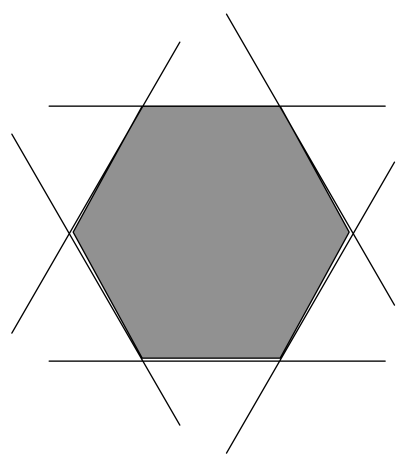
A first look at duality

It is time for our first look at one of the most important concepts in convex optimization (and all of convex analysis in general): duality. With the separating and supporting hyperplane theorems, we have had a glimpse of how hyperplanes and halfspaces interact with convex sets. In this section, we will take this even further, and show that for our closest point problem, the search for the optimal vector (points in $\mathbb{R}^N$) can be equivalently written as a search for a hyperplane (linear functional on $\mathbb{R}^N$).

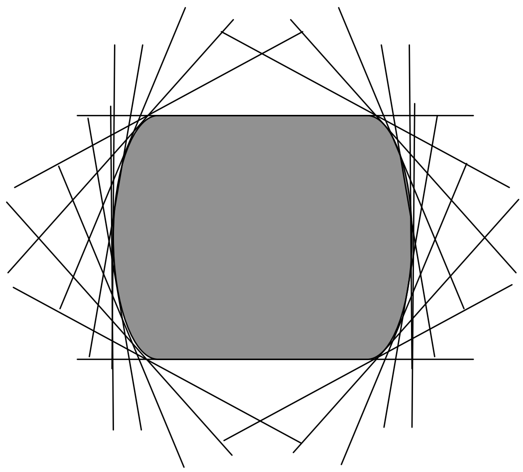
First, let me point out something which should almost be obvious at this point:

A closed convex set is equal to the intersection of (halfspaces defined by) its supporting hyperplanes.

A polygon is the intersection of a finite number of halfspaces:



A general closed convex set is the intersection of a possibly infinite number of halfspaces:



That this is true can almost immediately be deduced from work we have done already. Let $\mathcal{H}_{\mathcal{C}}$ be the intersection of the halfspaces defined by supporting hyperplanes:

$$\mathcal{H}_{\mathcal{C}} = \bigcap_{\boldsymbol{z}\in\mathrm{bd}\,\mathcal{C}} \; \bigcap_{(\boldsymbol{a},b)\in\mathcal{A}(\boldsymbol{z})} \{\boldsymbol{x} \; : \; \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \leq b\}$$

where $\mathcal{A}(\boldsymbol{z}) = \{\text{set of supporting hyperplanes at } \boldsymbol{z}\}$

$$= \{(\boldsymbol{a}, b) \; : \; \{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = b\} \text{ is a supporting hyperplane at } \boldsymbol{z}\}.$$

If $\boldsymbol{x} \in \mathcal{C}$, then $\boldsymbol{x}$ is also in every halfspace in the intersection above, simply by the definition of "supporting hyperplane", thus $\boldsymbol{x} \in \mathcal{C} \Rightarrow \boldsymbol{x} \in \mathcal{H}_{\mathcal{C}}$. If $\boldsymbol{x} \notin \mathcal{C}$, then there is at least one halfspace which does not contain $\boldsymbol{x}$; in particular, we can choose the hyperplane that supports $\mathcal{C}$ at $P_{\mathcal{C}}(\boldsymbol{x})$ with normal vector $\boldsymbol{x} - P_{\mathcal{C}}(\boldsymbol{x})$. Thus $\boldsymbol{x} \notin \mathcal{C} \Rightarrow \boldsymbol{x} \notin \mathcal{H}_{\mathcal{C}}$.

Moral: A (closed) convex set can be thought of as a collection of points, or a collection of hyperplanes.

To continue, it will be critical for you to appreciate that vectors play two distinct roles when we talk about convexity. The usual role (the "primal" role, if you will) is as a point (set of coordinates) in $\mathbb{R}^N$. The other role (the "dual" role) is as a linear functional: any $\boldsymbol{a} \in \mathbb{R}^N$ defines a functional $f(\boldsymbol{v}) = \langle \boldsymbol{v}, \boldsymbol{a} \rangle = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{v}$ over $\boldsymbol{v} \in \mathbb{R}^N$. When coupled with a scalar, these functionals can be used to define subsets of $\mathbb{R}^N$ (hyperplanes and halfspaces).

