $Q$ are orthogonalized to the obtained column. Note that the matrix $Q$ is initiated to $H$. Therefore, the corresponding row of the matrix $R$ is obtained from $Q$. Algorithm 1 depicts a MATLAB-like algorithmic description of the Stable GS (STGS) algorithm by introducing a minor update of the Classical GS (CLGS) algorithm.

**Algorithm 1 Stable Gram-Schmidt QRD algorithm.**

1: INPUT: $H$
2: OUTPUT: $\hat{y}$, $R$
3: Initialization: $Q \leftarrow H$, $R \leftarrow I_{n_T}$
4: for $i = 1, \cdots, n_T$ do
5: $\hat{R}_{i,i} \leftarrow \sqrt{Q_{i,i}^H Q_{i,i}}$
6: $Q_{:,i} \leftarrow Q_{:,i} / \hat{R}_{i,i}$
7: for $j = i+1, \cdots, n_T$ do
8: $R_{j,i} \leftarrow Q_{:,j}^H Q_{:,i}$
9: $Q_{:,j} \leftarrow Q_{:,j} - R_{j,i} Q_{:,i}$
10: end for
11: end for
12: $\hat{y} \leftarrow Q^H y$

1Conventional MATLAB notation are employed in the introduced algorithms i.e. $A_{m,n}$ corresponds to the $(n, m)$-th entry of the matrix $A$, $A_{:,n}$ corresponds to the $n$-th column-vector of $A$ and $A_{m:N, n}$ corresponds to entries from $n$ to $N$ of the $n$-th column-vector of $A$.

B. Householder Reflections

Householder (HH) constitutes another classical QRD technique that is used to obtain the upper triangular matrix $R$ from which the matrix $Q$ can be obtained if required. The idea behind the HH technique is to obtain the matrix $R$ using a reflection matrix. This reflection matrix, also known as Householder matrix, is used to cancel all the elements of a vector except its first element which is assigned the norm of the vector. Therefore, the columns of the matrix $H$ are treated iteratively to obtain the $R$ matrix. Algorithm 2 depicts a MATLAB-like algorithmic description of the HH technique.

The $\text{diag}\{\cdot\}$ function returns a matrix with the input vector on the main diagonal and the $\text{triu}\{\cdot\}$ function extracts the upper triangular part of an input matrix, excluding main diagonal.

C. Givens Rotations

Givens Rotations (GR) can also be employed to factorize the matrix $H$ [11]. This technique is usually used in embedded systems because of its numerical stability [12]. The general principle of this technique is to cancel the elements of the matrix $H$ so that a triangular form is obtained. Therefore, GR technique is the only technique which is not iterative and, hence, has a parallelism capabilities that will be explored later on.

Algorithm 3 depicts a MATLAB-like algorithmic description of the GR technique, where the rotation matrix $\Theta$ is
Algorithm 2  Householder triangulation technique.

1: INPUT: $H$
2: OUTPUT: $\tilde{y}$, $R$
3: Initialization: $Q \leftarrow H, R \leftarrow I_{n_T}$
4: for $i = 1, \ldots, n_T$ do
5:     $[\alpha, \beta] \leftarrow \text{house}(Q_{i:n_R,i})$
6:     $Q_{i:n_R,i+1:n_T} \leftarrow Q_{i:n_R,i+1:n_T} - \alpha vv^H Q_{i:n_R,i+1:n_T}$
7: end for
8: $R \leftarrow \text{diag}\{\alpha\} + \text{triu}\{Q_{1:n_T,1:n_T}\}$
9: for $i = n_T, \ldots, 1$ do
10:    $Q_{i:n_R,i:n_T} \leftarrow Q_{i:n_R,i:n_T} - vv^H Q_{i:n_R,i:n_T}$
11: end for
12: $\tilde{y} \leftarrow Q^H y$

Algorithm 3  Givens rotations algorithm.

1: INPUT: $H$
2: OUTPUT: $\tilde{y}$, $R$
3: Initialization : $Q \leftarrow H, R \leftarrow I_{n_T}$
4: for $i = 1, \ldots, n_T$ do
5:     for $j = i + 1, \ldots, n_R$ do
6:         $\alpha \leftarrow \frac{Q_{i,j}}{\|Q_{i,j}\|}$
7:         $\beta \leftarrow \frac{Q_{i,i}}{\|Q_{i,i}\|}$
8:         $\Theta \leftarrow \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$
9:         $[R_{i,i:n_T}; R_{j,i:n_T}] \leftarrow \Theta[R_{i,i:n_T}; R_{j,i:n_T}]$
10:    $[Q_{i:n_R,i}; Q_{j:n_R,j}] \leftarrow [Q_{i:n_R,i}; Q_{j:n_R,j}]\Theta^H$
11: end for
12: end for
13: $\tilde{y} \leftarrow Q^H y$

calculated such that the elements of $H$ are cancelled column by column, and an upper triangular matrix $R$ is finally obtained.

