



CHALMERS

Chalmers Publication Library

Structured Sparse Approximation via Generalized Regularizers: With Application to V2V Channel Estimation

This document has been downloaded from Chalmers Publication Library (CPL). It is the author's version of a work that was accepted for publication in:

IEEE Global Telecommunications Conference (GLOBECOM).

Citation for the published paper:

Beygi, S. ; Ström, E. ; Mitra, U. (2014) "Structured Sparse Approximation via Generalized Regularizers: With Application to V2V Channel Estimation". IEEE Global Telecommunications Conference (GLOBECOM).

Downloaded from: <http://publications.lib.chalmers.se/publication/202317>

Notice: Changes introduced as a result of publishing processes such as copy-editing and formatting may not be reflected in this document. For a definitive version of this work, please refer to the published source. Please note that access to the published version might require a subscription.

Chalmers Publication Library (CPL) offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all types of publications: articles, dissertations, licentiate theses, masters theses, conference papers, reports etc. Since 2006 it is the official tool for Chalmers official publication statistics. To ensure that Chalmers research results are disseminated as widely as possible, an Open Access Policy has been adopted. The CPL service is administrated and maintained by Chalmers Library.

(article starts on next page)

Structured Sparse Approximation via Generalized Regularizers: With Application to V2V Channel Estimation

Sajjad Beygi, Erik G. Ström*, Urbashi Mitra

School of Electrical Engineering, University of Southern California (USC), Los Angeles, CA

*Department of Signals and Systems, Chalmers University of Technology, Gothenburg, Sweden

Abstract—In this paper, we consider the estimation of a signal that has both group- and element-wise sparsity (joint sparsity); motivated by channel estimation in vehicle-to-vehicle channels. A general approach for the design of separable regularizing functions is proposed to adaptively induce sparsity in the estimation. A joint sparse signal estimation problem is formulated via these regularizers and its optimal solution is computed based on proximity operations. Our optimization results are quite general and they can be applied in the context of hierarchical sparsity models as well. The proposed recovery algorithm is a nested iterative method based on the alternating direction method of multipliers (ADMM). Due to regularizer separability, key operations can be performed in parallel. V2V channels are estimated by exploiting the joint sparsity (group/element-wise) exhibited in the delay-Doppler domain. Simulation results reveal that the proposed method can achieve as much as a 10 dB gain over previously examined methods.

I. INTRODUCTION

The sparse approximation of signal \mathbf{x} using the noisy measured signal \mathbf{y} is often obtained as the solution of

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \Phi_{\text{fid}}(\mathbf{y}, \mathbf{x}) + \sum_k \lambda_k \phi_{\text{reg}}^{(k)}(\mathbf{x}) \right\},$$

where $\phi_{\text{reg}}^{(k)}(\cdot)$ regularizes the solution by enforcing certain prior sparsity constraints on \mathbf{x} ; $\Phi_{\text{fid}}(\cdot)$ measures the violation of the relation between \mathbf{x} and its observation \mathbf{y} , and λ_k , the penalty parameters, are positive constants that weigh the terms in the minimization. Assuming a mixture model relationship between the measurement and desired signal, we have $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$, where \mathbf{A} is a known over-complete dictionary, \mathbf{x} is a signal which has sparse structure, and \mathbf{n} is additive Gaussian noise. There are different types of sparsity structures:

Element-wise sparsity occurs if \mathbf{x} has only a few non-zero entries. For this form of sparsity, basis pursuit (BP) [1] and the least absolute shrinkage and selection operator (LASSO) [2], *etc.*, have been considered. In all of these methods, the fidelity function is the l_2 norm on the error $(\mathbf{y} - \mathbf{A}\mathbf{x})$ and the sparsity constraint l_0 is relaxed yielding the sparse approximation problem given by,

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_e \|\mathbf{x}\|_1, \quad (P_1)$$

where $\|\mathbf{x}\|_1$ is the sparsity-inducing regularizer. Alternatively, there is *group-wise* or *block-wise* sparsity. In this case, the elements of \mathbf{x} can be clustered in groups where a few groups have non-zero values. One can again use the l_2 norm on the

error and relax the sparsity measure to l_1 which is applied to groups, yielding the following sparse approximation problem,

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_g \sum_{i=1}^{N_g} \|\mathbf{x}_i\|_2, \quad (P_2)$$

where N_g is the number of groups, the group vectors \mathbf{x}_i contains disjoint elements from \mathbf{x} , and $\sum_{i=1}^{N_g} \|\mathbf{x}_i\|_2$ is the group sparsity-inducing term. Intuitively, we can say that the term $\sum_{i=1}^{N_g} \|\mathbf{x}_i\|_2$ is equivalent to the l_1 -norm of (the square-root of) the energy in each group. In order to exploit the group-sparsity structure of the signal being sensed, a number of variants of LASSO have become popular, *e.g.*, group LASSO [3] and block sparsity methods [4]–[6].

Finally, there is *joint element and group-wise* sparsity structure, where the corresponding optimization problem includes the combination of regularizers in (P_1) and (P_2) , *e.g.*, sparse group/block LASSO [7], [8]. We observe that for both (P_1) and (P_2) , the objective functions are convex resulting in straightforward optimization. On the other hand, non-convex regularizer functions are advantageous in that they usually yield sparser solutions for a given residual energy [9]–[11]. Generally, convex approaches are based on sparsity-promoting convex penalty functions (*e.g.*, the l_1 norm), while non-convex approaches are based on non-convex penalty functions (*e.g.*, the l_p pseudo-norm with $p < 1$).

The contributions of our work can be stated as follows: we propose a sparse approximation method with two regularization terms to control the element-wise sparsity and group-wise sparsity similar to the sparse-group-LASSO, but we utilize the advantage of non-convex regularization functions. We propose a general method for designing the regularizer term with desired properties. We show that the joint sparse signal estimation problem using our designed regularizers can be performed via simple proximity operations. Furthermore, the optimal solution is evaluated by applying the group and element-wise proximity operators in a nested fashion. The proposed algorithm is based on an alternating direction method of multipliers [16]. Our numerical results show a significant improvement for V2V channels estimation using our proposed structured signal estimator relative to the prior work in the literature.

