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

Anders Eriksson, anders.eksson@qut.edu.au
Trung Pham, Tat-Jun Chin, Ian Reid, trung.pham,tat-jun.chin,ian.reid@adelaide.edu.au

Abstract

Sparsity, or cardinality, as a tool for feature selection is extremely common in a vast number of current computer vision applications. The $k$-support norm is a recently proposed norm with the proven property of providing the tightest convex bound on cardinality over the Euclidean norm unit ball. In this paper we present a re-derivation of this norm, with the hope of shedding further light on this particular surrogate function. In addition, we also present a connection between the rank operator, the nuclear norm and the $k$-support norm. Finally, based on the results established in this re-derivation, we propose a novel algorithm with significantly improved computational efficiency, empirically validated on a number of different problems, using both synthetic and real world data.

Sparsity and Convex Envelope Relaxation

Example: Sparse coding aims at finding a sparse representation $x$ of the input signal $y$.

Cardinality constraint: $\|x\|_k = \{x \in \mathbb{R}^m | \|x\|_k \leq k, \|x\|_\infty \leq 1 \}$

Convex hull: $\text{conv}(\mathbb{R}^m) = \{x \in \mathbb{R}^m | \|x\|_k \leq k, \|x\|_\infty \leq 1 \}$

Spectral $k$-support norm

Cardinality constraint: $\|x\|_{k,\ell} = \{x \in \mathbb{R}^m | \|x\|_k \leq k, \|x\|_\ell \leq 1 \}$

Convex hull: $\text{conv}(\mathbb{R}^m) = \{x \in \mathbb{R}^m | \|x\|_k \leq k, \|x\|_\ell \leq 1 \}$

Sparse coding.

$y \in \mathbb{R}^m$  \hspace{1cm} $D \in \mathbb{R}^{m \times n}$  \hspace{1cm} $x \in \mathbb{R}^n$

Cardinality with $\ell_1$ constraint and no $\ell_1$ convex envelop.

Cardinality with $\ell_2$ constraint and no $\ell_2$ convex envelop.

Cardinality with $\ell_\infty$ constraint and no $\ell_\infty$ convex envelop.

Convex hull: $\text{conv}(\mathbb{R}^m) = \{x \in \mathbb{R}^m | \|x\|_k \leq k, \|x\|_\infty \leq 1 \}$

Sparsity Regularized Parameter Estimation

Consider the optimisation problem:

$$\min_{x \in \mathbb{R}^d} f(x) + \frac{1}{2}\|x\|_2^2.$$  

First-order accelerated proximal gradient methods requires computing proximal operator, $(f(x) + \frac{1}{2}\|x\|_2^2)$.

$\text{prox}_{\|x\|_2^2}(v) = \arg \min_{x \in \mathbb{R}^d} \frac{1}{2}\|x\|_2^2$.

1. Elementwise non-expansive: $|v| \leq |x|$, $\forall i \in [1, \ldots, d]$.  
2. Sign-preserving: $\text{sign}(v) = \text{sign}(x)$, $\forall i \in [1, \ldots, d]$.  
3. Order-preserving: $\|v\| \geq \|x\|$. Previous methods compute proximal operators in $O(dk + \log(d))$ [1] and $O(d + k\log(d))$ [2] steps.

Efficiently Solving the Proximity Operator for the $k$-support Norm

We showed that it is much more convenient to work with the dual $k$-support norm.

$$\|x\|_{k^*} = \frac{1}{2}\sum_{i=1}^k |x_i|^2$$  

$$x = \text{prox}_{\|x\|_2^2}(x)$$

$$\text{prox}_{\|x\|_{k^*}}(v) = \arg \min_{x \in \mathbb{R}^d} \frac{1}{2}\|y - v\|_2^2 + \frac{1}{2}\|\|x\|_{k^*}\|^2$$  

(1)

Assume that $v$ is all positive and sorted, then

(2)

$$\text{prox}_{\|x\|_{k^*}}(v) = \arg \min_{x \in \mathbb{R}^d} \frac{1}{2}\|y - v\|_2^2 + \frac{1}{2}\|\|x\|_{k^*}\|^2$$  

(1)

$$\text{prox}_{\|x\|_{k^*}}(v) = \arg \min_{x \in \mathbb{R}^d} \frac{1}{2}\|y - v\|_2^2 + \frac{1}{2}\|\|x\|_{k^*}\|^2$$  

(2)

Here $E \in \mathbb{R}^{d \times d}$ with $|E|_{ij} =$

$$Z^* = [x, \gamma^*]$$

$Z^*$ active constraint set at optimal

Solving prox$\text{prox}_{\|x\|_{k^*}}(v)$ can hence be done in $O(d \log d)$ time. The sorting of $v$ takes $O(d \log d)$. At most $k$ and $d - k$ binary searches are conducted on $\gamma^*$ and $\gamma^*$, each with a complexity of $O(d \log d)$, respectively.

Experimental Results

Proximal operator with $k$-support norm: Computational efficiency comparison


Image denoising with $k$-support norm.

Left: Original image. Middle: Noisy image. Right: Recovered image using the spectral $k$-support norm.

More results can be found in the main paper.

References
