QUASI-SPARitest SOLUTIONS FOR QUANTIZED COMPRESSED SENSING
BY GRADUATED-NON-CONVEXITY BASED REWEIGHTED $\ell_1$ MINIMIZATION

Ines Elleuch*, Fatma Abdelkefi*, Mohamed Siala*, Ridha Hamila†, Naofal Al-Dhahir‡

* MEDIATRON Laboratory, SUP’COM, University of Carthage, Tunisia
† Electrical Engineering Department, Qatar University, Qatar
‡ Electrical Engineering Department, University of Texas at Dallas, USA

ABSTRACT

In this paper, we address the problem of sparse signal recovery from scalar quantized compressed sensing measurements, via optimization. To compensate for compression losses due to dimensionality reduction and quantization, we consider a cost function that is more sparsity-inducing than the commonly used $\ell_1$-norm. Besides, we enforce a quantization consistency constraint that naturally handles the saturation issue. We investigate the potential of the recent Graduated-Non-Convexity based reweighted $\ell_1$-norm minimization for sparse recovery over polyhedral sets. We demonstrate by simulations, the robustness of the proposed approach towards saturation and its significant performance gain, in terms of reconstruction accuracy and support recovery capability.

Index Terms— Quantized Compressed Sensing, Concave Approximation, Graduated-Non-Convexity, Reweighted $\ell_1$, Support Recovery

1. INTRODUCTION AND RELATED WORKS

Recovery of the sparsest signal representation from an underdetermined system of linear equations, which arises in Compressed Sensing (CS) theory, is a well-studied problem for which convex relaxation by $\ell_1$-norm minimization provides (1) an exact solution via Basis Pursuit (BP), provided the sparsest solution is unique, and (2) a stable solution via Basis Pursuit De-Noising (BPDN), whenever the measurements are corrupted by noise of bounded energy [1].

When addressing the practical situation of quantized measurements, the saturation phenomenon introduces large and potentially unbounded errors, and the stability property of BPDN does no longer hold. Besides, the convexification approach via the $\ell_2$-norm leads to a significantly bad approximation of the $\ell_0$ counting “norm”. However, in the literature on Quantized Compressed Sensing (QCS), optimization based reconstruction is commonly addressed from an $\ell_1$-minimization perspective [2]-[7].

Fortunately, the literature on CS provides more sparsity-inducing alternatives to the $\ell_1$-norm cost function, enjoying nice properties (differentiability, concavity, etc) and amenable to use tractable reconstruction algorithms. For instance, the Maximally Sparse Convex (MSC) approach in [8], deals with sparsity-penalized least squares using a non-convex sparsity-inducing regularizer that ensures the convexity of the total cost function. The Iterative-Reweighted-Least-Squares method (IRLS) presented in [9] is nothing but a Majorization-Minimization (MM) scheme that employs a logarithmic approximation of the “$\ell_0$-norm” upper-bounded by a quadratic surrogate function. The reweighted-$\ell_1$ minimization of [10], is also a MM procedure using an $\epsilon$-regularized logarithmic approximation and a linear surrogate function. In [11], the author proposes an exponential concave approximation that more closely approximate the $\ell_0$-norm and uses a linear surrogate function, yielding another reweighted-$\ell_1$ scheme.

Recently, a Graduated-Non-Convexity (GNC) approach [12], also known as continuation approach, which is an iterative optimization scheme involving a family of approximation functions with gradual approximation accuracy / smoothness trade-off, has been proposed to solve the sparse recovery problem [13, 14, 15]. The GNC methodology has the merit of converging to a local solution of the $\ell_0$-norm itself. In [14], a class of differentiable approximation functions, amenable to use a gradient projection procedure at each iteration, is proposed. More recently, the concave approximation of [11] has been integrated into a GNC-based procedure in [13, 15], to solve the sparse recovery problem under a linear constraint and over a polyhedral set, respectively.

