To Convexify or Not? Regression with Clustering Penalties on Graphs

Marwa El Halabi, Luca Baldassarre, and Volkan Cevher
Laboratory for Information and Inference Systems (LIONS), EPFL

***Abstract*—We consider minimization problems that are compositions of convex functions of a vector $\mathbf{x} \in \mathbb{R}^N$ with submodular set functions of its support (i.e., indices of the non-zero coefficients of $\mathbf{x}$). Such problems are in general difficult for large N due to their combinatorial nature. In this setting, existing approaches rely on "convexifications" of the submodular set function based on the Lovász extension for tractable approximations. In this paper, we first demonstrate that such convexifications can fundamentally change the nature of the underlying submodular regularization. We then provide a majorization-minimization framework for the minimization of such composite objectives. For concreteness, we use the Ising model to motivate a submodular regularizer, establish the total variation semi-norm as its Lovász extension, and numerically illustrate our new optimization framework.**

I. Introduction

We consider the following optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^N} f(\mathbf{x}) + \lambda R(\mathrm{supp}(\mathbf{x})), \tag{1}$$

where f is a closed, convex function, R is a set function, $\mathrm{supp}(\mathbf{x}) = \{i : x_i \neq 0\}$ is the support function, and $\lambda \geq 0$ is the regularization parameter. Formulation and analysis of (1) are important in several applications from compressive sensing to data-mining, and from medical imaging to array signal processing.

Finding even local optimal solutions to (1) is generally difficult when N is large, since the problem includes a combinatorial component. To be able to proceed further, we assume that f has L-Lipschitz continuous gradient (i.e, $\forall\ \mathbf{x}, \mathbf{y} \in \mathbb{R}^N, \|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|$) and that R is a *submodular* set function (cf., Sect. II).

While R leads to computational difficulties, its presence is key in many problems revolving around *model-based* sparsity [1]. As an example, consider a compressive sensing scenario where we observe compressive samples $\mathbf{y} \in \mathbb{R}^m$ of a "clustered" sparse vector $\mathbf{x} \in \mathbb{R}^N$, through a dimensionality reducing matrix $\mathbf{A}$ [2]. In this case, $f(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2/2$. Then, the clustering model can be naturally encoded on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ via a submodular function based on Ising model

$$R_{\text{ISING}}(\mathcal{S}) = \frac{1}{2}\left(|\mathcal{E}| \ - \sum_{(i,j)\in\mathcal{E}} s_i s_j\right), \tag{2}$$

where $\mathbf{s} \in \mathbb{R}^{\mathbf{N}}$ is an indicator vector for a set $\mathcal{S}$ such that $s_i = 1$ if $i \in \mathcal{S}$ and $s_i = -1$, otherwise. There are many other structured models that can be captured in this fashion: cf., [3] for a review.

The prevailing approach in circumventing the difficulty of solving (1) is to convexify the set function so that the overall problem is convex. When R is submodular and monotone (i.e., $\forall \mathcal{S} \subseteq \mathcal{T}, R(\mathcal{S}) \leq R(\mathcal{T})$), this convexification is achieved through the Lovász extension [4]. While convexity by itself does not necessarily imply efficient optimization, the convexification of monotone submodular functions leads to tractable convex optimization. In fact, [5] shows that the proximal operator of this convexification is equivalent to solving an unconstrained submodular minimization problem, where the best provable complexity is $\mathcal{O}(N^5T + N^6)$ (T is the function evaluation complexity) [6]. In this setting, the minimum norm-point algorithm [4], which does not have a worst case complexity, usually scales as $\mathcal{O}(N^2)$ in practice.

Theoretical justifications of the convexification approach are often based on the success story of the ℓ_0 and ℓ_1 equivalence [7]. For instance, we can use $R(\mathcal{S}) = |\mathcal{S}|$ to penalize the sparsity of the solution, which leads to an NP-hard problem. However, the convex envelope of R on the unit ℓ_∞-ball in this case is the ℓ_1-norm of the vector $\mathbf{x}$, yielding the LASSO problem, whose solution quality is well-understood theoretically. As a result, the authors in [5] propose a structured regression framework based on the convexification of submodular monotone functions, which has been quite popular in the literature.

We illustrate that the ℓ_0-ℓ_1 type of equivalence does not necessarily hold for general submodular set functions. One must be very careful in stating the equivalence of solutions of the problem (1) to the one based on convexifications of R. Similarly, it is important to be aware of the fact that support consistency results [5] are typically given with respect to the convexification rather than to the discrete structured sparsity model.

