PROXIMAL OPERATOR FOR THE K-NORM

Ender Erkaya, Alper T. Erdogan

Department of Electrical and Electronics Engineering
Koc University
Istanbul, Turkey

ABSTRACT

K-norm corresponds to the sum of the K largest absolute entries of its argument vector. It can be considered as the generalization of the ℓ_∞-norm, which corresponds to the special case of $K = 1$, and the ℓ_1-norm, which corresponds to the case where K is selected as the dimension of the input argument. In this article, we derive the proximal operator for the K-norm and show that it is a hybrid of soft-thresholding and clipping functions, which are proximal operators of ℓ_1-norm and ℓ_∞-norm respectively.

***Index Terms*—** K-Norm, Proximal Operator, Clipping, Soft-Thresholding, ℓ_∞-norm, ℓ_1-norm.

1. INTRODUCTION

The development of low complexity convex optimization algorithms for non-differentiable cost functions has been at the center of both signal processing and machine learning research. The momentum is mainly due to optimization settings promoting sparsity through the use of ℓ_1-norm.

The proximal approach offers fast and low complexity solutions for non-differentiable optimization settings [1, 2]. It has been successfully applied especially to ℓ_1-norm based problem formulations due to the existence of the convenient soft-thresholding operator as the proximal operator for the ℓ_1-norm. Similarly, we can obtain an explicit form for the proximal operator of the ℓ_∞-norm, which turns out to be a clipping function whose level is dependent on its argument [3].

K-norm is defined as the sum of the K largest absolute elements. It actually defines a norm family based on the different choices of K. ℓ_1-norm is a special case with $K = n$, i.e., the dimension of the argument vector, and ℓ_∞-norm is another special case with $K = 1$.

The use of K-norm can be motivated from different directions. For example, in applications where the argument vector to ℓ_∞ norm is noisy, the actual peak value becomes ambiguous. In such a case, the set of large magnitude entries may represent potential actual peak locations and their average (sum with a proper scaling factor) can provide an estimate for the actual peak value. This perspective is potentially useful for applications using ℓ_∞-norm for blind (see for example [4, 5, 6, 7] and compressed training based [8] equalization. In [9], k-norm cost function is used to extend the multi-class support vector machine approach to handle the class ambiguity. An alternative perspective on K-norm is regarding potential applications where it is more sensible to reduce the sum of the higher values than just the peak value, which can, for example, offer a balancing option between approaches minimizing the peak error versus the total error. As a practical tool, K-norm was used to reduce the complexity of Peak-to-Average Ratio minimization problem for multicarrier systems [10].

K-norm is a special case of ordered (partial) ℓ_1-norm [11] where K non-zero norm-weights are set as equal. In reference [11], a proximity algorithm for the sorted ℓ_1 norm has already been introduced. In [12], the proximal operator for k-norm has been derived based on Moreau's decomposition which requires projection onto dual norm unity ball. In this article, we apply an alternative approach to derive the proximal operator for the K-norm and show that it has an explicit element-wise operation form which is the sum of soft-thresholding and clipping operators as shown in Fig. 1. We note that the soft thresholding, the proximal operator for the ℓ_1-norm, tends to equalize values around zero (to zero) while shifting the rest, and the clipping, the proximal operator for the ℓ_∞-norm tends to equalize the values around the peak value to a constant while leaving the remaining entries unchanged. Whereas the K-norm's proximal operator in Fig. 1 performs this equalization over an adjustable range based on the selection of the K parameter and the proximal operator's weighting parameter.

The article is organized as follows: in Section 2 is a reminder about the basic definition of the proximal operator. The proximal operator for the K-norm is derived in Section 3. Section 4 provide numerical examples illustrating the use of proximal operators.

