Structure Learning in Random Fields for Heart Motion Abnormality Detection (addendum)

Mark Schmidt, Kevin Murphy
Computer Science Dept.
University of British Columbia
Vancouver, BC

Glenn Fung, Rómer Rosales
IKM CAD and Knowledge Solutions USA, Inc.
Siemens Medical Solutions
Malvern, PA 19355

Abstract

In this addendum, we characterize the solutions to the projection problems arising in the constrained formulations of the $\ell_{1.\infty}$-regularized and $\ell_{1,2}$-regularized optimization problems. Subsequently, these can be used to derive efficient algorithms for computing these projections, suitable for use in spectral projected gradient methods that solve these problems.

1. $\ell_{1,\infty}$ Projection

In the main paper, to formulate the $\ell_{1,\infty}$-regularized optimization as a constrained problem, we introduce an additional variable α_g for each group g. We then replace each norm $||\mathbf{w}_g||_\infty$ with the variable α_g, and optimize subject to the constraint that $\alpha_g \geq ||\mathbf{w}||_\infty$. This leads to a constrained optimization of the form

$$\min_{\mathbf{w}} f(\mathbf{w}) + \sum_g \alpha_g \quad s.t. \quad \forall_g \alpha_g \geq ||\mathbf{w}_g||_\infty.$$

At a minimizer of this problem, the constraint $\alpha_g \geq ||\mathbf{w}||_\infty$ holds with equality (if it does not, then the objective could be decreased by decreasing α_g, contradicting that we are at a minimizer). Hence, this problem has the same minimum as our original problem.

The non-differentiable constraint $\alpha_g \geq ||\mathbf{w}||_\infty$ can be re-written as a set of inequality constraints of the form $\forall_{i \in g} - \alpha_g \leq w_i \leq \alpha_g$. Further, since the projection separates across groups, we simply need to solve a (potentially small) projection problem for each group in order to compute the projection across groups. In the main paper, we used this representation and an interior-point method to solve the projection problem for each group. Here, we discuss directly solving the individual projection problems that arise for each group. These take the form

$$\min_{\mathbf{w},\alpha} \left|\left| \begin{bmatrix} \mathbf{w}^* \\ \alpha^* \end{bmatrix} - \begin{bmatrix} \mathbf{w} \\ \alpha \end{bmatrix} \right|\right|_2 \quad s.t. \quad \alpha \geq ||\mathbf{w}||_\infty. \tag{1}$$

If $\mathbf{w}$ is simply a scalar, then the projection (1) can be computed by considering three cases:

$$\pi(w,\alpha) = \begin{cases} (w,\alpha), & \text{if } \alpha \geq |w| \\ (\frac{\alpha+|w|}{2} sgn(w), \frac{\alpha+|w|}{2}), & \text{if } \alpha < |w|, \frac{\alpha+|w|}{2} > 0 \\ (0,0), & \text{if } \alpha < |w|, \frac{\alpha+|w|}{2} \leq 0 \end{cases}$$

The first scenario simply returns the input variables since the constraint was already satisfied. The second scenario moves both α and $|w|$ to their average (assuming it is positive), the closest point satisfying the constraint. The final scenario is the case where the average is negative, and in this case the origin is the closest point.

Proof. We use w^* and α^* to refer to the input values, and w and α to refer to the optimal values. If $\alpha^* \geq |w^*|$, then the constraints are already satisfied and the solution (w^*, α^*) has an objective value of 0, the minimum possible value of a norm. This proves the first case, and the remainder of the proof focuses on the case where $\alpha^* < |w^*|$. In this case, $|sgn(w^*)| \geq |sgn(w)|$, since if $sgn(w^*) = -sgn(w)$ we could decrease the objective by setting w to 0. Using this, we have that $(w - w^*)^2 = (|w| - |w^*|)^2$. Using this property and that the objective is non-negative (so squaring it does not change the location of its solution), we can re-write the objective as $(|w| - |w^*|)^2 + (\alpha - \alpha^*)^2$. In the case where $\alpha^* < |w^*|$, we can not have $\alpha < \alpha^*$ or $|w| > |w^*|$, since these will have greater objective values than the solutions $(sgn(w^*)\alpha^*, \alpha^*)$ and $(w^*, |w^*|)$, respectively. Similarly, we can not have $\alpha > |w^*|$ or $|w| < \alpha^*$ in the solution, since these will have greater objective values than the solutions $(w^*, |w^*|)$ and $(sgn(w^*)\alpha^*, \alpha^*)$, respectively. With these constraints that $\alpha^* \leq \alpha \leq |w^*|$ and $\alpha^* \leq |w| \leq |w^*|$, we have that $\alpha = |w|$ in the solution, since if $\alpha > |w|$ then we could decrease the objective by decreasing α or increasing $|w|$. Using this, we can eliminate $|w|$ and re-write the problem as $\min_{w,\alpha}(\alpha - |w^*|)^2 + (\alpha - \alpha^*)^2$ subject to $\alpha \geq 0$ (where the constraint is needed because we have made α equal to a norm). Differentiating the objective and setting its derivative to zero, the unconstrained solution to this problem is $\alpha = (|w^*| + \alpha^*)/2$. Thus, if the constraint is not active we have that $\alpha = (|w^*| + \alpha^*)/2$ and $w = sgn(w^*)(|w^*| + \alpha^*)/2$ (from $\alpha = |w|$ and $|sgn(w^*)| \geq |sgn(w)|$, giving $w = sgn(w^*)\alpha$). The constraint becomes active if $(|w^*| + \alpha^*)/2 \leq 0$, and if the constraint is active then we have the trivial solution $(0, 0)$. □

