Efficient Projections onto L-1 ball

Ying Zhang

Department of Information Engineering
The Chinese University of Hong Kong

September 22, 2015
1 Motivation: Projected subgradient method
2 Projection onto L-1 Ball
3 Projection on Positive Simplex
Consider the simplest problem in convex optimization,

$$\min_{w} L(w),$$

where $L(w)$ is convex.
Consider the simplest problem in convex optimization,

$$\min_{w} L(w),$$

where $L(w)$ is convex.

**Subgradient descent**

Iteratively compute

$$w^{t+1} = w^t - \eta^t \Delta^t,$$

$\Delta^t$ is the (sub)gradient of $L(w)$ at $w^t$. 
To solve the problem

\[ \min_w L(w) \quad (1) \]

\[ s.t. \quad w \in \mathcal{X}. \quad (2) \]

where \( \mathcal{X} \) is a convex set.
More challenging...

To solve the problem

\[
\min_w \ L(w) \quad (1)
\]

s.t. \quad w \in \mathcal{X}. \quad (2)

where \( \mathcal{X} \) is a convex set.

Projected subgradient descent

Iteratively compute

\[
w^{t+1} = \Pi_{\mathcal{X}}(w^t - \eta^t \Delta^t),
\]

\( \Delta^t \) is the (sub)gradient of \( L(w) \) at \( w^t \).
The algorithm introduced in this presentation is used in the projection when $\mathcal{X} = \{x, \|x\|_1 \leq z\}$.

Meaning we want our solution located inside a L-1 ball (to encourage sparsity, possibly).
The Optimization Problem

In this presentation, we try to solve the following problem,

\[
\min_w \|w - v\|_2^2 \tag{3}
\]

\[
s.t. \|w\|_1 \leq z, \tag{4}
\]

where \(v\) is a given vector and \(z\) is a scaler.

The problem is convex and thus ‘easy’ to solve. This paper tries to provide an efficient algorithm.

Reference

To narrow down the scenario

Lemma

Let $\mathbf{w}^*$ be an optimal solution. Then

$$w_i^* v_i \geq 0, \forall i.$$  

By this lemma, we can only focus on the problems with input $v_i \geq 0$.

If $\mathbf{\hat{v}}$ is the original vector with negative components, we can firstly solve the problem with $v_i = |\mathbf{\hat{v}}_i|$ to obtain the optimal solution $\mathbf{w}^*$, and then the optimal solution for the original solution is

$$\mathbf{\hat{w}}_i^* = \text{sign}(\mathbf{\hat{v}}_i) w_i^*.$$  

It can be proved by math contradiction.
Figure: An illustration
To further narrow down the scenario

To avoid being trivial, we focus on the scenario where \( \|v\| > z \), meaning the original point is outside the ball. Otherwise, \( w^* = v \)

Then the optimal solution \( w^* \) must be on the ball, \( i.e., \|w^*\|_1 = z \).

Furthermore, since \( v_i \geq 0 \), we can have \( w^*_i \geq 0 \). In this way, we only need to search the optimal solution in the region of \( \{w \geq 0, w_i = z \} \).
The reduced optimization problem: projection onto a positive simplex

With the above understanding, we need to solve

\[
\min_w \|w - v\|_2^2
\]

s.t.

\[
\sum_i w_i = z,
\]

\[
w_i \geq 0,
\]
The reduced optimization problem: projection onto a positive simplex

With the above understanding, we need to solve

\[
\begin{align*}
\min_w & \quad \|w - v\|_2^2 \\
\text{s.t.} & \quad \sum_i w_i = z, \\
& \quad w_i \geq 0,
\end{align*}
\]

(5) (6) (7)

As a recall, the original problem is

\[
\begin{align*}
\min_w & \quad \|w - v\|_2^2 \\
\text{s.t.} & \quad \|w\|_1 \leq z,
\end{align*}
\]

(8) (9)
The problem is used to solve the original problem and itself is important, and is considered in:

Reference:
A necessary condition of optimality

The Lagrangian of the problem is given by

$$L(w, \theta, \Delta)_{\Delta_i \geq 0} = \frac{1}{2} \|w - v\|_2^2 + \theta \left( \sum w_i - z \right) - \Delta^T w.$$  

Due to the convexity, the optimal solution $w^*$ will satisfy the KKT condition,

- Stationarity: $\frac{dL(w^*, \theta, \Delta)}{dw_i} = w_i^* - v_i + \theta^* - \Delta_i^* = 0$;
- Complementary slackness: If $w_i > 0$, $\Delta_i^* = 0$.