To clarify the subtleties, we first establish that the convex envelope of the Ising penalty (2) is the zero function. The Lovász extension is then the second most natural convexification; we provide a novel elementary proof that the Lovász extension of the Ising penalty is the anisotropic total variation semi-norm. We then give an example where none of the solutions of (1) and its convexification coincide for any non-zero regularization parameter λ. To tackle (1), we provide an efficient majorization-minimization scheme that is guaranteed to converge. Numerical results show that solutions returned

This work was supported in part by the European Commission under Grant MIRG-268398, ERC Future Proof, SNF 200021-132548, SNF 200021-146750 and SNF CRSII2-147633

978-1-4673-3146-3/13/$31.00 ©2013IEEE

by this algorithm upon convergence can have significant recovery benefits as compared to the convex solutions.

II. Preliminaries

We denote scalars by lowercase letters, e.g., λ, vectors by lowercase boldface letters, e.g., $\mathbf{x}$, matrices by boldface uppercase letters, e.g., $\mathbf{A}$, and sets by uppercase script letters, e.g., $\mathcal{V}$. We denote the ground set of N indices by $\mathcal{V} = \{1, \ldots, N\}$.

a) Submodularity: A set function $F : 2^{\mathcal{V}} \to \mathbb{R}$ is *submodular* iff $R(\mathcal{S}) + R(\mathcal{T}) \geq R(\mathcal{S} \cup \mathcal{T}) + R(\mathcal{S} \cap \mathcal{T})$ for all $\mathcal{S}, \mathcal{T} \subseteq \mathcal{V}$.

b) Lovász extension: Given a submodular function R such that $R(\emptyset) = 0$, we define its Lovász extension r as follows. Given $\mathbf{x} \in \mathbb{R}^N$, we sort its components in decreasing order $x_{j_1} \geq \cdots \geq x_{j_N}$ and define $r(\mathbf{x}) =$

$$\sum_{k=1}^{N-1} R(\{j_1, \ldots, j_k\})(x_{j_k} - x_{j_{k+1}}) + R(\mathcal{V})x_{j_N} \quad (3)$$

We can treat R as a function on the boolean hypercube $\{0,1\}^N$, and r forms its *convex closure* on $[0,1]^N$ [8].

c) Convex biconjugate of the Ising model: In the sequel, we focus on the Ising model example, as defined in (2). [5] shows that the convex biconjugate [9] of $R(\text{supp}(\mathbf{x}))$ on the unit ℓ_∞-ball, for R any submodular function, is $R^{**}(\mathbf{x}) = \min_{\delta \in [0,1]^N, \delta \geq |\mathbf{x}|} r(\delta)$. Thus, when R is non-decreasing, its convex envelope on the unit ℓ_∞-ball is $r(|\mathbf{x}|)$, but in the Ising penalty case, which is not monotonic, the biconjugate is the zero function. As a result, we consider another convexification: the Lovász extension. This convex extension is tight in the sense that it forms the convex closure, i.e., the largest convex real function on $[0,1]^N$ that always lower bounds R.

d) The Lovász extension of the Ising model: Let $G(\mathcal{V}, \mathcal{E})$ be a graph with one node for each of the support indices and whose edge set $\mathcal{E}$ contains the edges connecting neighboring variables. Let $\delta(\mathcal{S}) = \{(i,j) \in \mathcal{E} : i \in \mathcal{S}, j \notin \mathcal{S}\}$ be the cut-set induced by $\mathcal{S}$. Since $s_i s_j = -1$ if $(i,j) \in \delta(\mathcal{S})$ and $s_i s_j = 1$ otherwise, we have that $R_{\text{ISING}}(\mathcal{S}) = |\delta(\mathcal{S})|$. R_{ISING} favors sets whose elements are clustered on the given graph. Cut functions are submodular [4], therefore R_{ISING} is submodular.