Moving to the case where $\mathbf{w}$ is a 2-vector with $|w_1| \geq |w_2|$ (they can be permuted if this is not the case), we now must consider four cases:

$$\pi(w_1, w_2, \alpha) = \begin{cases} (w_1, w_2, \alpha), \\ \text{if } \alpha \geq ||\mathbf{w}||_\infty \\ \\ (\frac{\alpha+|w_1|}{2} sgn(w_1), w_2, \frac{\alpha+|w_1|}{2}), \\ \text{if } \alpha < ||\mathbf{w}||_\infty, \frac{\alpha+|w_1|}{2} \geq |w_2| \\ \\ (\frac{\alpha+|w_1|+|w_2|}{3} sgn(w_1), \frac{\alpha+|w_1|+|w_2|}{3} sgn(w_2), \frac{\alpha+|w_1|+|w_2|}{3}), \\ \text{if } \alpha < ||\mathbf{w}||_\infty, \frac{\alpha+|w_1|}{2} < |w_2|, \frac{\alpha+|w_1|+|w_2|}{3} > 0 \\ \\ (0, 0, 0), \\ \text{if } \alpha < ||\mathbf{w}||_\infty, \frac{\alpha+|w_1|+|w_2|}{3} \leq 0 \end{cases}$$

The only surprising aspect of moving from a scalar to a 2-vector is that we move to the average of α and $|w_1|$ if this average is larger than $|w_2|$ (w_2 moves nowhere in this case), and only if this is not satisfied do we consider moving to the average of α, $|w_1|$, and $|w_2|$.

Proof. If $\alpha^* \geq ||\mathbf{w}^*||_\infty$, then the answer is (w_1^*, w_2^*, α^*), since this obtains the minimum objective value of 0. We now focus on the case where $\alpha^* < ||\mathbf{w}^*||_\infty$. By similar reasoning to the scalar case, we will have that $|sgn(w_1^*)| \geq |sgn(w_1)|$, $|sgn(w_2^*)| \geq |sgn(w_2)|$, and $\alpha = |w_1|$. In the solution, we must have that either $\alpha \geq |w_2^*|$ or $\alpha < |w_2^*|$. If $\alpha \geq |w_2^*|$, then the optimal w_2 is w_2^*, while finding the optimal w_1 and α reduces to solving the scalar problem with inputs w_1^* and α^*. If $\alpha < |w_2^*|$, by similar reasoning to the scalar case we will have that $\alpha = |w_2|$ and can formulate the problem as a minimization in terms of α with a non-negativity constraint. The optimal solution will be $(0, 0, 0)$ if this constraint is active, and otherwise it will be the unconstrained solution $\alpha = (|w_1^*| + |w_2^*| + \alpha^*)/3$, where the corresponding values of w_1 and w_2 are $w_1 = sgn(w_1^*)(|w_1^*| + |w_2^*| + \alpha^*)/3$ and $w_2 = sgn(w_2^*)(|w_1^*| + |w_2^*| + \alpha^*)/3$. What remains to be shown is when will $\alpha \geq |w_2^*|$ be satisfied. This can occur only if α in the solution to the scalar problem with inputs w_1^* and α^* satisfies $\alpha \geq |w_2^*|$. □

The generalization of this idea to a p-vector is straightforward, where for p dimensions we must check only $p + 2$ cases. This can be shown recursively, where if we have a p-vector $\mathbf{w}$ we will either move to the average of its absolute values and α, or we will set the smallest element of $\mathbf{w}$ to its input value and solve a $(p - 1)$-dimensional problem. Algorithm 1 outlines a procedure for computing the projection with a general p-vector, at a cost of $\mathcal{O}(p \log p)$, with the dominant cost coming from sorting the absolute values of $\mathbf{w}$.

