ON QUANTIZED COMPRESSED SENSING WITH SATURATED MEASUREMENTS VIA CONVEX OPTIMIZATION

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 multi-bit scalar quantized compressed sensing measurements, where the saturation issue is taken into account. We propose a convex optimization approach, where saturation errors are jointly estimated with the sparse signal to be recovered. In the proposed approach, saturated measurements, even though over-identified, are considered as outliers and the associated errors are handled as non-negative *sparse corruptions* with partial support information. We highlight the theoretical recovery guarantee of the proposed approach and we demonstrate, via simulation results, its reliability in cancelling out the effect of the outlying saturated measurements.

Index Terms— Multi-Bit Quantized Compressed Sensing, Saturation, Sparse Corruptions, Sign Constraint, Convex Optimization

1. INTRODUCTION

Classical Compresses Sensing (CS) enables high-dimensional K-sparse signals $\mathbf{x} \in \mathbb{R}^N$ to be recovered from significantly fewer real-valued, possibly noisy, linear measurements $\mathbf{y} \in \mathbb{R}^M$. In this context, convex optimization using ℓ_1 minimization provides an appealing theoretical framework where the sparsest signal is exactly recovered in the noiseless case and stably recovered in case of additive noise with bounded energy, using the Basis Pursuit (BP) and the Basis Pursuit Denoising (BPDN) decoders [1], respectively. In practice, CS measurements are unavoidably distorted during the quantization step involved in the acquisition process. Using BPDN for signal recovery would amount to consider the unrealistic assumptions of infinite range (unsaturated) quantizer and Gaussian quantization noise.

Recently, convex optimization for Quantized Compressed Sensing (QCS) has been investigated by considering different data-fidelity constraints yielding to different adaptations of BP. By ignoring the saturation phenomena, the quantization noise is rather uniformly distributed and as suggested in [2] and studied in [3], the ℓ_2 norm fidelity in BPDN is replaced by an ℓ_∞-norm of the residual error, yielding the Quantization Consistency (QC) constraint. In [4] the authors proposed the Basis Pursuit De-Quantizer of moment p (BPDQ$_p$), the quantization noise is assumed to be of bounded ℓ_p norm for $2 \leq p \leq \infty$. The optimal moment p is finite and increases with the oversampling ratio m/K, so that BPDQ$_p$ data-fidelity gets closer to the QC constraint. In [5], the authors incorporate an additional gaussian noise and propose an ℓ_1-regularized maximum likelihood decoder that outperforms the LASSO decoder (the equivalent unconstrained formulation of BPDN) for coarsely quantized (unsaturated) measurements.

In practice, due to the quantizer finite dynamic range, a fraction of S measurements may saturate, which leads to large and potentially unbounded errors, and implies a substantial departure from the above assumptions. The QC approach of [3], accounts for both quantization and saturation consistency. From another perspective, saturated measurements can be seen as outliers that would significantly deteriorate the recovery performance of reconstruction approaches based on residual norm minimization. In this scope, the authors in [6], suggest to discard potentially saturated measurements and to use BPDN on the remaining measurements. To prevent wasting valuable measurements, the saturation consistency method of [6] decouples the measurements and integrates a saturation consistency constraint. These two approaches lead to a *swamping* problem consisting of classifying non-outlying artificially saturated measurements as outliers. These measurements do not contribute in the denoising process, they are either accidentally discarded or only involved in a consistency constraint.

In this paper, we provide a robust reconstruction method that blindly reaches the oracle-assisted rejection approach performance, where only effectively saturated measurements (true outliers) are removed. We propose to jointly estimate the saturation noise and to remove it from the measurements for a *clean* signal recovery. When the sparsity level K 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.

perfectly known, we have integrated the joint estimation approach within a greedy reconstruction procedure in [7], without theoretically investigating the recovery guarantee. In this paper, we study the convex optimization approach, suitable for the practical case of unknown K, and we provide theoretical recovery guarantees.

