# Solving Projection Problems using Lagrangian Duality

Aaron Mishkin
amishkin@cs.stanford.edu

**Abstract**

We discuss strategies for evaluating Euclidean projection operators for convex sets using Lagrangian duality. First, a recipe for approaching projection problems is given. Our recipe avoids simple mistakes and structures the analysis. Then we show how to apply the recipe to compute projection operators for the unit balls associated with the $\ell_2$, $\ell_\infty$, and $\ell_1$ norms. Each norm ball illustrates a different aspect of the recipe and showcases how it may be used in practice.

## 1 Introduction

We consider the problem of projecting a point $y \in \mathbb{R}^n$ onto a convex set $\mathcal{C} \subset \mathbb{R}^n$,

$$\min_x \left\{ \frac{1}{2}\|x - y\|_2^2 : x \in \mathcal{C} \right\}. \tag{1}$$

This problem can be evaluated in closed-form when $C$ has a simple structure, but often the projection requires solving a "simple" sub-problem. These sub-problems typically correspond to maximizing the Lagrange dual function to obtain optimal dual parameters. We introduce our notation before discussing this procedure in detail.

Given a set $\mathcal{E} \subseteq \mathbb{R}^n$, the indicator for $\mathcal{E}$ is the function

$$\mathbb{1}_{\mathcal{E}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{E} \\ +\infty & \text{otherwise.} \end{cases}$$

The *domain* of a function is the set of inputs for which it is finite,

$$\text{dom }(f) = \{x : f(x) < \infty\}.$$

For any constrained optimization problem $\min \{f(x) : x \in \mathcal{E}\}$, the *essential objective*,

$$f_{\text{ess}}(x) = f(x) + \mathbb{1}_{\mathcal{E}}(x),$$

has domain $\text{dom }(f_{\text{ess}}) = \mathcal{E} \cap \text{dom }(f)$. Minimizing the essential objective is equivalent to solving the original constrained optimization problem. Given a constraint $f_i(x) \leq b_i$, we say that it is tight or active at $x$ if $f_i(x) = b$. The constraint is violated if $f_i(x) > b$ and it strictly satisfied if $f_i(x) < b$. A strictly satisfied constraint is also called inactive.

## 2 Recipe for Solving Projections

Now we lay our procedure for computing the projection operator of a convex set. Some of the steps in the following list may not be necessary for every projection problem, but approaching the computation in a structured manner prevents mistakes and eases the analysis. We develop each step in as much detail as possible so that the approach is clear.

1. **Put into Standard Form**: re-write the projection problem so that it is in the standard form for convex optimization problems. This prevents sign mistakes and other tricky-to-diagnose issues. After doing this, the problem should resemble the following:

$$\min_x \frac{1}{2}\|x-y\|_2^2 \quad \text{s.t.} \quad \begin{aligned} & f_i(x) \le 0 \text{ for all } i \in \{1,\ldots,m\}, \\ & h_j(x) = 0 \text{ for all } j \in \{1,\ldots p\}, \end{aligned} \tag{2}$$

where $f_i$ and $h_j$ are convex functions forming the constraints.

2. **Form the Lagrangian**: the Lagrangian function is given by

$$\mathcal{L}(x,\lambda,\nu) = \frac{1}{2}\|x-y\|_2^2 + \underbrace{\sum_{i=1}^m \lambda_i f_i(x)}_{\text{inequality penalties}} + \underbrace{\sum_{j=1}^p \nu_j h_j(x).}_{\text{equality penalties}} \tag{3}$$

Note that the choice of notation (i.e. using $\lambda, \nu$) is arbitrary. You can use any names for these variables, but Greek letters are conventional. One advantage to putting the constraints in standard form is that the sign of the inequality penalty terms is unambiguous. A way to sanity check these signs is to convince yourself that

$$\sup_{\lambda \ge 0, \nu} \mathcal{L}(x,\lambda,\nu) = f_{\text{ess}}(x), \tag{4}$$

meaning the supremum evaluates to $+\infty$ when $x \notin \mathcal{C}$, i.e., when the primal variables are infeasible.