Notation: Let $\mathbf{v}, \mathbf{x} \in \mathbb{R}^n$

$\mathbf{x} = [\mathbf{v}]^+$	Rectified Linear Unit Operator: $x_i = v_i$ if $v_i > 0$ and $x_i = 0$ otherwise.
$\mathbf{x} = [\mathbf{v}]^-$	Negative Rectified Linear Unit Operator: $x_i = v_i$ if $v_i < 0$ and $x_i = 0$ otherwise.
$\mathbf{1}$	Vector of all 1's (of appropriate size)
Δ	Unit Simplex in $\mathbb{R}^n$: $\Delta = \{\mathbf{x} \vert \sum_{i=1}^n x_i = 1, \mathbf{x} \in \mathbb{R}_+^n\}$

Table 1. Notation Table

2. PROXIMAL OPERATORS

The proximal operator for the convex function $f(\mathbf{x})$, with $\mathbf{x} \in \mathbb{R}^n$ is defined as [1]

$$prox_{\lambda f}(\mathbf{v}) = \arg\min_{\mathbf{x}} \; f(\mathbf{x}) + \frac{1}{2\lambda}\left\|\mathbf{x} - \mathbf{v}\right\|_2^2. \tag{1}$$

$\mathbf{x}^*$ is the solution to the above optimization problem, output of the proximal operator, if and only if thecondition

$$\mathbf{0} \in \lambda\partial f(\mathbf{x}^*) + (\mathbf{x}^* - \mathbf{v}), \tag{2}$$

holds. Furthermore, $\mathbf{x}^*$ is the unique minimizer of the problem due to strong convexity coming from $\left\|\mathbf{x} - \mathbf{v}\right\|_2^2$.

3. PROXIMAL OPERATOR OF THE K-NORM

We define K-norm as $\left\|\mathbf{x}\right\|_{[K]} = \sum_{i=1}^{K} |x|_{[i]}$ where $|x|_{[i]}$ is the ith largest magnitude element of $\mathbf{x}$ or ith index value of descendingly sorted absolute form of $\mathbf{x}$. Then, the proximal operator of $f = \left\|.\right\|_{[K]}$ for $\mathbf{v} \in \mathbb{R}^n$ can be written as

$$prox_{\lambda f}(\mathbf{v}) = \operatorname*{argmin}_{\mathbf{x}} \quad \lambda\left\|\mathbf{x}\right\|_{[K]} + (1/2)\left\|\mathbf{x} - \mathbf{v}\right\|_2^2.$$

Before we introduce the expression for this proximal operator, we provide the following definitions:

- $C_\lambda(\mathbf{v}) = [\mathbf{v} - \lambda\mathbf{1}]^- + \lambda\mathbf{1} + [-\mathbf{v} - \lambda\mathbf{1}]^+$ is the elementwise clipping operation.
- $S_\lambda(v) = [\mathbf{v} - \lambda\mathbf{1})]^+ - [-\mathbf{v} - \lambda\mathbf{1}]^+$ is the elementwise soft-thresholding operation.
- $prox_{\lambda\ell_\infty}(\cdot)$ is the proximal operator for the ℓ_∞-norm [**?**].

The following theorem offers an explicit form for the proximal operator of the K-norm:

Theorem: The solution $\mathbf{x}^* = prox_{\lambda\|.\|_{[K]}}(\mathbf{v})$ is given by the elementwise clipping combined subthresholding operator:

$$\mathbf{x}^* \;\; = \;\; C_\mu(\mathbf{v}) + S_{\mu+\lambda}(\mathbf{v}), \tag{3}$$

which is illustrated in Fig. 1, where

- $\mu = \left\|prox_{K\lambda\ell_\infty}(\mathbf{v}^{(s)}_{\tau+1:N})\right\|_\infty$ is clipping parameter,
- $\mathbf{v}^{(s)} = \mathbf{Q}\,|\mathbf{v}|$ such that $v_i^{(s)} \geq v_j^{(s)}$ for all $i, j \in \{1, 2, ..., n\}$, is the sorted absolute vector of $\mathbf{v}$, and $\mathbf{Q}$ is the permutation matrix corresponding to this sorting,
- $\tau \in \{1, \ldots, K-1\}$ is the number of components above $\mu + \lambda$, which satisfy the condition $v_\tau^{(s)} - \lambda - \left\|prox_{(K-\tau)\lambda\ell_\infty}(\mathbf{v}^{(s)}_{\tau+1:N})\right\|_\infty > 0\})$.