In this paper, we address the problem of sparse recovery from scalar quantized and possibly saturated CS measurements, under a Quantization Consistency (QC) constraint, by considering the GNC-based reweighted $\ell_1$ minimization approach of [13, 15]. The contribution of this paper is threefold. First, the proposed method is essentially a reweighted $\ell_1$-minimization technique, thus computationally attractive, and more sparsity-inducing than the commonly used $\ell_1$-minimization in QCS. Second, the QC constraint yields a built-in capability to handle the saturation phenomenon, in the same way of [3, 7]. Third, as the proposed algorithm is

This paper was made possible by grant number NPRP 06-070-2-024 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.
essentially minimizing the $\ell_0$-norm, even though the sparsest solution is not unique, we study the uniqueness of its support, via simulations. The reason behind this motivation is that the feasible solution space defined by the QC constraint is not an arbitrary polyhedron. It should be noted that the problem of support recovery from QCS measurements has been mostly addressed in the case of one-bit CS [16, 17], under additional constraints on the design of the measurement matrix. In [16], the case where the measurements are corrupted by Gaussian noise and outliers is also considered. For the case of multi-bit QCS, the authors in [18] derived lower bounds on the required number of measurements for exact support recovery from scalar quantized and noisy measurements using a Maximum-Likelihood decoder. In [19], support recovery for $\Sigma\Delta$-QCS is addressed using an $\ell_1$ minimization approach.

The paper is organized as follows. Section II presents the observation model and the optimization problem to be considered throughout the paper. Section III describes the algorithmic framework to solve the proposed optimization problem. Section IV provides simulations results that demonstrate the significant performance gain of the proposed approach, both in terms of quality of the estimation and support recovery capabilities. Finally, Section V concludes the paper.

2. OBSERVATION MODEL AND PROBLEM FORMULATION

2.1. QCS Observation Model

In this paper, we assume that CS measurements are scalar quantized, yielding the observation model

$$z = \Phi x, \quad y = Q_b(z), \quad (1)$$

where $x \in \mathbb{R}^N$ is a $K$-sparse signal, $\Phi \in \mathbb{R}^{M \times N}$ is the measurement matrix with $M < N$, and $Q_b(\cdot)$ is the quantization function that maps real-valued vectors, element-wise, to a set of $2^b$ output levels $\{q_i, i = 1, \ldots, 2^b\}$, according to a set of thresholds $\{\tau_i, i = 1, \ldots, 2^{b-1}\}$, with $b$ denoting the quantizer precision in terms of bit-depth. Formally, we have

$$y_m = \begin{cases} q_1 & \text{if } z_m < \tau_1, \\ q_i & \text{if } z_m \in [\tau_{i-1}, \tau_i), \quad i = 2, \ldots, 2^{b-1}, \\ q_{2^b} & \text{if } z_m \geq \tau_{2^{b-1}}. \end{cases} \quad (2)$$

We will consider a uniform midrise quantizer i.e. the thresholds are equally spaced with step size $\delta$, and the levels are centered between the thresholds. The dynamic range of the quantizer is $[-g, g]$ where $g \triangleq 2^{b-1}\delta$ is the saturation level. Measurements of magnitude larger than $g - \delta$ saturate at $\pm (g - \frac{\delta}{2})$.

2.2. Concave Approximation of the $\ell_0$-norm

The continuous $\ell_0$ approximation function, adopted in this paper, was originally proposed in [11]. By introducing a non-negative parameter $\sigma$, we define:

$$F_{\sigma}(x) = \sum_{i=1}^{N} f_{\sigma}(x_i) = \sum_{i=1}^{N} (1 - \frac{1}{e^{\frac{-|x_i|}{\sigma}}}) \quad (3)$$

The above approximation enjoys the following properties:

- $F_{\sigma}(x) \leq ||x||_0$,
- $\lim_{\sigma \to 0^+} F_{\sigma}(x) = ||x||_0$,
- $F_{\sigma}(x) = F_{\sigma}(-x)$,
- $F_{\sigma}$ is non-decreasing and concave in the positive orthant.

As shown in Fig 1, this function better approximates the $\ell_0$-norm and is likely to be more sparsity-inducing than the commonly used $\ell_1$-norm.