D. Parallel Givens Rotations

The Parallel GR (PGR) technique allows independent operations in the QR decomposition process so that the processing speed is increased. An example of the PGR is explained in the case of a $4 \times 4$ real matrix. The conventional GR technique obtains the following matrix:

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{3,1} & r_{4,1} \\ 0 & r_{2,2} & r_{3,2} & r_{4,2} \\ 0 & 0 & r_{3,3} & r_{4,3} \\ 0 & 0 & 0 & r_{4,4} \end{bmatrix},$$

where the entries $r_{i,j}$ represent the elements of $H$ that are modified during the QRD. The notation $k$ represents the order in which the corresponding elements of the matrix $H$ are cancelled out and such that $\{r_{i,j} \mid 1 \leq i, j \leq n_T\} \subset \mathbb{C}$ and $\{r_{i,j} \mid 1 \leq i \leq n_T\} \subset \mathbb{R}$. Hence, we remark that 6 sequential and dependent steps are necessary for the cancellation stage for a $4 \times 4$ real matrix.

Equation (7) illustrates the PGR technique to obtain the triangular matrix $R$ from $H$:

$$R = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{3,1} & r_{4,1} \\ 0 & r_{2,2} & r_{3,2} & r_{4,2} \\ 0 & 0 & r_{3,3} & r_{4,3} \\ 0 & 0 & 0 & r_{4,4} \end{bmatrix},$$

where 2 elements are simultaneously cancelled at the step 3 due to their independence. Hence, a parallelization gain is obtained which increases with the size of the matrix $H$.

Fig. 1, which consists of 8 pipes, shows the advantages of the PGR. The white blocks represent a free tube while the dark blocks represent a tube being doing the calculation. This calculation may correspond to the calculation of the rotation matrix, to the cancellation of the imaginary part of an element of $R$, of the real and imaginary parts of an element of $R$, or the calculation of $Q$. The potential parallelism has been evoked and can be associated to the QRD complexity reduction.

**Fig. 1:** Parallelism illustration with a QRD process on a $4 \times 4$ matrix made of real entries.

E. Reduced Complexity of Parallel Givens Rotations

The main idea of the Reduced Complexity PGR (RCPGR) technique is to perform the computational operations in parallel, for both the real and imaginary parts of $H$, and the matrix $Q$ is not explicitly obtained.

**Lemma 1:** Consider a modified $n_R \times (n_T+1)$ input matrix constructed as $[H \ y]$ and with $n_R \geq n_T$, the QRD process on the $n_T$ first rows results in a $n_T \times (n_T+1)$ triangular matrix corresponding to $[R \ y]$, with $\tilde{y} = Q^H y$.

**Proof:** Let us consider the initialization step of the QRD algorithm by introducing the $n_R \times (n_T+1)$ matrix $Q = [H \ y]$.
A. Computational Complexities

Consequently GS technique. By processing the QRD from columns like description. This considerably reduces the computational complexity as will be seen in the following section. Lemma 1 implies the inputs of the detection stage, i.e., $\tilde{y}$ and $R$, are obtained without explicitly calculating the unitary matrix $Q$. This technique is depicted in Algorithm 4 in a MATLAB-like description. This considerably reduces the computational complexity as will be seen in the following section.

IV. SIMULATION RESULTS

A. Computational Complexities

Computational complexities\(^2\) of the previously evoked QRD techniques have been theoretically obtained and are displayed in Table I. Herein, we consider the computational complexity to decompose a $n_R \times n_T$ complex matrix.

The corresponding computational complexities are plotted in Figure 2 and the $y$ received vector computation is also taken into account, which is not the case in general. These simulation results roughly match the classical theoretical complexities [13], but they are more realistic in any Digital Signal Processor (DSP) fashion programming. GR are more complex in terms of the required number of operations compared to STGS: for a $8 \times 8$ complex matrix, the additional complexity is 164%. This difference is due to the need for explicitly computing the $Q$ matrix. Also, results show that RCPGR achieves a computational complexity reduction of 39\% and 77\% compared to STGS and GR, respectively, all for a $8 \times 8$ complex matrix.

Results on computational complexities have been presented in the paper and offer strictly the same performance. This point is verified in the Subsection IV-C with the LTE-A parameters background.

B. Algorithmic Parallelism

Let us define the parallelism gain as the ratio of the number of parallel operations by the total number of operations. The parallelism gain of the summarized QRD algorithms is plotted in Figure 3, with in reference the STGS-based QRD and the proposed RCPGR-based QRD, as a function of the size of the decomposed complex matrix; the size of the channel matrix ranges from $2 \times 2$ to $8 \times 8$, which are the available modes in the LTE-A standard.

The proposed RCPGR-based QRD is more complex, in terms of computational complexity, compared to all the classical QRD. However, thanks to the parallelism potentiality of the algorithm, the proposed technique is the best in terms of execution time, which is an essential point for implementation considerations. Performances are shown to be strictly the same in Section IV-C.