This paper is organized as follows: In Section II, we state some required definitions from convex analysis and we propose the general method to design a regularizer to induce the sparsity. In Section III, the sparse signal estimation algorithm is given. We provide the simulation results of application of

This research was funded in part by one or all of these grants: ONR N00014-09-1-0700, NSF CNS-0832186, NSF CCF-1117896, CCF-0917343, CNS-1213128, AFOSR FA9550-12-1-0215, DOT CA-26-7084-00, Barbro Osher Pro Suecia Foundation, Ericsson's Research Foundation FOSTIFT-13:038, and Adlerbert Research Foundation.

our proposed method for V2V channel estimation in Section IV. Finally, Section V concludes the paper.

Notation: We denote a scalar by x , a column vector by \mathbf{x} , and its i -th element with $[\mathbf{x}]_i$. Similarly, we denote a matrix by \mathbf{X} and its (i, j) -th element by $[\mathbf{X}]_{i,j}$. The transpose and conjugate transpose of \mathbf{X} is denoted by \mathbf{X}^T and \mathbf{X}^H , respectively, and \mathbb{R} and \mathbb{C} denote the set of real and complex numbers. We denote the sub-gradient of function F by ∂F and the gradient of function F by ∇F . The element-wise (Schur) product is denoted by \odot .

II. REGULARIZER: CONSTRUCTIVE METHOD

In this section, we first review some definitions from convex and variational analysis. Then, we state our general approach to design the regularizer based on the relationship between the *proximity operator* and the *Moreau envelope* [12]. From a signal processing perspective, proximity operators have a very natural interpretation in terms of de-noising [9]. The denoising problem can be formulated as

$$\operatorname{argmin}_{\mathbf{a} \in \mathbb{R}^N} \left\{ \frac{1}{2} \|\mathbf{b} - \mathbf{a}\|_2^2 + \lambda F(\mathbf{a}) \right\},$$

where $\mathbf{b} \in \mathbb{R}^N$ and F is a regularization function. Such a formulation derives, in particular, from a Bayesian approach to de-noising in the presence of Gaussian noise and of a *prior* with a log-concave density function. Proximity operators have very attractive properties that make them particularly well suited for iterative minimization algorithms.

Definition 1. (see [12], p. 20) Let $F : \mathbb{R}^N \rightarrow \mathbb{R}$ be a real-valued function. The *proximity operator* of F for every $\mathbf{b} \in \mathbb{R}^N$ and $\lambda > 0$ is defined as

$$P_{\lambda, F}(\mathbf{b}) := \operatorname{argmin}_{\mathbf{a}} \left\{ \frac{1}{2} \|\mathbf{b} - \mathbf{a}\|_2^2 + \lambda F(\mathbf{a}) \right\} \quad (1)$$

Remark 1. The proximity operator of F admits a unique solution, if the objective function $J(\mathbf{a}) = \frac{1}{2} \|\mathbf{b} - \mathbf{a}\|_2^2 + \lambda F(\mathbf{a})$ is a strictly convex function. Furthermore, $P_{\lambda, F}(\mathbf{a})$ is characterized by the inclusion that $\forall (\mathbf{a}^*, \mathbf{b})$, $\mathbf{a}^* = P_{\lambda, F}(\mathbf{b}) \iff \mathbf{b} - \mathbf{a}^* \in \lambda \partial F(\mathbf{a}^*)$.

Remark 2. Note that F does not need to be a convex or differentiable function to satisfy the condition in Remark 1.

Definition 2. (see [12], p. 20) Let $F : \mathbb{R}^N \rightarrow \mathbb{R}$ be a lower semi-continuous (lsc) real-valued function and $\lambda > 0$, the *Moreau envelope* function, $e_{\lambda, F}(\mathbf{b}) : \mathbb{R}^N \rightarrow \mathbb{R}$ is defined by

$$e_{\lambda, F}(\mathbf{b}) := \inf_{\mathbf{a}} \left\{ \frac{1}{2\lambda} \|\mathbf{b} - \mathbf{a}\|_2^2 + F(\mathbf{a}) \right\}.$$

Remark 3. Consider a general function $f : \mathbb{R} \rightarrow \mathbb{R}$, $e_{\lambda, f}(b)$ approximates f from below, namely we have $e_{\lambda, f}(b) \leq f(b)$ for $\forall b \in \mathbb{R}$. For smaller and smaller λ , $e_{\lambda, f}(b)$ approximates $f(b)$ better and better.

One of the well-known regularization functions is the absolute value function, *i.e.*, $f(a) = |a|$, which in vector form becomes the l_1 -norm, $F(\mathbf{a}) = \|\mathbf{a}\|_1 = \sum_{i=1}^N |[\mathbf{a}]_i|$. For this

case, using simple calculus operations, we can derive the proximity operator and Moreau envelope for $f(a) = |a|$ as

$$P_{\lambda, f}(b) = \operatorname{sign}(b) \max(0, |b| - \lambda) \quad (2)$$

and

$$e_{\lambda, f}(b) = \begin{cases} \frac{|b|^2}{2\lambda}, & \text{if } |b| \leq \lambda \\ |b| - \frac{\lambda}{2}, & \text{if } |b| > \lambda \end{cases} \quad (3)$$

The function $e_{\lambda, f}(b)$ defined in (3) is called the *Huber loss function* (*e.g.*, [9], [13]). Also, $P_{\lambda, f}(b)$ in (2) is well-known as the *soft thresholding operator* [2], [9].

Remark 4. Consider that $F(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ is a separable, *i.e.*, $F(\mathbf{b}) = \sum_{i=1}^N f([\mathbf{b}]_i)$. Then, $[P_{\lambda, F}(\mathbf{b})]_i = P_{\lambda, f}([\mathbf{b}]_i)$ and $e_{\lambda, F}(\mathbf{b}) = \sum_{i=1}^N e_{\lambda, f}([\mathbf{b}]_i)$. For example, the proximity operator of $F(\mathbf{b}) = \|\mathbf{b}\|_1$ is

$$[P_{\lambda, F}(\mathbf{b})]_i = \operatorname{sign}([\mathbf{b}]_i) \max\{0, |[\mathbf{b}]_i| - \lambda\}.$$

Definition 3. (see [12], p. 473) The function F^* is called the convex conjugate of convex function $F : \mathbb{R}^N \rightarrow \mathbb{R}$, if $F^*(\mathbf{p}) = \sup_{\mathbf{a} \in \mathbb{R}^N} (\mathbf{p}^T \mathbf{a} - F(\mathbf{a}))$.

