FAST ALTERNATING PROJECTED GRADIENT DESCENT ALGORITHMS FOR RECOVERING SPECTRALLY SPARSE SIGNALS

Myung Cho¹, Jian-Feng Cai³, Suhui Liu², Yonina C. Eldar⁴ and Weiyu Xu¹

¹ Dept. of ECE, University of Iowa, Iowa City, IA, 52242

² Dept. of Mathematics, University of Iowa, Iowa City, IA, 52242

³ Dept. of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China

⁴ Dept. of EE, Technion–Israel Institute of Technology, Haifa, Israel

ABSTRACT

We propose fast algorithms that speed up or improve the performance of recovering spectrally sparse signals from underdetermined measurements. Our algorithms are based on a non-convex approach of using alternating projected gradient descent for structured matrix recovery. We apply this approach to two formulations of structured matrix recovery: Hankel and Toeplitz mosaic structured matrix, and Hankel structured matrix. Our methods provide better recovery performance, and faster signal recovery than existing algorithms, including atomic norm minimization.

Index Terms— compressed sensing, atomic norm, sparse recovery, spectral estimation, matrix completion

1. INTRODUCTION

Compressed sensing is a signal processing paradigm of reducing the number of measurements needed for accurate signal recovery [1, 2]. Compressed sensing has successfully found applications in various areas such as medical imaging, fluorescence microscopy, face recognition, and radar [3].

Compressed sensing has also found applications in superresolution and line spectral estimation, which aims at estimating spectral information from few observations. To recover spectrally sparse signals from uniformly sampled observations, we can apply many conventional techniques, such as Prony's method [4], ESPRIT [5], matrix pencil method [6], and the Tufts and Kumaresan approach [7]. Recently, in [8], Candès and Fernandez-Granda proposed a total variation minimization approach to use a set of *uniformly* spaced time samples to recover sparse continuous-valued frequencies. Tang et al. introduced off-the-grid compressed sensing that employs atomic norm minimization to recover sparse continuous-valued frequencies from few randomly chosen nonuniformly-spaced time samples in [9]. The total variation minimization or atomic norm minimization employ semidefinite programs (SDP) for recovery. Solving these SDP's is

of high complexity, making it difficult to efficiently recover spectrally sparse signals of high dimensions. Recently, there have been additional efforts to improve the speed or performance of recovering spectrally sparse signals, including grid discretization [10], structured Hankel matrix completion [11], iterative reweighting [12, 13], and alternating descent conditional gradient for solving atomic norm minimization [14].

In this paper, we are interested in improving the speed and performance of recovering spectrally sparse signals, especially targeting recovery of high-dimensional spectrally sparse signals. Towards this end, we propose to study a nonconvex optimization approach to structured matrix completion. In particular, we consider two non-convex optimization formulations: low-rank Hankel and Toeplitz Mosaic (HTM) matrix completion, and low-rank Hankel matrix completion initially introduced in [11]. We suggest general projected Wirtinger gradient [15] descent methods for directly solving these two non-convex optimization formulations, without resorting to convex relaxations. Numerical results show that we can improve the performance or the speed of recovery using projected Wirtinger gradient descent, compared with atomic norm minimization [9] and nuclear norm minimization for Hankel matrix completion [11].

The remainder of this paper is organized as follows. In Section 2, we briefly review atomic norm minimization [9] and Hankel matrix completion [11]. In Section 3, we introduce low-rank Hankel and Toeplitz Mosaic (HTM) matrix completion. Thereafter, in Section 4, we propose projected Wirtinger gradient descent to directly solve the HTM and Hankel completion problems. Finally, in Section 5, we demonstrate the efficacy of our algorithms in terms of signal recovery performance as well as the computational complexity.

Notations: We denote the set of complex numbers as \mathbb{C} . We reserve calligraphic uppercase letters for index sets. When we use an index set \mathcal{K} as the subscript of a vector x, i.e., $x_{\mathcal{K}}$, it represents the part of the vector x over index set \mathcal{K} . We use the superscripts H, T, and * for matrix conjugate transpose, transpose, and conjugate respectively. For

The work of W. Xu was supported by Simons Foundation, Iowa Energy Center, KAUST, NIH 1R01EB020665-01.

