Performance Analysis of Modified Gram-Schmidt Cholesky Implementation on 16 bits-DSP-chip

Rabah Maoudj, Luc Fety, Christophe Alexandre

International Journal of Computing and Digital Systems

Performance Analysis of Modified Gram-Schmidt Cholesky Implementation on 16 bits-DSP-chip

Rabah Maoudj[1], Luc Fety[2] and Christophe Alexandre[3]

Lab. CEDRIC / LAETITIA, CNAM, Paris, France

[1]rabah.maoudj@cnam.fr, [2]luc.fety@cnam.fr, [3]christophe.alexandre@cnam.fr

Received 19 Oct. 2012, Revised 20 Nov. 2012, Accepted 15 Dec. 2012

Abstract: This paper focuses on the performance analysis of a linear system solving based on Cholesky decomposition and QR factorization, implemented on 16bits fixed-point DSP-chip (TMS320C6474). The classical method of Cholesky decomposition has the advantage of low execution time. However, the modified Gram-Schmidt QR factorization performs better in term of robustness against the round-off error propagation. In this study, we have proposed a third method called Modified Gram-Schmidt Cholesky Decomposition. We have shown that it provides a compromise of the two performance criterias cited above. A joint theoretical and experimental analysis of global performance of the three methods has been presented and discussed.

Keywords: fixed-point; cholesky; qr; system solving; dsp; signal processing.

I. INTRODUCTION

This paper presents the implementation results on the 16bit-fixed-point DSP chip (TMS 320C6474 which is optimized for 16bits arithmetic operations), [1] of a linear system solving performed by Cholesky decomposition and modified Gram-Schmidt process.

The algorithms that contain a matrix inversion or a least-squares problem are nowadays frequently used in a vast range of domains. But wireless telecommunications is the main user of linear system solving, as attest the several works published in the IEEE. For the field of wireless telecommunications, linear system solving are found mainly in the algorithmic part of the propagation channel equalization and interference cancellation for MIMO (Multi Input Multi Output) systems where the least square problem is often encountered [2, 3] as result of solving MMSE (Minimum Mean Square Estimation) or LSE (Least Square Estimation) algorithms.

In the case of least square problem, the most popular techniques are the classical Cholesky decomposition and QR factorization as will be detailed in section 2. We can show that the lower triangular matrix in the classical Cholesky decomposition can be obtained by the QR factorization and vice versa. This property can be used for performing a solving system using only the upper triangular matrix issued from QR factorization without using the orthogonal matrix Q as in the classical Cholesky method. We call this method Gram-Schmidt-Cholesky (GS-Cholesky). In this work, we try to evaluate the efficiency versus the round-off error of this method compared to the classical method of Cholesky.

Essentially it is shown that the QR factorization using the modified Gram-Schmidt process (MGS) is less sensitive to the round-off error than the classical Gram-Schmidt process (GS) when using the fixed-point calculation format [4, 5]. Therefore, the study will focus on evaluating the robustness of methods versus their round-off errors in intermediate calculations.

In the second section, we give an overview of the classical Cholesky decomposition and MGS-QR factorization algorithms that are implemented along with the analytical aspects. In the third section, the 16-bits implementation of solving systems based on Cholesky decomposition, QR factorization and GS-Cholesky are detailed. In the fourth section, a theoretical study followed by an experimental evaluation is performed. The performance in sense of time execution and robustness against the round-off error in intermediate calculations due to 16-bit resolution are discussed.

II. LINEAR SYSTEM SOLVING BASED ON CHOLESKY DECOMPOSITION AND QR FACTORIZATION

The Cholesky decomposition is among the well-known methods and most popular for the linear system solving. This method is applied only if the matrix is Hermitian positive definite which is often encountered when dealing with least

square problems. The classic Cholesky decomposition can be achieved in two ways as we shall develop below.

A. Least square problem

The least square problem results often from an over-determined linear system [6]. It means that the number of equations M, is greater than the number N of unknowns. This situation occurs in the case of pilot-aided propagation channel estimation with assumption of low rank channel [2, 7]. The following mathematical expressions can summarize the situation.

