Compressive Sensing Least Square problem solution suitable for implementation

Andjela Draganić, *Member, IEEE,* Irena Orović, *Member, IEEE* and Srdjan Stanković, *Senior Member, IEEE*

Abstract — **Compressive Sensing approach allows reconstruction of under-sampled sparse signals, by using different optimization techniques. These techniques solve undetermined systems of equations which may be recast as least square problems. Since there is a growing need for real-time hardware implementations of the reconstruction methods, it is important for these methods to be fast enough and not be computationally demanding. Here, we will focus on QR decomposition based approach for solving least square problems. Least square problem solution is defined in such way that does not require Q matrix, obtained as a result of QR decomposition of the measurement matrix, to be used in calculation and leads to the lower computational complexity.**

Keywords —**Compressive Sensing, classical Gram-Schmidt, Givens rotations, least square problem, modified Gram-Schmidt, QR decomposition.**

I. INTRODUCTION

Processing and analysis of discrete signals can be done in different domains (e.g. time, frequency, time-frequency). Signals can have different representations in different domains - e.g. in one domain signal can cover whole interval while in transform domain can express sparsity property. Sparsity of the signal means that only few non-zero components of the signal exist in the sparse domain, i.e. signal energy is concentrated within only those small number of non-zero coefficients [1]-[3]. Sparse signal can be sampled or reconstructed from an incomplete set of samples, by using Compressive Sensing (CS) approach and under certain conditions.

CS [4]-[5] is newly used approach for signal sampling and for its recovering from small amount of available samples. It allows signal to be samples by acquiring significantly smaller number of samples, compared to the Shannon-Nyquist theorem. In some cases and from some reasons we have left with reduced set of signal samples (e.g. signal samples are omitted as a result of noise). In such cases, CS allows recovering signal with high accuracy from this incomplete set. In order to recover the signal, a least square problem has to be solved [6]-[10]. It is set of linear equations with more equations than unknowns. Therefore, different optimization algorithms for solving least square problems are used [11]-[16].

When considering hardware implementation of the least square problems, the most demanding task is matrix inversion. Different methods for matrix inversion problems solving, suitable for hardware implementations, have been proposed in the literature [7]. In this paper, we will focus on the QR decomposition – a method for factorization of the matrix into two matrices, an upper triangular and an orthogonal one [17]. The inversion is then reduced to the inversion of the triangular matrix. QR decomposition method is widely used for solving least-squares problems in various applications. Several methods for QR matrix factorization exist, and they are used depending on the nature of the matrix and computational requirements. The least square problem is defined in a way that does not require matrix Q to be calculated and stored. The problem is recast to the triangular matrix R calculation and its inversion, which is less computationally demanding compared to the calculation both, Q and R matrices.

The paper is organized as follows: In Section II theoretical background on the CS approach is given. Least square problem is introduced in this section as well. Methods for solving least square problems are considered in the next section. In section IV, CS least square problem is simplified by using the QR decomposition. Block scheme for hardware architecture are given in this part. Conclusion is given in the section V.

II. COMPRESSIVE SENSING

Traditional signal sampling is done according to the Nyquist sampling theorem with the sampling rate at least twice higher than the maximal signal frequency. As an alternative way of signal sampling, the CS strategy has been recently introduced allowing the signal to be sampled with much fewer samples if the certain conditions are satisfied.

Authors are with the University of Montenegro, Faculty of Electrical Engineering, Dzordza Vasingtona bb, 81000 Podgorica, Montenegro. Corresponding author is Andjela Draganic (phone: +382 63 218 048, e-mail: andjelad@ac.me).

One of the required conditions for CS application is related to the nature of the signal. If the signal, in sensing domain, has dense representation and, at the same time, the sparse representation in some other domain, it is said that signal is sparse. Hence, the sparsity is the first CS condition, and it assures that the signal energy is concentrated in small number of coefficients. The second condition is related to the sampling procedure, which needs to assure incoherent measurements. It was shown that random selection of signal samples can provide incoherence allowing signal reconstruction using small number of samples. Reconstruction of the full signal from an incomplete set of samples is done by using optimization techniques [15]. The CS procedure will be shortly described in the sequel, and optimization problem will be introduced.