It should be noted that the idea of joint estimation using an optimization based approach, has been investigated for QCS in different settings. In [8], the authors proposed to jointly estimate unquantized noisy measurements under a QC constraint, using Bayesian inference. In the context of one-bit CS, the authors in [9] and [10], consider joint estimation of the sparse sign-flip positions caused by measurement noise, in non-Bayesian and Bayesian frameworks, respectively. Differently from [8]-[10], our joint estimation approach considers S extra unknowns instead of M.

The rest of the paper is organized as follows. Section II presents the observation model. Section III links the rejection approach to the optimal reconstruction strategy when the outlying saturated measurements are correctly identified. Section IV proposes the convex optimization-based recovery approach and provides its theoretical recovery guarantee. Section V demonstrates via simulation results, the performance gain of the proposed method, mainly for coarse quantization. Finally, Section VI concludes the paper.

In the sequel, bold-face lowercase and capital letters represent vectors and matrices, respectively. For an index set T, T^c denotes its complement. Notations $\mathbf{x}_T$ and $\mathbf{\Phi}_T$ stand for the sub-vector of elements of $\mathbf{x}$, and the sub-matrix of columns of $\mathbf{\Phi}$, indexed by T, respectively. The best K-term approximation of a $\mathbf{x}$, obtained by keeping the K components of largest magnitude and zeroing the others, is denoted $\mathbf{x}^{(K)}$. The support of $\mathbf{x}$, denoted $\mathrm{supp}(\mathbf{x})$, corresponds to the ordered set of indices of its nonzero entries. The ℓ_1-norm of $\mathbf{x} \in \mathbb{R}^N$, is defined as $\|\mathbf{x}\|_1 \triangleq \sum_{n=1}^{N} |x_n|$. The sign operator sign and the inequality symbol $\succeq$, are applied on vectors, component-wise.

2. QUANTIZED COMPRESSED SENSING MODEL

We consider that the acquisition of CS measurements involves a b-bit uniform midrise quantizer operator $\mathcal{Q}_b$ with quantization interval δ, 2^b quantization levels and a saturation level $g = 2^{b-1}\delta$, such that the QCS model can be written by considering a two-component noise

$$\mathbf{y} = \mathcal{Q}_b(\mathbf{z}) = \mathcal{Q}_b(\mathbf{\Phi x}) = \mathbf{\Phi x} + \mathbf{e} + \mathbf{n}, \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ with $M < N$ is the measurement matrix and $\mathbf{e}$ and $\mathbf{n} \in \mathbb{R}^M$ account for the saturation errors and the quantization errors, respectively. At a given saturation rate, $\mathbf{e}$ would only corrupt a fraction E of the unquantized measurements $\mathbf{z}$. and we have $e_m = 0$ or $|e_m| > \frac{\delta}{2}$. The only information available on the support $\mathcal{E}$ of $\mathbf{e}$, is that it is included within the known support of potentially saturated measurements defined as $\mathcal{S} \triangleq \{m \in \{1, \ldots, M\} : |y_m| = g - \frac{\delta}{2}\}$. Moreover, the vector $\mathbf{e}$ satisfies a sign property of the form $\mathrm{sign}(\mathbf{e}_{\mathcal{E}}) = -\mathrm{sign}(\mathbf{y}_{\mathcal{E}})$.

By exploiting partial support information on the corruption term $\mathbf{e}$, the model in (1) can be adjusted as follows

$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{\Theta s} + \mathbf{n} \quad \text{and} \quad \mathbf{s}_{\mathrm{supp}(\mathbf{s})} \succeq \tfrac{\delta}{2}, \tag{2}$$

where $\mathbf{\Theta} = \mathbf{I}_{\mathcal{S}}\mathbf{\Lambda} \in \mathbb{R}^{M \times S}$, $\mathbf{s} = \mathbf{\Lambda e}_{\mathcal{S}} \in \mathbb{R}^S$, $\mathbf{\Lambda} \in \mathbb{R}^{S \times S}$ is a diagonal matrix whose diagonal elements are given by $-\mathrm{sign}(\mathbf{y}_{\mathcal{S}})$.