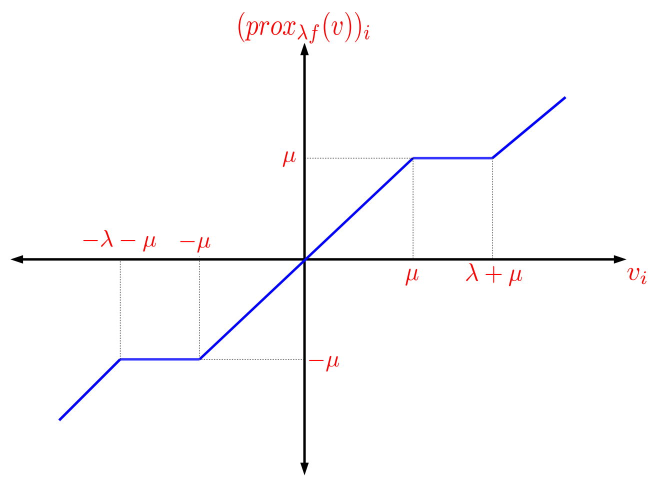

Fig. 1. $(prox_{\lambda f}(\mathbf{v}))_i = C_\mu(\mathbf{v}_i) + S_{\lambda+\mu}(\mathbf{v}_i)$

Proof: For the optimality condition in (2), we first need to obtain the expression for the subdifferential set of the K-norm. We start by providing the following definitions:

- We can write the K-norm function $f(\mathbf{x})$ as $f(\mathbf{x}) = \max_{i=1,\ldots,P} f_i(\mathbf{x})$, i.e., the maximum of some linear functionals, where $f_i(\mathbf{x}) = \mathbf{q}_i^T\mathbf{x}$, and $\mathbf{q}_i$'s are distinct vectors with exactly K non-zero entries, and the non-zero entries take their values from the set $\{-1, 1\}$. There are $P = 2^K\binom{n}{k}$ such $\mathbf{q}_i$ vectors.
- At the optimal point $\mathbf{x}^*$, $f(\mathbf{x}^*) = f_i(\mathbf{x}^*)$ for $i \in \mathcal{W}^*$ where $\mathcal{W}^* \subset \{1, \ldots, P\}$, and the subdifferential set is $\partial f(\mathbf{x}^*) = \mathbf{Co}\bigcup_{i\in\mathcal{W}^*}\{\mathbf{q}_i\}$.
- We refer to a set $\mathcal{T}_i^* \subset \{1, \ldots, n\}$ as an *active set* of the optimal point if and only if $\sum_{l\in\mathcal{T}_i^*}|x_l^*| = f(\mathbf{x}^*)$, and $card(T_i)^* = K$. Let $\{\mathcal{T}_i^*, i = 1, \ldots, M\}$ represent the set of all active sets for the optimal point, where M is the number of distinct active sets.
- We define the set $\mathcal{I}^*$ as the union of all active sets, i.e., $\mathcal{I}^* = \bigcup_{i=1}^M \mathcal{T}_i^*$. We note that $C_{\mathcal{I}^*} \triangleq card(\mathcal{I}) \geq K$.

- We define the set $\mathcal{J}_2^*$ as the subset of $\mathcal{I}^*$

$$\mathcal{J}_2^* = \{l||x_l^*| = \min_{i\in\mathcal{I}} |x_i^*|, l \in \mathcal{I}^*\}, \quad (4)$$

  and $\mathcal{J}_1^*$ as its complement in $\mathcal{I}^*$, i.e., $\mathcal{J}_1^* = \mathcal{I} \cap \bar{\mathcal{J}}_2^*$. We also define $\tau \triangleq card(\mathcal{J}_1^*)$ and $L \triangleq card(\mathcal{J}_2^*)$ for $m = 1, 2$. Note that $\mathcal{J}_1^*$ contains the indexes that are contained in all active sets, i.e., $\mathcal{T}_i^*$'s, i.e., $\mathcal{J}_1^* \subset \mathcal{T}_i^*$ for all $i = 1, \ldots, M$.