We assume a linear system given by,

$$Ax - b = \varepsilon \tag{1.}$$

With, $A \in C^{M\times N}$, $b \in C^{M\times 1}$ and $\varepsilon \in C^{M\times 1}$

The two most popular methods to resolve the equation (1.) for $min\|\varepsilon\|^2$ are given by the following sections 1 and 2.

1) Normal equation method

The normal equation method is based on the search of the minimum modulus of the residual error ε as expressed in the following equation,

$$min\|\varepsilon\|^2 \Rightarrow \frac{d(Ax-b)^H(Ax-b)}{dx} = 0$$

Therefore, $$A^H Ax = A^H b \tag{2.}$$

If we set, $\Gamma = A^H A$ and $\beta = A^H b$ then,

$$\Gamma x = \beta \;\; and \;\; x = \Gamma^{-1}\beta \tag{3.}$$

With,
$\Gamma \in C^{N\times N}$ being Hermitian positive definite matrix and $\beta \in C^{N\times 1}$.

Γ can be decomposed by the Cholesky method, and the equation (3.) becomes,

$$x = (LL^H)^{-1}\beta \tag{4.}$$

With, $\in C^{N\times N}$, a lower triangular matrix

2) QR factorization

It consists in a QR decomposition of A. Then equation (1.) becomes,

$$QRx = b + \varepsilon \Rightarrow x = R^{-1}Q^H b\,,\; \varepsilon = 0 \tag{5.}$$

With,

$A = QR$

$R \in C^{N\times N}$, an upper triangular matrix
$Q \in C^{M\times N}$, an orthogonal matrix

From section ***A*** and section ***B*** we have,

$$\Gamma = A^H A = L^H L = RR^H \Rightarrow L = R^H$$

Then, Q can be obtained from the Cholesky decomposition by the following system solving,

$$A = QL^H \Rightarrow Q = AL^{-H} \tag{6.}$$

The operation in equation (6.) is commonly called Cholesky orthogonalization of matrix A.
We conclude in this case that Cholesky decomposition is a way to obtain QR decomposition and vice versa. This observation confirms that the QR factorization requires more execution time than the Cholesky decomposition since QR can be considered as a Cholesky decomposition, plus construction of matrix Q.
In the rest of the paper, we focus on the experimental comparison of robustness versus round-off errors introduced by the intermediate calculations in 16-bit fixed-point of the Cholesky decomposition. For this purpose, we evaluated the classical method and the Gram-Schmidt process, by taking Gram-Schmidt QR decomposition as reference.

B. Classical Cholesky decomposition

This method needs to compute initially the Hermitian matrix $\Gamma = A^H A$ before performing the classical Cholesky decomposition process as given below.
Let's denote γ_{ij} the elements of matrix Γ, and l_{ij} the elements of matrix L which is the Cholesky decomposition of Γ with $(1 \le i,j \le N)$. Then the elements l_{ji} are obtained by the following algorithm.

for $i = 1$ *to* N
  $l_{ii} = \left(\gamma_{ii} - \sum_{k=1}^{i-1} {l_{ik}}^2\right)^{\frac{1}{2}}$
  for $j = i + 1$ *to* N
    $l_{ji} = \frac{1}{l_{ii}}\left(\gamma_{ij} - \sum_{k=1}^{i-1} l_{ik} l_{jk}\right)$
  end for
end for

C. Modified Gram-Schmidt process

This second method does not require to pre-compute the Hermitian matrix Γ, but uses the matrix A directly as input. Matrix Γ is implicitly obtained in a sequential manner in the process of QR decomposition. This decomposition uses the Gram-Schmidt process as given by the following algorithm.

We denote, r_{ij} the elements of matrix R and q_{ij} the elements of matrix Q.

$$
\begin{aligned}
&Q = A \\
&for\ i = 1\ to\ N \\
&\quad r_{ii} = \left(\sum_{m=1}^{M} {q_{mi}}^{H} q_{mi}\right)^{\frac{1}{2}} \\
&\quad for\ j = i + 1\ to\ N \\
&\quad\quad r_{ij} = \frac{1}{r_{ii}}\left(\sum_{m=1}^{M} {q_{mj}}^{H} q_{mi}\right)^{H} \\
&\quad\quad for\ m = 1\ to\ M \\
&\quad\quad\quad q_{mj} = q_{mj} - \frac{1}{r_{ii}}\left(q_{mi} r_{ij}\right) \\
&\quad\quad end\ for \\
&\quad end\ for \\
&end\ for
\end{aligned}
$$

The number of mathematical operations required in floating point, for Cholesky and QR decompositions, associated to the solving systems proposed in (4.) and (5.), is given in the following tables.