Let x be discrete signal that can be represented by using transform domain Ψ as [1], [3]:

$$x = \Psi s , \qquad (1)$$

where s represents transform domain coefficients. By randomly selecting M signal samples, the measurement vector y is formed as:

$$y = \Phi x = \Phi \Psi s \qquad (2)$$

where Φ $M \times N$ performs random selection of the rows of the $N \times N$ matrix Ψ. The system of equations (2) is an undetermined system, since there is larger number of unknowns (N) than equations (M). In order to find the sparsest solution s among very large number of possible solutions, optimization problems are used (e.g. l_1 minimization): $\min \|s\|_{\ell_1}$ $s.t.$ $y = \Phi \Psi s$. If the number of detected signal components K is smaller than the number of available signal samples M (i.e. $K<M$), than the problem (2) can be solved in the least square sense as [12]:

$$\hat{s} = (\mathbf{A}_{\mathbf{CS}}^{\mathrm{T}} \mathbf{A}_{\mathbf{CS}})^{-1} (\mathbf{A}_{\mathbf{CS}}^{\mathrm{T}} y) . \qquad (3)$$

Matrix $\mathbf{A}_{\mathbf{CS}}$ denotes $M \times K$ CS matrix, and it is obtained by selecting rows that correspond to the M available measurements and columns that correspond to the K signal frequencies from the original transform matrix Ψ. Note that, starting from the incomplete set signal samples y, the overdetermined system (3) is obtained. Vector $\hat{s}$ is a solution of the optimization problem, while $\mathbf{A}_{\mathbf{CS}}^{\mathrm{T}}$ denotes Hermitian transpose of the CS matrix.

III. LEAST SQUARE PROBLEMS AND QR DECOMPOSITION BASED SOLUTION

A. Methods for least square problem solving

The most challenging part for hardware implementation of the CS procedure, is the least square problem. Although the matrix $\mathbf{A}_{\mathbf{CS}}$ is not full matrix, but random partial transform matrix, it can still have large dimensions (depending on the number of available samples). Inversion, as well as multiplication of such matrices is a demanding task. Therefore, different methods that facilitate solving least square problems are defined:

1. Normal equations method by using Cholesky factorization;
2. Singular value decomposition (SVD);

3. Transformation to a linear system;
4. QR decomposition.

Each of the mentioned methods has its drawbacks and advantages. The Cholesky factorization is the fastest method for solving least square problems, but it is numerically unstable. The method based on SVD requires more computational work, and cannot work with the rank deficient matrices. The third procedure is the fastest but least accurate. Although the method based on SVD is more stable and robust than the QR approach, QR method can be applied on rank-deficient matrices and, therefore, we will focus on QR decomposition for solving least square problem.

QR decomposition (QR factorization) decomposes matrix $\mathbf{A}$ into two matrices: an orthogonal matrix $\mathbf{Q}$ and upper triangular matrix $\mathbf{R}$. Consider matrix $A \in \mathbb{R}^{m \times n}$, with full rank n. QRD of the matrix $\mathbf{A}$ results in:

$$\mathbf{A} = \mathbf{QR} , \qquad (4)$$

where $\mathbf{Q} \in \mathbb{R}^{n \times m}$ is an orthogonal matrix, with $\mathbf{Q}^T \mathbf{Q} = \mathbf{QQ}^T = \mathbf{I}$, $\mathbf{Q}^T = \mathbf{Q}^{-1}$. Matrix $\mathbf{R} \in \mathbb{R}^{n \times n}$ is an upper triangular matrix.

Several algorithms for QR decomposition calculation exist: Householder reflections/transformations [18], Gram Schmidt (GS) decomposition [19] and Givens rotations (GR) [20]. All of the methods have been considered for hardware realization, and all of the methods have their advantages and disadvantages. Householder transformation is more complex for hardware realization compared to the rest of the algorithms and cannot be efficiently parallelized. Therefore, we will focus on the GS and GR methods. Regarding the computational complexity, GS is numerically equivalent to the GR method.

Gram-Schmidt method allows parallel computations with high processing throughput. Therefore, GS-QR decomposition is commonly used in applications where the parallelization of computations is desirable. There are two variants of the GS method:

-Classical Gram-Schmidt method (CGS);
-Modified Gram-Schmidt method (MGS).

Both, CGS and MGS may produce a set of vectors which is far from orthogonal [19]. Also, in some cases the orthogonality can be completely lost. CGS algorithm requires re-orthogonalization, which increases computational complexity of the algorithm. The MGS algorithm never requires re-orthogonalization and therefore has much better numerical properties. It is important to note that there are applications where the orthogonality of computed vectors does not play a crucial role. In such applications CGS algorithm can be used.