3. RECOVERY WITH PERFECT IDENTIFICATION OF SATURATED MEASUREMENTS

Imposing saturation consistency helps improving the performance over the rejection approach [6]. However, it is not clear whether it reaches the performance of the oracle assisted rejection scheme where only effectively saturated measurements $\mathbf{y}_{\mathcal{E}}$ are discarded. Indeed, it has been shown in [11], that the optimal recovery strategy to reconstruct a sparse signal $\mathbf{x}$ from its sparsely corrupted measurements $\mathbf{y} = \mathbf{\Phi x} + \mathbf{\Theta s}$, with general $\mathbf{\Theta}$ and known corruption term support Σ, is the *cancel-then-recover* approach, initially proposed in [12]. This approach begins by projecting the measurements onto the orthogonal complement of the subspace spanned by the corruption term, using matrix $\mathbf{P}_\Sigma = \mathbf{I} - \mathbf{\Theta}_\Sigma \mathbf{\Theta}_\Sigma^\dagger$, where $(\cdot)^\dagger$ denotes the Moore–Penrose pseudoinverse operator. As $\mathbf{P}_\Sigma \mathbf{\Theta}_\Sigma = \mathbf{0}$, the corruption term is canceled out from the measurements, and standard recovery methods could be applied to recover the sparse signal $\mathbf{x}$ from the clean measurements $\mathbf{P}_\Sigma \mathbf{y}$. In our specific setting, the projection matrix, involved in the cancelation approach, acts simply by zeroing entries of $\mathbf{y}$ and rows of $\mathbf{\Phi}$ indexed by $\mathcal{E}$, which turns to reject effectively saturated measurements. Hence, the rejection approach with correct identification and removal of effectively saturated measurements is equivalent to the *cancel-then-recover* approach of [11].

4. PROPOSED CONVEX OPTIMIZATION BASED RECOVERY FOR QCS

The model in (2) incorporates a non-convex constraint on the saturation errors. We propose to perform robust sparse recovery jointly with saturation error estimation using a convex optimization approach, where the sign property is exploited to convexify the saturation error constraint. Formally, we solve the following optimization problem

$$\{\widehat{\mathbf{x}}, \widehat{\mathbf{s}}\} \in \underset{\substack{\tilde{\mathbf{x}} \in \mathbb{R}^N, \\ \tilde{\mathbf{s}} \in \mathbb{R}^S.}}{\operatorname{argmin}} \|\tilde{\mathbf{x}}\|_1 \quad s.t. \quad \begin{cases} \|\mathbf{y} - \mathbf{\Theta}\tilde{\mathbf{s}} - \mathbf{\Phi}\tilde{\mathbf{x}}\|_2 \le \epsilon & \text{(3a)} \\ \tilde{\mathbf{s}} \succeq 0 & \text{(3b)} \end{cases}$$

4.1. Key Underlying Virtues

The key advantage of the proposed problem formulation is that all M *cleaned* measurements $\mathbf{y} - \mathbf{\Theta}\tilde{\mathbf{s}}$ contribute to the denoising constraint (3a). Moreover, $\mathbf{s}$ could be seen as a non-negative E-sparse vector with a potentially high fraction of sparsity $\frac{E}{S}$. By considering the equivalent bi-objective problem formulation of (3) with respect to the residual norm constraint (3a), and by fixing $\tilde{\mathbf{x}}$, the solution $\widehat{\mathbf{s}}$ could be interpreted as a Non Negative ℓ_1-regularised Least Squares (LS) solution from pseudo-observation $\mathbf{y} - \mathbf{\Phi}\widehat{\mathbf{x}}$. Fortunately, it has been shown that Non Negative Least Squares (NNLS) is sparsity promoting, albeit no explicit ℓ_1-norm minimization is used, especially for high levels of sparsity [13]. Hence, artificially saturated measurements would be detected thanks to the built-in $\tilde{\mathbf{s}}$ sparsity promoting property of (3).