The number of square roots is not included in the tables because it is the same for different methods, and equal to N. It should be noted that, multiplication and addition are complex operations. All the division operations used in the different algorithms are divisions of a complex number by a real number.

Table 1. Cost in number of multiplication and addition operation

Multiplications and additions	***Cholesky***	***QR***	***GS-Cholesky***
$A^H A$	$\frac{MN(N+1)}{2}$	-	-
$A^H b$	MN	-	MN
Classical Cholesky factorization.	$\frac{N(N-1)^2}{2}$	-	-
Modified Gram Schmidt process	-	MN^2	MN^2
$x = (LL^H)^{-1}A^H b$	$N(N-1)$	-	$N(N-1)$
$x = R^{-1}Q^H b$	-	$\frac{N(2M+N-1)}{2}$	-

Table 2. Cost in number of division operation

Divisions	***Cholesky***	***QR***	***GS-Cholesky***
$A^H A$	-	-	-
$A^H b$	-	-	-
Classical Cholesky factorization.	$\frac{N(N-1)}{2}$	-	-
Modified Gram Schmidt process	-	$N\left(M + \frac{N-1}{2}\right)$	$N(N-1)$
$x = (LL^H)^{-1}A^H b$	$2N$	-	$2N$
$x = R^{-1}Q^H b$	-	N	-

III. DSP IMPLÉMENTATION

This section is dedicated to the practical 16bit fixed-point implementation of the system solving algorithms described in the last section. In order to compare results of robustness of these various algorithms, we give the exact code embedded in the DSP by flowcharts 1 to 3 shown below.

The main functions are given first, followed by the final DSP implementation of each algorithm.

The $Q_{x/y}$ notation of fixed point format is widely used in fixed-point arithmetic [8] and briefly explained in [9].

We note that $Q_{x/y}$ do not mean the classical representation $Q_{a.b}$ (where a is the number of integer bits and b the number of fractional bits). The $Q_{x/y}$ representation means here that x is the number of fractional bits and y the number of total available bits including signed bit.

These two representations are related by this expression,

$$y = 1 + a + b \ \text{ and } \ x = b$$

The initial and final vectors and matrices are expressed on $Q_{15/16}$ format. Intermediate vectors and matrices are in $Q_{Z/16}$ format in order to prevent over flow in multiplication process.

In the following flowcharts, the real positive number Z is calculated by the following equation:

$$Z = 15 - round(log_2(M) + 0.5) \qquad (7.)$$

In the following, Figures 1 and 3 show respectively the classical Cholesky decomposition and Modified Gram-Schmidt factorization processes as implemented on the DSP. Once these processes are performed the triangular matrices resulted are carried to the final step of the system solving where the solution is obtained as depicted by figure 2.