The method based on Givens rotation algorithm and the coordinate rotation digital computer (CORDIC) algorithm can reduce hardware area and can be easily parallelized compared to the Householder reflections. However, this method has longer clock latency in the QRD procedure. Here, we will present and compare the Gram-Schmidt and the Givens rotation methods.

B. GS and GR methods

Let us summarize both versions of the Gram-Schmidt algorithm. Having the starting matrix $\mathbf{A}=\{a_1, a_2, ..., a_n\}$, a set of orthogonal vectors $\{q_1, q_2, ..., q_n\}$ can be obtained as:

$$q_1 = a_1 , \qquad (5)$$

$$q_j = a_j - \sum_{i=1}^{j-1} p_{i,j} q_i, \;\; for \;\; j = 2,...,n . \qquad (6)$$

Vectors $\{q_1, q_2, ..., q_n\}$ are the columns of the matrix $\mathbf{Q}$, while $p_{i,j}$ are obtained as:

$$p_{i,j} = \left(q_i^T a_j \right) / \left(q_i^T q_i \right) . \qquad (7)$$

The MGS algorithm is similar to the CGS, with slight modification of the relation (7):

$$p_{i,j} = \left(q_i^T a_j^{(i)} \right) / \left(q_i^T q_i \right), \qquad (8)$$

where

$$a_j^{(i)} = a_j - \sum_{k=1}^{i-1} p_{k,i} q_k . \qquad (9)$$

The second commonly used method for QR decomposition is based in GR. In recent years, QR decomposition based on this algorithm attracts much attention, because of good numerical properties and possibility to be easily parallelized. The GR method is based on set of plane rotations of the original matrix $\mathbf{A}$. Matrix $\mathbf{G}$, used for performing plane rotations, is defined as follows:

$$G(i,j,\theta) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & \cdots & C & \cdots & S & \cdots & 0 \\ \vdots & \cdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & \cdots & -S & \cdots & C & \cdots & 0 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{matrix} \\ \\ i \\ \\ j \\ \\ \\ \end{matrix} \qquad (10)$$

where coefficients C and S are defined as follows:

$$C = \cos(\theta_{i,j}), \;\; S = -\sin(\theta_{i,j}) . \qquad (11)$$

Multiplication of the matrix $\mathbf{A}$ with the matrix $\mathbf{G}$ annuls (i,j) element of the matrix $\mathbf{A}$. An upper triangular matrix $\mathbf{R}$ can be obtained through successive multiplication of the rotation matrix $\mathbf{G}$:

$$(\mathbf{G}_{n,n-1})...(\mathbf{G}_{n2}...\mathbf{G}_{32})(\mathbf{G}_{n1}...\mathbf{G}_{21})\mathbf{A} = \mathbf{R} , \qquad (12)$$

i.e.

$$\mathbf{Q}^T\mathbf{A} = \mathbf{R} . \qquad (13)$$

IV. QR BASED SOLUTION OF THE OPTIMIZATION PROBLEM

Block scheme for CS procedure can be represented as in the Fig. 1. At the input of the block scheme, signal $\mathbf{x}$,

transform matrix $\mathbf{\Psi}$ and matrix that models random selection of the samples $\mathbf{\Phi}$ is fed. At the output of the first block, CS matrix $\mathbf{A_{CS}}$ and vector of measurements $\mathbf{y}$ is obtained. The outputs of the Block 1 are then fed to the input of the Block 2 – block for optimization problem solving. It contains matrix transpose, matrix-matrix multiplication, inversion of the matrix and matrix-vector multiplication parts. The most numerically demanding part of the architecture is the matrix inversion part.
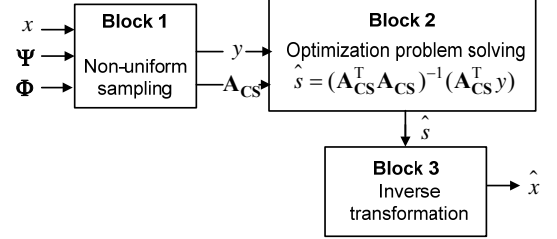

Fig.1. Block scheme for CS reconstruction

Despite the dimensionality reduction, product of the $\mathbf{A_{CS}}$ matrix and its transpose can still has large dimensions and can be computationally demanding. Therefore, the modification of the least square is proposed. By applying QR decomposition, the problem of inversion can be recast as inversion of triangular matrix. Inversion of triangular matrices is far less demanding for hardware implementation, compared to the inversion of the full matrix such as $\mathbf{A_{CS}}$.