4.2. Theoretical Recovery Guarantee

In order to study the stability of the solution of (3), we propose the following reformulation of the problem. Let $\mathbf{D} = [\mathbf{\Phi} \ \ \mathbf{\Theta}] \in \mathbb{R}^{M\times(N+S)}$, $\mathbf{w} = \begin{bmatrix}\mathbf{x}\\ \mathbf{s}\end{bmatrix} \in \mathbb{R}^{(N+S)}$ and $T = [N+1, \ldots, N+S]$. Then, our observation model of (2) could be recast to the Justice Pursuit (JP) model [14] that leverages the sparsity of the signal and the corruptions, as follows

$$\mathbf{y} = \mathbf{D}\mathbf{w} + \mathbf{n} = [\mathbf{\Phi} \ \mathbf{\Theta}] \begin{bmatrix}\mathbf{x}\\ \mathbf{s}\end{bmatrix} + \mathbf{n}, \tag{4}$$

where $\mathbf{w}$ is $(K + E)$-sparse. Consequently, the proposed optimization problem (3) could be reformulated as

$$\min_{\tilde{\mathbf{w}}\in\mathbb{R}^{(N+S)}} \|\tilde{\mathbf{w}}_{T^c}\|_1 \quad s.t. \quad \begin{cases} \|\mathbf{D}\tilde{\mathbf{w}} - \mathbf{y}\|_2 \leq \epsilon & \text{(5a)} \\ \tilde{\mathbf{w}}_T \succeq 0 & \text{(5b)} \end{cases}$$

where the support of its solution contains the smallest number of new additions to its positive part on T. Surprisingly, the subproblem (5a) is simply the *innovative* BPDN (*i*BPDN) problem studied in [15], where PKS information on sparse signals (here $\mathbf{w}$) has been incorporated into BPDN.

The JP model and the *i*BPDN formulation provide the key foundation to prove recovery guarantee for our proposed method. Firstly, if $\mathbf{\Phi}$ entries are drawn according to $\mathcal{N}(0, \frac{1}{M})$ and since $\mathbf{\Theta}$ has orthonormal columns, matrix $\mathbf{D}$ is shown to satisfy the RIP condition [Theorem 1, 14]. Secondly, under mild conditions on the RIP constant of $\mathbf{D}$, and given that $||\mathbf{n}||_2 \leq \epsilon$, the iBPDN has the $\ell_2 - \ell_1$ instance optimality [Theorem 1, 15] meaning that:

$$||\widehat{\mathbf{x}} - \mathbf{x}||_2 \leq C_0 K^{-1/2} ||\mathbf{x} - \mathbf{x}^{(K)}||_1 + C_1\epsilon,$$

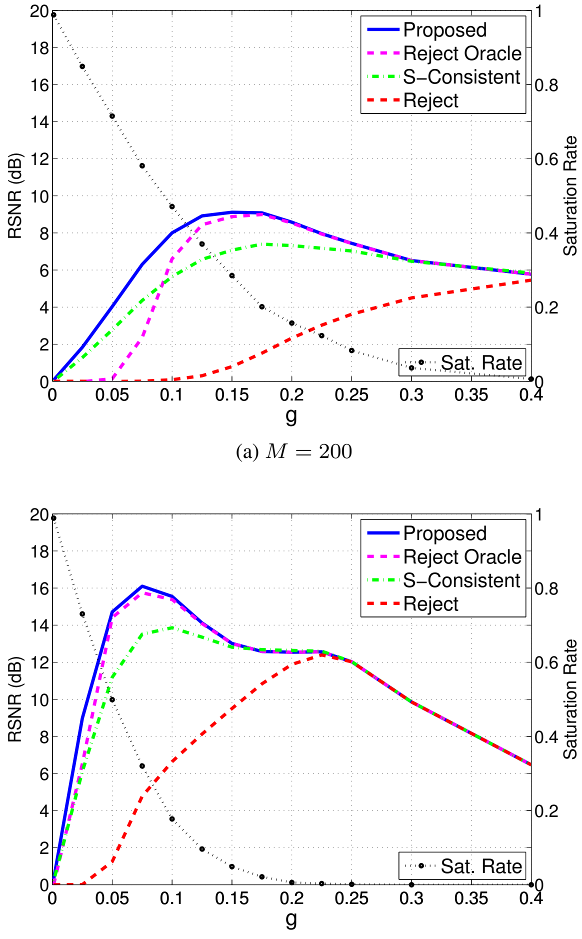
where C_0 and C_1 are small parameters. Motivated by these results, the stability of the proposed recovery approach is guaranteed, regardless of the additional convex sign constraint.

