A note on duality for the weighted sparse group Lasso

M. Massias

November 21, 2019

Notation The positive part function is $(\cdot)_+ = \max(\cdot, 0)$, which acts entry-wise when applied to vectors. For two vectors of same length, inequality, multiplication and division are meant entry-wise. In the sequel $w \in \mathbb{R}^n_{++}$ is a fixed vector of strictly positive weights. The soft-thresholding of $x \in \mathbb{R}^n$ with weights w, is $\mathrm{ST}(x, w) \in \mathbb{R}^n$, whose i-th entry is $(x_i - w_i)_+$. Let $\|x\|_{1,w} \triangleq \sum_{i=1}^n w_i |x_i|$ denote the weighted ℓ_1-norm on $\mathbb{R}^n$ (it is a norm since all w_i are positive).

Proposition 1. *The dual norm of $\|\cdot\|_{1,w}$ is the weighted ℓ_∞-norm with inverse weights, $\|x\|_{\infty,1/w} \triangleq \max_i |x_i| / w_i$.*

Proof Recall that the dual norm at x is given by $\sup\limits_{\|u\|_{1,w} \leq 1} x^\top u$. By Hölder's inequality,

$$x^\top u = (x/w)^\top (wu) \leq \|x/w\|_\infty \|wu\|_1 = \|x\|_{\infty,1/w} \|u\|_{1,w} \quad . \tag{1}$$

It remains to see that the value $\|x\|_{\infty,1/w}$ is attained, for u vanishing everywhere except $u_j = (\operatorname{sign} x_j)/w_j$ where $j = \arg\max_i |x_i|/w_i$. ■

1 Weighted epsilon norm

Definition 2. *Let $x \in \mathbb{R}^n$. Let $R, \alpha \in \mathbb{R}_+$. The weighted epsilon norm of x is defined as the smallest positive solution in ν of*

$$\sum_{i=1}^n (|x_i| - \nu\alpha w_i)_+^2 = (\nu R)^2 \quad . \tag{2}$$

Remark 3. *Equation (2) can be rewritten $\|\mathrm{ST}(x, \nu\alpha w)\|_2 = \nu R$. With the notation $h(x, \nu) \triangleq \|\mathrm{ST}(x, \nu\alpha w)\|_2 - \nu R$, Equation (2) is also equivalent to $h(x, \nu) = 0$. When $x \neq 0$, it is clear that the solution is strictly positive, and if $\alpha \neq 0$ it is then the solution of:*

$$\sum_{i=1}^n \left(\frac{|x_i|}{\alpha\nu} - w_i \right)_+^2 = \frac{R^2}{\alpha^2} \quad . \tag{3}$$

Remark 4. *The cases $R = 0$ or $\alpha = 0$ are trivial:*

- *if $R = 0$, the smallest value making $\sum_{i=1}^n (|x_i| - \nu\alpha w_i)_+^2$ vanish is $\nu = \max_i \frac{x_i}{\alpha w_i} = \frac{1}{\alpha}\|x\|_{\infty,1/w}$,*
- *if $\alpha = 0$, $\nu = \frac{\|x\|_2}{R}$.*

In the following, we consider $\alpha > 0$ and $R > 0$.

Proposition 5. *The solution of Equation (2) is unique.*

Proof The function $\nu \mapsto h(x, \nu)$ is strictly decreasing, because its first term is non-increasing and its second term is strictly decreasing. It is continuous, with $h(x, 0) = \|x\|_2 \geq 0$ and $\lim_{\nu\to\infty} h(x, \nu) = -\infty$. Hence, $h(x, \cdot)$ takes the value 0 exactly once. ■

Remark 6. *For $x \neq 0$, it is trivial to see that $\nu(x) < \frac{\|x\|_{\infty,1/w}}{\alpha}$, because $h(x, \frac{\|x\|_{\infty,1/w}}{\alpha}) < 0$.*