3. **Compute the Dual Problem**: the dual problem is given by minimizing the Lagrangian over the primal variables $x$,

$$d(\lambda,\nu) = \inf_x \mathcal{L}(x,\lambda,\nu). \tag{5}$$

The domain of the dual problem is $\{(\lambda,\nu) \in \mathbb{R}^{m\times p} : d(\lambda,\nu) > -\infty\}$, which is reversed compared to a convex function. This is because the dual is an infimum over a set of affine functions and so is always concave.

One approach to computing the dual problem is to compute the sub-differential of the Lagrangian and check for the inclusion of 0:

$$0 \in \partial_x \mathcal{L}(x,\lambda,\nu) \iff 0 \in (x-y) + \sum_{i=1}^m \lambda_i \partial_x f_i(x) + \sum_{j=1}^p \nu_j \partial_x h_j(x). \tag{6}$$

Any solution to this equation will be a global minimizer of the Lagrangian. If the $f_i$ and $h_j$ are linear functions in $x$, then $\mathcal{L}(x,\lambda,\nu)$ is a convex quadratic for any choice of $\lambda, \nu$ and the dual problem can be computed in closed-form. The solution is found by solving (6) (which is made easier since the sub-differential is composed only of the gradient). Substituting the solution $x(\lambda,\nu)$ into $\mathcal{L}$ gives the dual problem.

4. **Remove Inactive Constraints**: sometimes it is possible to show that particular inequality constraints must be inactive at the primal solution $x^*$. For example, the box constraint $x_i \in [-1,1]$ is equivalent to $-(x_i+1) \le 0$ and $x-1 \le 0$. Clearly both inequalities cannot be tight at the same time. For any dual optimal $\lambda^*$, if constraint $i$ satisfies $f_i(x^*) < 0$, then the KKT conditions — specifically complementary slackness — require

$$\lambda_i^* f_i(x^*) = 0 \iff \lambda_i^* = 0.$$

That is, we can eliminate this particular constraint from the Lagrangian. Another way to see this is to recall that $\lambda^*, \nu^*$ maximize the Lagrangian evaluated at $x^*$:

$$\begin{aligned}
\sup_{\lambda \geq 0, \nu} d(\lambda, \nu) &= \sup_{\lambda \geq 0, \nu} \mathcal{L}(x^*, \lambda, \nu) \\
&= \sup_{\lambda, \nu} \left\{ \frac{1}{2} \|x^* - y\|_2^2 + \sum_{i=1}^m \lambda_i f_i(x^*) + \sum_{j=1}^p \nu_j h_j(x^*) \right\} \\
&= \sup_{\lambda_l \geq 0, l \neq i} \sup_{\nu} \left\{ \frac{1}{2} \|x^* - y\|_2^2 + \sum_{l=1, l \neq i}^m \lambda_l f_l(x^*) + \sum_{j=1}^p \nu_j h_j(x^*) + \sup_{\lambda_i \geq 0} \{\lambda_i f_i(x^*)\} \right\} \\
&= \sup_{\lambda_l \geq 0, l \neq i} \sup_{\nu} \left\{ \frac{1}{2} \|x^* - y\|_2^2 + \sum_{l=1, l \neq i}^m \lambda_l f_l(x^*) + \sum_{j=1}^p \nu_j h_j(x^*) \right\},
\end{aligned}$$

since $f_i(x^*) < 0$.

5. **Maximize the Dual**: now we maximize the dual problem. We always minimize the negative dual,

$$\max_{\lambda \geq 0, \nu} d(\lambda, \nu) = - \min_{\lambda \geq 0, \nu} \left[ -d(\lambda, \nu) \right],$$

instead of maximizing the dual function. Since $d(\lambda, \nu)$ is concave, $-d(\lambda, \nu)$ is convex and this is a convex minimization problem. We prefer the minimization formulation because convex functions are familiar; while all of our tools for convex minimization can be applied to concave maximization, it is easy to make mistakes since we are less used to concave functions.

Minimizing the negative dual is usually made easier by simplifying the expression for $d(\lambda, \nu)$ after substituting in $x(\lambda, \nu)$. A special situation occurs again when $f_i, h_j$ are linear, in which case the dual function is a concave quadratic. If there are no inequality constraints ($m = 0$), this quadratic function can be maximized in closed form by setting the gradient to 0 and solving for $\nu$.