- We also define an indicator function

$$\gamma_i^*[j] = \begin{cases} 1 & \text{if } j \in \mathcal{T}_i^* \\ 0 & \text{otherwise,} \end{cases}$$

  i.e., it indicates whether a given index j is included in the active set $\mathcal{T}_i^*$. We observe that $\sum_{j\in\mathcal{I}^*} \gamma_i^*[j] = K$, and $\gamma_i^*[j] = 1$ for all $j \in \mathcal{J}_1^*$ and $i = 1, \ldots, M$.

Based on these definitions, we can write the subdifferential of the K-Norm at $\mathbf{x}^*$ as follows:

- if $\mathbf{x}^* = \mathbf{0}$, then $\mathcal{W}^* = \{1, \ldots, P\}$ and the subdifferential set is given by the polytope $\partial f(\mathbf{0}) = \mathcal{P}_{n,K} = \{\sum_{i=1}^P \beta_i \mathbf{q}_i | \boldsymbol{\beta} \in \Delta\}$,
- if $\mathbf{x}^* \neq \mathbf{0}$ then we have two different conditions to consider
  i. When $1 \leq \|\mathbf{x}^*\|_0 \leq K - 1$: In this case, $card(\mathcal{J}_2^*) = n - \tau$ and the components corresponding to J_2^* are equal to zero. The corresponding subdifferential set is given by

$$\partial f(\mathbf{x}) = \{\sum_{j\in\mathcal{J}_1^*} sign(x_j^*)\mathbf{e}_j + \mathbf{s} | \mathbf{s} \in \mathcal{P}_{n,K-\tau}\} \quad (5)$$

  where $\mathcal{P}_{n,K-\tau} = \{\sum_{i=1}^P \beta_i \bar{\mathbf{q}}_i | \boldsymbol{\beta} \in \Delta\}$ and $\bar{\mathbf{q}}_i$'s are distinct vectors with only $K - \tau$ non-zero entries (with values -1 or $+1$) and zero for indexes in $\mathcal{J}_1^*$.

  ii. When $\|\mathbf{x}^*\|_0 \geq K$, we have

$$\begin{aligned}
\partial f(\mathbf{x}^*) &= \{\sum_{i=1}^M \beta_i \sum_{j\in T_i^*} sign(x_j^*)\mathbf{e}_j | \boldsymbol{\beta} \in \Delta\} \\
&= \{\sum_{i=1}^M \beta_i \sum_{j\in\mathcal{I}^*} \gamma_i^*[j] sign(x_j^*) | \boldsymbol{\beta} \in \Delta\} \\
&= \{\sum_{j\in\mathcal{J}_1^*} \sum_{i=1}^M \beta_i \gamma_i^*[j] sign(x_j^*) + \\
&\qquad \sum_{j\in\mathcal{J}_2^*} \sum_{i=1}^M \beta_i \gamma_i^*[j] sign(x_j^*) | \boldsymbol{\beta} \in \Delta\} \\
&= \{\sum_{j\in\mathcal{J}_1^*} sign(x_j^*) \sum_{j\in\mathcal{J}_2^*} \alpha_j(\boldsymbol{\beta}) sign(x_j^*) | \boldsymbol{\beta} \in \Delta\}, \quad (6)
\end{aligned}$$

  where, in the last line, we used $\sum_{i=1}^M \beta_i \gamma_i^*[j] = 1$ for $j \in \mathcal{J}_1^*$ and defined $\alpha_j(\boldsymbol{\beta}) \triangleq \sum_{i=1}^M \beta_i \gamma_i^*[j]$.

Based on the subdifferential set expressions, we are ready to analyze the optimality condition in (2):

For $\mathbf{x}^* = 0$, the optimality condition in (2) implies that there exists $\mathbf{s}^* \in \mathcal{P}_{n,K}$

$$\mathbf{0} = \lambda \mathbf{s} - \mathbf{v}^* \Leftrightarrow \mathbf{v} = \lambda \mathbf{s}^*. \quad (7)$$