In the following, we propose a theorem to build the connections between the regularizer, proximity operator, and Moreau envelope. This idea was first suggested in [13] for a particular form of regularizer (*i.e.*, Huber loss function), but *herein* we extend it for a general regularizer.

Theorem 1. Define the two following auxiliary functions,

$$\eta_1(\mathbf{a}) = \frac{1}{2} \|\mathbf{a}\|_2^2 + \lambda F(\mathbf{a}), \text{ and } \eta_2(\mathbf{a}) = \frac{1}{2} \|\mathbf{a}\|_2^2 - \lambda G(\mathbf{a}).$$

Assume that $\eta_1(\mathbf{a})$ is strictly convex. Then,

- 1) $G(\mathbf{a})$ is the Moreau envelope of $F(\mathbf{a})$, *i.e.*, $G(\mathbf{a}) = e_{\lambda, F}(\mathbf{a})$, if and only if $\eta_2 = \eta_1^*$, where η_1^* is the convex conjugate of η_1 .
- 2) The proximity operator $P_{\lambda, F}(\mathbf{a}) = \nabla \eta_2(\mathbf{a})$, where $\nabla \eta_2(\mathbf{a})$ is the gradient of $\eta_2(\mathbf{a})$ and $G(\mathbf{a}) = e_{\lambda, F}(\mathbf{a})$.

Proof: see Appendix A.

Remark 5. Based on the results in Theorem 1, if we specify one of the functions $e_{\lambda, F}(\mathbf{a})$ or $P_{\lambda, F}(\mathbf{a})$, then we can compute the other function accordingly.

Here, we consider a specific function for the Moreau envelope and we construct the proximity operator for de-noising using the specified Moreau envelope. In this way, we do not need to specify the regularizer, instead we work only with the Moreau envelope which intuitively can be considered as the smooth approximation of the regularizer, as mentioned earlier in Remark 3. Furthermore we can anticipate the structure of the reconstruction method. Our goal here is to design an adaptive non-convex regularizer such that we end up with a simple proximity operator in the joint sparse signal estimation algorithm. Based on our knowledge about the Huber loss function, we know that its proximity operator is a simple thresholding algorithm. Thus, as an example, we can consider a generalized version of the Huber loss function as follows:

$$e_{\lambda, f_H}(b; p) = \begin{cases} \frac{b^2}{2\lambda}, & \text{if } |b| \leq \lambda \\ \frac{|b|^p}{p\lambda^{p-1}} + \frac{\lambda}{2} - \frac{\lambda}{p}, & \text{if } |b| > \lambda \end{cases} \quad (4)$$

where $0 < p \leq 1$ and $\lambda > 0$ are constants. Now using Theorem 1, we can compute the proximity operator function corresponding to $e_{\lambda, f_H}(b; p)$ as, $P_{\lambda, f_H}(b; p) = \nabla \eta_2(b)$, where $\eta_2(b) = \frac{b^2}{2} - \lambda e_{\lambda, f_H}(b; p)$ is a convex continuous function.

Lemma 1. For $e_{\lambda, f_H}(b; p)$ defined in (4), we have

$$P_{\lambda, f_H}(b; p) = b \cdot \max \left\{ 0, 1 - \left(\frac{|b|}{\lambda} \right)^{p-2} \right\}. \quad (5)$$

where $0 < p \leq 1$.

The proof of Lemma 1 is elementary and thus omitted.

Using Remark 4 and Lemma 1, for $F_H(\mathbf{b}) = \sum_{i=1}^N f_H(|\mathbf{b}|_i)$, we have $[P_{\lambda, F_H}(\mathbf{b}; p)]_i = |\mathbf{b}|_i \cdot \max \left\{ 0, 1 - \left(\frac{|\mathbf{b}|_i}{\lambda} \right)^{p-2} \right\}$.

Remark 6. For $p = 1$, the proximity operator defined above becomes soft thresholding [2] and corresponds to l_1 -norm regularization. For $0 < p < 1$, it is possible to compute the regularization function $f_H(\cdot; p)$ using the auxiliary function η_2 , but since we can compute the proximity operator without knowing the regularization function, we do not need to find the explicit regularization function. Furthermore, we can infer some of properties of the function $f_H(\cdot; p) : \mathbb{R} \rightarrow \mathbb{R}$, by just using its relationship with the auxiliary functions η_1 and η_2 , and their conjugate relationship. As in the above example of the generalized version of the Huber loss function, one can easily check that $f_H(\cdot; p)$ is continuous, absolute non-decreasing¹, differentiable except at 0 with $\partial f_H(0; p) = [-1, 1]$, and concave on $(-\infty, 0)$ and $(0, \infty)$. As a simple trick introduced in Remark 3, one can easily verify all these properties by $f_H(\mathbf{a}) \approx \lim_{\lambda \rightarrow 0} e_{\lambda, f_H}(\mathbf{a})$.

III. RECONSTRUCTION ALGORITHM: JOINT SPARSITY

Assume that we have the observation vector $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ where $\mathbf{A} \in \mathbb{C}^{M \times N}$ is a sensing matrix and $\mathbf{n} \in \mathbb{C}^M$ is an additive white Gaussian noise. Furthermore, assume that the magnitude vector of the elements in vector $\mathbf{x} \in \mathbb{C}^N$ has the joint group and element-wise sparsity pattern. Our goal in this section, is to estimate the vector \mathbf{x} from an observation vector \mathbf{y} . To this aim, we need regularization terms such that they enforce the sparsity, while simultaneously encouraging group sparsity. This goal can be achieved by solving following optimization problem

$$\underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_g \phi_g(\|\mathbf{x}\|; p) + \lambda_e \phi_e(\|\mathbf{x}\|; q), \quad (6)$$

where $\phi_g(\cdot; p)$ is regularization term to induce the group sparsity, $\phi_e(\cdot; q)$ is the term to induce the element-wise sparsity, and $\|\mathbf{x}\|$ is the (real-valued) vector found by taking the element-wise magnitude of \mathbf{x} . In general, the penalty parameters can be selected from a given range in a cross-validation manner, by varying one of the parameters and keeping others fixed [7].

¹A absolute non-decreasing function $F : \mathbb{R}^N \rightarrow \mathbb{R}$ is a function such that $F(\mathbf{a}) = G(\|\mathbf{a}\|)$, where $G : \mathbb{R} \rightarrow \mathbb{R}$ is a non-decreasing function.