If there are inequality constraints, maximization is made more difficult by the requirement that $\lambda \geq 0$. In the worst case, we can find the dual solution via the convex minimization problem

$$\min_{\lambda, \nu} -d(\lambda, \nu) \quad \text{s.t.} \quad \lambda \geq 0.$$

If $d(\lambda, \nu)$ is differentiable in both $\lambda$ and $\nu$, then first-order necessary conditions for a dual pair $(\lambda^*, \nu^*)$ to be optimal are

$$\begin{aligned}
\langle \nabla_\lambda \left[ -d(\lambda^*, \nu^*) \right], u - \lambda^* \rangle \geq 0 \\
\nabla_\nu \left[ -d(\lambda^*, \nu^*) \right] = 0,
\end{aligned} \tag{7}$$

where the first equation must hold for all $u \geq 0$. Note that the first line of (7) *does not* hold if $d$ is non-smooth, while the second becomes inclusion of 0 in the sub-differential. If $\lambda_i^* > 0$ for $i \in \{1, \ldots, m\}$, then the subdifferential of the dual function must include 0 at this point,

$$\lambda_i^* > 0 \implies 0 \in \partial_{\lambda_i} d(\lambda^*, \nu^*).$$

Another way to see this is that $\lambda_i^* > 0$ means the constraint $f_i(x^*) \leq 0$ is tight (this follows from complementary slackness) and optimality conditions for the inequality constraint behave like those for an equality constraint. Often this fact can be used to argue by contradiction that $\lambda_i^* = 0$ for some values of $y$.

6. **Recover the Projection**: now we have solved the dual problem and obtained a dual-optimal pair $(\lambda^*, \nu^*)$. The primal-optimal point $x^*$ is given by $x(\lambda^*, \nu^*)$; and the final step of the projection is to recover $x^*$ by solving this equation.

# 3 Examples

Now we consider applying our recipe to several examples. In particular, we compute projection operators for the $\ell_2$, $\ell_\infty$, and $\ell_1$ norm unit balls. Each of these norm balls illustrates a different aspect of the recipe.

## 3.1 Euclidean Ball: Warm Up

One of the simplest projections to compute is that for the Euclidean ball $\mathcal{C} = \{x : \|x\|_2 \leq 1\}$. The naive projection problem for this constraint set is

$$\min_x \frac{1}{2}\|x - y\|_2^2 \quad \text{s.t.} \quad \|x\|_2 \leq 1.$$

This formulation is not ideal since the $\ell_2$ norm is not differentiable everywhere. Minimization the Lagrangian of this problem will require working with subdifferentials. An equivalent and more convenient definition is $\mathcal{C} = \left\{x : \frac{1}{2}\|x\|_2^2 \leq \frac{1}{2}\right\}$, for which we obtain the projection problem

$$\min_x \frac{1}{2}\|x - y\|_2^2 \quad \text{s.t.} \quad \frac{1}{2}\|x\|_2^2 \leq \frac{1}{2}. \tag{8}$$

The Lagrangian associated with Equation 8 is

$$\mathcal{L}(x, \lambda) = \frac{1}{2}\|x - y\|_2^2 + \frac{\lambda}{2}\left(\|x\|_2^2 - 1\right).$$

This formulation also saves us from carrying around factors of 2 (and the possibility of making arithmetic mistakes) by choosing to work with half-Euclidean norms (squared), rather than the squared-norms themselves.

The Lagrangian is a convex quadratic which can be minimized in closed-form by taking the derivative and setting it to 0,

$$\nabla_x \mathcal{L}(x, \lambda) = x - y + \lambda x = 0 \iff x(\lambda) = \frac{y}{1 + \lambda}.$$

Substituting this into the Lagrangian and doing some algebra gives the following dual function:

$$\begin{aligned} d(\lambda) &= \frac{1}{2}\left\|\frac{y}{1+\lambda} - y\right\|_2^2 + \frac{\lambda}{2}\left(\left\|\frac{y}{1+\lambda}\right\|_2^2 - 1\right) \\ &= \frac{\lambda(\lambda+1)}{2(\lambda+1)^2}\|y\|_2^2 - \frac{\lambda}{2} \\ &= \frac{\lambda}{2(\lambda+1)}\|y\|_2^2 - \frac{\lambda}{2}. \end{aligned}$$

The dual problem corresponds to minimizing $-d(\lambda)$, giving

$$\min_{\lambda \geq 0} -d(\lambda) = \min_{\lambda \geq 0} \left\{\frac{\lambda}{2} - \frac{\lambda}{2(\lambda+1)}\|y\|_2^2\right\}$$