QR decomposition of the CS matrix $\mathbf{A_{CS}}$, results in:

$$\mathbf{A_{CS}} = \mathbf{Q_{CS}}\mathbf{R_{CS}}, \qquad (14)$$

where $\mathbf{Q_{CS}}$ is an orthogonal, while $\mathbf{R_{CS}}$ is an upper triangular matrix. By substituting (14) in equation (3), the least square problem can be written as:

$$\hat{s} = \left((\mathbf{Q_{CS}}\mathbf{R_{CS}})^T (\mathbf{Q_{CS}}\mathbf{R_{CS}}) \right)^{-1} (\mathbf{A_{CS}}^T y) , \qquad (15)$$

i.e.

$$\begin{aligned} \hat{s} &= (\mathbf{R_{CS}}^T\mathbf{R_{CS}})^{-1}(\mathbf{A_{CS}}^T y) = \\ &= \mathbf{R_{CS}}^{-1}(\mathbf{R_{CS}}^{-1})^T(\mathbf{A_{CS}}^T y) \end{aligned} . \qquad (16)$$

Instead of multiplying matrices $\mathbf{A_{CS}}^T$ and $\mathbf{A_{CS}}$, and finding QR decomposition of the obtained matrix [21]: $\tilde{s} = \mathbf{R_{CS}}^{-1}\mathbf{Q_{CS}}^T y$, the $\mathbf{A_{CS}}$ matrix is firstly QR decomposed. It is important to note that $\mathbf{R_{CS}}$ matrix has $K{\times}K$ non-zero elements (where K is number of signal component), while matrix $\mathbf{Q_{CS}}$ is of $M{\times}M$ size. Number of available samples M, required for successful reconstruction, in some cases can be much larger than the number of components K. Consequently, matrix $\mathbf{Q_{CS}}$ can be much larger than $\mathbf{R_{CS}}$. The block scheme of the optimization problem solving is part is shown in Fig. 2.
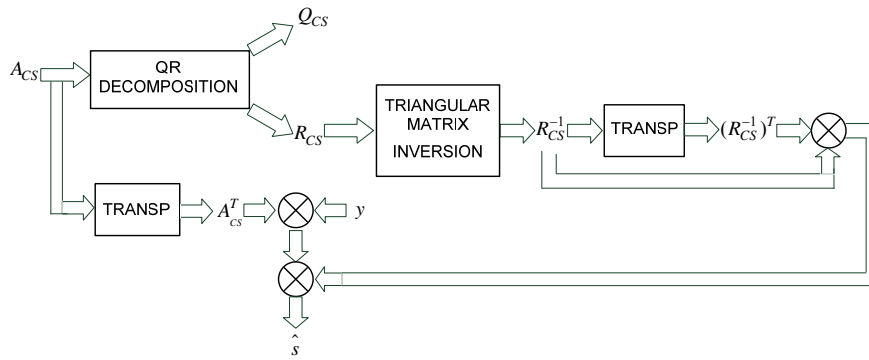
Fig.2. Block for optimization problem solving

V. CONCLUSION

The methods for solving least square problem, found in most of the CS reconstruction algorithms, are discussed in this paper. The focus was on the QR method. It is used for matrix decomposition and facilitates hardware implementations, leading to the reduced matrix size and lowering the computational complexity. The modified version of the least square algorithm is proposed in the paper as well. The proposed form uses only original transform matrix and its corresponding upper triangular matrix, avoiding matrix **Q** in the computations. This leads to lower computational complexity, as matrix **Q** can have larger dimensions compared to the triangular matrix **R**. Block schemes for the CS procedure and optimization problem are shown in the paper as well.

ACKNOWLEDGMENT

REFERENCES

[1] S. Stankovic, I. Orovic, E. Sejdic, "Multimedia Signals and Systems," Springer-Verlag, New York, 2012.

[2] S. Stankovic, LJ. Stankovic, I. Orovic, "Relationship between the Robust Statistics Theory and Sparse Compressive Sensed Signals Reconstruction," IET Signal Processing, Special issue on Compressive Sensing and Robust Transforms, vol. 8, no. 3, pp. 223 - 229, ISSN: 1751-9675, May, 2014.