Here, we consider the generalized penalty functions as:

$$\phi_g(\|\mathbf{x}\|; p) = \sum_{i=1}^{N_g} f_H(\|\mathbf{x}_i\|_2; p), \quad (7)$$

$$\phi_e(\|\mathbf{x}\|; q) = \sum_{i=1}^N f_H(\|\mathbf{x}\|_i; q). \quad (8)$$

Note that we do not know the explicit regularization function, $f_H(\cdot; p)$, but we know its Moreau envelope function as defined in (4). We will see that this information is enough to design the reconstruction algorithm.

Definition 4 (Group). The i th group vector is defined as $\mathbf{x}_i = \mathbf{W}_i \mathbf{x}$, where \mathbf{W}_i is a matrix with entries 0 or 1 such that in each column or row there is at most one 1, and also for $i \neq j$ (two distinct groups), we have $\mathbf{W}_i^T \mathbf{W}_j = \mathbf{0}$.

The optimization problem in (6) can be rewritten using an auxiliary variable \mathbf{z} as follows,

$$\begin{aligned} \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_g \phi_g(\|\mathbf{z}\|; p) + \lambda_e \phi_e(\|\mathbf{z}\|; q) \\ \text{s.t. } \mathbf{z} = \mathbf{x} \end{aligned} \quad (9)$$

The objective function in the optimization problem (9) consists of the three terms along with a linear constraint. The first term only depends on the variable \mathbf{x} and the second and third terms only depend on the variable \mathbf{z} . The problem formulation is therefore well suited to be addressed using alternating direction method of multipliers (ADMM) [16]. ADMM will find the optimal solution to (9), since it can be shown that the objective function is strictly convex with the regularizers in (7) and (8).

For the optimization problem given in (9), we form the augmented Lagrangian

$$\begin{aligned} L_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\theta}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_g \phi_g(\|\mathbf{z}\|; p) + \lambda_e \phi_e(\|\mathbf{z}\|; q) \\ + \operatorname{Re}(\langle \boldsymbol{\theta}, \mathbf{x} - \mathbf{z} \rangle) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}\|_2^2, \end{aligned} \quad (10)$$

where $\boldsymbol{\theta}$ is the dual variable, $\rho > 0$ is the augmented Lagrangian parameter, and $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \mathbf{x}_2^H \mathbf{x}_1$. For the optimization problem in (9), ADMM consists of the following iterations,

- **update-x:** $\mathbf{x}^{n+1} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \operatorname{Re}(\langle \boldsymbol{\theta}^n, \mathbf{x} - \mathbf{z}^n \rangle) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{z}^n\|_2^2$.
- **update-z:** $\mathbf{z}^{n+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \lambda_g \phi_g(\|\mathbf{z}\|; p) + \lambda_e \phi_e(\|\mathbf{z}\|; q) + \operatorname{Re}(\langle \boldsymbol{\theta}^n, \mathbf{x}^{n+1} - \mathbf{z} \rangle) + \frac{\rho}{2} \|\mathbf{x}^{n+1} - \mathbf{z}\|_2^2$.
- **update-dual variable:** $\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n + \rho(\mathbf{x}^{n+1} - \mathbf{z}^{n+1})$.

Deriving a closed form expressions for update-x is straightforward, $\mathbf{x}^{n+1} = \mathbf{A}_0(\rho \mathbf{z}^n - \boldsymbol{\theta}^n) + \mathbf{x}_0$, where $\mathbf{A}_0 = (\rho \mathbf{I} + \mathbf{A}^H \mathbf{A})^{-1}$ and $\mathbf{x}_0 = \mathbf{A}_0 \mathbf{A}^H \mathbf{y}$. If we pull the linear terms into the quadratic ones in the objective function of update-z and ignoring additive terms, independent of \mathbf{z} , then

we can express this step as

$$\mathbf{z}^{n+1} = \underset{\mathbf{z} \in \mathbb{C}^N}{\operatorname{argmin}} \sum_{i=1}^{N_g} \left(\frac{1}{2} \|\mathbf{x}_i^{n+1} + \boldsymbol{\theta}_{\rho i}^n - \mathbf{z}_i\|_2^2 + \lambda_{\rho g} f_H(\|\mathbf{z}_i\|; p) + \lambda_{\rho e} \sum_{j=1}^N f_H(\|[\mathbf{z}_i]_j\|; q) \right), \quad (11)$$

where $\lambda_{\rho g} = \lambda_g/\rho$, $\lambda_{\rho e} = \lambda_e/\rho$, $\mathbf{x}_i^{n+1} = \mathbf{W}_i \mathbf{x}^{n+1}$, $\boldsymbol{\theta}_{\rho i}^n = \mathbf{W}_i \boldsymbol{\theta}^{n+1}/\rho$, and $\mathbf{z}_i = \mathbf{W}_i \mathbf{z}$. We can solve (11) by solving for each group \mathbf{z}_i^{n+1} in parallel. To this end, we introduce the following notation and lemma. The vector $\mathbf{z} \in \mathbb{C}^N$ can be written as $\mathbf{z} = |\mathbf{z}| \odot \Phi(\mathbf{z})$, where the n th element of $\Phi(\mathbf{z})$ is $\exp(j\varphi([\mathbf{z}]_n))$, and $\varphi(z)$ is the phase of z , i.e., $z = |z| \exp(j\varphi(z))$.