We compute the derivative of the dual problem as follows:

$$\begin{aligned} \frac{\partial}{\partial \lambda}[-d(\lambda)] &= \frac{1}{2} - \frac{1}{2}\left[\frac{1}{\lambda+1} - \frac{\lambda}{(\lambda+1)^2}\right]\|y\|_2^2 \\ &= \frac{1}{2} - \frac{1}{2(1+\lambda)^2}\|y\|_2^2, \end{aligned}$$

which is well-defined for all $\lambda \geq 0$. The first-order conditions for a scalar $\lambda^*$ to be optimal for the dual problem are

$$\begin{aligned} &\left\langle \frac{\partial}{\partial \lambda}[-d(\lambda^*)], z - \lambda^* \right\rangle \geq 0 \quad \forall\, z \geq 0 \\ \iff &\left(\frac{1}{2} - \frac{1}{2(1+\lambda^*)^2}\|y\|_2^2\right)(z - \lambda^*) \geq 0 \quad \forall\, z \geq 0. \end{aligned}$$

Taking $z > \lambda^*$ and rearranging gives the condition

$$\|y\|_2 \leq (1 + \lambda^*) \tag{9}$$

Now we do a case analysis on $y$.

- **Case 1**: $\|y\|_2 > 1$. Then $\lambda^* > 0$ by (9) and we deduce that the dual problem is stationary at $\lambda^*$, meaning
$$\frac{\partial}{\partial \lambda}[-d(\lambda^*)] = \frac{1}{2} - \frac{1}{2(1+\lambda^*)^2}\|y\|_2^2 = 0 \iff \lambda^* = \|y\|_2 - 1.$$
Note this is feasible since $\|y\|_2 > 1$.
- **Case 2**: $\|y\|_2 \leq 1$. Suppose by way of contradiction that $\lambda^* > 0$. Then taking $z = 0$ in the necessary conditions and dividing by $-\lambda^* < 0$ gives
$$\left(\frac{1}{2} - \frac{1}{2(1+\lambda^*)^2}\|y\|_2^2\right) \leq 0 \iff \|y\|_2 \geq (1 + \lambda^*)^2,$$
which cannot hold since $\|y\|_2 \leq 1$. We conclude $\lambda^* = 0$ by contradiction. This makes sense because $y$ is in the interior of the constraint set.

We have shown that $\lambda^* = 0$ if $\|y\|_2 \leq 1$ (i.e. $y \in \mathcal{C}$) and $\lambda^* = \|y\|_2 - 1$ otherwise. Plugging this into $x(\lambda)$ gives the projection to be

$$x^* = \begin{cases} \frac{y}{\|y\|_2} & \text{if } \|y\| > 1 \\ y & \text{otherwise.} \end{cases}$$

This result matches our expectations for such a simple projection. Working it out through Lagrangian duality is not the easiest way to derive this projection operator, but it is a good exercise.

## 3.2 Infinity-Norm Ball: Constraint Elimination

Consider projecting onto the $\ell_\infty$ unit ball,

$$\min_x \frac{1}{2}\|x - y\|_2^2 \quad \text{s.t.} \quad \|x\|_\infty \leq 1.$$

There are several ways to formulate the constraints. One approach is to observe that

$$\mathcal{C} = \{x : \|x\|_\infty \leq 1\} = \left\{x : \max_{i \in \{1,\ldots,n\}} |x_i| \leq 1\right\} = \{x : -1 \leq x_i \leq 1 \, \text{for all} \, i \in \{1, \ldots, n\}\}.$$

Using this reformulation of the constraints gives the following optimization problem in standard form:

$$\min_x \frac{1}{2}\|x - y\|_2^2 \quad \text{s.t.}\, x_i - 1 \leq 0,\, -(x_i + 1) \leq 0, \tag{10}$$

for which the Lagrangian function is

$$\mathcal{L}(x, \lambda, \nu) = \frac{1}{2}\|x - y\|_2^2 + \langle \lambda, x - 1 \rangle - \langle \nu, x + 1 \rangle.$$

This is a convex quadratic since the constraint functions are linear. It can be minimized by setting the gradient to zero, we which do as follows:

$$\begin{aligned} \nabla_x \mathcal{L}(x\lambda, \nu) &= (x - y) + \lambda - \nu = 0 \\ \iff x(\lambda, \nu) &= y - \lambda + \nu. \end{aligned}$$