2.3. Optimization Problems Formulation

We propose to find a sparsest consistent solution by considering the QC-constrained optimization problem:

$$\min_{x \in \mathbb{R}^N} ||x||_0 \quad s.t. \quad Q_b(\Phi x) = y, \quad (4)$$

denoted $\ell_0$-QC. As we are dealing with a scalar quantization scheme, the QC constraint is obviously a linear inequality constraint. To be precise, let $\Phi_+$ and $\Phi_-$ denote row submatrices of $\Phi$, corresponding to the positively saturated and the negatively saturated measurements, respectively and let $\Phi_0$ be its row submatrix corresponding to the remaining measurements. Then, it can be easily verified that the QC constraint is equivalent to the following linear inequality:

$$Q_b(\Phi x) = y \quad \Leftrightarrow \quad \overline{\Phi}_0 x \leq \overline{y}, \quad (5)$$

where $\overline{\Phi}_0 = [\Phi_0^T, -\Phi_0^T, -\Phi_0^T]^T \in \mathbb{R}^{(2M-S)\times N}$ with $S$ being the number of saturated measurements, and $\overline{y}$ is the vector of thresholds with respect to the measurements. Then, the $\ell_0$-QC problem in (4) could be written as:

$$\min_{x \in \mathbb{R}^N} ||x||_0 \quad s.t. \quad \overline{\Phi}_0 x \leq \overline{y}, \quad (6)$$

where the problem at hand is to recover a sparsest solution over a polyhedral set.

By considering the approximation function $F_{\sigma}$, the $\ell_0$-QC problem could be approximated by:

$$\min_{x \in \mathbb{R}^N} F_{\sigma}(x) \quad s.t. \quad \overline{\Phi}_0 x \leq \overline{y}, \quad (7)$$

Fig. 1: $\ell_0$ approximations. For $f_{\sigma}(\cdot), \sigma = 0.1, 0.2, 0.5, 1.$
3. GRADUATED-NON-CONVEXITY BASED REWEIGHTED $\ell_1$ MINIMIZATION FOR QCS

3.1. QCS by Reweighted $\ell_1$ Minimization

In order to take advantage of the concavity of $F_\sigma$ in the positive orthant, we introduce the variable conversion suggested in [15], $u = [x^+, x^-] \in \mathbb{R}^N_+$, where $x^+ = \max(x, 0)$ and $x^- = -\min(x, 0)$. Then, it can be verified that the approximate problem (7) can be equivalently formulated as:

$$
\min_{u \in \mathbb{R}^N} F_\sigma(u) \quad \text{s.t.} \quad u \in \mathcal{P},
$$

where the objective function is concave and bounded from below over the feasible solution set

$$
\mathcal{P} = \{u \in \mathbb{R}^N, \quad [\overline{\Phi}, -\overline{\Phi}]u \leq \mathbf{y}, \quad u \succeq 0\}.
$$

A common way to solve a concave optimization problem, is to apply a Majorize-Minimize (MM) iterative procedure where the surrogate function is constructed from the linearization of the objective function around the current iterate. More precisely, to solve (8), we proceed iteratively by majorizing the objective function around the current iterate. More precisely, to solve (8), we proceed iteratively by majorizing $F_\sigma(u)$ by $\langle \nabla F_\sigma(u^k), u - u^k \rangle$, based on its concavity. Hence solving (8) amounts to iteratively solve the linear program:

$$
u^{k+1} = \arg\min_{\mathbf{u} \in \mathbb{R}^N} \langle \nabla F_\sigma(u^k), \mathbf{u} \rangle \quad \text{s.t.} \quad \mathbf{u} \in \mathcal{P}.
$$

Problem (9) could be cast into the following reweighted $\ell_1$-norm minimization:

$$
x^{k+1} = \arg\min_{x \in \mathbb{R}^N} \| Wx \|_1 \quad \text{s.t.} \quad \overline{\Phi} \mathbf{x} \leq \mathbf{y},
$$

where $W = \text{diag}(\nabla F_\sigma(|x|^k))$.

