Convexity of Orthonormal Regularizer and Exclusive Lasso

Mark Schmidt

November 8, 2014

1 Orthonormal Regularizer

Consider the regularizer
$$r(w) = \sum_{i,j} k_{ij} |w_i^T w_j|,$$
on a set of vectors w_i with $k_{ij} \geq 0$, which combines ℓ_2-regularization of each w_i with a regularizer that encourages orthogonality between each w_i and w_j. Surprisingly, this is convex if the matrix K containing elements k_{ii} on the diagonals and k_{ij} on the off-diagonals is positive semi-definite.

Proof. We will show r satisfies the definition of convexity,
$$r(\theta x + (1-\theta)y) \leq \theta r(x) + (1-\theta) r(y),$$
or equivalently
$$\theta r(x) + (1-\theta) r(y) - r(\theta x + (1-\theta)y) \geq 0. \tag{1.1}$$
We have that
$$\theta r(x) = \sum_i k_{ii} \theta \|x_i\|^2 + \sum_{i \neq j} k_{ij} \theta |x_i^T x_j|,$$
$$(1-\theta) r(y) = \sum_i k_{ii} (1-\theta) \|y_i\|^2 + \sum_{i \neq j} k_{ij} (1-\theta) |y_i^T y_j|.$$

$$\begin{aligned}
r(\theta x + (1-\theta)y) &= \sum_i k_{ii} \|\theta x_i + (1-\theta) y_i\|^2 + \sum_{i \neq j} k_{ij} |(\theta x_i + (1-\theta) y_i)^T (\theta x_j + (1-\theta) y_j)| \\
&= \sum_i k_{ii} \left(\theta^2 \|x_i\|^2 + 2\theta(1-\theta) x_i^T y_i + (1-\theta)^2 \|y_i\|^2\right) \\
&+ \sum_{i \neq j} k_{ij} |\theta^2 x_i^T x_j + \theta(1-\theta) x_i^T y_j + \theta(1-\theta) y_i^T x_j + (1-\theta)^2 y_i^T y_j|.
\end{aligned}$$

If we just focus on the terms in (1.1) that depend on k_{ii} we get

$$\begin{aligned}
&\sum_i k_{ii} \left[\theta \|x_i\|^2 + (1-\theta) \|y_i\|^2 - (\theta^2 \|x_i\|^2 + 2\theta(1-\theta) x_i^T y_i + (1-\theta)^2 \|y_i\|^2)\right] \\
&= \sum_i k_{ii} \left[\theta(1-\theta) \|x_i\|^2 + \theta(1-\theta) \|y_i\|^2 - 2\theta(1-\theta) x_i^T y_i\right) \\
&= \theta(1-\theta) \sum_i k_{ii} \|x_i - y_i\|^2.
\end{aligned}$$

If we just focus on the terms in (1.1) that depend on k_{ij} for $i \neq j$ we get

$$
\begin{aligned}
&\sum_{i\neq j} k_{ij}\left[\theta|x_i^T x_j| + (1-\theta)|y_i^T y_j| - |\theta^2 x_i^T x_j + \theta(1-\theta)x_i^T y_j + \theta(1-\theta)y_i^T x_j) + (1-\theta)^2 y_i^T y_j|\right] \\
&\geq \sum_{i\neq j} k_{ij}\left[\theta|x_i^T x_j| + (1-\theta)|y_i^T y_j| - \theta^2|x_i^T x_j| - \theta(1-\theta)|x_i^T y_j + y_i^T x_j| - (1-\theta)^2|y_i^T y_j|\right], \\
&= \sum_{i\neq j} k_{ij}\left[\theta(1-\theta)|x_i^T x_j| + \theta(1-\theta)|y_i^T y_j| - \theta(1-\theta)|x_i^T y_i + y_i^T x_j|\right] \\
&= \theta(1-\theta)\sum_{i\neq j} k_{ij}\left[|x_i^T x_j| + |y_i^T y_j| - |x_i^T y_i + y_i^T x_j|\right] \\
&\geq -\theta(1-\theta)\sum_{i\neq j} k_{ij}\left|x_i^T x_j + y_i^T y_j - x_i^T y_i - y_i^T x_j\right| \\
&= -\theta(1-\theta)\sum_{i\neq j} k_{ij}\left|(x_i - y_i)^T(x_j - y_j)\right| \\
&\geq -\theta(1-\theta)\sum_{i\neq j} k_{ij}\|x_i - y_i\|\|x_j - y_j\|.
\end{aligned}
$$

where we use the triangle inequality ($-|x+y| \geq -|x| - |y|$), then a variant on the reverse triangle inequality, and then Cauchy-Schwartz. To derive the variant on the reverse triangle inequality use the triangle inequality to give

$$|c+d| = |-c-d| = |(a+b-c-d) - (a+b)| \leq |a+b-c-d| + |a+b| \leq |a+b-c-d| + |a| + |b|,$$

which implies

$$|a| + |b| - |c+d| \geq -|a+b-c-d|.$$

Combining terms in both k_{ii} and k_{ij} we get

$$
\begin{aligned}
&\theta(1-\theta)\left[\sum_i k_{ii}\|x_i - y_i\|^2 - \sum_{i\neq j} k_{ij}\|x_i - y_i\|\|x_j - y_j\|\right] \\
&= \theta(1-\theta)\sum_{ij} \bar{k}_{ij}\|x_i - y_i\|\|x_j - y_j\| \\
&= \theta(1-\theta)\sum_{ij} \bar{k}_{ij} v_i v_j \\
&= \theta(1-\theta) v^T K v \\
&\geq 0,
\end{aligned}
$$

where $\bar{k}_{ii} = k_{ii}$ and $\bar{k}_{ij} = -k_{ij}$, v is a vector with elements $\|x_i - y_i\|$, and the inequality holds because we assumed K was positive semi-definite. □

2 Exclusive Lasso

Now consider the regularizer

$$r(w) = \sum_{ij} k_{ij}\|w_i \circ w_j\|_1,$$

where $\circ$ is element-wise multiplication. This is similar to the above except that it encourages each w_i and w_j to use different features rather than being orthogonal (and still uses ℓ_2-regularization of the individual w_i).