Theorem 7. *$\nu(\cdot)$ is a norm.*

Proof It is easy to check that $\nu(x) = 0 \Leftrightarrow x = 0$ and $x \mapsto \nu(x)$ is positively homogeneous. It remains to show that $x \mapsto \nu(x)$ satisfies the triangular inequality. Let $x, y \in \mathbb{R}^n$.

$$\begin{aligned}
\|\mathrm{ST}(x + y, \alpha(\nu(x) + \nu(y))w)\|_2 &= \|(|x + y| - \alpha(\nu(x) + \nu(y))w)_+\|_2 \\
&\leq \|(|x| + |y| - \alpha(\nu(x) + \nu(y))w)_+\|_2 \\
&\leq \|(|x| - \alpha\nu(x)w + |y| - \alpha\nu(y)w)_+\|_2 \\
&\leq \|(|x| - \alpha\nu(x)w)_+\|_2 + \|(|y| - \alpha\nu(y)w)_+\|_2 \\
&= \nu(x)R + \nu(y)R,
\end{aligned}$$

where we use the fact that $0 \leq u \leq v \Rightarrow \|u\|_2 \leq \|v\|_2$, and the triangular inequality for $|\cdot|$ and $(\cdot)_+$. We just showed that $h(x+y, \nu(x)+\nu(y)) \leq 0$. Since $h(x+y, \cdot)$ is decreasing and $h(x + y, \nu(x + y)) = 0$, this means $\nu(x + y) \geq \nu(x) + \nu(y)$. ■

Lemma 8. $\forall x \in \mathbb{R}^n$, $\exists!(x^R, x^\alpha) \in \mathbb{R}^n$, $\|x^R\| = R\nu(x)$, $\|x^\alpha\|_{\infty,1/w} = \alpha\nu(x), x = x^R + x^\alpha$.

Proof Let $x \in \mathbb{R}^n$. We assume x has all components positive, but the proof is similar in the general case. Define x^R and x^α by

$$x_i^R = \begin{cases} x_i - \nu(x)\alpha w_i \ , & \text{if} \quad x_i - \nu(x)\alpha w_i > 0 \ , \\ 0 \ , & \text{otherwise} \ . \end{cases} \tag{4}$$

$$x_i^\alpha = \begin{cases} \nu(x)\alpha w_i \ , & \text{if} \quad x_i - \nu(x)\alpha w_i > 0 \ , \\ x_i \ , & \text{otherwise} \ . \end{cases} \tag{5}$$

It is clear that $x = x^R + x^\alpha$. Additionally, $\|x^R\|^2 = \sum_i (x_i - \nu(x)\alpha w_i)_+^2 = (\nu(x)R)^2$ and $\|x^\alpha\|_{\infty,1/w} = \alpha\nu(x)$ by Remark 6.

Now we show the uniqueness. Let v such that $\|v\|_{\infty,1/w} = \alpha\nu(x)$.

$$\begin{aligned} \|x - v\|^2 &= \|x^R + x^\alpha - v\|^2 \\ &= \|x^R\|^2 + 2(x^R)^\top(x^\alpha - v) + \|x^\alpha - v\|^2 \end{aligned} \tag{6}$$

But $(x^R)^\top(x^\alpha - v)$ is positive, as a sum of n terms which are either 0 (if $x_i - \nu(x)\alpha w_i \leq 0$), or positive (because $\|v\|_{\infty,1/w} = \alpha\nu(x)$). Hence

$$\|x - v\|^2 \geq \|x^R\|^2 + \|x^\alpha - v\|^2 = (R\nu(x))^2 + \|x^\alpha - v\|^2 \ , \tag{7}$$

and if $v + (x - v)$ is a decomposition of x satisfying the imposed conditions, i.e. $\|x - v\| = R\nu(x)$, then we must have $\|x^\alpha - v\| = 0$. ■

Proposition 9 (Unit ball of weighted epsilon norm)**.** *The unit ball of the weighted epsilon norm is given by:*

$$\{x \in \mathbb{R}^n \,:\, \nu(x) \leq 1\} = \{u + v \,:\, u, v \in \mathbb{R}^n, \|u\| \leq R, \|v\|_{\infty,1/w} \leq \alpha\} \tag{8}$$

Proof The left-right inclusion is proved by Lemma 8.