\(^2\)The assumptions are the following: a real product is denoted 1 MUL, a real addition as 0 MUL (Multiply Accumulate (MAC) operation), a real division as 16 MUL (conditional add-subtract algorithm) and a real square root as 32 MUL (dynamic number between $\frac{1}{4}$ and 1 and Taylor series development).

| QRD       | MUL |
|-----------|-----|
| STGS      | $4n_R^2n_T + 34n_Rn_T^2 + 32n_T$ |
| $HH$      | $8n_R^2n_T + 6n_Rn_T^2 + \frac{2}{3}n_R^3 + 3n_T^2 + \frac{295}{6}n_T$ |
| GR        | $17n_R^2n_T - n_Rn_T^2 + 134n_Rn_T^2 - \frac{7}{2}n_R^3 - 73n_T^2 - \frac{212}{7}n_T$ |
| PGR       | $17n_R^2n_T - n_Rn_T^2 + 134n_Rn_T^2 - \frac{7}{2}n_R^3 - 73n_T^2 - \frac{212}{7}n_T$ |
| RCPGR     | $n_R^2n_T + 7n_Rn_T^2 + 154n_Rn_T^2 - \frac{7}{2}n_R^3 - 81n_T^2 - \frac{236}{7}n_T$ |

Fig. 2: Computational complexities of the summarized techniques as a function of the size of the to decompose complex matrix.
Possible computational complexity. The ML detector is also considered since it provides performance lower bound, but for an exponential computational complexity. The ML is practically infeasible in the LTE-A background. For example, by considering a QPSK modulation on each transmit antenna and for a 8 × 8 system configuration, 4^8 = 65536 Euclidean distances computations are necessary to decode 16 bits only. The QRD-M performances have been shown to be close to that of the optimum ML, for a polynomial mean complexity and with a parallel preprocessing using the RCPGR algorithm.

V. CONCLUSIONS

In this paper, several QR decomposition techniques have been addressed. The computational complexities of those algorithms as well as their methodology have been investigated. Our attention in this paper has also been paid to parallelism capabilities of the channel factorization where the parallelization gain of the Givens rotations technique, i.e., RCPGR, has been outlined. The performances of the detection algorithms can be improved by using signal ordering, which was the topic of several studies in the literature.

ACKNOWLEDGMENT

This work was supported by ST-ERICSSON.

REFERENCES

[1] E. Telatar, “Capacity of multi-antenna Gaussian channels,” European Trans. on Telecommunications, vol. 10, pp. 585-595, Dec. 1999.
[2] K.F. Lee and D.B. Williams, “A space-frequency transmitter diversity technique for OFDM systems,” Global Telecomm. Conf., vol. 3, pp. 1473-1477, 2000.
[3] D. Wubben, R. Bohneke, V. Khun, and K.-D. Kammeyer, “MMSE extension of V-BLAST based on sorted QR decomposition,” in Proc. IEEE Vehicular Tech. Conf., pp. 508-512, Oct. 2003.
[4] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, “Closest point search in lattices,” IEEE Transactions on Information Theory, vol. 48, no. 8, pp. 2201-2214, Nov. 2002.
[5] K. J. Kim, J. Yue, R. A. Itis, and J. D. Gibson, “A QRD-Kalman filter based detection and channel estimation algorithm for MIMO-OFDM systems,” IEEE Transactions on Wireless Communications, vol. 4, no. 2, pp. 710-721, Mar. 2005.
[6] A. Lenastra, H. Lenstra, and L. Lovasz, “Factoring polynomials with rational coefficients,” Mathematische Annalen, vol. 261, pp. 515-534, 1982.
[7] H. Shin and J.H. Lee, “Capacity of multiple-antenna fading channels: spatial fading correlation, double scattering, and keyhole,” IEEE Trans. on Information Theory, vol.49, no.10, pp. 2636-2647, Oct. 2003.
[8] P.W. Wohltiansky, G.J. Foschini, G.D. Golden and R.A. Valenzuela, “V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel,” in Signals, Systems, and Electronics, Int. Symp. on, pp. 295-300, Sep. 1998.
[9] D. Wubben, R. Bohneke, V. Khun, and K.-D. Kammeyer, “Efficient algorithm for decoding layered space-time codes,” Electronics Letters, vol. 37, no. 22, pp. 1348-1350, Oct. 2001.
[10] W. Van Etten, “Maximum likelihood receiver for multiple channel transmission systems,” IEEE Trans. on Commun., pp. 276-283, Feb. 1976.
[11] L. Hogben, “Handbook of linear algebra,” CRC Press, 2002.
[12] P. Luethi, A. Burg, S. Haene, D. Perels, N. Felber, and W. Fichtner, “VLSI Implementation of a High-Speed Iterative Sorted MMSE QR Decomposition,” IEEE Int. Symp. on Circuits and Systems, pp. 1421-1424, May 2007.
[13] N.J. Higham, “Accuracy and Stability of Numerical Algorithms,” Siam, 2002.