When $\mathbf{x}^* \neq 0$, based on (6) and (2), we analyze different cases outlined in the subdifferential derivations:

i. When $1 \|\mathbf{x}^*\|_0 \leq K - 1$: for $j \in \mathcal{J}_1^*$ we obtain $x_j^* - v_j + \lambda sign(x_j^*) = 0$. from which we obtain $\mathbf{x}_j^* = v_j - \lambda sign(x_j^*)$. Here we observe that $sign(x_j^*)$ must be equal to $sign(v_j)$. Because, if x_j^* had opposite sign, that would cause extra cost on the objective function compared to the equal signed version due to the quadratic term $\|\mathbf{x} - \mathbf{v}\|_2^2$. In other words, there always exists an equal signed version of x_j^* which yields same cost on f and less cost on $\|\mathbf{x} - \mathbf{v}\|_2^2$ relative to the opposite signed version. As a result, the above expression simplifies to

$$|x_j^*| = |v_j| - \lambda, \text{for } j \in \mathcal{J}_1^*. \quad (8)$$

   for $j \in \mathcal{J}_2$: $v_j = \lambda s_j^*$ for some $\mathbf{s} \in \mathcal{P}_{n,K-\tau}$.

ii. When $\|\mathbf{x}\|_0 \geq K$:

If the component index satisfies $j \notin \mathcal{I}^*$, then we have $x_j^* - v_j = 0$, or equivalently,

$$x_j^* = v_j \quad (9)$$

If the component index satisfies $j \in \mathcal{J}_1^*$, then we have

$$x_j^* - v_j + \lambda sign(x_j^*) = 0, \quad (10)$$

or equivalently, multiplying both sides with $sign(x_j^*)$, we obtain

$$|x_j^*| = v_j sign(x_j^*) - \lambda. \quad (11)$$

Based on the previous argument $sign(x_j^*) = sign(v_j)$, and therefore, we obtain the expression in (8).

As the final case, for the components with $j \in \mathcal{J}_2^*$, the optimality condition in (2) and the subdifferential set in (6) implies that there exists $\boldsymbol{\beta}^* \in \Delta$ such that

$$x_j^* = v_j - \lambda \alpha_j(\boldsymbol{\beta}^*) sign(x_j^*).$$

Due to the same reasoning as the previous case, we have $sign(x_j^*) = sign(v_j)$, and therefore,

$$|x_j^*| = |v_j| - \lambda \alpha_j(\boldsymbol{\beta}^*). \quad (12)$$

For a pair of indexes $j, j' \in \mathcal{J}_2^*$, due to the definition of $\mathcal{J}_2^*$, we have $|x_j^*| = |x_{j'}^*|$, which implies

$$|v_j| - \lambda \alpha_j(\boldsymbol{\beta}^*) = |v_{j'}| - \lambda \alpha_{j'}(\boldsymbol{\beta}^*), \quad (13)$$

or equivalently,

$$|v_j| = |v_{j'}| - \lambda\alpha_{j'}(\boldsymbol{\beta}^*) + \lambda\alpha_j(\boldsymbol{\beta}^*). \quad (14)$$

Summing both sides for all $j' \in \mathcal{J}_2^*$, we obtain

$$L|v_j| = \sum_{j' \in \mathcal{J}_2^*} |v_{j'}| - \lambda \sum_{j' \in \mathcal{J}_2'} \alpha_{j'}(\boldsymbol{\beta}^*) + L\lambda\alpha_j(\boldsymbol{\beta}^*).$$

Here, the second sum expression on the right simplifies to

$$\sum_{j' \in \mathcal{J}_2'} \alpha_{j'}(\boldsymbol{\beta}^*) = \sum_{j' \in \mathcal{J}_2'} \sum_{i=1}^{M} \beta_i^* \gamma_i^*[j']$$
$$= \sum_{i=1}^{M} \beta_i^* \sum_{j' \in \mathcal{J}_2'} \gamma_i^*[j'] = \sum_{i=1}^{M} \beta_i^*(K-\tau) = (K-\tau).$$

Therefore, we can write $|v_j| = \bar{v} - \lambda\frac{K-\tau}{L} + \lambda\alpha_j(\boldsymbol{\beta}^*)$, where $\bar{v} \triangleq \frac{1}{L}\sum_{j' \in \mathcal{J}_2^*} |v_{j'}|$. Manipulating this expression, we obtain an expressing for $\alpha_j(\boldsymbol{\beta}^*)$ as

$$\alpha_j(\boldsymbol{\beta}^*) = \frac{|v_j| - \bar{v}}{\lambda} + \frac{K-\tau}{L}. \quad (15)$$