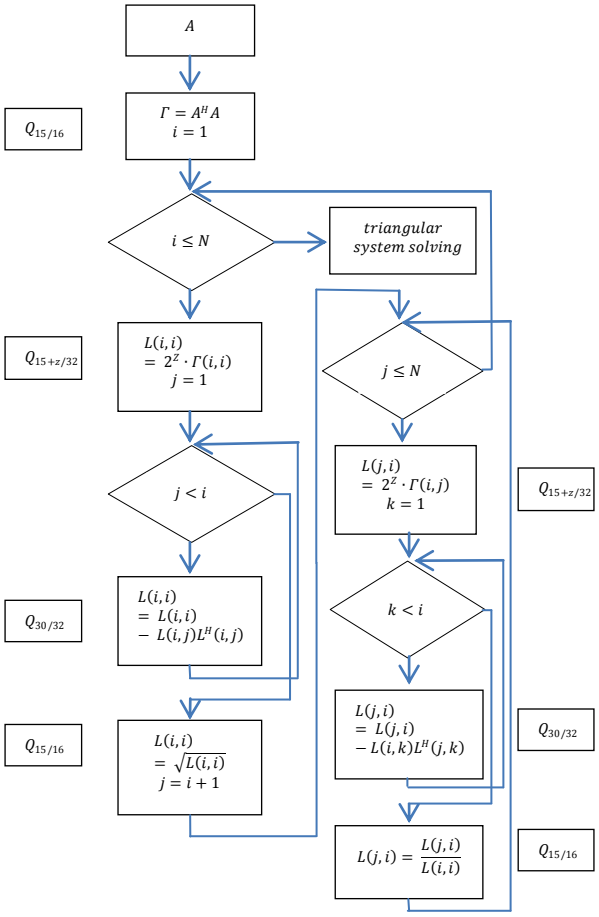

Figure 1. Flowchart (1) of Classical Cholesky decomposition

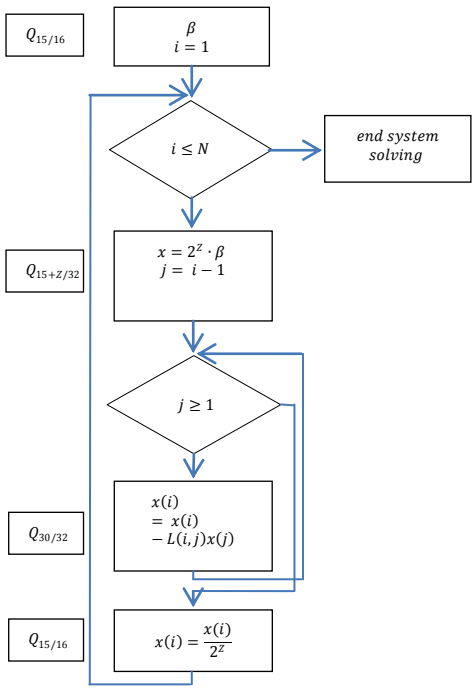

Figure 2. Flowchart (2) of triangular system solving

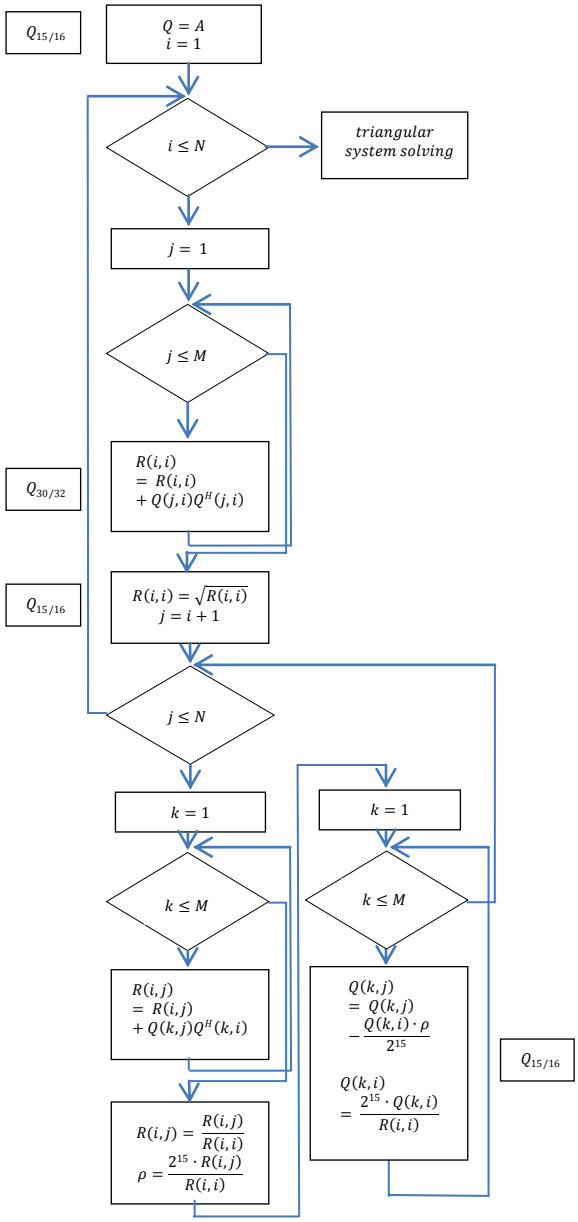

Figure 3. Flowchart (3) of modified Gram-Schmidt QR factorization

Finally, the three system solving implemented in the DSP are given as follows,

A. System solving using classical Cholesky method, as depicted in fig. (4-a)

B. System solving using GS-Cholesky method, as depicted in fig. (4-b)

C. System solving using QR method, as depicted in fig. (4-c)

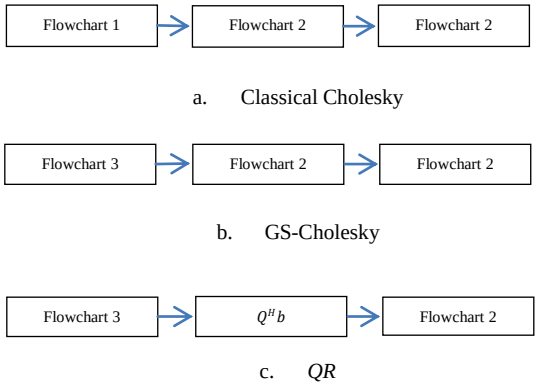

Figure 4. System solving algorithms

IV. NUMERICAL RESULTS

A. Theoretical analysis of round-off errors

In this section, we assume that the absolute value of all real and imaginary parts are less than 1. All calculations are performed without overflow. The additions and subtractions are made with $t + Z$ bits ($t = 15$). It means that the following error model for the arithmetic operations used to process the various algorithms given in last section is accurate [10]. Let's denote $x, y,$ and w to be complex variables with absolute values of their real and imaginary components inferior to 1. In addition $\varepsilon_x, \varepsilon_y, \varepsilon_w$ are the corresponding complex errors.

1) Addition and subtraction

$$w + \varepsilon_w = (x + \varepsilon_x) + (y + \varepsilon_y), \quad |\varepsilon_x| \le 2^{-t-1}, |\varepsilon_y| \le 2^{-t-1}$$