3.2. QCS by GNC-based Optimization

Solving the $\ell_0$-QC problem by considering a single approximation function $F_\sigma$ is not very judicious. Indeed, the choice of $\sigma$ would either favor approximation accuracy at the price of extra local minima or vice versa.

3.2.1. GNC General Framework

Graduated-Non-Convexity is a deterministic global optimization methodology that addresses non-convex cost functions by considering a family of approximation functions scaled by a parameter $\sigma$ that monotonically controls the tradeoff between approximation accuracy and smoothness. When $\sigma$ grows towards $\infty$, the family should embed a convex approximation function. Then, as $\sigma$ decreases, it leads continuously to the original cost function with an increasing “non-convexity” rate i.e. an increasing number of local minima. The motivation behind such a construction is essentially to avoid these annoying local minima by: (1) gradually decreasing the scale parameter so that the global minimizers of two consecutive problems are sufficiently close, more precisely, the former should fall within the locally convex region around the latter, the so-called basin of attraction, and (2) each approximation problem is initialized with the minimizer of the previous one, so that a descent method would converge to one of its global minimizer. Clearly, the initialization issue and the decreasing rule for $\sigma$ are crucial for the effectiveness of such an optimization methodology.

3.2.2. GNC-based scheme for $\ell_0$-QC

$F_\sigma$ minimization is equivalent to the convex $\ell_1$-minimization for $\sigma$ sufficiently large and to the $\ell_0$-minimization for $\sigma$ sufficiently small [13]. Then, $\ell_0$-QC could be solved using a GNC-based procedure involving the family of approximation functions $F_\sigma$. Moreover, a solution of $\ell_1$-QC is potentially a good starting point. The decreasing rule for $\sigma$ is inspired from the exposition suggested in [13], with respect to the behaviour of the function $F_\sigma$. To be more precise,

$$
f_\sigma(x) \simeq \begin{cases} 
1, & \text{if } x \gg \sigma, \\
0, & \text{if } x \ll \sigma.
\end{cases}
$$

Let $x^j$ be the current iterate, i.e a solution of $F_{\sigma^{j-1}}$ minimization, and let $I = \{i \mid |x_i^j| < \sigma^{j-1}\}$. It is reasonable to take $I$ as an estimation of the zero components of $x^j$. To ensure a gradual decrease of $\sigma$, yielding a small number of GNC iterations, we propose the following adaptive update rule:

$$
\sigma^j = \text{mean}(x_I) + \text{std}(x_I),
$$

where std denotes the standard deviation, so as to ask whether the most significant components of $x_I$ are really zero components. The QCS reconstruction scheme, involving a GNC-based outer loop and a reweighted $\ell_1$ minimization inner loop, is summarized in Algorithm 1.

---

**Algorithm 1 GNC-based Reweighted $\ell_1$ for QCS**

**Require:** $\Phi, \mathbf{y}, \epsilon_1, \epsilon_2, \sigma_0, \sigma_{\text{min}}$.

1: compute $\overline{\Phi}$ according to (5).
2: $x^0 = \arg\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{s.t.} \quad \overline{\Phi}x \leq \mathbf{y}$.
3: $j = 0$.
4: while $d_1 > \epsilon_1$ and $\sigma_1 > \sigma_{\text{min}}$ do $\triangleright$ GNC outer loop

5: \hspace{1em} $j = j + 1, k = 0, \overline{x}^0 = x^j$.
6: \hspace{1em} while $d_2 > \epsilon_2$ do $\triangleright$ reweighted $\ell_1$ inner loop

7: \hspace{2em} $k = k + 1$.
8: \hspace{2em} $W = \text{diag}(\nabla F_{\sigma^{j-1}}(|\overline{x}^{k-1}|))$.
9: \hspace{2em} $\overline{x}^k = \arg\min_{x \in \mathbb{R}^N} \| Wx \|_1 \quad \text{s.t.} \quad \overline{\Phi}x \leq \mathbf{y}$.
10: \hspace{2em} $d_2 = \| \overline{x}^k - \overline{x}^{k-1} \| / \| \overline{x}^{k-1} \|$. 
11: \hspace{1em} end while