Plugging (15) in (12), for $j \in \mathcal{J}_2^*$, we get

$$x_j^* = sign(v_j)(\bar{v} - \frac{\lambda(K-\tau)}{L}). \quad (16)$$

As a summary, the components corresponding to $\mathcal{J}_2^*$ are clipped to the same magnitude level $\mu = \bar{v} - \frac{\lambda(K-\tau)}{L}$. Since $0 \leq \alpha_j \leq 1$, based on (15), we can write

$$\mu \leq |v_j| \leq \mu + \lambda \quad \text{for } j \in \mathcal{J}_2^*. \quad (17)$$

Now, using the definitions of $\mathcal{J}_1^*$, $\mathcal{J}_2^*$ and equalities (9,8,16), $\forall j_1 \in \mathcal{J}_1^*, j_2 \in \mathcal{J}_2^*$ and $j_3 \notin I$ we have the following relations:

$$|x_{j_1}^*| > |x_{j_2}^*| = \mu = \bar{v} - \frac{(K-\tau)\lambda}{L} > |x_{j_3}^*| \quad (18)$$

$$|v_{j_1}| - \lambda > \bar{v} - \frac{(K-\tau)\lambda}{L} > |v_{j_3}| \quad (19)$$

We can summarize the characterization of the optimal point as

$$\mathbf{x}_j^* = \begin{cases} v_j & |v_j| < \mu, \\ \mu sign(v_j) & \mu \leq |v_j| \leq \mu + \lambda, \\ v_j - \lambda sign(v_j) & |v_j| > \mu + \lambda, \end{cases} \quad (20)$$

which corresponds to the mapping in Fig. 1, i.e., $C_\mu(\mathbf{v}) + S_{\mu+\lambda}(\mathbf{v})$ operator. As a result, what remains is to identify τ and L which in turn determines μ. For this purpose, we first define the vector $\mathbf{v}^{(s)}$ as the vector obtained by taking the absolute values of the elements of $\mathbf{v}$ and then sorting them in descending order, we also define $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_n]^T$ as the sorting index vector, i.e.. $v_m^{(s)} = v_{\eta_m}$ Note that the first τ elements of $\boldsymbol{\eta}$, is equal to the set $\mathcal{J}_1^*$ and its next L elements is equal to the set $\mathcal{J}_2^*$.

As an important observation, the optimality condition in (19) for indexes $\eta_{\tau+1}, \ldots, \eta_n$, i.e., the components that do not correspond to $\mathcal{J}_1^*$ is exactly the same as the condition obtained for the ℓ_∞-norm proximal operator [**?**], where λ is replaced with $(K-\tau)\lambda$. Therefore, we can write

$$\begin{aligned} \mathbf{x}_{\eta_{\tau+1},\ldots,\eta_n} &= prox_{(K-\tau)\lambda\ell_\infty}(\mathbf{v}_{\eta_{\tau+1},\ldots,\eta_n}) \\ &= prox_{(K-\tau)\lambda\ell_\infty}(\mathbf{v}^{(s)}_{\tau+1,\ldots,n}). \end{aligned}$$

In the above expressions, τ, the cardinality of $\mathcal{J}_1^*$ needs to be determined based on the optimality conditions. In fact, if we define

$$\mu(i) = \|prox_{(K-\tau)\lambda\ell_\infty}(\mathbf{v}^{(s)}_{\tau+1,\ldots,n})\|_\infty. \quad (21)$$

where i needs to be selected in $\{1, \ldots, K-1\}$, for the correct choice $i = \tau$, we satisfy ,from (18), the condition that we can write $|x_{\eta_i}^*| > |x_{\eta_{i+1}}^*|$ which in turn implies, $|v_{\eta_i}| - \lambda > \mu$, or equivalently $|v_i^{(s)}| - \lambda > \mu$. In fact, if we start with $i = K-1$ and reduce it till we achieve this condition, which will in fact be an optimality certificate as also confirmed with Lemma 2.3 in [11] ■

