The k-Support Norm and Convex Envelopes of Cardinality and Rank

Anders Eriksson[1], Trung Thanh Pham[2], Tat-Jun Chin[2] and Ian Reid[2]
[1]School of Electrical Engineering and Computer Science, Queensland University of Technology.
[2]School of Computer Science, The University of Adelaide.

The use of sparsity in the field of Computer Vision can perhaps best be motivated by two main principles, computational efficiency and model simplicity. Selecting as few variables or features as possible will hopefully yield both a cheap and accurate model of the problem at hand.

Sparsity is typically obtained by regularizing the goodness of fit with the cardinality, (denoted card$(\cdot)$ or $||\cdot||_0$) of the variables. However, as this typically leads to non-convex optimization problems with high computational demands, the standard approach is to replace and relax these regularizers with convex surrogates for cardinality. The most popular such surrogate function is unquestionably the l_1-norm. The use of which is regularly justified by saying that the l_1-norm makes up the *convex envelope* of the cardinality function. However, as it was pointed out in [1], this is an argument that must be used with care. Correctly stated, we have that the l_1-norm is only the convex envelope of $||\cdot||_0$ on the bounded domain, $\{x \in \mathbb{R}^d \mid ||x||_\infty \leq 1\}$, the l_∞-norm unit ball.

It was argued in [1] that in certain instances it might be reasonable to not only expect that each individual entry of the entering variables be bounded, but to have bounds on their Euclidean norm as well. This led to the proposal of the k-support norm, a norm that was shown to provide the tightest convex relaxation of cardinality on the Euclidean-norm unit ball. It was also shown that this norm leads to improved learning guarantees as well as better algorithmic stability. However, as [1] did not address computational complexity issues, the proposed algorithm, employing an exhaustive search method, has proven painfully slow for even modest sized problems.

The authors of [2] presented an improved algorithm for computing the solution of k-support norm regularized optimization problems. They proposed the use of binary search as a replacement for certain subproblems in the original algorithm of [1]. Despite reporting significant speed-ups over [1], this method was still exhaustive in nature and could for larger problems and/or certain parameter choices still prove to be quite inefficient.

In this paper we attempt to make progress towards shedding further light on a number of different aspects of the k-support norm. In our opinion, there are three main contributions made here. Firstly, we present a slightly different derivation of the k-support norm with a stronger emphasis on the concept of convex envelopes. By doing so we hope to provide a different perspective to previous work and also make the connection to the rank operator on matrices perhaps more obvious. The k-support norm $||\cdot||_k^{sp}$ was defined in [1] as the gauge function with its unit ball coinciding with the convex hull of $\mathcal{C}_k^{(2)}$.

$$\mathcal{C}_k^{(2)} = \left\{ x \in \mathbb{R}^d \mid ||x||_0 \leq k,\ ||x||_2 \leq 1 \right\}, \tag{1}$$

In this paper we present an alternate way of deriving this convex envelope by finding the Fenchel biconjugate of the indicator function of $\mathcal{C}_k^{(2)}$. We obtain

$$f^*(y) = \sup_x y^T x - \chi_{\mathcal{C}_k^{(2)}}(x) = \sup_{\substack{||x||_0 \leq k \\ ||x||_2 \leq 1}} y^T x = \sqrt{\sum_{i=1}^{k} (|y|_i^{\downarrow})^2} = ||y||_k^{(2)}, \tag{2}$$

where $||\cdot||_k^{(p)} : \mathbb{R}^n \mapsto \mathbb{R}$ is the vector k-norm, defined as the l_p-norm of the k largest component values, in magnitude, of any vector in $\mathbb{R}^d$, also known as a symmetric gauge norm. Then its biconjugate becomes

$$f^{**}(x) = \sup_y x^T y - ||y||_k^{(2)} = \chi_{||x||_k^{(2)*} \leq 1}(x). \tag{3}$$

Secondly, we show that there is an equivalence between cardinality, the rank operator and the nuclear norm and k-support norm on domains shared between the elements of vectors and singular values of matrices. With a similar argument to that of above, obtaining convex envelopes of rank on different domains can then be attempted. On the domain of the Frobenius norm unit ball we have the following result.

The convex envelope of the indicator function $\chi_{\mathcal{D}_k^{(F)}}$ of the set $\mathcal{D}_k^{(F)} = \{X \in \mathbb{R}^{m \times n} \mid rank(X) \leq k,\ ||X||_F \leq 1\}$, becomes $f^{**}(X) = \chi_{||X||_{*k}^{sp}}$. Where $||X||_{*k}^{sp}$ denotes a *spectral* k-support norm. Let σ denote the vector of $\min(m,n)$ singular values of X, then the spectral k-support norm is given by $||X||_{*k}^{sp} = ||\sigma||_k^{sp}$. An interesting intuitive verification of this result might be reached with the following observation. On the domain of the matrix operator norm unit ball the convex envelope of rank is given by the nuclear norm, the l_1-norm of the singular values. Similarly, on the domain of the matrix Frobenius norm unit ball the convex envelope of rank is given by the spectral k-support norm, the k-support norm of the singular values. These relationships and more are summarised in table 1.

Concept:	cardinality	
Elements:	vectors, $x \in \mathbb{R}^d$	
Domain:	$\|\|x\|\|_\infty \leq 1$	$\|\|x\|\|_2 \leq 1$
Convex Surrogate:	$\|\|x\|\|_1$	$\|\|x\|\|_k^{sp}$
Concept:	rank	
Elements:	matrices, $X \in \mathbb{R}^{m \times n}$	
Domain:	$\|\|X\|\| = \|\|\sigma\|\|_\infty \leq 1$	$\|\|X\|\|_F = \|\|\sigma\|\|_2 \leq 1$
Convex Surrogate:	$\|\|X\|\|_* = \|\|\sigma\|\|_1$	$\|\|X\|\|_{*k}^{sp} = \|\|\sigma\|\|_k^{sp}$

Table 1: Summary of the convex surrogates and different domains discussed in this paper.

Our final contribution is a proposed algorithm for solving optimization problems involving the k-support norm. We show how any dichotomic divide and conquer method can be used to find minimizers to a wide class of problems. This proposed algorithm is empirically proven to be orders of magnitude faster than the existing state-of-the-art approaches.

Acknowledgements

This research was supported by the Australian Research Council through the Discovery Early Career Researcher Award project DE130101775, the ARC Centre of Excellence in Robotic Vision CE140100016 and the Laureate Fellowship FL130100102.

[1] Andreas Argyriou, Rina Foygel, and Nathan Srebro. Sparse prediction with the k-support norm. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1457–1465. Curran Associates, Inc., 2012.

[2] Hanjiang Lai, Yan Pan, Canyi Lu, Yong Tang, and Shuicheng Yan. Efficient k-support matrix pursuit. In *European Conference on Computer Vision*, 2014.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.