Lemma 2. For any $\mathbf{c} \in \mathbb{C}^N$

$$\underset{\mathbf{z} \in \mathbb{C}^N}{\operatorname{argmin}} \|\mathbf{c} - \mathbf{z}\|_2^2 = \Phi(\mathbf{c}) \odot \underset{|\mathbf{z}| \in \mathbb{R}^N}{\operatorname{argmin}} \|\mathbf{c} - |\mathbf{z}|\|_2^2.$$

Proof : The function $J(\mathbf{z}) = \|\mathbf{c} - \mathbf{z}\|_2^2 = \|\mathbf{c}\|_2^2 + \|\mathbf{z}\|_2^2 - 2\Re\{\mathbf{c}^H \mathbf{z}\} = \|\mathbf{c}\|_2^2 + \|\mathbf{z}\|_2^2 - 2\Re\{\mathbf{c}^H \mathbf{z}\}$ is minimized when $\Re\{\mathbf{c}^H \mathbf{z}\}$ is maximized. Now,

$$\begin{aligned} \Re\{\mathbf{c}^H \mathbf{z}\} &= \sum_{n=1}^N |[c]_n| |[z]_n| \cos(\varphi([c]_n) - \varphi([z]_n)) \\ &\leq \sum_{n=1}^N |[c]_n| |[z]_n| = |\mathbf{c}|^T |\mathbf{z}| \end{aligned}$$

with equality if and only if $\Phi(\mathbf{z}) = \Phi(\mathbf{c})$, which in turn implies that $\|\mathbf{c} - \mathbf{z}\|_2^2 = \|\mathbf{c} - |\mathbf{z}|\|_2^2$. Hence,

$$\begin{aligned} \underset{|\mathbf{z}| \odot \Phi(\mathbf{z}) \in \mathbb{C}^N}{\operatorname{argmin}} \|\mathbf{c} - \mathbf{z}\|_2^2 &= \underset{|\mathbf{z}| \odot \Phi(\mathbf{c}) \in \mathbb{C}^N}{\operatorname{argmin}} \|\mathbf{c} - |\mathbf{z}|\|_2^2 \\ &= \Phi(\mathbf{c}) \odot \underset{|\mathbf{z}| \in \mathbb{R}^N}{\operatorname{argmin}} \|\mathbf{c} - |\mathbf{z}|\|_2^2 \end{aligned}$$

and the lemma follows. \square

Since the two last terms in (11) are independent of the phase of \mathbf{z}_i , we use Lemma 2 to write the i th group problem in (11) as

$$\begin{aligned} \mathbf{z}_i^{n+1} &= \Phi(\mathbf{x}_i^{n+1} + \boldsymbol{\theta}_{\rho i}^n) \odot \underset{\mathbf{a} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{x}_i^{n+1} + \boldsymbol{\theta}_{\rho i}^n - \mathbf{a}_i\|_2^2 \\ &\quad + \lambda_{\rho g} f_H(\|\mathbf{a}_i\|; p) + \lambda_{\rho e} \sum_j f_H(\|[\mathbf{a}_i]_j\|; q), \end{aligned} \quad (12)$$

where $\mathbf{a}_i = \mathbf{W}_i \mathbf{a}$. As is clear, the optimization problem in (12) involves only real-valued vectors. In following, we state a theorem that enables us to find the optimal solution of optimization problem in (12) using proximity operator of function f_H .

Theorem 2. Consider a function f_H that satisfies in Equation (4). Then, the solution of optimization problem,

$$\hat{\mathbf{a}} = \underset{\mathbf{a} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{b} - \mathbf{a}\|_2^2 + \lambda_{\rho g} g(\mathbf{a}; p) + \lambda_{\rho e} E(\mathbf{a}; q),$$

where $g(\mathbf{a}; p) = f_H(\|\mathbf{a}\|_2; p)$ and $E(\mathbf{a}; q) = \sum_j f_H(\|[\mathbf{a}]_j\|; q)$, is given as

$$\hat{\mathbf{a}} = P_{\lambda_{\rho g}, g}(P_{\lambda_{\rho e}, E}(\mathbf{b})). \quad (13)$$

Furthermore, the proximity operators of functions g and E can be evaluated as follows, $P_{\lambda_{\rho g}, g}(\mathbf{b}) = \frac{P_{\lambda_g, f_H}(\|\mathbf{b}\|_2)}{\|\mathbf{b}\|_2} \mathbf{b}$ and $[P_{\lambda_{\rho e}, E}(\mathbf{b})]_j = P_{\lambda_{\rho e}, f_H}(b[j])$.

The proof is given in Appendix B. Thus we can determine a closed form solution for the optimization problem in the \mathbf{z} -update step in Equation (12) using the proximity operators of the univariate functions f_H . This update rule is a direct consequence of Theorem 2. Therefore, update- \mathbf{z} can be performed as follows,

$$\mathbf{z}_i^{n+1} = P_{\lambda_{\rho g}, g}(P_{\lambda_{\rho e}, E}(|\mathbf{x}_i^{n+1} + \boldsymbol{\theta}_{\rho i}^n|)) \odot \Phi(\mathbf{x}_i^{n+1} + \boldsymbol{\theta}_{\rho i}^n).$$

Remark 7. Note that for $p = q = 1$, Theorem 2 includes the recovery algorithm for joint sparse problem with l_1 regularizer i.e. sparse-group Lasso [7], as a particular case.

In following, we state our proposed algorithm to solve the optimization problem given in Eq. (9).

Proposed ADMM Algorithm for Sparse Approximation

Input: Signal vector \mathbf{y} , matrix \mathbf{A} , ρ , and λ_g, λ_e .

Initialize: $\mathbf{z}^0 = \boldsymbol{\theta}^0 = \mathbf{0}$.

Pre-computation: $\mathbf{A}_0 = (\rho \mathbf{I} + \mathbf{A}^H \mathbf{A})^{-1}$ and $\mathbf{x}^0 = \mathbf{A}_0 \mathbf{A}^H \mathbf{y}$.

For $n = 0$ to END

$\mathbf{x}^{n+1} = \mathbf{A}_0 (\rho \mathbf{z}^n - \boldsymbol{\theta}^n) + \mathbf{x}^0$

$\mathbf{z}_i^{n+1} = P_{\lambda_{\rho g}, g}(P_{\lambda_{\rho e}, F}(|\mathbf{x}_i^{n+1} + \boldsymbol{\theta}_{\rho i}^n|)) \odot \Phi(\mathbf{x}_i^{n+1} + \boldsymbol{\theta}_{\rho i}^n)$,
for $i = 1, \dots, N_g$

$\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n + \rho (\mathbf{x}^{n+1} - \mathbf{z}^{n+1})$.

End

Output: Vector \mathbf{x} .