 $x = [x_1, x_2, ..., x_{2n-1}]^T$, the Hankel matrix over x, denoted Hank(x), is the Hankel matrix with first column equal to $[x_1, ..., x_n]^T$ and last row equal to $[x_n, ..., x_{2n-1}]$. The Hermitian Toeplitz matrix over x, denoted Toep(x), has its first column equal to x. We write the Frobenius norm of a matrix $A \in \mathbb{C}^{m \times n}$ as $||A||_F \triangleq \sqrt{\sum_{i=1}^n \sum_{j=1}^n |A_{ij}|^2}$, where A_{ij} is the element of A in its *i*-th row and *j*-th column.

2. ATOMIC NORM MINIMIZATION AND MATRIX COMPLETION

Let x^{\bullet} be a spectrally sparse signal expressed as a sum of k complex exponentials as follows:

$$x_l^{\bullet} = \sum_{j=1}^k c_j^{\bullet} e^{i2\pi f_j^{\bullet} l} = \sum_{j=1}^k |c_j^{\bullet}| a(f_j^{\bullet}, \phi_j^{\bullet})_l, \quad l \in \mathcal{N},$$
(2.1)

where $f_j^{\bullet} \in [0,1]$ represents an unknown frequency, $c_j^{\bullet} = |c_j^{\bullet}| e^{i\phi_j^{\bullet}}$ is its coefficient, $\phi_j^{\bullet} \in [0,2\pi]$ is its phase, and $\mathcal{N} = \{l : 1 \leq l \leq 2n-1\}$ is the set of time indices. Here, $a(f_j^{\bullet}, \phi_j^{\bullet}) \in \mathbb{C}^{|\mathcal{N}|}$ is a *frequency-atom*, with *l*-th element $a(f_j^{\bullet}, \phi_j^{\bullet})_l = e^{i(2\pi f_j^{\bullet} l + \phi_j^{\bullet})}$. When the phase is 0, we denote the frequency-atom simply by $a(f_j)$. We assume that the signal is observed over the time index set $\mathcal{M} \subseteq \mathcal{N}$, $|\mathcal{M}| = m \leq 2n-1$, where *m* observations are chosen randomly. Our goal is to recover the true signal with the smallest possible number of observations. Recovering the true signal is not trivial because the frequencies are in a continuous domain, and their phases and magnitudes are also unknown.

In [9], the authors propose the following atomic norm minimization to recover a spectrally sparse signal x^{\bullet} using randomly chosen time samples $\mathcal{M} \subseteq \mathcal{N}$:

$$\begin{array}{l} \underset{x}{\text{minimize }} \|x\|_{\mathcal{A}} \\ \text{subject to } x_{\mathcal{M}} = x_{\mathcal{M}}^{\bullet}, \end{array}$$
(2.2)

where $||x||_{\mathcal{A}}$ represents the atomic norm of a signal x, defined as $||x||_{\mathcal{A}} = \inf\{\sum_{j} |c_{j}| : x = \sum_{j} c_{j}a(f_{j})\}$. The atomic norm minimization (2.2) can be cast as an SDP as follows [9, (II.6)]:

$$\begin{array}{l} \underset{u,x,t}{\text{minimize}} \quad \frac{1}{2|\mathcal{N}|} \operatorname{Tr}(\operatorname{Toep}(u)) + \frac{1}{2}t \\ \text{subject to} \quad U_1 = \begin{bmatrix} \operatorname{Toep}(u) & x \\ x^H & t \end{bmatrix} \ge 0, \\ x_{\mathcal{M}} = x_{\mathcal{M}}^{\bullet}, \end{array}$$
(2.3)

where $Tr(\cdot)$ is the trace function, and Toep(u) is a Hermitian Toeplitz matrix. We refer the reader to [9] for details. The atomic norm minimization requires a certain minimum separation between frequencies for successful recovery.