Observe that $\nu - \lambda$ will span $\mathbb{R}^n$ despite the non-negativity constraints on these variables. This is necessary to ensure that $x \in \mathcal{C}$. Finally, we form the dual problem by substituting $x(\lambda, \nu)$ into $\mathcal{L}(x, \lambda, \nu)$:

$$\begin{aligned} d(\lambda, \nu) &= \frac{1}{2}\|y - \lambda + \nu - y\|_2^2 + \langle \lambda, y - \lambda + \nu - 1 \rangle - \langle \nu, y - \lambda + \nu + 1 \rangle \\ &= \frac{1}{2}\|\nu - \lambda\|_2^2 + \langle \lambda - \nu, y - (\lambda - \nu) \rangle - \langle \lambda + \nu, 1 \rangle \\ &= -\frac{1}{2}\|\nu - \lambda\|_2^2 - \langle \nu - \lambda, y \rangle + \langle \lambda + \nu, 1 \rangle \\ &= -\frac{1}{2}\|(\nu - \lambda) + y\|_2^2 - \langle \nu + \lambda, 1 \rangle - \frac{1}{2}\|y\|_2^2. \end{aligned}$$

Following the recipe, we can try to eliminate constraints or proceed directly to maximizing the dual problem. First we show how to directly solve the dual, which illustrates how optimality conditions can be used to deduce when certain dual parameters must be 0.

### 3.2.1 Direct Solution

The dual function is separable over dimension, meaning

$$\begin{aligned} d(\lambda, \nu) &= -\frac{1}{2}\|(\nu - \lambda) + y\|_2^2 - \langle \nu + \lambda, 1 \rangle - \frac{1}{2}\|y\|_2^2 \\ &= \sum_{i=1}^{n} -\frac{1}{2}\left((\nu_i - \lambda_i) + y_i\right)^2 - \nu_i - \lambda_i + \frac{1}{2}\|y\|_2^2. \end{aligned}$$

This means we can drop the subscript $i$ and focus our analysis on a generic one-dimensional problem. We also drop the constant term $\frac{1}{2}\|y\|_2^2$ since it does not affect the solution. The full $n$-dimensional solution can be obtained by combining the solutions to the $n$ one-dimensional problems.

The one-dimensional dual problem is

$$\max_{\lambda \geq 0, \nu \geq 0} d(\lambda, \nu) = -\min_{\lambda \geq 0, \nu \geq 0} \frac{1}{2}\left((\nu - \lambda) + y\right)^2 + \nu + \lambda.$$

This is differentiable for all $\lambda, \nu$ and , computing the derivatives of $d(\lambda, \nu)$, we find that the first-order necessary conditions for dual optimality are

$$\begin{aligned} \frac{\partial}{\partial \lambda}\left[-d(\lambda^*, \nu^*)\right](u - \lambda^*) &= (\lambda^* - \nu^* - y + 1)(u - \lambda^*) \geq 0 \\ \frac{\partial}{\partial \nu}\left[-d(\lambda^*, \nu^*)\right](v - \nu^*) &= (\nu^* - \lambda^* + y + 1)(u - \nu^*) \geq 0, \end{aligned}$$

for all $u, v \geq 0$. Choosing $u > \lambda^*$ and $v > \nu^*$, we deduce

$$-(y + 1) \leq \nu - \lambda \leq 1 - y. \tag{11}$$

We again resort to a case analysis on $y$.

- **Case 1**: $y > 1$. Then $\nu^* - \lambda^* < 0$ by (11), which can only occur if $\lambda^* > 0$. Since the constraint on $\lambda^*$ is not tight, the dual objective must be stationary at $\lambda^*$:

$$\frac{\partial}{\partial \lambda}[-d(\lambda^*, \nu^*)] = 0 \iff \lambda^* = \nu^* + y - 1.$$

Now, suppose that $\nu^* > 0$. The dual must also be stationary in $\nu^*$, which gives

$$\nu^* - \lambda^* + y + 1 = 0.$$

However, plugging $\lambda^*$ into this expression yields $2 = 0$, which is a contradiction. We shown that $\lambda^* = y - 1, \nu^* = 0$ in this case.