IV. NUMERICAL RESULTS

Here, we consider the application of our proposed method for estimation of a vehicle-to-vehicle (V2V) channel. To simulate the V2V channel, the statistical parameter values for different scatterers are selected to be those reported in Table I in [14], which are determined from experiment and measurement data analysis. Furthermore, the channel is defined as that in [15]. The pilot data matrix, \mathbf{A} , is a $M \times N$ matrix where here we consider $M = 2040$, $N = 4M$, more details about the structure of \mathbf{A} can be found in [15]. The pilot symbol are drawn from a zero-mean Gaussian distribution. The length of horizon is considered from -500 m to 500 m, the road width is 24 m, and the length of the diffuse strip around the road 18 m. It is assumed that the number of mobile discrete (MD) scatterers is $N_{MD} = 6$, and their speed are chosen randomly from the interval $[80, 140]$ km/h; we have $N_{SD} = 4$ static discrete (SD) scatterers and $N_{DI} = 1000$ diffuse (DI) scatterers. The locations of the transmitter and receiver are chosen randomly in a geometry with distance 50 m from each other. The speed of the transmit and receive vehicles are assumed as $v_T = 100$ km/h and $v_R = 125$ km/h, respectively.

In Fig. 1, a typical structure of the V2V channel in the delay-Doppler domain is depicted [15]. It can be observed from Fig. 1, that most of the large coefficients in the delay-Doppler representation are concentrated in the area close to zero delay (Region R_1) or some constant Doppler value (Region R_2), thus we should consider our groupings such that in each group,

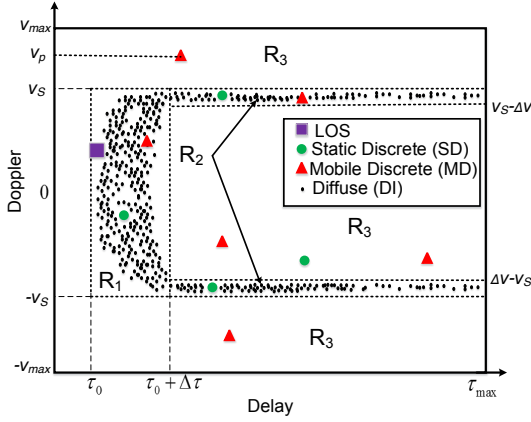


Fig. 1. Delay-Doppler domain representation of V2V channel. The delay-Doppler spreading of diffuse components is confined to a U-shape area.

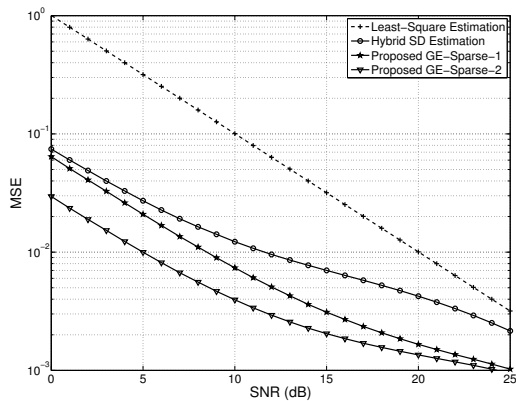


Fig. 2. Comparison of MSE v.s SNR for Least-Square (LS) method, Hybrid Sparse/Diffuse (SD) method, and proposed method with following conditions: $p = q = 1$ (GE-Sparse-1), and $p = q = 0.2$ (GE-Sparse-2).

either the all coefficients are large or all the coefficients are close to zero (group-wise sparsity).

First, we estimate the channel by computing the least-squares estimator, where we do not utilize any structural information about the channel coefficients. In the second step, we use the prior information about the channel geometric structure in the delay-Doppler domain, computed from the LS solution using an energy detection algorithm (for more details see, *e.g.*, [20]). We can consider a group as a collection of coefficients with similar Doppler values and similar magnitudes. The group size in our simulation was common for all groups and equal to 100, so $N_g = \frac{N}{100}$. We consider two sets of regularizers, $p = q = 1$ (soft-thresholding) and a $p = q = 0.2$ and set $\lambda_e = \lambda_g/5 = 5$. These values were found by trial and error. Note that we do not optimize over the values of p and q , this problem is beyond the scope of this paper, but we show by simulation that we get better results with non-convex regularizers, *i.e.* $0 < p, q < 1$ than that obtained by convex regularizers. Furthermore, our approach can be used for other non-convex regularizers like SCAD [19].

Fig. 2 depicts the result for Monte Carlo simulations for the estimation of the channel using the proposed estimator for the different cases specified above, and our prior method in [15] based on a hybrid sparse/diffuse model [17],

[18], and the least-squares estimator. The MSE is defined as $E\{(\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2)/\|\mathbf{x}\|_2^2\}$, where $\hat{\mathbf{x}}$ is the estimated channel vector. We added a white Gaussian noise vector \mathbf{n} , whose variance was adjusted to achieve a prescribed receive signal-to-noise ratio (SNR) defined as $\text{SNR} = E\{\|\mathbf{y} - \mathbf{n}\|_2^2\}/E\{\|\mathbf{n}\|_2^2\}$. Results show that our joint sparse signal estimation algorithm using the regularization term for $p = q = 0.2$ results in better performance, almost 2dB to 3dB compared to $p = q = 1$ at low SNR. It is clear that there is at least a 10dB to 15dB performance enhancement compared to the least-squares estimator, because we consider the structural information of the channel in different regions relative to other prior information.

V. CONCLUSIONS

In this work, we proposed a method to design the regularizer given that Moreau envelope of the regularizer is known. Using this algorithm, we designed an adaptive Huber regularizer function to induce sparsity in the joint sparse signal estimation problem. Furthermore, we showed that the optimal solution for the joint sparse signal estimation can be evaluated using a nested proximity operations. Our result showed that within a group, joint sparsity can be induced by first apply the element proximity operator and then the group proximity operator. Our simulation results reveal that exploiting the joint sparsity structure with non-convex regularizers yields improvements compared to the convex regularizers. In particular, for the V2V channel estimation problem, we showed using simulation that structured estimators yield significantly improved channel estimates over unstructured approaches.