For the right-left inclusion, let $u, v \in \mathbb{R}^n$ s.t. $\|u\| \leq R, \|v\|_{\infty,1/w} \leq \alpha$.

$$\begin{aligned} h(u + v, 1) &= \Big(\sum_{i=1}^n (|u_i + v_i| - \alpha w_i)_+^2 \Big)^{\frac{1}{2}} - R \\ &\leq \Big(\sum_{i=1}^n (|u_i|_+)^2 \Big)^{\frac{1}{2}} + \Big(\sum_{i=1}^n (|v_i| - \alpha w_i)_+^2 \Big)^{\frac{1}{2}} - R \\ &= \|u\| - R \\ &\leq 0. \end{aligned} \tag{9}$$

Hence $\nu(u + v) \leq 1$.

■

Theorem 10. *The dual norm of $\nu(\cdot)$ is $R\|\cdot\| + \alpha\|\cdot\|_{1,w}$.*

Proof The value of the dual norm at $y \in \mathbb{R}^n$ is

$$\begin{aligned} \max_{\nu(x)\leq 1} y^\top x &= \max_{\substack{\|u\|\leq R \\ \|v\|_{\infty,1/w}\leq\alpha}} y^\top(u + v) \\ &= \max_{\|u\|\leq R,} y^\top u + \max_{\|v\|_{\infty,1/w}\leq\alpha} y^\top v \\ &= R\,\|y\| + \alpha\|y\|_{1,w} \ . \end{aligned}$$

■

2 Computing the weighted epsilon norm

Without loss of generality we assume that for all $i = 1, \dots, n$, $x_i \geq 0$. Let $(x_{(1)}, \dots, x_{(n)})$ be a reordering of $(x_1, \dots, x_n)$ such that $x_{(i)}/w_{(i)}$ is a non-increasing sequence.

Theorem 11. *With the convention* $x_{(n+1)}/w_{(n+1)} = 0$, *and for* $a_j \triangleq \sum_{i=1}^{j} \left(\frac{x_{(i)} w_{(j)}}{x_{(j)}} - w_{(i)} \right)^2$, *there exists an index* $i_0 \in [\![1, n]\!]$ *such that*

$$R^2/\alpha^2 \in [a_{i_0}, a_{i_0+1}[\quad . \tag{10}$$

Furthermore, this index can be computed in $\mathcal{O}(n \log n)$ *operations.*

Proof Since the intervals $[x_{(i+1)}/w_{(i+1)}, x_{(i)}/w_{(i)}[$ form a partition of $[0, x_{(1)}/w_{(1)}[$, and $\alpha\nu(x) < \|x\|_{\infty,1/w} = x_{(1)}/w_{(1)}$ (Remark 6), there exists an index $i_0 \in [\![1, n]\!]$ such that

$$\nu(x) \in \left[\frac{x_{(i_0+1)}}{\alpha w_{(i_0+1)}}, \frac{x_{(i_0)}}{\alpha w_{(i_0)}} \right[\quad . \tag{11}$$

Since the function $f : \nu \mapsto \sum_1^n (\frac{x_{(i)}}{\alpha\nu} - w_{(i)})_+^2$ is decreasing,

$$f(\nu(x)) \in \left[f\Big(\frac{x_{(i_0)}}{\alpha w_{(i_0)}} \Big), f\Big(\frac{x_{(i_0+1)}}{\alpha w_{(i_0+1)}} \Big) \right[\quad . \tag{12}$$