Then,

$$|\varepsilon_w| \le 2^{-t-1}$$

2) Multiplication

$$w + \varepsilon_w = (x + \varepsilon_x) \times (y + \varepsilon_y)$$

Then,

$$|\varepsilon_w| \le 2^{-t-1}$$

In the case of scalar product of two vectors the model is,

$$w + \varepsilon_w = \sum_{k=1}^{K} (x_k + \varepsilon_{xk}) \times (y_k + \varepsilon_{yk})$$

Then,

$$|\varepsilon_w| \le 2^{-t-1}$$

3) Division

$$w + \varepsilon_w = \frac{(x + \varepsilon_x)}{(y + \varepsilon_y)} = (x + \varepsilon_x) \times \frac{1}{(y + \varepsilon_y)}$$

By Taylor development of $\frac{1}{(y+\varepsilon_y)}$ around y, we have

$$\frac{1}{(y + \varepsilon_y)} \approx \frac{1}{y} - \frac{\varepsilon_y}{y} \Rightarrow \varepsilon_w \approx \frac{1}{y}(\varepsilon_x + x\varepsilon_y)$$

Then,

$$|\varepsilon_w| \le \frac{1}{|y|} 2^{-t}$$

4) Square root

$$w + \varepsilon_w = \sqrt{(x + \varepsilon_x)}$$

By Taylor development of $(x + \varepsilon_x)^{\frac{1}{2}}$ around x,

$$\sqrt{(x + \varepsilon_x)} \approx \sqrt{x} - \frac{\varepsilon_x}{\sqrt{x}} \Rightarrow \varepsilon_w \approx \frac{\varepsilon_x}{\sqrt{x}}$$

Then,

$$|\varepsilon_w| \le \frac{1}{|\sqrt{x}|} 2^{-t-1}$$

Now we apply this error model to the Cholesky decomposition and QR factorization.

5) Cholesky decomposition

According to the flowchart 1, Γ is obtained by,

$$\Gamma + E_\Gamma = (A + E_A)^H (A + E_A)$$

With,

$$|\varepsilon_{A(i,j)}| \le 2^{-t-1}$$

We denote by: E_Γ, and E_A are the error matrices corresponding to Γ and A respectively.