APPENDIX A

PROOF OF THEOREM 1

Part 1), using the definition of convex conjugate of a function, we know that $\eta_1^*(\mathbf{a}) = \sup_{\mathbf{b}} (\mathbf{a}^T \mathbf{b} - \frac{1}{2} \|\mathbf{b}\|_2^2 - \lambda F(\mathbf{b}))$. Now, we first assume that $\eta_2 = \eta_1^*$, namely $\sup_{\mathbf{b}} (\mathbf{a}^T \mathbf{b} - \frac{1}{2} \|\mathbf{b}\|_2^2 - \lambda F(\mathbf{b})) = \frac{1}{2} \|\mathbf{a}\|_2^2 - \lambda G(\mathbf{a})$, which can be rewritten as, $G(\mathbf{a}) = \frac{1}{2\lambda} \|\mathbf{a}\|_2^2 - \sup_{\mathbf{b}} (\frac{1}{\lambda} \mathbf{a}^T \mathbf{b} - \frac{1}{2\lambda} \|\mathbf{b}\|_2^2 - F(\mathbf{b}))$. If we negate the objective in the sup operator, we can change it to inf operator. Finally we use the equality $\frac{1}{2\lambda} \|\mathbf{a} - \mathbf{b}\|_2^2 = \frac{1}{2\lambda} \|\mathbf{a}\|_2^2 - \frac{1}{\lambda} \mathbf{a}^T \mathbf{b} + \frac{1}{2\lambda} \|\mathbf{b}\|_2^2$. Since all of our steps are if and only if, the reverse direction is also true and the proof is completed. Part 2), is a standard result of convex duality (see [12], page 476, Proposition 11.3). Note that it is easy to show that $\eta_2(\mathbf{a})$ is continuous and strictly convex using its connection to the Moreau envelope, which results in the singleton value for the sub-gradient, namely $\partial \eta_2(\mathbf{a}) = \{\nabla \eta_2(\mathbf{a})\}$.

APPENDIX B

PROOF OF THEOREM 2

We first prove a lemma, which is useful in the proof of the main theorem.

Lemma 3. Consider that $g(\mathbf{a}; p) = f_H(\|\mathbf{a}\|_2; p)$. Then,

$$P_{\lambda, g}(\mathbf{b}) = \frac{P_{\lambda, f_H}(\|\mathbf{b}\|_2)}{\|\mathbf{b}\|_2} \mathbf{b} \text{ for } \|\mathbf{b}\|_2 > 0$$

and $P_{\lambda, g}(\mathbf{0}) = \mathbf{0}$.

Proof: Let us start with the definition of proximity operator for $g(\mathbf{a}; p)$, namely $P_{\lambda, g}(\mathbf{b}) = \underset{\mathbf{a}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{b} - \mathbf{a}\|_2^2 + \lambda g(\mathbf{a}; p) \right\}$. Since $g(\mathbf{a}; p) = f_H(\|\mathbf{a}\|_2; p)$ is absolute and non-decreasing, the minimizer will have the form $P_{\lambda, g}(\mathbf{b}) = \mathbf{a} = r\mathbf{b}$ for some $r \in [0, 1]$. Thus, the optimization problem becomes, $\underset{r \in [0, 1]}{\operatorname{argmin}} \lambda f_H(r\|\mathbf{b}\|_2; p) + \frac{1}{2} \|\mathbf{b} - r\mathbf{b}\|_2^2$. By considering $t = r\|\mathbf{b}\|_2$, it can be rewritten as,

$$\underset{t \in [0, \|\mathbf{b}\|_2]}{\operatorname{argmin}} \lambda f_H(t; p) + \frac{1}{2} \|\mathbf{b}\|_2 - t \|^2 = P_{\lambda, f_H}(\|\mathbf{b}\|_2; p).$$

The last equality is due to the definition of the proximity operator. Therefore,

$$P_{\lambda, g}(\mathbf{b}) = \frac{P_{\lambda, f_H}(\|\mathbf{b}\|_2)}{\|\mathbf{b}\|_2} \mathbf{b} = \max \left\{ 0, 1 - \lambda^{2-p} \|\mathbf{b}\|_2^{p-2} \right\} \mathbf{b},$$

and the proof of the Lemma is completed.

Proof of Theorem 3: Let us define $J(\mathbf{a}) = \frac{1}{2} \|\mathbf{b} - \mathbf{a}\|_2^2 + \lambda_{\rho g} g(\mathbf{a}; p) + \lambda_{\rho e} E(\mathbf{a}; q)$. Assume $\mathbf{v} = P_{\lambda_{\rho e}, E}(\mathbf{b})$ and $\mathbf{u} = P_{\lambda_{\rho g}, g}(\mathbf{v})$. To prove the claim of Theorem 2, we just need to show that \mathbf{u} is the minimizer of $J(\mathbf{a})$. To prove this claim, we consider two cases: I: $\mathbf{u} \neq \mathbf{0}$, and II: $\mathbf{u} = \mathbf{0}$.

Case (I): $\mathbf{u} \neq \mathbf{0}$. Since $\mathbf{u} = P_{\lambda_{\rho g}, g}(\mathbf{v})$, by Lemma 3, it means that there is some $r > 0$ (i.e., $r = 1 - \lambda_{\rho g}^{2-p} \|\mathbf{v}\|_2^{p-2}$), and $\mathbf{u} = r\mathbf{v}$ such that $\mathbf{u} = \underset{\mathbf{a}}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\mathbf{a} - \mathbf{v}\|_2^2 + \lambda_{\rho g} f_H(\|\mathbf{a}\|_2; p) \right\}$. Therefore, \mathbf{u} satisfies first order optimality for this objective function, namely

$$\mathbf{u} - \mathbf{v} + \lambda_{\rho g} f'_H(\|\mathbf{u}\|_2; p) \frac{\mathbf{u}}{\|\mathbf{u}\|_2} = \mathbf{0}, \quad (14)$$

where $f'_H(\cdot; p)$ is the derivative of $f_H(\cdot; p)$. By definition of the proximity operator given in Eq. (1), we know that

$$P_{r\lambda_{\rho e}, E}(r\mathbf{b}; q) = \underset{\mathbf{a}}{\operatorname{argmin}} r\lambda_{\rho e} E(\mathbf{a}; q) + \frac{1}{2} \|\mathbf{a} - r\mathbf{b}\|_2^2. \quad (15)$$

Based on Lemma 2, we know that $P_{r\lambda_{\rho e}, E}(r\mathbf{b}; q) = rP_{\lambda_{\rho e}, E}(\mathbf{b}; q)$. Besides, $\mathbf{v} = P_{\lambda_{\rho e}, E}(\mathbf{b}; q)$, so $P_{r\lambda_{\rho e}, E}(r\mathbf{b}; q) = r\mathbf{v} = \mathbf{u}$. Therefore, $r\mathbf{v}$ is the minimizer of the objective function in (15), and by the first order optimality condition for the objective function in (15), we have $\mathbf{0} \in r\lambda_{\rho e} \partial E(\mathbf{u}; q) + \mathbf{u} - r\mathbf{b}$. If we substitute $r\mathbf{v} = \mathbf{u}$, then we have

$$\mathbf{0} \in \lambda_{\rho e} \partial E(\mathbf{u}; q) + \mathbf{v} - \mathbf{b}. \quad (16)$$

Summing Equations (14) and (16), we have

$$\mathbf{0} \in \lambda_{\rho e} \partial E(\mathbf{u}; q) + \lambda_{\rho g} f'_H(\|\mathbf{u}\|_2; p) \frac{\mathbf{u}}{\|\mathbf{u}\|_2} + \mathbf{u} - \mathbf{b},$$

which is the first order optimality of \mathbf{u} for the objective function $J(\mathbf{a})$. Thus the proof for the case (I) is completed.