- **Case 2**: $y < -1$. This case is symmetric to the above and left as an exercise.
- **Case 3**: $y \in [-1, 1]$. Suppose $\lambda^* > 0$. Stationarity of the dual again implies

$$\lambda^* = \nu^* + y - 1 > 0,$$

which, since $y \leq 1$, forces $\nu^* > 0$. Then the dual must also be stationary in $\nu^*$ and we return to the same contradiction as in Case 1. We have shown that $\lambda^* = 0$. It is an exercise for the reader to to show that $\nu^* = 0$ as well (proceed again by contradiction).

We can now compute the projection. Plugging $\lambda^*, \nu^*$ into the expression for $x(\lambda, \nu)$ gives the optimal primal variables as:

$$x_i^* = \begin{cases} 1 & \text{if } y_i > 1 \\ -1 & \text{if } y_i < -1 \\ y_i & \text{otherwise.} \end{cases}$$

That is, we clip each value of $y_i$ to the interval $[-1, 1]$.

#### 3.2.2 Constraint Elimination

The last section ignored the fact that only one of the two constraints on $x_i$ can be active at $x_i^*$. As a result, we were forced to argue by contradiction to show which constraints were inactive and when. That is, $x_i^* = 1$ or $x_i^* = -1$, but not both! Considering the one-dimensional component of the dual function,

$$d(\lambda, \nu) - \frac{1}{2}\left((\nu - \lambda) + y\right)^2 - \nu - \lambda,$$

we again proceed by case analysis, but a much simpler one this time.

- **Case 1**: $y \geq 0$. Then $x^* \geq 0$ as well (this can be verified by checking the primal objective). It follows that $x^* \neq -1$ and the constraint associated with $\nu$ must be inactive. We deduce from complementary slackness that $\nu^* = 0$. The dual problem reduces to

$$\min_{\lambda \geq 0} \frac{1}{2}\|y - \lambda\|^2 + \lambda,$$

for which the necessary optimality conditions are

$$(\lambda^* - y + 1)(u - \lambda^*) \geq 0, \tag{12}$$

for $u \geq 0$. In particular, choosing $u > \lambda^*$ gives

$$\lambda^* \geq y - 1 \tag{13}$$

If $y > 1$, then $\lambda^* > 0$ by (13) which, moreover, must hold with equality. If $y \leq 1$, then assuming $\lambda^* > 0$ produces a contradiction with (12); to see this, take $u = 0$ and observe

$$\lambda^* \leq y - 1 \leq 0.$$

Thus, $\lambda^* = 0$ if $y^* \leq 1$ and $\lambda^* = y - 1$ otherwise.

- **Case 2**: $y < 0$. This case is symmetric to the first case and is left to the reader.

Note that this approach gives the same solution as the direct analysis.

### 3.3 L1-Norm Ball: Algorithmic Solutions

The final problem we consider is that of projecting onto the $\ell_1$ unit ball,

$$\min_x \frac{1}{2}\|x - y\|_2^2 \quad \text{s.t.} \quad \|x\|_1 \leq 1.$$

The optimization problem in standard form is

$$\min_x \frac{1}{2}\|x - y\|_2^2 \quad \text{s.t.} \quad \|x\|_1 - 1 \leq 0, \tag{14}$$

with Lagrangian function,

$$\mathcal{L}(x, \lambda) = \frac{1}{2}\|x - y\|_2^2 + \lambda\left(\|x\|_1 - 1\right).$$

The Lagrangian is convex, but non-smooth in $x$, so we check for inclusion of 0 in the sub-differential,

$$0 \in \partial\mathcal{L}(x, \lambda) \iff (y - x) \in \lambda\partial\|x\|_1,$$

Like the dual function in the previous section, this equation is separable over dimension. As a result, we may check the optimality conditions independently for each $i \in \{1, \ldots, n\}$ as follows:

$$y_i - x_i \in \begin{cases} \{-\lambda\} & \text{if } x_i < 0 \\ [-\lambda, \lambda] & \text{if } x_i = 0 \\ \{\lambda\} & \text{if } x_i > 0. \end{cases}$$