Therefore,

$$|\varepsilon_{\Gamma(i,j)}| \le 2^{-t-1}, \quad 1 \le i, j \le N$$

The diagonal elements of matrix are given by the following equation,

$$l_{ii} + \varepsilon_{l_{ii}} = \left(\gamma_{ii} + \varepsilon_{\gamma_{ii}} - \sum_{k=1}^{i-1} (l_{ik} + \varepsilon_{l_{ik}})^2 \right)^{\frac{1}{2}}$$

For $i = 1$, the round-off error is $|\varepsilon_{l_{ii}}| \le \frac{1}{|l_{ii}|} 2^{-t-1}$

For $i > 1$, $|\varepsilon_{l_{ii}}| \le \frac{1}{|l_{ii}|} (2^{-t-1} + 2\sum_{k=1}^{i-1} \varepsilon_{l_{ik}})$,

$\varepsilon_{l_{ik}}$ is calculated iteratively as shown below. The non-diagonal element l_{ji} of the matrix L is given by,

$$l_{ji} + \varepsilon_{l_{ji}} = \frac{1}{l_{ii} + \varepsilon_{l_{ii}}}\left(\gamma_{ij} + \varepsilon_{\gamma_{ij}} - \sum_{k=1}^{i-1}\left(l_{ik} + \varepsilon_{l_{ik}}\right)\left(l_{jk} + \varepsilon_{l_{jk}}\right)\right)$$

$$\varepsilon_{l_{ji}} \approx \frac{1}{l_{ii}}\left(\varepsilon_{\gamma_{ij}} + \left(\gamma_{ij} - \sum_{k=1}^{i-1} l_{ik} l_{jk}\right)\varepsilon_{l_{ii}}\right)$$

Then,

$$\left|\varepsilon_{l_{ji}}\right| \le \frac{1}{|l_{ii}|}\left(2^{-t-1} + \left|\frac{\varepsilon_{l_{ii}}}{l_{ii}}\right|\right)$$

The upper limit of the round-off error of the components of the matrix L is too difficult. But if we assume the calculation with only two recursive iterations, we find the following result.

$$\left|\varepsilon_{l_{ii}}\right| \le \frac{2^{-t-1}}{|l_{ii}|}\left(1 + 2\frac{1}{|l_{i-1i-1}|}\left(1 + \left|\frac{1}{l_{i-1i-1}}\right|\right)\right)$$

And,

$$\left|\varepsilon_{l_{ji}}\right| \le \frac{1}{|l_{ii}|}\left(2^{-t-1} + \left|\frac{\varepsilon_{l_{ii}}}{l_{ii}}\right|\right)$$

6) QR Factorization

According to the flowchart 3, the diagonal elements of matrix R can be written as,

$$r_{ii} + \varepsilon_{r_{ii}} = \left(\sum_{m=1}^{M}\left(q_{mi} + \varepsilon_{q_{mi}}\right)^{H}\left(q_{mi} + \varepsilon_{q_{mi}}\right)\right)^{\frac{1}{2}}$$

Then the round-off error $\varepsilon_{r_{ii}}$ can be bounded by,

$$\left|\varepsilon_{r_{ii}}\right| \le \frac{1}{r_{ii}} 2^{-t-1}$$

For the non-diagonal element of the matrix R, the round-off error is given as follows,

$$r_{ij} + \varepsilon_{r_{ij}} = \frac{1}{r_{ii} + \varepsilon_{r_{ii}}}\left(\sum_{m=1}^{M}\left(q_{mj} + \varepsilon_{q_{mj}}\right)^{H}\left(q_{mi} + \varepsilon_{q_{mi}}\right)\right)^{H}$$

$$\varepsilon_{r_{ij}} \approx \frac{1}{r_{ii}}\left(\sum_{m=1}^{M}\left({q_{mj}}^{H}\varepsilon_{q_{mj}} + q_{mi}\,\varepsilon_{mi}\right) + \varepsilon_{r_{ii}}\sum_{m=1}^{M}{q_{mj}}^{H} q_{mi}\right)$$

Then,

$$\left|\varepsilon_{r_{ij}}\right| \approx \left|\frac{\varepsilon_{r_{ii}}}{r_{ii}}\right| \le \frac{1}{|r_{ii}|^2} 2^{-t-1}$$

Finally we calculate the round-off error of matrix Q.