12: $x^j = \overline{x}^k, d_1 = \| x^j - x^{j-1} \| / \| x^{j-1} \|$.
13: $I = \{i \mid |x_i^j| < \sigma^{j-1}\}, \sigma^j = \text{mean}(x_I) + \text{std}(x_I)$.
14: end while
**RSNR (dB)**

**False Non Zeros**

**Successful Support Recovery Probability**

<table>
<thead>
<tr>
<th>Sparsity Level of the Reconstruction (%)</th>
<th>10</th>
<th>15</th>
<th>30</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>20</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>30</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>40</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>50</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>70</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

- For each realization of the $K$-sparse signal $x$, the support is selected uniformly at random from all possible supports of cardinality $K$, the non-zero entries are drawn independently from $\mathcal{N}(0, 1)$, and $x$ is scaled to unit $\ell_2$-norm.
- For the sake of comparison, we consider the $\epsilon$-support of the reconstructed signal $\hat{x}$, defined as $\epsilon$-supp$(\hat{x}) = \{i \mid |x_i| > \epsilon \}$ where $\epsilon$ is a small value used to prune out negligible non-zeros, and we consider the following performance metrics:
  - The Reconstruction Signal-to-Noise-Ratio $\text{RSNR} = -20 \log_{10}(\|x - \hat{x}\|_2)$, where $\hat{x}$ is the reconstructed signal;
  - The number of False Non-Zeros (FNZ) and the number of False Zeros (FZ);
  - The probability of successful support recovery i.e. the probability of the event $\text{SE} = 0$, where $\text{SE} \triangleq \text{FZ} + \text{FNZ}$ stands for Support Error.

For all experiments, we set $\epsilon = 10^{-7}$, $\epsilon_1 = 10^{-4}$, $\epsilon_2 = 10^{-3}$, $\sigma^b = 2 \max(|x|)$ and $\sigma_{min} = 10^{-7}$. We use an oracle bound for the $\ell_2$-norm constraint within the $\ell_1-\ell_2$/SC method, namely $\|y_T - z_T\|_2$, where $T$ is the support of non-saturated measurements i.e. $T = \{i \mid |y_i| < g - \frac{\delta}{2}\}$.

In the first experiment, we set $K = 10$, $M = 250$, $N = 500$, $b = 4$ and we tune the saturation rate $g$ over the range $[0, 0.4]$. For the proposed $\ell_0$-QC approach, the maximal inner iterations number is set to 10, in order to tradeoff computational complexity at sub-optimal saturation rates. Figure 2 depicts the RSNR performance and the saturation rate, indicated on the left vertical axis and the right vertical axis, respectively. Results are averaged over 100 trials. As shown, all the methods reach their optimal performances at non-zero saturation rates. The proposed approach provides a substantial reconstruction performance gain (15 dB) with a higher saturation rate at its optimal operating point.

In the second experiment, we examine the impact of the bit-depth on the reconstruction performance. Figure 3 reports the average performance over 500 trials, at the optimal saturation rate for each method, in terms of (a) RSNR and (b) the number of FNZ. For the proposed $\ell_0$-QC method, we also depict (c) the successful support recovery rate and (d) the histograms of support sizes (sparsity levels) of the recovered signals. As expected, the proposed approach shows an interesting support recovery capability. Indeed, the search space included by the QC constraint is a $N$-dimensional polyhedron included within the intersection of many slabs (a slab is the region between two parallel hyperplanes) defined by the unsaturated measurements. For a given bit-depth, a high saturation rate induces a small step size $\delta$, and consequently thinner slabs are involved in the definition of the polyhedral solution set. Thus, even though the polyhedron is unbounded, it is likely to intersect a single $K$-dimensional subspace, given a small $\delta$. Of course, the step size $\delta$ could not be arbitrary small to ensure a sufficient number of unsaturated measurements and equivalently a sufficient number of slabs. As illus-