Inspired by the matrix pencil approach [6, 16], the authors of [11] developed the Enhanced matrix completion (EMaC) to recover spectrally sparse signals from randomly chosen time samples. The EMaC formulation is stated as follows [11, (13)]:

$$\begin{array}{l} \underset{x}{\text{minimize }} \|\text{Hank}(x)\|_{*} \\ \text{subject to } x_{\mathcal{M}} = x_{\mathcal{M}}^{\bullet}, \end{array}$$

$$(2.4)$$

where $||M||_*$ represents the nuclear norm of a matrix M, which is the sum of the singular values of M. This convex program can be further as an SDP:

$$\begin{array}{l} \underset{x,Q_{1},Q_{2}}{\text{minimize}} \quad \frac{1}{2} \operatorname{Tr}(Q_{1}) + \frac{1}{2} \operatorname{Tr}(Q_{2}) \\ \text{subject to} \quad U_{2} = \begin{bmatrix} Q_{1} & \operatorname{Hank}(x) \\ \operatorname{Hank}(x)^{H} & Q_{2} \end{bmatrix} \geq 0, \qquad (2.5) \\ x_{\mathcal{M}} = x_{\mathcal{M}}^{\bullet}, \end{array}$$

where the matrices Q_1 and Q_2 are Hermitian matrices.

When the underlying frequencies satisfy the separation condition, atomic norm minimization has better signal recovery performance than EMaC. However, EMaC provides robust signal recovery even when the separation condition is not satisfied. Moreover, solving the SDP for atomic norm minimization and EMaC can be slow, especially for problems of large dimensions. In the next section, we propose new optimization formulations and algorithms which can provide better recovery performance of spectrally sparse signals in faster speed, compared with atomic norm minimization and nuclear norm minimization for Hankel matrix completion.

3. HANKEL AND TOEPLITZ MOSAIC MATRIX COMPLETION

In this section, we introduce a new optimization formulation, called Hankel and Toeplitz Mosaic (HTM) matrix completion, to recover spectrally sparse signals. Our HTM matrix completion is formulated as follows:

minimize rank(U₃)
subject to
$$U_3 = \begin{bmatrix} \text{Toep}(z) & \text{Hank}(x) \\ \text{Hank}(x)^H & \text{Toep}(z)^T \end{bmatrix} \ge 0,$$
 (3.1)
 $x_{\mathcal{M}} = x_{\mathcal{M}}^{\bullet},$

where rank(\cdot) denotes the matrix rank. One can relax (3.1) to its nuclear norm minimization:

$$\begin{array}{ll} \underset{z,x}{\text{minimize } 2\text{Tr}(\text{Toep}(z))} \\ \text{subject to } & U_3 = \begin{bmatrix} \text{Toep}(z) & \text{Hank}(x) \\ \text{Hank}(x)^H & \text{Toep}(z)^T \end{bmatrix} \ge 0, \qquad (3.2) \\ & x_{\mathcal{M}} = x_{\mathcal{M}}^{\bullet}. \end{array}$$

The difference between (2.5) and (3.2) is that (3.2) further imposes that the diagonal matrices are Hermitian Toeplitz matrices.

The following theorem states a relation between (3.2) and atomic norm minimization. We omit its proof due to space limitations.

Theorem 3.1. Let x^{\dagger} be an optimal solution to (3.2), and suppose that $Hank(x^{\dagger})$ can be decomposed as VDV^{T} , where D is a diagonal matrix and V has frequency atoms as its columns. Then the optimal objective value of (3.2) is the smallest atomic norm among all signals x satisfying the observation constraints; namely (3.2) is equivalent to atomic norm minimization.

4. PROJECTED WIRTINGER GRADIENT DESCENT FOR HTM AND HANKEL MATRIX COMPLETION

We now propose a projected Wirtinger gradient descent algorithm for various structured matrices, especially, HTM matrix completion and Hankel matrix completion. We first introduce the method for HTM matrix completion, which solves (3.1).