5. SIMULATION RESULTS

In this section, we demonstrate the performance gain of our proposed convex optimization based recovery method, in comparaison with the state-of-the-art saturation consistency and rejection approaches of [6], in terms of recovery efficiency and robustness to saturation. We also consider the oracle assisted rejection approach, where effectively saturated measurements are assumed to be identified by an "oracle", as the reference approach for performance comparaison. Furthermore, we consider oracle values for ϵ, $\epsilon = \|\mathbf{y}_{\mathcal{S}^c} - \mathbf{z}_{\mathcal{S}^c}\|_2$ or $\epsilon = \|\mathbf{y}_{\mathcal{E}^c} - \mathbf{z}_{\mathcal{E}^c}\|_2$ where applicable, for a fair comparaison of the intrinsic performance of each method, avoiding performance degradation due to sub-optimal tuning of parameter ϵ. The simulations were conducted using the general-purpose convex optimization package CVX [16], to implement all the aforementioned methods. The $M \times N$ measurement matrix $\mathbf{\Phi}$ is generated from an i.i.d. Gaussian distribution with mean zero and variance $1/M$. The K-sparse signal $\mathbf{x}$, with support selected uniformly at random in $\{1, \ldots, N\}$, is drawn from an i.i.d Gaussian distribution and then normalized to have unit ℓ_2-norm. For all experiments, we set $N = 1000$ and $K = 20$ and measure the reconstruction performance by the Reconstruction Signal-to-Noise-Ratio $\text{RSNR} \triangleq -20\log_{10}(\|\mathbf{x} - \widehat{\mathbf{x}}\|_2)$, where $\widehat{\mathbf{x}}$ is the reconstructed signal. Results are averaged over 100 Monte Carlo trials.

In the first experiment, we consider a coarse quantization with a bit-depth $b = 2$ and we vary the saturation level g over the range $[0, 0.4]$, under two measurement regimes $M = 200, 700$. Figure 1 depicts the RSNR of all methods. Solid, dashed and dashed dot lines, follow the scale on the left vertical axis, while dot lines are associated with the right vertical axis. The saturation rate $\frac{S}{M}$, averaged over 1000 trials, depends not only on g and δ but also on M. Indeed, column normalization (in expectation) of $\mathbf{\Phi}$ imposes an M-dependent dynamic range for the measurements. All the curves meet as the saturation rate is effectively zero. Indeed, in the unsaturated quantizer regime, each method reduces to the basic BPDN method. The proposed approach and the consistency approach achieve their optimal RSNR performances at a nonzero saturation rate, which confirms the benefit of saturated measurements in sparse recovery from quantized measurements. The optimal operating point for each approach represents the best tradeoff between the fraction of outlying measurements (sparsity of the corruption term) and the precision of the rest of measurements (quantization noise level). Only the proposed approach reaches the oracle assisted rejection approach performance at the same operating point for each parameter settings. This demonstrates the efficiency of the proposed method over the saturation consistency method, in terms of robustness against saturation.

In the second experiment, we vary the bit-depth and report the maximal RSNR performances under optimal operating conditions on g for each method, in Figure 2. The per-



(a) $M = 200$

(b) $M = 700$

Fig. 1: RSNR versus the saturation level g, with $N = 1000$, $K = 20$ and $b = 2$.