Proposition 1. *The Lovász extension of R_{ISING} is the anisotropic discrete Total Variation semi-norm $\|\mathbf{x}\|_{TV} = \sum_{(i,j)\in\mathcal{E}} |x_i - x_j|$.*

Proof: Let $n = |\mathcal{E}|$ and $\forall k \in \mathcal{V}$ and $\ell \in [1, n]$, let $\sigma_k(e_\ell) = 1$ if $e_\ell \in \mathcal{E}$ is cut by $\{j_1, \ldots, j_k\}$ and $\sigma_k(e_\ell) = 0$ otherwise, then:

$$\begin{aligned} r_{\text{ISING}}(\mathbf{x}) &= \sum_{k=1}^{N-1} |\delta(\{j_1, \ldots, j_k\})|(x_{j_k} - x_{j_{k+1}}) \\ &= \sum_{k=1}^{N-1} \sum_{\ell=1}^{n} \sigma_k(e_\ell)(x_{j_k} - x_{j_{k+1}}) \\ &= \sum_{\ell=1}^{n} \sum_{k \in [s,t]} (x_{j_k} - x_{j_{k+1}}) := \|\mathbf{x}\|_{TV} \ , \end{aligned}$$

where the first equality follows from the definition of the Lovász extension (3) since $R_{\text{ISING}}(\mathcal{V}) = 0$, and the third equality holds since for any $\ell \in [1, n]$, there exists a range of indices $[s, t]$, such that e_ℓ is cut by $\{j_1, \ldots, j_k\}$ for $k \in [s, t]$. Indeed, let $e_\ell = (i, j)$, then $j_s = i$ and $j_{t+1} = j$, $\sigma_k(e_\ell) = 1$ for $k \in [s, t]$ and 0 otherwise. ■

III. To convexify or not to convexify?

We now elucidate how convexification can radically alter the solutions of (1). For concreteness, we use R_{ISING} as defined in (2). In this case, since $R_{\text{ISING}}(\mathcal{V}) = 0$, the *globally optimal* minimizer of (1) for $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$ for any λ is simply the least squares solution (assuming it has full support) $\hat{\mathbf{x}} = \mathbf{A}^\dagger \mathbf{y}$, where $\mathbf{A}^\dagger$ is the pseudo-inverse. Now, instead we substitute the Lovász extension of R_{ISING} and consider the convex problem

$$\min_{x \in \mathbb{R}^p} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_c \|\mathbf{x}\|_{TV}, \quad (4)$$

for a regularization parameter λ_c. It is clear that unless λ_c is zero, the solutions of (1) and (4) are *never* equal.

Such considerations are also discussed in [10], which show that while many group-based discrete structured sparsity models have convex relaxations (many of which are exactly the Lovász extensions), their relaxed solutions do not necessarily correspond to the solution we seek. The work [10] goes one step further to show that if we could use convex relaxations in these problems, then we can have polynomial time algorithms for the weighted maximum coverage problem, which is NP-hard.

Furthermore, while the Lovász extension (3) is defined over the entire real domain, it is tight only on the unit hypercube and most penalties $R(\text{supp}(\mathbf{x}))$ are not constrained there. Also note that $R(\text{supp}(\mathbf{x}))$ is symmetric around the origin, while the Lovász extension is not, which makes it vulnerable to sign flip errors. On the other hand, composing the Lovász extension with the absolute value is symmetric, but the resulting function is only guaranteed to be convex in the case of monotonic submodular functions. For example, $\||x|\|_{TV}$ is not convex. In the numerical experiments, we exploit this weakness to show that the TV norm has poor performance when we perturb the signal with random sign flips.

IV. Our optimization framework

We propose an iterative majorization-minimization scheme for solving (1) in Algorithm 1. Given its Lipschitz constant L, we have the following bound on $f(\mathbf{x})$:

$$f(\mathbf{x}) \leq f(\mathbf{x}') + \langle \nabla f(\mathbf{x}'), \mathbf{x} - \mathbf{x}' \rangle + \frac{L}{2}\|\mathbf{x} - \mathbf{x}'\|_2^2 := Q(\mathbf{x}, \mathbf{x}') \quad (5)$$

for all $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^N$. With this majorizer, we obtain an easy-to-deal-with convex quadratic upper bound on $f(\mathbf{x})$.

It turns out that this bound corresponds to a *modular* set function M, which satisfies the submodularity definition in Section II with *equality* for all sets $\mathcal{S}$ and $\mathcal{T}$. To see this, we compute the optimal minimizer of $Q(\mathbf{x}, \mathbf{x}')$

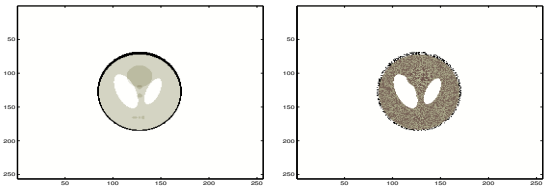

Fig. 1: *Shepp-Logan phantom: Original (Left) and* Dirty *(with randomized signs)*

for any given support $\mathcal{S}$. Simple calculus shows that the minimizer is given by $\widehat{\mathbf{x}}_{\mathcal{S}^c} = 0$ and

$$\widehat{\mathbf{x}}_{\mathcal{S}} = \mathbf{x}'_{\mathcal{S}} - \frac{1}{L}\nabla f(\mathbf{x}')_{\mathcal{S}}. \tag{6}$$