Algorithm 1 $\ell_{1,\infty}$ Projection

if $\alpha \geq ||\mathbf{w}||_\infty$ **then**
    return {input value satisfies constraints}
end if
$sorted := \{sort(|\mathbf{w}|), 0\}$ {sort absolute values in descending order, append a zero}
s = 0
for k = 1:p **do**
    $s := s + sorted\{k\}$
    $\alpha := (s + \alpha)/(k + 1)$ {trial value for α}
    if $\alpha > 0$ and $\alpha > sorted\{k + 1\}$ **then**
        $\mathbf{w}(|\mathbf{w}| \geq sorted\{k\}) := sgn(\mathbf{w}(|\mathbf{w}| \geq sorted\{k\})\alpha$ {elements with magnitudes greater than α set to $sgn(w_i)\alpha$}
        return
    end if
end for
$\alpha = 0$
$\mathbf{w} = \mathbf{0}$ {return zero}

2. $\ell_{1,2}$ Projection

We now turn to the task of writing the $\ell_{1,2}$ penalty formulation as a constrained optimization, and deriving an efficient means of computing the projection onto the corresponding constraint set. Fortunately, having gone through the exercise above, this turns out to be a very simple task.

As before, we will replace each individual norm $||\mathbf{w}_g||_2$ with a new variable α_g, and minimize subject to the constraint that $\alpha_g \geq ||\mathbf{w}_g||_2$ in order to give a constrained problem with the same solution:

$$\min_{\mathbf{w}} f(\mathbf{w}) + \sum_g \alpha_g \quad s.t. \quad \forall_g \alpha_g \geq ||\mathbf{w}_g||_2. \tag{2}$$

This type of constraint is known as a second-order cone constraint, and as before the projection will separate into solving a problem for each group of the form

$$\min_{\mathbf{w},\alpha} \left|\left| \begin{bmatrix} \mathbf{w}^* \\ \alpha^* \end{bmatrix} - \begin{bmatrix} \mathbf{w} \\ \alpha \end{bmatrix} \right|\right|_2 \quad s.t. \quad \alpha \geq ||\mathbf{w}||_2.$$

Defining the signum of a vector as $sgn(\mathbf{w}) = \mathbf{w}/||\mathbf{w}||_2$, where by convention we will define $sgn(\mathbf{0}) = \mathbf{0}$, the solution of this problem is simply

$$\pi(\mathbf{w}, \alpha) = \begin{cases} (\mathbf{w}, \alpha), & \text{if } \alpha \geq ||\mathbf{w}||_2 \\ (sgn(\mathbf{w})\frac{||\mathbf{w}||_2+\alpha}{2}, \frac{||\mathbf{w}||_2+\alpha}{2}), & \text{if } \alpha < ||\mathbf{w}||_2, \frac{||\mathbf{w}||_2+\alpha}{2} > 0 \\ (0, 0), & \text{if } \alpha < ||\mathbf{w}||_2, \frac{||\mathbf{w}||_2+\alpha}{2} \leq 0 \end{cases}$$

This projection is thus easily computed in $\mathcal{O}(p)$.

Proof. In the case where $\alpha^* \geq ||\mathbf{w}^*||_2$ the solution is simply $(\mathbf{w}^*, \alpha^*)$, so we turn to the case where $\alpha^* < ||\mathbf{w}^*||_2$. In this case, by similar reasoning to the $\ell_{1,\infty}$ case, we will have that $\alpha = ||\mathbf{w}||_2$ in the solution. Further, we have that $\mathbf{w} = \lambda\mathbf{w}^*$ for some non-zero scalar λ because for any $\mathbf{w}^*$ the closest point to a hyper-sphere centered at the origin will have the same direction as the vector $\mathbf{w}^*$. By co-linearity of $\mathbf{w}$ and $\mathbf{w}^*$, we have $||\mathbf{w} - \mathbf{w}^*||_2^2 = (||\mathbf{w}||_2 - ||\mathbf{w}^*||_2)^2$. Using this property and $\alpha = ||\mathbf{w}||_2$, we can re-write our objective as $\min_\alpha (\alpha - ||\mathbf{w}^*||_2)^2 + (\alpha - \alpha_*)^2$ subject to $\alpha \geq 0$. This has the solution $\alpha = (||\mathbf{w}^*||_2 + \alpha^*)/2$ if the constraint is inactive, and zero if the constraint is active. When the constraint is active we set $\mathbf{w}$ to $\mathbf{0}$, while when it is inactive we set $\mathbf{w}$ to the projection of $\mathbf{w}^*$ onto the hyper-sphere of radius α, giving $\mathbf{w} = sgn(\mathbf{w}^*)\alpha$. □