Case (II): $\mathbf{u} = \mathbf{0}$ or $\|\mathbf{u}\|_2 = 0$. Since $\mathbf{0} = P_{\lambda_{\rho g}, g}(\mathbf{v})$, using Lemma 3, $\mathbf{v} = \mathbf{0}$ or $P_{\lambda_{\rho g}, f_H}(\|\mathbf{v}\|_2) = \mathbf{0}$. Therefore using Lemma 2, $\|\mathbf{v}\|_2 \leq \lambda_{\rho g}$. Here we need to show the first-order optimality condition of $\mathbf{u} = \mathbf{0}$ for the objective function $J(\mathbf{a})$, i.e.,

$$\begin{aligned} \mathbf{0} &\in \{ \lambda_{\rho e} \partial E(\mathbf{u}; q) + \lambda_{\rho g} f'_H(\|\mathbf{u}\|_2; p) \partial \|\mathbf{u}\|_2 + \mathbf{u} - \mathbf{b} \}_{\mathbf{u}=\mathbf{0}} \\ &= \lambda_{\rho e} \partial E(\mathbf{0}; q) + \lambda_{\rho g} \partial f_H(\|\mathbf{0}\|_2; p) \partial \|\mathbf{0}\|_2 - \mathbf{b}, \end{aligned}$$

Note that $\partial(\|\mathbf{0}\|) = \{ \mathbf{c} \in \mathbb{R}^N, \|\mathbf{c}\|_2 \leq 1 \}$ and $\partial f_H(\|\mathbf{0}\|_2; p) \in [-1, +1]$. Thus, the first order optimality condition is equivalent to show that there exist a $\boldsymbol{\theta}_1 \in [-1, 1]^N$ (equivalent to the term $\partial E(\mathbf{0}; q)$) and a $\boldsymbol{\theta}_2$ with $\|\boldsymbol{\theta}_2\|_2 \leq 1$ (equivalent to the term $\partial f_H(\|\mathbf{0}\|_2; p) \partial \|\mathbf{0}\|_2$) such that $\mathbf{b} = \lambda_{\rho g} \boldsymbol{\theta}_1 + \lambda_{\rho e} \boldsymbol{\theta}_2$. Since $\mathbf{v} = P_{\lambda_{\rho e}, E}(\mathbf{b}; q)$, using the first-order optimality of \mathbf{v} for objective function of proximity operator, we have $\mathbf{0} \in \lambda_{\rho e} \partial E(\mathbf{v}; q) + \mathbf{v} - \mathbf{b}$. Since $\|\mathbf{v}\|_2 \leq \lambda_{\rho g}$, there exist $\boldsymbol{\theta}_1 \in [-1, 1]^N$ and $\boldsymbol{\theta}_2 = \frac{\mathbf{v}}{\lambda_{\rho g}}$ such that $\mathbf{b} = \lambda_{\rho e} \boldsymbol{\theta}_1 + \lambda_{\rho g} \boldsymbol{\theta}_2$, which completes the proof.

REFERENCES

- [1] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.
- [2] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statist. Soc Series B*, pp. 267–288, 1996.
- [3] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statist. Soc Series B*, vol. 68, no. 1, pp. 49–67, 2001.
- [4] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, 2009.
- [5] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, 2010.
- [6] M. Stojnic, " ℓ_2/ℓ_1 -optimization in block-sparse compressed sensing and its strong thresholds," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 350–357, 2010.
- [7] J. Friedman, T. Hastie and R. Tibshirani, "A note on the group lasso and a sparse group lasso," *Technical report*, Department of Statistics, Stanford University, 2010.
- [8] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. Eldar, "C-HiLasso: A collaborative hierarchical sparse modeling framework," *IEEE Trans. Signal Process.*, vol. 59, pp. 4183–4198, Sept 2011.
- [9] F. Bach, et al., "Optimization with sparsity-inducing penalties," *Foundations and Trends in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2012.
- [10] R. Chartrand and B. Wohlberg, "A nonconvex ADMM algorithm for group sparsity with sparse groups," in *Proceedings IEEE Int. Conf. Acoustics, Speech and Signal Process.*, pp. 6009–6013, IEEE, 2013.
- [11] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Letter*, vol. 14, no. 10, pp. 707–710, 2007.
- [12] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Berlin: Springer-Verlag, 2004.
- [13] J. Yang, et al. "A fast algorithm for edge-preserving variational multi-channel image restoration." *SIAM J. on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.
- [14] J. Karedal, et al., "A geometry-based stochastic MIMO model for vehicle-to-vehicle communications," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3646–3657, 2009.
- [15] S. Beygi, E. G. Ström, and U. Mitra, "Geometry-based stochastic modeling and estimation of vehicle to vehicle channels," *Proceedings IEEE Int. Conf. Acoustics, Speech, Signal Process.*, (ICASSP), 2014.
- [16] S. Boyd, et al., "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.
- [17] N. Michelusi, U. Mitra, A. Molisch, and M. Zorzi, "UWB sparse/diffuse channels, part I: Channel models and bayesian estimators," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5307–5319, 2012.
- [18] N. Michelusi, U. Mitra, A. Molisch, and M. Zorzi, "UWB sparse/diffuse channels, part II: Estimator analysis and practical channels," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5320–5333, 2012.
- [19] F. Jianqing and L. Runze, "Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties," *J. American Statist. Association*, no. 96, pp. 1348–1360, 2001.
- [20] F. Marvasti, et al., "A unified approach to sparse signal processing," *EURASIP J. Adv. Sig. Proc.*, p.p. 1–44, 2012.