---

**Fig. 2**: Average RSNR versus the saturation level $g$, with $K = 10$, $M = 250$, $N = 500$ and $b = 4$.

**Fig. 3**: Reconstruction performance versus the bit-depth $b$, with $K = 10$, $M = 250$ and $N = 500$.

### 4. SIMULATION RESULTS

In this section, we demonstrate the performance gain of the proposed $\ell_0$-QC recovery approach over two benchmark $\ell_1$ recovery approaches that handle saturation errors. More precisely, we consider the $\ell_1$-QC method [3] and the Satura-

![Diagram](image-url)
trated in Figure 3d, the proposed approach provide a quasi-
sparest consistent solution of support size $\ell_0$ (x is a solution
of $\ell_0$-QC) with an increasing probability with the bit-
precision. Support sizes slightly greater than $K$ (11, 12) re-
fect local minima caused by sub-optimal $\sigma$-initialization and
update rule, and stopping rules. Support sizes less than $K$
(7 – 9) are related to very small non-zero components
in $x$, with comparison to $\delta$. It is also to be mentioned that
the proposed method requires around 3.2 GNC iterations on
average, for all bit-depth settings.

5. CONCLUSION

We presented an efficient method to recover a sparse signal
from scalar QCS measurements, by considering the set of
sparest consistent solutions. A quasi-sparest solution was
estimated using an iterative GNC-based reweighted $\ell_1$ mini-
mization approach that consider a concave approximation of
the $\ell_0$-norm, with an increasing accuracy. We demonstrated the per-
formance gain of the proposed approach, with comparison
to two benchmark $\ell_1$ minimization based QCS sparse re-
covety methods, both in terms of reconstruction accuracy and
probability of successful support recovery, via simulations.

REFERENCES


ing with quantized measurements,” IEEE Sig. Proces.

algorithmic approaches to quantized compressive sens-
1866, 2011.

ing compressed sensing: When oversampling and non-
gaussian constraints combine,” IEEE Trans. Inf. Theory,


R. G. Baraniuk, “Democracy in action: Quantization,

[7] A. Moshtaghpour, L. Jacques, V. Cambareri, K. De-
graux, and C. D. Vleeschouwer, “Consistent basis pur-
suit for signal and matrix estimates in quantized com-
pressed sensing,” IEEE Sig. Process. Letters, vol. 23,

tion by maximally sparse convex optimization,” IEEE
Trans. Sig. Process., vol. 62, no. 5, pp. 1078–1092,
2014.

methodology for best basis selection,” IEEE Trans. Sig-

ing sparsity by reweighted $\ell_1$ minimization,” Journal of
Fourier analysis and applications, vol. 14, no. 5-6, pp.

cave minimization,” in Applied Mathematics and Paral-


sparse optimization over polyhedral sets,” in Model.
p. 369–379.

approach for overcomplete sparse decomposition based
on smoothed norm,” IEEE Trans. Signal Processing,

[15] M. Malek-Mohammadi, A. Koochakzadeh, M. Babaie-
Zadeh, M. Jansson, and C. R. Rojas, “Successive con-
cave sparsity approximation: Near-oracle performance
in a wide range of sparsity levels,” arXiv preprint arXiv:
1505.06841, 2015.

[16] J. Haupt and R. Baraniuk, “Robust support recovery us-
ing sparse compressive sensing matrices,” in 45th An-

bit compressed sensing: Provable support and vector
recovery,” in 30th Inter. Conf. Mach. Learn. (ICML),
2013.

bounds for sparsity pattern recovery with quantized no-
sy random projections,” IEEE Journal of Selected
Topics in Signal Processing, vol. 6, no. 1, pp. 43–57,
2012.

and ¨O. Yılmaz, “Sobolev duals for random frames
and $\Sigma\Delta$ quantization of compressed sensing mea-
surements,” Foundations of Computational mathematics,

[20] M. Grant and S. Boyd, “CVX: Matlab software for dis-
ciplined convex programming, version 2.0,” available