The matrix Q is obtained by performing the following equation,

$$q_{mj} + \varepsilon_{q_{mj}} = q_{mj} + \varepsilon_{q_{mj}} - \frac{1}{r_{ii} + \varepsilon_{r_{ii}}}\left(\left(q_{mi} + \varepsilon_{q_{mi}}\right)\left(r_{ij} + \varepsilon_{r_{ij}}\right)\right)$$

$$\varepsilon_{q_{mj}} \approx \varepsilon_{q_{mj}} + \frac{1}{r_{ii}}\left(q_{mi}\,\varepsilon_{r_{ij}} + r_{ij}\,\varepsilon_{q_{mi}} + q_{mi}\,r_{ij}\,\varepsilon_{r_{ii}}\right)$$

Then,

$$\left|\varepsilon_{q_{mj}}\right| \le \left(\frac{3}{{r_{ii}}^2} + 1\right) 2^{-t-1}$$

7) Observation

From the sections e and f, we observe that the round-off error is greater in the case of Cholseky decomposition than in the case of QR factorization. This round-off error depends on the values of l_{ii} or r_{ii}, but the Cholesky decomposition is more sensitive to l_{ii} than QR factorization to r_{ii}.

B. Simulation results

The fixed point implementation is done according to flowcharts 1, 2, and 3. The target is TMS 320C6474 with 1GHz clock rate. The experimental results depicted by figures 5, 6 and 7 are obtained using Monte Carlo statistical method of solving system $Ax = b$ with well-conditioned $(A^H A)$ matrices i.e.,(condition number $(A^H A) \approx 30$). Simulations are performed for a fixed row number $M = 16$ and several values of column number $N = 4, 6, 8, 10, 12, 14$. Figure 5 depicts the relative norm of rounding errors on the triangular matrix L, $\frac{\|L - L_0\|_2}{\|L_0\|_2}$. The reference triangular matrix L_0 is obtained with full precision floating-point (IEEE float-point) format. Figure 6 depicts the square root of the mean square error $\|Ax - b\|_2$. These experimental results confirm the theoretical ones and show that the MGS-QR is less sensitive to the round off errors than Classical Cholesky and MGS-Cholesky. However MGS-Cholesky is more robust than Classical Cholesky. This classification criteria based on round-off error immunity is inverted when applying the time execution criteria. Classical Cholesky requires twice less execution time than MGS-QR and around 1.5 less than MGS-Cholesky as depicted in figure 7.

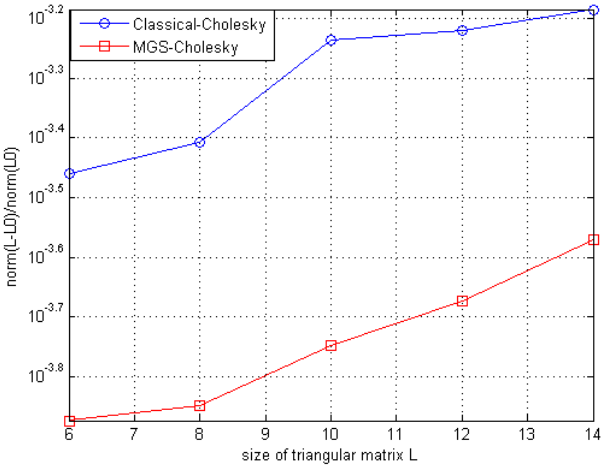

Figure 5. 2-norm of error triangular matrix $\boldsymbol{L}$, $\frac{\|L - L_0\|_2}{\|L_0\|_2}$