By substituting $\widehat{\mathbf{x}}_{\mathcal{S}}$ back into the upper bound, we obtain

$$\begin{aligned} f(\mathbf{x}) + R(\mathcal{S}) &\leq C - \frac{L}{2}\sum_{i\in\mathcal{S}}(x'_i - \frac{1}{L}\nabla f(\mathbf{x}')_i)^2 + R(\mathcal{S}) \\ &:= M(\mathcal{S}) + R(\mathcal{S}), \end{aligned} \tag{7}$$

where C is a constant. In (7), we deliberately kept the continuous variable $\mathbf{x}$ on the left hand side, while the right hand side only depends on $\mathcal{S}$. Note that this inequality relies on the explicit mapping $\mathcal{S} \to \widehat{\mathbf{x}}$ of any given set $\mathcal{S}$ to a continuous solution $\widehat{\mathbf{x}}$ through (6).

Minimization of the set majorizer in (7) is efficient, because the upper bound $M + R$ is submodular. While general submodular minimization algorithms may not scale gracefully with the problem size, some special cases of submodular functions have very efficient algorithms. For instance, if the submodular regularizer R is graph-representable, the optimization can be done efficiently via min s-t-cut algorithms. For the Ising model based on a lattice, we can solve the min s-t-cut in $\mathcal{O}(N^{1.5}\log(N)\log(U))$ (U is the max arc weight) [11]. Based on the above, we have the following guarantee

Proposition 2. *Algorithm 1 converges.*

Proof: Based on the definitions above, we have

$$\begin{aligned} f(\mathbf{x}^{i+1}) + R(S^{i+1}) &\leq f(\widehat{\mathbf{x}}_{\mathcal{S}^{i+1}}) + R(\mathcal{S}^{i+1}) && (8)\\ &\leq Q(\widehat{\mathbf{x}}_{\mathcal{S}^{i+1}}, \mathbf{x}^i) + R(\mathcal{S}^{i+1}) && (9)\\ &\leq Q(\widehat{\mathbf{x}}_{\mathcal{S}^{i}}, \mathbf{x}^i) + R(\mathcal{S}^{i}) && (10)\\ &\leq Q(\mathbf{x}^i, \mathbf{x}^i) + R(\mathcal{S}^i) && (11)\\ &= f(\mathbf{x}^i) + R(\mathcal{S}^i), && (12) \end{aligned}$$

where (8) holds because $\mathbf{x}^{i+1}$ minimizes $f(\mathbf{x})$ over the set $\mathcal{S}^{i+1}$, (9) follows from the bound (5), (10) holds because $\mathcal{S}^{i+1}$ minimizes $Q(\widehat{\mathbf{x}}_{\mathcal{S}}, \mathbf{x}^i) + R(\mathcal{S})$, while (11) is due to the fact that $\widehat{\mathbf{x}}_{\mathcal{S}^i}$ minimizes $Q(\mathbf{x}, \mathbf{x}^i)$ over the set $\mathcal{S}^i$. Hence, we have $f(\mathbf{x}^{i+1}) + R(\mathrm{supp}(\mathbf{x}^{i+1})) \leq f(\mathbf{x}^i) + R(\mathrm{supp}(\mathbf{x}^i))$. Therefore, at each iteration the objective value does not increase and it is always bounded below by 0, hence it converges. ∎

Once we find the optimal set $\mathcal{S}^*$, we minimize f over this set. Finally, as the iterations of the algorithm typically produce sparse iterates, we use block-coordinate descent to accelerate convergence in Algorithm 1.