4. NUMERICAL EXAMPLE

In order to illustrate the use of the proximal operator, we consider a Peak-to-Average Ratio problem where we demonstrate over an OFDM signal with 256 carriers[8]. We assumed tones $\{5, 25, 54, 102, 125, 131, 147, 200, 204, 209, 247\}$ are the reserved tones and data are loaded with QPSK signals. The problem reduces to the optimization scheme where we use k-norm: min $\|\boldsymbol{\Gamma}\rho + \gamma\|_{[K]}$. We use Alternating Direction Method(ADMM) algorithm[1], using the proximal operator derived in this article, choosing k=5. Fig. 2 demonstrates the convergence of objective function. As a result of 10th iteration using 5-norm, we reduced PAR value to 6.51dB from 8.47dB on average.

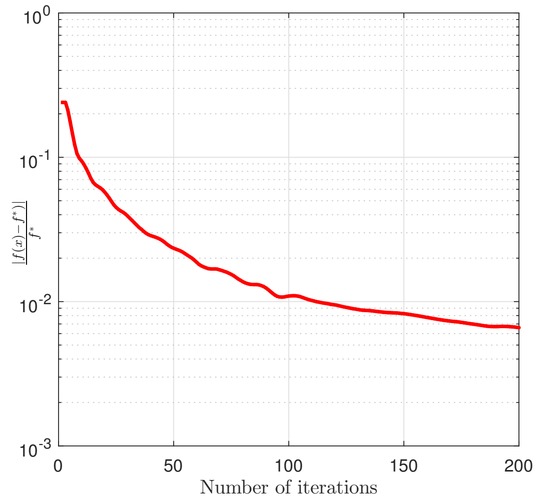

Fig. 2. The objective function convergence curve for ADMM algorithm

5. REFERENCES

[1] Neal Parikh, Stephen Boyd, et al., "Proximal algorithms," *Foundations and Trends® in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[2] Patrick Louis Combettes and Jean-Christophe Pesquet, "Proximal Splitting Methods in Signal Processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, R.S.; Combettes P.L.; Elser V.; Luke D.R.; Wolkowicz H. (Eds.) Bauschke, H.H.; Burachik, Ed., pp. 185–212. Springer, 2011.

[3] Ender Erkaya and Alper T. Erdogan, "A direct derivation for the ℓ_∞-norm proximal operator," in *Submitted to 26th European Signal Processing Conference(EUSIPCO 2018)*, 2018.

[4] S. Vembu, S. Verdu, R. Kennedy, and W. Sethares, "Convex cost functions in blind equalization," *IEEE Trans. on Signal Processing*, vol. 42, pp. 1952–1960, August 1994.

[5] Z. Ding and Z. Luo, "A fast linear programming algorithm for blind equalization," *IEEE TCOM.*, vol. 48, pp. 1432–1436, September 2000.

[6] Zhi-Quan Luo, Mei Meng, Kon Max Wong, and Jian-Kang Zhang, "A fractionally spaced blind equalizer based on linear programming," *IEEE Trans. on Signal Processing*, vol. 50, pp. 1650–1660, July 2002.

[7] Alper T. Erdogan and Can Kizilkale, "Fast and low complexity blind equalization via subgradient projections," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2513–2524, July 2005.

[8] B. B. Yilmaz and A. T. Erdogan, "Compressed training adaptive equalization: Algorithms and analysis," *IEEE Transactions on Communications*, vol. 65, no. 9, pp. 3907–3921, Sept 2017.

[9] Maksim Lapin, Matthias Hein, and Bernt Schiele, "Top-k multiclass svm," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, Cambridge, MA, USA, 2015, NIPS'15, pp. 325–333, MIT Press.

[10] Alper Tunga Erdogan, "A low complexity multicarrier par reduction approach based on subgradient optimization," *Signal Processing*, vol. 86, no. 12, pp. 3890–3903, 2006.

[11] Małgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Weijie Su, and Emmanuel J Candès, "Slope adaptive variable selection via convex optimization," *The annals of applied statistics*, vol. 9, no. 3, pp. 1103, 2015.

[12] Bin Wu, Chao Ding, Defeng Sun, and Kim-Chuan Toh, "On the moreau-yosida regularization of the vector k-norm related functions," vol. 24, pp. 766794, 05 2014.