Moreover, $f(\nu(x)) = \sum_1^n (\frac{x_{(i)}}{\alpha\nu(x)} - w_{(i)})_+^2 = \frac{R^2}{\alpha^2}$, and for every index $i \in [\![1, n]\!]$, $i \geq i_0 \implies \frac{x_{(i)}}{w_{(i)}} \leq \frac{x_{(i_0)}}{w_{(i_0)}} \implies \frac{x_{(i)}}{x_{(i_0)}/w_{(i_0)}} - w_{(i)} \leq 0$, hence

$$\begin{aligned} f\Big(\frac{x_{(i_0)}}{\alpha w_{(i_0)}} \Big) &= \sum_{i=1}^{n} \left(\frac{x_{(i)}}{x_{(i_0)}/w_{(i_0)}} - w_{(i)} \right)_+^2 \\ &= \sum_{i=1}^{i_0} \left(\frac{x_{(i)} w_{(i_0)}}{x_{(i_0)}} - w_{(i)} \right)^2 \\ &= a_{i_0} \quad , \end{aligned} \tag{13}$$

which proves Equation (10). If we define $S_j^{x^2} \triangleq \sum_1^j x_{(i)}^2$, $S_j^{xw} \triangleq \sum_1^j x_{(i)} w_{(i)}$ and $S_j^{w^2} \triangleq \sum_1^j w_{(i)}^2$, then we have:

$$a_{i_0} = \frac{w_{(i_0)}^2}{x_{(i_0)}^2} S_{i_0}^{x^2} - 2 \frac{w_{(i_0)}}{x_{(i_0)}} S_{i_0}^{xw} + S_{i_0}^{w^2} \quad . \tag{14}$$

Thus, Equation (14) enables us to compute i_0 in $\mathcal{O}(n \log(n))$ operations at most: sorting the values x_i/w_i, and then incrementally computing a_1, a_2, etc. until a value larger than R^2/α^2 is reached. Proposition 14 shows that the cost is in fact lower. ■

Remark 12. *In the definition of a_j, the sum could stop at $j-1$ since the last term is 0. We keep it in order to have only i_0 as index in Equation (14).*

Theorem 13. *With $\Delta \triangleq \alpha^2 (S_{i_0}^{xw})^2 - (\alpha^2 S_{i_0}^{w^2} - R^2) S_{i_0}^{x^2}$, we have*

$$\nu(x) = \begin{cases} \frac{\alpha^2 S_{i_0}^{xw} - \sqrt{\Delta}}{\alpha^2 S_{i_0}^{w^2} - R^2} , & \text{if} \quad (\alpha^2 S_{i_0}^{w^2} - R^2) \neq 0 \ , \\ \frac{S_{i_0}^{x^2}}{2\alpha S_{i_0}^{xw}} , & \text{otherwise} \ . \end{cases}$$

Proof By the proof of Theorem 11, i_0 is the largest index such that $x_{(i_0)}/\nu \geq \alpha w_{(i_0)}$, we have

$$\begin{aligned} R^2 &= \sum_{i=1}^{i_0} \left(\frac{x_{(i)}}{\nu} - \alpha w_{(i)} \right)^2 \\ &= \frac{1}{\nu^2} S_{i_0}^{x^2} - \frac{2}{\nu} \alpha S_{i_0}^{xw} + \alpha^2 S_{i_0}^{w^2} \ . \end{aligned}$$

or equivalently

$$\nu^2 \left(\alpha^2 S_{i_0}^{w^2} - R^2 \right) - 2\nu\alpha S_{i_0}^{xw} + S_{i_0}^{x^2} = 0 \ .$$