Let us define the set of all complex-valued matrices with rank no greater than k as follows:

$$\mathscr{R}^k_{\mathbb{C}} = \{ L \in \mathbb{C}^{2n \times 2n} \mid \operatorname{rank}(L) \le k \}.$$
(4.1)

We further denote the set of all complex-valued Toeplitz Hankel mosaic matrices that are consistent with the observed data as

$$\mathcal{H}_{HTM} = \left\{ \begin{bmatrix} \text{Toep}(z) & \text{Hank}(x) \\ \text{Hank}(x)^H & \text{Toep}(z)^T \end{bmatrix} \mid z \in \mathbb{C}^n, \\ x \in \mathbb{C}^{2n-1}, \ x_{\mathcal{M}} = x_{\mathcal{M}}^{\bullet} \right\}. \quad (4.2)$$

We then seek a matrix in \mathscr{H}_{HTM} with rank no larger than k by considering the problem:

$$\underset{L \in \mathscr{R}^k_{\mathbb{C}}, H \in \mathscr{H}}{\text{minimize}} \quad \frac{1}{2} \|L - H\|_F^2.$$
(4.3)

To solve (4.3), we use a projected gradient descent algorithm. We start with initial values H_0 and L_0 . By considering Wirtinger calculus [15] and applying gradient descent on the function with complex variables, we derive the update rule of our projected Wirtinger algorithm as follows:

$$L_{t+1} \in \mathcal{P}_{\mathscr{R}^k_{\mathbb{C}}}(L_t - \delta_1(L_t - H_t)),$$

$$H_{t+1} \in \mathcal{P}_{\mathscr{H}_{HTM}}(H_t - \delta_2(H_t - L_{t+1})), \qquad (4.4)$$

where t is the iteration number, $\delta_1 > 0$ and $\delta_2 > 0$ are step sizes, and $\mathcal{P}_{\mathscr{R}^k_{\mathbb{C}}}$ and $\mathcal{P}_{\mathscr{H}_{HTM}}$ are the projections onto $\mathscr{R}^k_{\mathbb{C}}$ and \mathscr{H}_{HTM} respectively.

The projection $\mathcal{P}_{\mathscr{R}^k_{\mathbb{C}}}(X)$, which is the projection onto the best rank-k approximation to X, is calculated through singular value decomposition. Let the singular value decomposition of X be $X = U\Lambda V^H$. Then, $\mathcal{P}_{\mathscr{R}^k_{\mathbb{C}}}(X) = U_k\Lambda_k V^H_k$,

Algorithm 1: Projected Wirtinger gradient descent					
method					
Input : $x_{\mathcal{M}}^{\bullet}$, MaxItr, ϵ_{err} , δ_1 , δ_2 , k					
Output: x					
1 Initialize: $L_0 \leftarrow 0, H_0 \leftarrow 0$					
2 for $t = 0$ to MaxItr do					
3 $L_{t+1} \leftarrow \mathcal{P}_{\mathscr{R}^k_{\mathbb{C}}}(L_t - \delta_1(L_t - H_t))$					
4 $H_{t+1} \leftarrow \mathcal{P}_{\mathscr{H}}(H_t - \delta_2(H_t - L_{t+1})) \triangleright \mathscr{H} \text{ is } \mathscr{H}_{HTM} \text{ for}$					
HTM completion, or \mathcal{H} is the set of Hankel matrices for					
Hankel matrix completion					
5 if $ H_{t+1} - H_t _F / H_t _F < \epsilon_{err}$ then					
6 break					
7 end					
s end					

where Λ_k is the diagonal matrix that only retains the k largest nonnegative singular values of X, and U_k and V_k are the matrices whose columns are the corresponding left and right singular vectors respectively.