Algorithm 1 Majorization-minimization algorithm

Input: $\mathbf{x}^0 \in \mathbb{R}^N$
while *not converged* **do**
  $\widehat{\mathbf{x}}_{\mathcal{S}} = \mathbf{x}^i_{\mathcal{S}} - \frac{1}{L}\nabla f(\mathbf{x}^i)_{\mathcal{S}}$
  $\mathcal{S}^{i+1} = \arg\min_{\mathcal{S}\in 2^{\mathcal{V}}} Q(\widehat{\mathbf{x}}_{\mathcal{S}}, \mathbf{x}^i) + R(\mathcal{S})$
  $\mathbf{x}^{i+1} = \arg\min_{\mathbf{x}:\mathrm{supp}(\mathbf{x})=\mathcal{S}^{i+1}} f(\mathbf{x})$
end while

TABLE I: *A summary of the regularizers used in experiments*

Model	Regularizer
Ising plus cardinality (IC)	$\lambda R_{\text{ISING}}(\mathrm{supp}(\mathbf{x})) + \tau\lvert\,\mathrm{supp}(\mathbf{x})\rvert$
Total Variation (TV)	$\lambda\lVert\mathbf{x}\rVert_{TV}$
Sparse TV (TV + L1)	$\lambda\lVert\mathbf{x}\rVert_{TV} + \tau\lVert\mathbf{x}\rVert_1$

V. Numerical experiments

We perform a compressive sensing experiment to higlight the differences between solutions of (1) and its convexifications. We take dimensionality reducing measurements of a structured sparse $\mathbf{x}$ via $\mathbf{A}$ and then seek to minimize $f(\mathbf{x}) = \frac{1}{2}\lVert\mathbf{y} - \mathbf{A}\mathbf{x}\rVert_2^2$ with a regularizer matched to the structure of $\mathbf{x}$. Our linear measurements $\mathbf{A}$ are randomly subsampled Fourier coefficients of $\mathbf{x}$, and hence, the Lipschitz constant of f is $L = 0.5$.

We compare the performance of three regularizers as summarized in Table I. Since the Ising model (2) by itself yields the least squares solution, we add a cardinality constraint to promote sparsity. The new regularizer is still submodular, and its convexification is the sparse total variation regularizer, i.e., the TV semi-norm plus the ℓ_1-norm. We also include the total variation regularizer alone to emphasize that its solutions are significantly different from regularization with the Ising model directly.

We consider the standard Shepp-Logan phantom image of size 256×256 pixels. The resulting image is sparse ($K = 8084$ non-zero pixels) with its coefficients forming constant value clusters, see Fig. 1(left). This image suits the TV models that encourage the signal coefficients to have constant values. We then randomly flip the signs of the coefficients, obtaining the *Dirty* phantom, Fig. 1(right). In this case, the TV models can enforce an incorrect structure as the true coefficient values are not smooth. However, the sign change does not affect the IC penalty, since it is agnostic to the coefficients values and only cares about whether they cluster.

We show recovery performance using $m = 1.5K$ samples, which is less than the theoretically minimum number of samples for ℓ_0 recovery (i.e., $m = 2K$). Hence, without the structured sparsity model, it is im-

TABLE II: *Relative Recovery Errors*

Model	Original phantom	*Dirty* phantom
TV	0.25	0.13
TV + L1	0.005	0.14
TV (debiased)	$\mathbf{8.8 * 10^{-12}}$	0.013
TV + L1 (debiased)	$\mathbf{8.8 * 10^{-12}}$	0.027
IC	$1.2 * 10^{-10}$	$\mathbf{1.4 * 10^{-11}}$