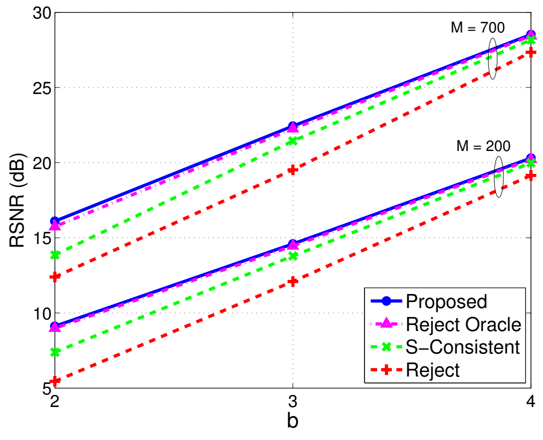


Fig. 2: RSNR versus the bit-depth b, with $N = 1000$, $K = 20$ and $M = 200, 700$.

formance gain of the proposed joint estimation approach is essentially significant under the coarse quantization regime ($b = 2$).

6. CONCLUSION

We presented a robust approach to recover sparse signals from their partially saturated QCS measurements. We capitalized on the structure of the saturation noise originated from its partial support information and its sign characterization. We formulated a convex optimization based recovery method that performs a joint saturation noise estimation and *clean* signal recovery. We provided the key foundation for the theoretical recovery guarantee of the proposed method. Simulation results confirmed its robustness against saturation compared to the saturation consistency approach, with better recovery SNR for coarse quantization, and same performances as the oracle-assisted rejection approach.

REFERENCES

[1] E. J. Candes, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[2] E. J. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] W. Dai and O. Milenkovic, “Information theoretical and algorithmic approaches to quantized compressive sensing,” *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1857–1866, 2011.

[4] L. Jacques, D. Hammond, and J. Fadili, “Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine,” *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 559–571, Jan. 2011.

[5] A. Zymnis, S. Boyd, and E. Candes, “Compressed sensing with quantized measurements,” *IEEE Signal Proces. Letters*, vol. 17, no. 2, pp. 149–152, 2010.

[6] J. N. Laska, P. T. Boufounos, M. A. Davenport, and R. G. Baraniuk, “Democracy in action: Quantization, saturation, and compressive sensing,” *Appl. Comput. Harmon. Anal.*, vol. 31, no. 3, pp. 429–443, 2011.

[7] I. Elleuch, F. Abdelkefi, M. Siala, R. Hamila, and N. Al-Dhahir, “On quantized compressed sensing with saturated measurements via greedy pursuit,” in *23rd European Signal Processing Conference (EUSIPCO)*, France, Aug. 2015, pp. 1706–1710.

[8] Z. Yang, L. Xie, and C. Zhang, “Variational bayesian algorithm for quantized compressed sensing,” *IEEE*

Trans. Signal Process., vol. 61, no. 11, pp. 2815–2824, 2013.

[9] M. Yan, Y. Yang, and S. Osher, "Robust 1-bit compressive sensing using adaptive outlier pursuit," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3868–3875, 2012.

[10] F. Li, J. Fang, H. Li, and L. Huang, "Robust onebit bayesian compressed sensing with sign-flip errors," *IEEE Signal Process. Letters*, vol. 22, no. 7, pp. 857–861, July 2015.

[11] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, 2012.

[12] M. A. Davenport, P. T. Boufounos, and R. Baraniuk, "Compressive domain interference cancellation," in *SPARS'09-Signal Processing with Adaptive Sparse Structured Representations*, 2009.

[13] M. Slawski and M. Hein, "Sparse recovery by thresholded non-negative least squares," in *Proceedings of Advances in Neural Information Processing Systems (NIPS)*, Spain, 2011.

[14] J. Laska, M. Davenport, and R. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *43d Asilomar Conference on Signals, Systems and Computers*, Nov. 2009, pp. 1556–1560.

[15] L. Jacques, "A short note on compressed sensing with partially known signal support," *Signal Processing*, vol. 90, no. 12, pp. 3308–3312, 2010.

[16] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0," available online at http://cvxr.com/cvx, Aug. 2012.