Suppose $|y_i| \leq \lambda$, but $x_i > 0$. Then $y_i - x_i < \lambda$, which is a contradiction. Thus, $x_i \leq 0$. A symmetric argument shows $x_i \geq 0$, which forces $x_i = 0$. Arguing in the same fashion (the remainder is left as an exercise), we find

$$x_i(\lambda) = \text{sign}(y_i)\left(|y_i| - \lambda\right)_+.$$

The dual function is thus

$$\begin{aligned} d(\lambda) &= \frac{1}{2}\|\text{sign}(y_i)\left(|y_i| - \lambda\right)_+ - y\|_2^2 + \lambda(\|\text{sign}(y_i)\left(|y_i| - \lambda\right)_+\|_1 - 1) \\ &= \frac{1}{2}\|\min\{|y|, \lambda\}\|_2^2 + \lambda(\|\left(|y_i| - \lambda\right)_+\|_1 - 1). \end{aligned}$$

This dual problem is not as nice as those for the $\ell_2$ or $\ell_\infty$ balls. It doesn't separate over dimension, since $\lambda$ is scalar, and maximizing the dual problem appears to be hard, since, although it is only a one-dimensional problem, it is non-smooth.

Rather than directly maximizing $d$, we can leverage the form of $x(\lambda)$ and the problem constraints to compute $\lambda^*$. If $y \notin \mathcal{C}$, then the solution $x(\lambda)$ will lie on the boundary of $\mathcal{C}$; this follows from optimality conditions for the primal problem. As a result, we argue that

$$\|x(\lambda)\|_1 = 1 \iff \sum_{i=1}^n \left(|y_i| - \lambda\right)_+ = 1.$$

In other words, we seek a root of $h(\lambda) = \sum_{i=1}^{n} (|y_i| - \lambda)_+ - 1$. To this end, observe that $h$ is continuous, piece-wise linear, and decreasing in $\lambda$. Moreover, $h(0) > 0$ since $\|y\|_1 > 1$ and $h(\max_i |y_i|) = -1$, which guarantees that $h$ attains a root on the interval $[0, \max_i |y_i|]$.

To find this root, we sort the set $\{|y_i|\}_{i=1}^{n}$ (in increasing order) to get an ordering $\left(y_{\sigma(1)}, \ldots, y_{\sigma(n)}\right)$. The root of $h$ must occur in one interval $[y_{\sigma(j)}, y_{\sigma(j+1)}]$, which can be found by evaluating $h$ at each point in the ordering and observing when the sign of $h$ changes. On this interval, $h(\lambda)$ is a linear function with form,

$$h(\lambda) = \sum_{i=j+1}^{n} |y_{\sigma(j+1)}| - (n-j)\lambda - 1.$$

Since $h$ is linear, we can easily compute a root as follows:

$$\lambda^* = \frac{\sum_{i=j+1}^{n} |y_{\sigma(j+1)}| - 1}{n-j}.$$

It remains to verify that $(x(\lambda^*), \lambda^*)$ satisfy the KKT conditions, which are necessary and sufficient for primal-dual optimality in this problem. This is a useful exercise for the reader.

# 4 Further Reading

Projections are a crucial part of practical first-order optimization algorithms for constrained problems. Such problems arise frequently in machine learning, where the constraints express limitations on the parameter space. For example, box constraints in the style of the $\ell_\infty$-norm ball appear in the standard formulation of support vector machines (Boser et al., 1992). Constraints using or related to the $\ell_1$-ball also appear frequently; Duchi et al. (2008) derive a more efficient version of the method presented here and discuss strategies for randomized evaluation of projections. Yu et al. (2012) consider projections onto intersections of norm balls.

# 5 Conclusion

We give a simple recipe for computing projection operators arising in convex optimization problems. Although our framework generalizes to non-Euclidean projections, we restrict our discussion to the Euclidean distance for simplicity. The recipe is illustrated in detail by the computation of projections for the $\ell_2$, $\ell_1$, and $\ell_\infty$ norm balls.

# References

[BGV92] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. "A training algorithm for optimal margin classifiers". In: *Proceedings of the fifth annual workshop on Computational learning theory*. 1992, pp. 144–152.

[Duc+08] John Duchi et al. "Efficient projections onto the l 1-ball for learning in high dimensions". In: *Proceedings of the 25th international conference on Machine learning*. 2008, pp. 272–279.

[YSF12] Adams Wei Yu, Hao Su, and Li Fei-Fei. "Efficient euclidean projections onto the intersection of norm balls". In: *arXiv preprint arXiv:1206.4638* (2012).