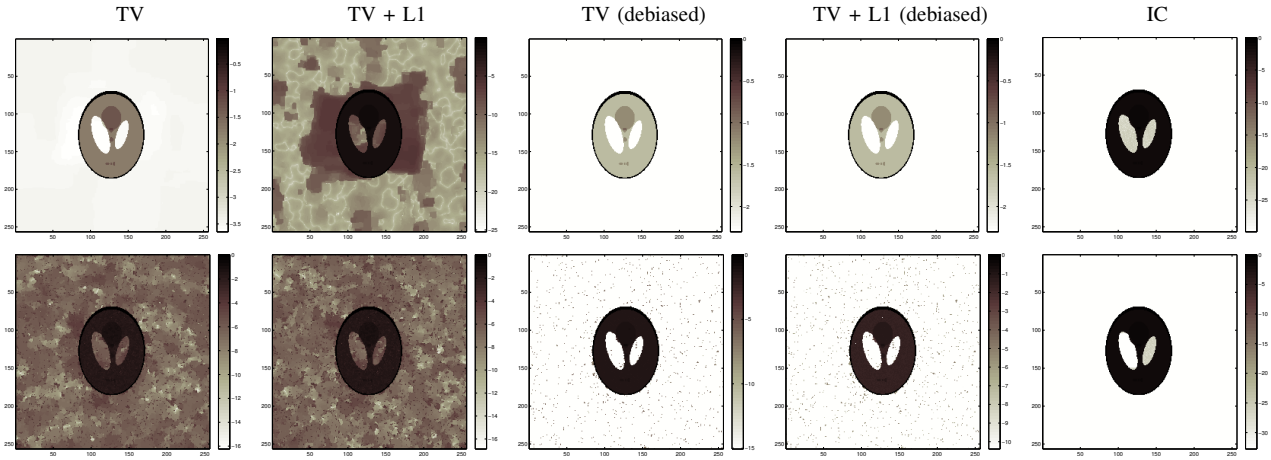

Fig. 2: *Clean Phantom recovery (Top): Debiased and IC perform well, TV and TV + L1 do not set the background exactly to zero. Dirty Phantom recovery (Bottom): TV does not perform well. Debiasing helps, but does not recover the correct support. The colorbar changes with each figure.*

possible to do tractable guaranteed recovery. We measure performance with the relative recovery error, $E = \|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2/\|\mathbf{x}\|_2^2$, where $\widehat{\mathbf{x}}$ is the estimated image and $\mathbf{x}$ the original one. The regularization parameters λ and τ have been tuned according to E to yield the best possible result for each model. The convex models might not yield exactly sparse solutions due to numerical issues. To level the playing field in favor of the convexifications, we adopt the debiasing heuristic of finding the support of the coefficients that are greater in magnitude than a threshold. We do this by visually inspecting the histogram of the solutions. The debiased estimate is then given by the least squares solution on the estimated support. Fig. 2 presents the recovered images for the original and *Dirty* phantom respectively, while Table II contains the relative recovery errors. In the figures, we use the log scale of the coefficients' absolute value to accentuate the errors.

The debiased estimates perform well in recovering the standard image, while the TV and TV + L1 penalties fail to set the background exactly to zero. Our method recovers the image with no need for debiasing since it correctly identifies the support of the signal. It is important to note that in the original phantom, the values of the pixels inside the eyes of the phantom are not exactly zero. Hence the Ising model actually perfectly recovers the entire support. As expected, on the Dirty Shepp-Logan, TV does not perform well. Debiasing helps in this case too, but it is not able to recover the correct support. We also note that our algorithm converges with 4 to 5 iterations, while TV and TV + L1 require thousands of iterations, because the proximal operator of TV cannot be computed in closed form.

VI. Conclusions

Model-based sparse models can make an enormous impact in diverse applications revolving around linear regression problems, such as simultaneously reducing the number of compressive samples in data acquisition while improving noise robustness. Unfortunately, the descriptions of many structured sparse models are inherently discrete, and lead to—seemingly—difficult combinatorial optimization problems. On this front, the prevailing approach to circumvent the computational difficulties has been to convexify the underlying discrete models. In this paper, we first show that convexifications can radically alter the underlying structure in regression problems, whose solutions should be interpreted with care. We then provide a fully discrete optimization framework that exploit the *structures* in the objectives. Our numerical results indicate that the direct discrete solutions can be significantly better than the solutions based on convex relaxations. The missing piece in our framework is the quantification of the solutions obtained by our proposed algorithm and how they relate to the global minimum of (1), which is an interesting research direction.

References

[1] R. Baraniuk, V. Cevher, M. Duarte, , and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[2] V. Cevher, M. F. Duarte, C. Hegde, and R. Baraniuk, "Sparse signal recovery using markov random fields," in *Advances in Neural Information Processing Systems*, 2008, pp. 257–264.

[3] R. G. Baraniuk, V. Cevher, and M. B. Wakin, "Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 959–971, 2010.

[4] S. Fujishige, *Submodular functions and optimization*. Elsevier Science, 2005, vol. 58.

[5] F. Bach, "Structured sparsity-inducing norms through submodular functions," *NIPS*, 2010.

[6] J. Orlin, "A faster strongly polynomial time algorithm for submodular function minimization," *Mathematical Programming*, vol. 118, no. 2, 2009.

[7] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[8] S. Dughmi, "Submodular functions: Extensions, distributions, and algorithms. a survey," *arXiv preprint arXiv:0912.0322*, 2009.

[9] J. Borwein and A. Lewis, *Convex analysis and nonlinear optimization: theory and examples*. Springer, 2006.

[10] L. Baldassarre, N. Bhan, V. Cevher, and A. Kyrillidis, "Group-sparse model selection: Hardness and relaxations," *arXiv preprint arXiv:1303.3207*, 2013.

[11] A. V. Goldberg and S. Rao, "Beyond the flow decomposition barrier," *J. ACM*, vol. 45, no. 5, pp. 783–797, Sep. 1998.