[3] E. J. Candès, and M. B. Wakin, "An Introduction To Compressive Sampling, " IEEE Signal Processing Magazine 2008, vol 25, No 2, pp. 21-30, 2008.

[4] I. Orović, S. Stanković, "Improved higher order robust distributions based on compressive sensing reconstruction," IET Signal Processing, vol.8, no.7, pp.738-748, September 2014.

[5] I. Orovic, S. Stankovic, T. Thayaparan, "Time-Frequency Based Instantaneous Frequency Estimation of Sparse Signals from an Incomplete Set of Samples," IET Signal Processing, Special issue on Compressive Sensing and Robust Transforms, vol. 8, no. 3, pp. 239 - 245, ISSN: 1751-9675, May, 2014.

[6] G. H. Golub, M. A. Saunders, "Linear least squares and quadratic programming," Technical report no. CS 134, Stanford University, May 1969.

[7] A. Björck, "Solving linear least squares problems by Gram-Schmidt orthogonalization," BIT Numerical Mathematics March 1967, vol. 7, No 1, pp 1-21.

[8] D. Manolakis, F. Ling, J. G. Proakis, "Efficient least-squares algorithms for finite memory adaptive filtering," IEEE International Symposium on Circuits Systems, Montreal, P.Q., Canada, May 1984

[9] A. Bjorck, "Solving linear least squares problems by Gram-Schmidt orthogonalization," BIT, vol. 7, pp. 1-21, Jan. 1967.

[10] W. M. Gentleman, "Least squares computation by Givens transformation without square roots," J. Inst. Math. Appl., vol. 12, pp. 329-336, 1973

[11] Lin Bai, P. Maechler, M. Muehlberghuber, H. Kaeslin, "High-speed compressed sensing reconstruction on FPGA using OMP and AMP," 19th IEEE International Conference on Electronics, Circuits and Systems (ICECS), 2012, pp.53-56, 9-12 Dec. 2012

[12] S. Stankovic, I. Orovic, LJ. Stankovic, "An Automated Signal Reconstruction Method based on Analysis of Compressive Sensed Signals in Noisy Environment," Signal Processing, vol. 104, Nov 2014, pp. 43 - 50, 2014.

[13] S. Stankovic, I. Orovic, M. Amin, "L-statistics based Modification of Reconstruction Algorithms for Compressive Sensing in the Presence of Impulse Noise," Signal Processing, vol.93, no.11, November 2013, pp. 2927-2931 , 2013

[14] LJ. Stankovic, S. Stankovic, M. Amin, "Missing Samples Analysis in Signals for Applications to L-estimation and Compressive Sensing," Signal Processing, vol. 94, Jan 2014, pp. 401-408, 2014.

[15] M. Elad, "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing", Springer 2010.

[16] M. A. T. Figueiredo, R. D. Nowak, S. J. Wright, "Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems," IEEE Journal of Selected Topics in Signal Processing, vol.1, no.4, pp.586-597, Dec. 2007.

[17] J.L.V.M. Stanislaus, T. Mohsenin, "High performance compressive sensing reconstruction hardware with QRD process," IEEE International Symposium on Circuits and Systems (ISCAS), pp.29-32, 20-23 May 2012.

[18] K.R. Liu, S.-F. Hsieh, K. Yao, "Systolic block Householder transformation for RLS algorithm with two-level pipelined implementation," IEEE Transactions on Signal Processing, vol.40, no. 4, pp. 946-958, Apr 1992

[19] L. Giraud, J. Langou, M. Rozloznik, "The loss of orthogonality in the Gram-Schmidt orthogonalization process," Numerical Methods and Computational Mechanics, vol. 50, no 7, October 2005, pp 1069–1075.

[20] F. Ling, D. Manolakis, and J. G. Proakis, "A flexible, numerically robust array processing algorithm and its relationship to the Givens transformation." in Proc. IEEE ICASSP'86, Tokyo, Japan, paper 1439-1446, Oct. 1978.

[21] A. Majumdar, N. Krishnan, S.R.B. Pillai, R. Velmurugan, "Extensions to Orthogonal Matching Pursuit for Compressed Sensing," 2011 National Conference on Communications (NCC), pp.1-5, 28-30 Jan. 2011.