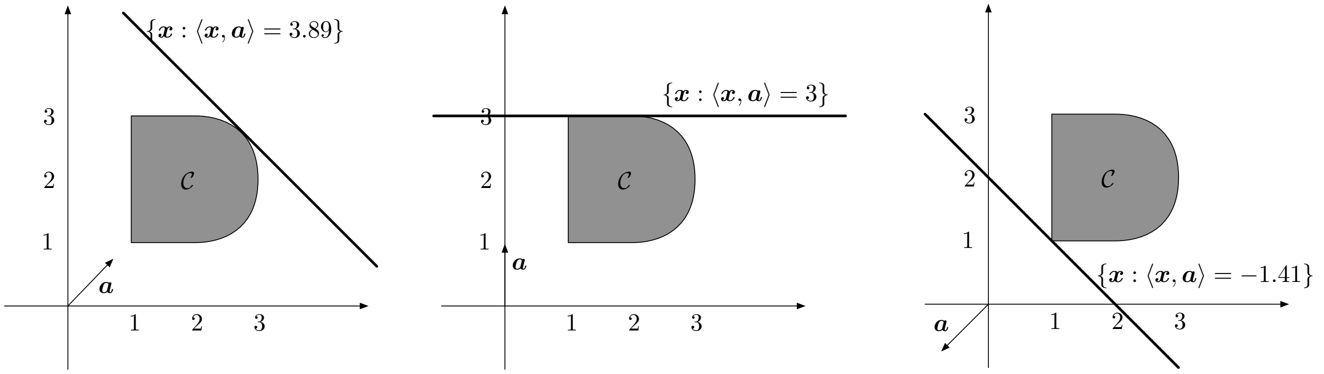
To develop this appreciation, let's look at another way to describe a convex set that is closely related to the infinite-intersection-of-hyperplanes interpretation. The **support functional** of a convex set $\mathcal{C}$ is

$$h(\boldsymbol{a}) = \sup_{\boldsymbol{x} \in \mathcal{C}} \ \langle \boldsymbol{x}, \boldsymbol{a} \rangle.$$

This is a function that takes linear functionals (which again are just vectors) as arguments, and returns the maximum[1] value of that linear functional over $\mathcal{C}$. Geometrically, we take take the halfspace $\{\boldsymbol{x} \ : \ \langle \boldsymbol{x}, \boldsymbol{a} \rangle \leq b\}$ and see how large we need to make b so that it contains $\mathcal{C}$ (of course, b might be negative).

[1] We are writing sup instead of max here since if $\mathcal{C}$ is not bounded, it very well may be that $h(\boldsymbol{a})$ is ∞. If $\mathcal{C}$ is closed and $h(\boldsymbol{a})$ is finite, then there is at least one point in $\mathcal{C}$ which achieves the maximum.

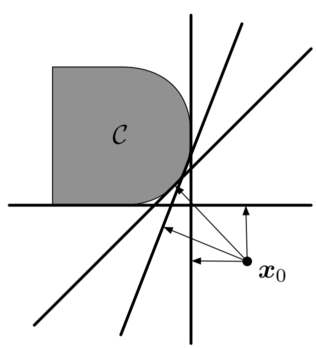
Examples:



$$\boldsymbol{a} = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix},\ h(\boldsymbol{a}) = 3.89 \qquad \boldsymbol{a} = \begin{bmatrix}0\\1\end{bmatrix},\ h(\boldsymbol{a}) = 3 \qquad \boldsymbol{a} = -\frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix},\ h(\boldsymbol{a}) = -1.41$$

Note that scaling $\boldsymbol{a}$ scales $h(\boldsymbol{a})$ by the same amount; for all $\beta \geq 0$, we can see that $h(\beta\boldsymbol{a}) = \beta\, h(\boldsymbol{a})$.

We can use the support functional to recast the closest point problem as a search for a linear functional (i.e. hyperplane or halfspace). We will see that the projection of $\boldsymbol{x}_0$ onto $\mathcal{C}$ can be found by searching over the hyperplanes which separate $\boldsymbol{x}_0$ from $\mathcal{C}$ and choosing the one which is the *maximum* distance from $\boldsymbol{x}_0$:



This maximal hyperplane will support $\mathcal{C}$ at the solution $P_{\mathcal{C}}(\boldsymbol{x}_0)$.