The projection $\mathcal{P}_{\mathscr{H}_{HTM}}(X)$ is carried out for Hankel and Toeplitz matrices separately due to its mosaic structure. More precisely, for any $X = \begin{bmatrix} X_{11} & X_{12} \\ X_{12}^* & X_{22} \end{bmatrix} \in \mathbb{C}^{2n \times 2n}$, we have

$$\mathcal{P}_{\mathcal{H}_{HTM}}(X) = \begin{bmatrix} \frac{1}{2} \mathcal{P}_T(X_{11} + X_{22}^T) & \mathcal{P}_H(X_{12}) \\ \mathcal{P}_H(X_{12})^H & \frac{1}{2} \left(\mathcal{P}_T(X_{11} + X_{22}^T) \right)^T \end{bmatrix}$$

Here \mathcal{P}_H is the projection onto the set of Hankel matrices whose anti-diagonals coincide with the observed data. From [17, (8)],

$$\mathcal{P}_{H}(Y) = \text{Hank}(z),$$

where
$$\begin{cases} z_{i} = y_{i}, & i \in \mathcal{M}, \\ z_{i} = \text{mean}\{Y_{ab}|a + b = i\}, & \text{otherwise.} \end{cases}$$
 (4.5)

Namely, for the missed measurement signal z_i , $i \in \mathcal{N} \setminus \mathcal{M}$, we calculate the average of the *i*-th anti-diagonal elements of Y. The operator \mathcal{P}_T is the projection onto the set of Toeplitz matrices, and is given by

$$\mathcal{P}_T(Y) = \text{Toep}(z),$$

where $z_i = \text{mean}\{Y_{ab}|a - b = i, a \ge b\}.$ (4.6)

Namely, for Toeplitz matrix projection, we compute the average of the *i*-th diagonal elements. The projected Wirtinger algorithm continues the projections as in (4.4) onto a low-rank matrix and Hankel Toeplitz mosaic matrix until it converges to a solution or the maximum number of iterations (MaxItr) is exhausted.

The projected Wirtinger for Hankel matrix completion is similar to the projected Wirtinger for HTM. It uses the projection onto the Hankel matrix, instead of Hankel Toeplitz mosaic matrix, according to (4.5). Based on Attouch and Bolte's theory [18, 19], we provide global convergence analysis of the algorithm in [17]. We refer the reader to our longer version [17] for details. We summarize our procedure in Algorithm 1.

5. NUMERICAL EXPERIMENTS

We compare the performance and complexity of our projected Wirtinger gradient descent methods for both HTM and Hankel matrix completion with the standard Atomic Norm Minimization (ANM) [9], and the Enhanced Matrix Completion (EMaC) [11]. We conducted our numerical experiments on HP Z220 CMT with Intel Core i7-3770 dual core CPU @3.4GHz clock speed and 16GB DDR3 RAM, using Matlab (R2013b release) on Windows 7 OS. We use CVX [20] to solve convex programs for ANM and EMaC. In all experiments, the phases and frequencies are sampled uniformly at random in $[0, 2\pi)$ and [0, 1] respectively. The amplitudes $|c_j|, j = 1, ..., k$, are chosen as $|c_j| = 1 + 10^{0.5m_j}$, where m_j is randomly drawn in the uniform distribution on [0, 1].

We carried out numerical experiments to compare the signal recovery performance of our projected Wirtinger method with ANM and EMaC. We abbreviate HTM and Hankel matrix completion using projected Wirtinger to HTM-PW and EMaC-PW respectively. We varied the number of measurements m from 20 to 127, and the sparsity k from 1 to 40. We obtained the probability of successful signal recovery over 100 trials for each parameter setup. We consider a recovery successful if $||x^{\bullet} - \hat{x}||_2 / ||x^{\bullet}||_2 \leq \epsilon_{succ}$, where $\epsilon_{succ} =$ 10^{-2} , \hat{x} is the estimated signal, and x^{\bullet} is the true signal. Fig. 1 demonstrates that the projected Wirtinger algorithms (Fig. 1(a) and (b)) improve the phase transition boundary over other algorithms (Fig. 1(c) and (d)). We provide Fig. 2 to more clearly show the advantage of our algorithms in signal recovery, where we use the success criterion $\epsilon_{succ} = 5 \times 10^{-3}$. In particular, when the sparsity k is 40, HTM-PW has 50% success rate with around 87 measurements out of 127 respectively, while the success rate of other algorithms with the same number of measurements is 0. The 50% success rate of other algorithms is achieved around 110 for EMaC, and 120 for ANM. Even though HTM-PW has the best phase transition boundary curve, in certain regions (upper orange color region) of its phase transition, HTM-PW has smaller success rate than other algorithms. It would be interesting to understand this phenomenon more deeply. We leave this for future work.