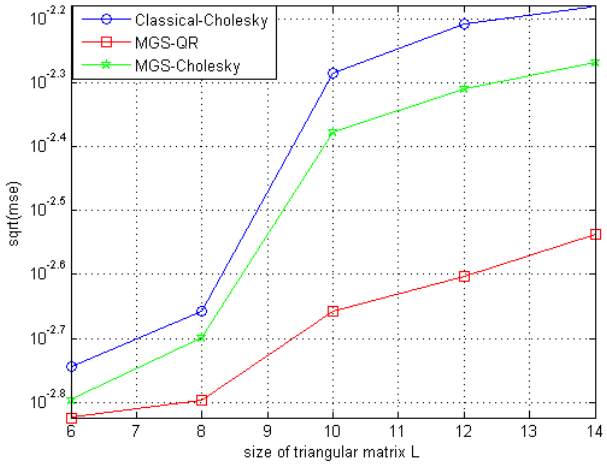

Figure 6. Mean square error, $\|\boldsymbol{Ax}-\boldsymbol{b}\|_2$

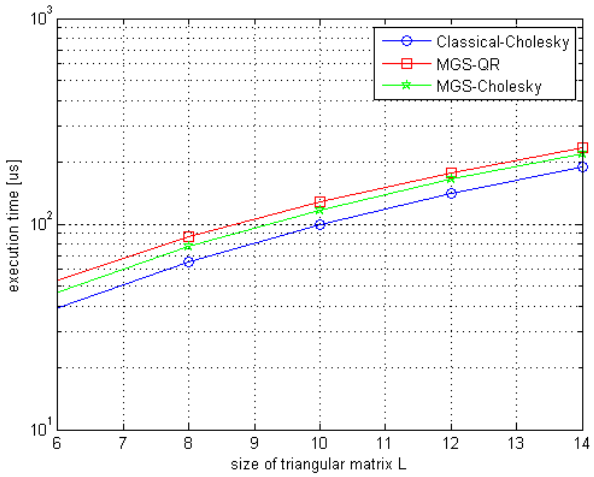

Figure 7. Execution time (us)

V. CONCLUSION

In this paper, we have analyzed, theoretically and experimentally using Matlab simulations, the solving system using classical Cholesky , MGS-QR and MGS-Cholesky. Both analyses are concordant. As expected through the theoretical analysis, the experimental analysis has confirmed that the MGS-QR is more robust than the classical Cholesky and MGS-Cholesky. However this robustness has a cost in terms of execution time. Classical Cholesky requires twice less execution time than MGS-QR and around 1.5 less than MGS-Cholesky. With the introduction of the MGS-Cholesky, we give an alternative to MGS-QR if we search a good robustness of the system solving against additive round-off error in the final solution with a slight degradation in execution time performance.

REFERENCES

[1] TMS320C6474 Multicore Digital Signal Processor Data Manual (Rev. H), April 2001.

[2] Rabah Maoudj, Christophe Alexandre, Denis Popielski, Michel Terré, "DSP Implementation of Interference Cancellation Algorithm for a SIMO System", ISWCS 2012, Paris 2012.

[3] Saqib Saleem, Qamar-ul-Islam, "Optimization of LSE and LMMSE Channel Estimation Algorithms based on CIR Samples and Channel Taps," IJCSI International Journal of Computer Science Issues, vol. 8, Issue 1, January 2011.

[4] A. Bjorck, "Numerics of Gram-Schmidt Orthogonalization," Linear Algebra and Its Applications, vol. 198, pp. 297–316, Feb. 1994.

[5] Chitranjan K. Singh, Sushma Honnavara Prasad, and Poras T. Balsara," VLSI Architecture for Matrix Inversion using Modified Gram-Schmidt based QR Decomposition", 6th International Conference on Embedded Systems, 20th International Conference on, pp. 836 - 841, Jan. 2007.

[6] Charles L. Lawson, Richard J. Hanson, "Solving Least Squares Problems", Printice-Hall, Inc., Englewood Cliffs, N. J, 1974.

[7] Chang Dong Lee, Jin Sam Kwak, and Jae Hong Lee, "Low-Rank Pilot-Symbol-Aided Channel Estimation for MIMO-OFDM Systems", IEEE VTC'2004, vol. 7, pp. 469-473. 2004.

[8] Israel Koren, "Computer Arithmetic Algorithms", Brookside Court Publishers, 1998.

[9] Ali Irturk, Shahnam Mirzaei and Ryan Kastner, "FPGA Implementation of Adaptive Weight Calculation Core Using QRD-RLS Algorithm", Technical repport CS2009-0937, Mar. 2009.

[10] J.H. Wilkinson, "The Algebraic Eigenvalue Problem", Oxford University Press Inc, New York, 1965.

.