We will now make this geometrically intuitive fact mathematically precise. Let $\mathcal{C}$ be a closed and convex set, let $\boldsymbol{x}_0$ be a point in $\mathbb{R}^N$, and let $\hat{\boldsymbol{x}} = P_{\mathcal{C}}(\boldsymbol{x}_0)$ be the projection of $\boldsymbol{x}_0$ onto $\mathcal{C}$. Then the distance from $\boldsymbol{x}_0$ to $\mathcal{C}$ can be written in two different ways:

$$\begin{aligned} d := \|\boldsymbol{x}_0 - P_{\mathcal{C}}(\boldsymbol{x}_0)\|_2 &= \min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x}_0 - \boldsymbol{x}\|_2 \\ &= \max_{\boldsymbol{a}\in\mathbb{R}^N} \ \langle \boldsymbol{x}_0, \boldsymbol{a}\rangle - h(\boldsymbol{a}) \ \text{ subject to } \ \|\boldsymbol{a}\|_2 \le 1. \end{aligned}$$

We will show this first for the case where $\boldsymbol{x}_0 = \mathbf{0}$; the result for general $\boldsymbol{x}_0$ will then follow easily just by shifting everything relative to $\boldsymbol{x}_0$. We want to show that the distance from $\mathcal{C}$ to the origin,

$$d := \min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x}\|_2,$$

can also be computed as

$$d = \max_{\|\boldsymbol{a}\|_2\le 1} -h(\boldsymbol{a}).$$

We will assume that $\mathbf{0} \notin \mathcal{C}$, as otherwise there is not much to talk about.

First, note that $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{a}\rangle = h(\boldsymbol{a})\}$ is a supporting hyperplane of $\mathcal{C}$, and this hyperplane separates $\mathcal{C}$ from $\mathbf{0}$ if and only if $h(\boldsymbol{a}) \le 0$. For if $h(\boldsymbol{a}) \le 0$, then $\langle \mathbf{0}, \boldsymbol{a}\rangle = 0 \ge h(\boldsymbol{a})$ and by definition $\langle \boldsymbol{x}, \boldsymbol{a}\rangle \le h(\boldsymbol{a})$ for all $\boldsymbol{x} \in \mathcal{C}$. Conversely, if $h(\boldsymbol{a}) > 0$, then the half space $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{a}\rangle \le h(\boldsymbol{a})\}$ contains both interior points of $\mathcal{C}$ and $\mathbf{0}$.

Now consider any supporting hyperplane which does separate $\mathbf{0}$ and $\mathcal{C}$:

$$\langle \boldsymbol{x}, \boldsymbol{a}\rangle \le h(\boldsymbol{a}) \ \text{ for all } \ \boldsymbol{x} \in \mathcal{C}, \quad \|\boldsymbol{a}\|_2 = 1, \quad \text{where } \ h(\boldsymbol{a}) \le 0.$$

(We may assume $\boldsymbol{a}$ is unit norm above; since $h(\beta\boldsymbol{a}) = \beta h(\boldsymbol{a})$ for $\beta \geq 0$, scaling $\boldsymbol{a}$ simply scales both sides of the inequality on the left by the same amount.) For any $\boldsymbol{x} \in \mathcal{C}$, we have

$$\|\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{v}\|_2=1} \langle \boldsymbol{x}, \boldsymbol{v} \rangle \geq \langle \boldsymbol{x}, -\boldsymbol{a} \rangle \geq -h(\boldsymbol{a}),$$

meaning that $\boldsymbol{x}$ is a distance of at least $-h(\boldsymbol{a})$ from the origin. This holds for every point in $\mathcal{C}$, so

$$d \geq -h(\boldsymbol{a}).$$

The inequality above holds uniformly over all unit norm $\boldsymbol{a}$ with $h(\boldsymbol{a}) \leq 0$, and extends trivially to $\boldsymbol{a}$ with $h(\boldsymbol{a}) > 0$. Thus

$$d \geq \max_{\|\boldsymbol{a}\|_2=1} -h(\boldsymbol{a}).$$

Now we show that there is a particular $\hat{\boldsymbol{a}}$ that achieves equality, $-h(\hat{\boldsymbol{a}}) = d$. We know by the projection theorem that there is a unique close point to the origin, which we call $\hat{\boldsymbol{x}}$. We know that $\|\hat{\boldsymbol{x}}\|_2 = d$ and this is the only point in $\mathcal{C}$ for which this is true. Now consider the closed ball of radius d around the origin:

$$\mathcal{B}_d = \{\boldsymbol{x} \ : \ \|\boldsymbol{x}\|_2 \leq d\}.$$

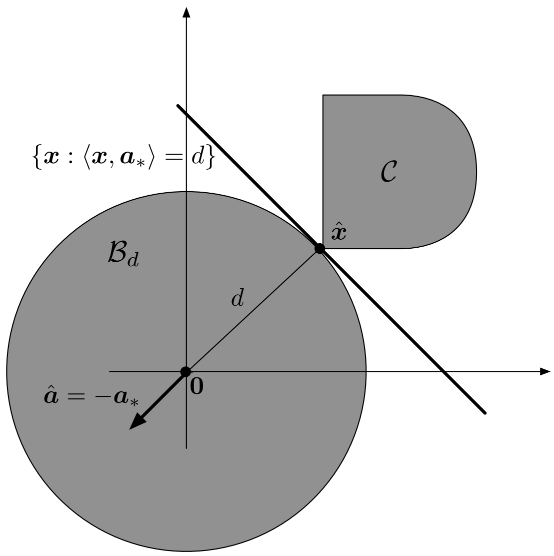
$\hat{\boldsymbol{x}}$ is also in $\mathcal{B}_d$ (on the boundary), and so there is a hyperplane which separates $\hat{\boldsymbol{x}}$ from the interior of $\mathcal{B}_d$. This hyperplane must also support $\mathcal{B}_d$ at $\hat{\boldsymbol{x}}$, so it can be expressed as