Proof. We have that

$$\theta r(x) = \sum_i k_{ii}\theta\|x_i\|^2 + \sum_{i\neq j} k_{ij}\theta\|x_i \circ x_j\|_1,$$

$$(1-\theta)r(y) = \sum_i k_{ii}(1-\theta)\|y_i\|^2 + \sum_{i\neq j} k_{ij}(1-\theta)\|y_i \circ y_j\|_1.$$

$$\begin{aligned}
r(\theta x + (1-\theta)y) &= \sum_i k_{ii}\|\theta x_i + (1-\theta)y_i\|^2 + \sum_{i\neq j} k_{ij}\|(\theta x_i + (1-\theta)y_i)\circ(\theta x_j + (1-\theta)y_j)\|_1 \\
&= \sum_i k_{ii}\left(\theta^2\|x_i\|^2 + 2\theta(1-\theta)x_i^T y_i + (1-\theta)^2\|y_i\|^2\right) \\
&+ \sum_{i\neq j} k_{ij}\left\|\theta^2(x_i\circ x_j) + \theta(1-\theta)(x_i\circ y_j) + \theta(1-\theta)(y_i\circ x_j) + (1-\theta)^2(y_i\circ y_i)\right\|_1.
\end{aligned}$$

The terms in k_{ii} are the same as before. The terms in k_{ij} are

$$\begin{aligned}
&\sum_{i\neq j} k_{ij}\left[\theta\|x_i\circ x_j\|_1 + (1-\theta)\|y_i\circ y_j\|_1 - \left\|\theta^2(x_i\circ x_j) + \theta(1-\theta)(x_i\circ y_j) + \theta(1-\theta)(y_i\circ x_j) + (1-\theta)^2(y_i\circ y_j)\right\|_1\right] \\
&\geq \sum_{i\neq j} k_{ij}\left[\theta\|x_i\circ x_j\|_1 + (1-\theta)\|y_i\circ y_j\|_1 - \theta^2\|x_i\circ x_j\|_1 - \theta(1-\theta)\|(x_i\circ y_j) + (y_i\circ x_j)\|_1 - (1-\theta)^2\|y_i\circ y_j\|_1\right] \\
&= \sum_{i\neq j} k_{ij}\left[\theta(1-\theta)\|x_i\circ x_j\|_1 + \theta(1-\theta)\|y_i\circ y_j\|_1 - \theta(1-\theta)\|(x_i\circ y_i) + (y_i\circ x_j)\|_1\right] \\
&= \theta(1-\theta)\sum_{i\neq j} k_{ij}\left[\|x_i\circ x_j\|_1 + \|y_i\circ y_j\|_1 - \|(x_i\circ y_i) + (y_i\circ x_j)\|_1\right] \\
&= \theta(1-\theta)\sum_{i\neq j} k_{ij}\left[\sum_m |x_{im}x_{jm}| + \sum_m |y_{im}y_{jm}| - \sum_m |x_{im}y_{jm} + y_{im}x_{jm}|\right] \\
&= \theta(1-\theta)\sum_{i\neq j} k_{ij}\sum_m \left[|x_{im}x_{jm}| + |y_{im}y_{jm}| - |x_{im}y_{jm} + y_{im}x_{jm}|\right] \\
&\geq -\theta(1-\theta)\sum_{i\neq j} k_{ij}\sum_m |x_{im}x_{jm} + y_{im}y_{jm} - x_{im}y_{jm} - y_{im}x_{jm}| \\
&= -\theta(1-\theta)\sum_{i\neq j} k_{ij}\sum_m |(x_{im} - y_{im})(x_{jm} - y_{jm})| \\
&\geq -\theta(1-\theta)\sum_{i\neq j} k_{ij}\|(x_i - y_i)\circ(x_j - y_j)\|_1.
\end{aligned}$$

□

Combining terms in both k_{ii} and k_{ij} we get

$$
\begin{aligned}
&\theta(1-\theta)\left[\sum_i k_{ii}\|(x_i-y_i)\circ(x_i-y_i)\|_1 - \sum_{i\neq j} k_{ij}\|(x_i-y_i)\circ(x_j-y_j)\|_1\right] \\
&= \theta(1-\theta)\sum_{ij}\bar{k}_{ij}\|(x_i-y_i)\circ(x_j-y_j)\|_1 \\
&= \theta(1-\theta)\sum_{ij}\bar{k}_{ij}\sum_m (x_{im}-y_{im})(x_{jm}-y_{jm}) \\
&= \theta(1-\theta)\sum_m\sum_{ij}\bar{k}_{ij}(x_{im}-y_{im})(x_{jm}-y_{jm}) \\
&= \theta(1-\theta)\sum_m\sum_{ij}\bar{k}_{ij}v_{im}v_{jm} \\
&= \theta(1-\theta)\sum_m v_{(m:)}^T K v_{(m:)} \\
&\geq 0.
\end{aligned}
$$

where $\bar{k}_{ii} = k_{ii}$ and $\bar{k}_{ij} = -k_{ij}$, $v_{mi} = (x_{im} - y_{im})$, and the inequality holds because K is positive semi-definite.