We can have $w_i^* = v_i - \theta^*$ if $w_i > 0$, i.e.,

$$w_i^* = (v_i - \theta^*).$$

The problem reduces to computing the dual optimal $\theta^*$. 
To compute the dual optimal

By the feasibility of $\mathbf{w}^*$, we can have $\sum w_i = \sum (v_i - \theta^*)^+ = z$. If there are exactly $\rho$ nonzero components in $\mathbf{w}^*$, we can have $\sum_{i=1}^{\rho} (v(i) - \theta^*) = z$ and

$$\theta^* = \frac{1}{\rho} \left( \sum_{i=1}^{\rho} v(i) - z \right),$$

where $v(i)$ is the $i^{th}$ largest component of $\mathbf{v}$. We are left with the problem of finding the value of $\rho$, the number of nonzero components in the optimal solution.
A delima?

So, can we find the number of nonzero components in the optimal solution without computing the optimal solution, in an efficient manner?
A delima?

So, can we find the number of nonzero components in the optimal solution without computing the optimal solution, in an efficient manner?

Yes!

Lemma

Let \( \mathbf{u} \) denote the vector obtained by sorting \( \mathbf{v} \) in a descending order with \( u_i = v(i) \). Then the value of \( \rho \) is given by

\[
\rho = \max \left\{ j : u_j - \frac{1}{j} \left( \sum_{i=1}^{j} u_j - z \right) > 0 \right\}.
\]
The Algorithm

**INPUT:** A vector \( \mathbf{v} \in \mathbb{R}^n \) and a scalar \( z > 0 \)

Sort \( \mathbf{v} \) into \( \mu : \mu_1 \geq \mu_2 \geq \ldots \geq \mu_p \)

Find \( \rho = \max \left\{ j \in [n] : \mu_j - \frac{1}{j} \left( \sum_{r=1}^{j} \mu_r - z \right) > 0 \right\} \)

Define \( \theta = \frac{1}{\rho} \left( \sum_{i=1}^{\rho} \mu_i - z \right) \)

**OUTPUT:** \( \mathbf{w} \) s.t. \( w_i = \max \{ v_i - \theta , 0 \} \)

**Figure:** Algorithm of projection onto a positive simplex

- non-iterative with complexity: \( O(n \log n) \).
Can it be possible that \( \rho = 0 \), i.e., \( u_j - \frac{1}{j} \left( \sum_{i=1}^{j} u_i - z \right) \leq 0 \), \( \forall j \)? In this case the algorithm will fail.
Can it be possible that $\rho = 0$, i.e., $u_j - \frac{1}{j} \left( \sum_{i=1}^{j} u_j - z \right) \leq 0, \forall j$? In this case the algorithm will fail.

Given a (sorted) optimal solution $w^* = (u - \theta)^+$ with $\rho^*$ positive components ($\rho^* > 0$), we have

$$w_{\rho^*} = u_{\rho^*} - \theta^* = u_{\rho^*} - \frac{1}{\rho^*} \left( \sum_{i=1}^{\rho^*} u_i - z \right) > 0.$$ 

**How $\rho$ is computed:**

$$\rho = \max \left\{ j : u_j - \frac{1}{j} \left( \sum_{i=1}^{j} u_j - z \right) > 0 \right\} = \max J$$
Will it always work?

Can it be possible that $\rho = 0$, i.e., $u_j - \frac{1}{j} \left( \sum_{i=1}^{j} u_i - z \right) \leq 0, \forall j$? In this case the algorithm will fail.

Given a (sorted) optimal solution $w^* = (u - \theta)^+$ with $\rho^*$ positive components ($\rho^* > 0$), we have

$$w_{\rho^*} = u_{\rho^*} - \theta^* = u_{\rho^*} - \frac{1}{\rho^*} \left( \sum_{i=1}^{\rho^*} u_i - z \right) > 0.$$

How $\rho$ is computed:

$$\rho = \max \left\{ j : u_j - \frac{1}{j} \left( \sum_{i=1}^{j} u_i - z \right) > 0 \right\} = \max \mathcal{J}$$

$\rho^* \in \mathcal{J}$, we can have $\rho \geq \rho^*$, then the corner case never happens.
The proof is conducted by math contradiction. Can you find any intuitions?
The proof is conducted by math contradiction. Can you find any intuitions?
The formula of computing $\rho$ and the formula of computing the optimal solution are in similar forms:

- $\rho = \max \left\{ j : u_j - \frac{1}{j} \left( \sum_{i=1}^{j} u_j - z \right) \right\}$.
- $w_i = u_i - \theta = u_i - \frac{1}{\rho} \left( \sum_{i=1}^{\rho} u_i - z \right)$.
Conclusion

- Projection onto positive simplex;
- Projection onto L-1 ball;
- Projected subgradient method with 1-norm constraint.
The End