We assess the computational complexity of our algorithms in terms of the average execution time by averaging over 10 trials. We provide results in Table 1 when the signal dimension (2n-1) varies from 101 to 1401, $m = \lfloor (2n-1)/2 \rfloor$, and $k = \lfloor 0.1(2n-1) \rfloor$. We stopped our projected Wirtinger algorithms when $||H_{t+1} - H_t||_F / ||H_t||_F \le 10^{-6}$ or the maximum number of iterations (MaxItr) is exhausted. We set Max-Itr to 1000. The success rate was similar to that shown in



Fig. 1. Phase transition for successful signal recovery when the signal dimension 2n - 1 = 127.



Fig. 2. The probability of signal recovery when (2n - 1, k) = (127, 40).

 Table 1. Execution Time (Unit: seconds)

2n - 1	Project Wirtinger		CVX		
	HTM	EMaC	ANM	EMaC	
101	3.7	0.1	27.1	5.7	
201	7.1	0.2	787.7	51.6	
401	309.7	0.9	N/A ^a	N/A	
601	733.4	1.4	N/A	N/A	
1001	3612.4	6.6	N/A	N/A	
1401	8999.2	10.1	N/A	N/A	

^a Not Available (Out of Memory)

Fig. 1. Table 1 clearly shows that the speed of our projected Wirtinger methods outperform those of ANM and EMaC.

6. REFERENCES

- D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [2] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [3] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*, Cambridge University Press, 2012.
- [4] R. Prony, "Essai experimental et analytique," J. de lÉcole Polytechnique (Paris), vol. 1, no. 2, pp. 24–76, 1795.
- [5] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.
- [6] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially dampled/undampled sinusoids in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 5, pp. 814–824, 1990.
- [7] D. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," in *Proceedings of the IEEE*, 1982, pp. 975– 989.
- [8] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.
- [9] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Transactions on Information The*ory, vol. 59, no. 11, pp. 7465–7490, 2013.
- [10] G. Tang, B. N. Bhaskar, and B. Recht, "Sparse recovery over continuous dictionaries: Just discretize," in *Proceedings* of Asilomar Conference on Signals, Systems, and Computers, 2013, pp. 1043–1047.
- [11] Y. Chen and Y. Chi, "Robust spectral compressed sensing via structured matrix completion," *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6576–6601, 2014.
- [12] M. Cho, K. V. Mishra, J.-F. Cai, and W. Xu, "Block iterative reweighted algorithms for super-resolution of spectrally sparse signals," *IEEE Signal Processing Letters*, vol. 22, no. 12, pp. 2319–2323, 2015, To appear.
- [13] Z. Yang and L. Xie, "Enhancing sparsity and resolution via reweighted atomic norm minimization," arXiv preprint arXiv:1408.5750, 2014.
- [14] N. Boyd, G. Schiebinger, and B. Recht, "The alternating descent conditional gradient method for sparse inverse problems," arXiv preprint arXiv:1507.01562, 2015.
- [15] R. F. Fischer, Precoding and signal shaping for digital transmission, John Wiley & Sons, 2005.
- [16] Y. Hua, "Esitmating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Transactions on Signal Processing*, vol. 40, no. 9, pp. 2267–2280, 1992.
- [17] J.-F. Cai, S. Liu, and W. Xu, "Projected Wirtinger gradient descent for low-rank hankel matrix completion in spectral compressed sensing," arXiv preprint arXiv:1507.03707, 2015.

- [18] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-lojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.
- [19] J. Bolte, S. Sabach, and M. Teboulle, "Proximal alternating linearized minimization for nonconvex and nonsmooth problems," *Mathematical Programming*, vol. 146, no. 1-2, pp. 459– 494, 2014.
- [20] M. Grant and S. P. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sept. 2012.