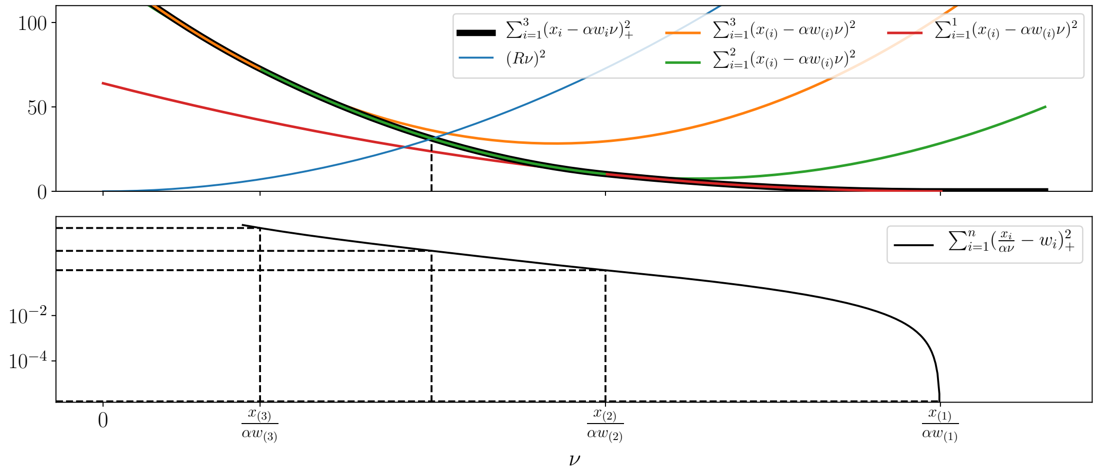
■

Proposition 14. *We have the following lower bound for $\nu(x)$: $\nu(x) \geq \|x\|_{\infty,1/(R+\alpha w)}$, hence when computing i_0 we can ignore indices such that x_i/w_i is smaller than $\alpha \|x\|_{\infty,1/(R+\alpha w)}$ (hence sort a smaller vector).*

Proof Let $k \in \arg\max_i \frac{x_i}{R+\alpha w_i}$ *i.e.*, $\|x\|_{\infty,1/(R+\alpha w)} = \frac{x_k}{R+\alpha w_k}$. We have

$$R^2 = \sum_{i=1}^{n} \left(\frac{x_i}{\nu(x)} - \alpha w_i \right)_+^2 \geq \left(\frac{x_k}{\nu(x)} - \alpha w_k \right)_+^2 \ . \tag{15}$$

Assume $\nu(x) < \|x\|_{\infty,1/(R+\alpha w)}$, then by definition of k, one has $\nu(x) < \frac{x_k}{R+\alpha w_k}$ or equivalently $(\frac{x_k}{\nu(x)} - \alpha w_k) = (\frac{x_k}{\nu(x)} - \alpha w_k)_+ > R > 0$. This leads to a contradiction in Equation (15), so $\nu(x) \geq \|x\|_{\infty,1/(R+\alpha w)}$. ■



Algorithm 1 COMPUTATION OF WEIGHTED EPSILON NORM

input : x, w, α, R

if $\alpha = 0$ **then**
  Trivial case
if $R = 0$ **then**
  Trivial case
Sort $(x_1, \dots, x_n)$ and $(w_1, \dots, w_n)$ by decreasing values of $\frac{x_i}{w_i}$
$S_1^{x^2} = x_{(1)}^2$ $S_1^{w^2} = w_{(1)}^2$ $S_1^{xw} = x_{(1)} w_{(1)}$
for $i = 2, \dots, n+1$ **do**
  $S_i^{x^2} = S_{i-1}^{x^2} + x_{(i)}^2$
  $S_i^{w^2} = S_{i-1}^{w^2} + w_{(i)}^2$
  $S_i^{xw} = S_{i-1}^{xw} + x_{(i)} w_{(i)}$
  if $R^2/\alpha^2 \leq a_i \triangleq \frac{w_{(i)}^2}{x_{(i)}^2} S_i^{x^2} - 2\frac{w_{(i)}}{x_{(i)}} S_i^{xw} + S_i^{w^2}$ **then**
    $i_0 = i - 1$
    break
if $\alpha^2 S_{i_0}^{w^2} - R^2 \neq 0$ **then**
  $\nu = \frac{\alpha^2 S_{i_0}^{xw} - \sqrt{\Delta}}{\alpha^2 S_{i_0}^{w^2} - R^2}$
else
  $\nu = \frac{S_{i_0}^{x^2}}{2\alpha S_{i_0}^{xw}}$
return ν