$$\{\boldsymbol{x} \ : \ \langle \boldsymbol{x}, \boldsymbol{a}_* \rangle = d\} \quad \text{with } \|\boldsymbol{a}_*\|_2 = 1, \tag{1}$$

with $\langle \boldsymbol{x}, \boldsymbol{a}_* \rangle \leq d$ for all $\boldsymbol{x} \in \mathcal{B}_d$ and $\langle \boldsymbol{x}, \boldsymbol{a}_* \rangle \geq d$ for all $\boldsymbol{x} \in \mathcal{C}$. This second statement means that

$$d = \min_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}, \boldsymbol{a}_* \rangle = -\max_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}, -\boldsymbol{a}_* \rangle = -h(\hat{\boldsymbol{a}}), \quad \hat{\boldsymbol{a}} = -\boldsymbol{a}_*.$$

The following picture might help you track the argument above:



Note also that normal vector $\hat{\boldsymbol{a}}$ must also be **aligned** with (points in the same direction as) the error $\mathbf{0} - \hat{\boldsymbol{x}}$. We know that $\hat{\boldsymbol{x}}$ is in the supporting hyperplane described by (1), and its norm is also d since it is the closest point. By the Cauchy-Schwarz inequality,

$$\langle \hat{\boldsymbol{x}}, \boldsymbol{a}_* \rangle = d, \ \|\boldsymbol{a}_*\|_2 = 1, \ \|\hat{\boldsymbol{x}}\|_2 = d \quad \Rightarrow \quad \boldsymbol{a}_* = \frac{\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2},$$

and so

$$\hat{\boldsymbol{a}} = -\boldsymbol{a}_* = \frac{-\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2^2}.$$

We have shown that we can write the "closest point to the origin" problem in two ways:

$$\min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{a}\|_2\leq 1} -h(\boldsymbol{a}).$$

For the general closest point problem, we use a simple change of variable:

$$\begin{aligned}
\min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x}-\boldsymbol{x}_0\|_2 &= \min_{\boldsymbol{x}'\in\mathcal{C}-\boldsymbol{x}_0} \|\boldsymbol{x}'\|_2 \\
&= \max_{\|\boldsymbol{a}\|_2\leq 1} - \sup_{\boldsymbol{x}'\in\mathcal{C}-\boldsymbol{x}_0} \langle \boldsymbol{x}', \boldsymbol{a}\rangle \\
&= \max_{\|\boldsymbol{a}\|_2\leq 1} - \sup_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}-\boldsymbol{x}_0, \boldsymbol{a}\rangle \\
&= \max_{\|\boldsymbol{a}\|_2\leq 1} \langle \boldsymbol{x}_0, \boldsymbol{a}\rangle - \sup_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}, \boldsymbol{a}\rangle \\
&= \max_{\|\boldsymbol{a}\|_2\leq 1} \langle \boldsymbol{x}_0, \boldsymbol{a}\rangle - h(\boldsymbol{a}).
\end{aligned}$$

It is also true that $\hat{\boldsymbol{a}}$ which maximizes the dual will be aligned with the error $\hat{\boldsymbol{x}} - \boldsymbol{x}_0$ of the primal; with

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x}-\boldsymbol{x}_0\|_2, \quad \hat{\boldsymbol{a}} = \arg\max_{\|\boldsymbol{a}\|_2\leq 1} \langle \boldsymbol{x}_0, \boldsymbol{a}\rangle - h(\boldsymbol{a}),$$

we have

$$\hat{\boldsymbol{a}} = \frac{\boldsymbol{x}_0 - \hat{\boldsymbol{x}}}{\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2}.$$

Note that for this problem, the uniqueness of the primal solution $\hat{\boldsymbol{x}}$ also implies the uniqueness of the dual solution $\hat{\boldsymbol{a}}$.

Summary: Closest Point Duality

Let $\mathcal{C}$ be a closed convex set, and let $\boldsymbol{x}_0 \in \mathbb{R}^N$ be a fixed point. On the previous pages, we have established:

1. For all $\boldsymbol{x} \in \mathcal{C}$ and $\boldsymbol{a}$ with $\|\boldsymbol{a}\|_2 \leq 1$,
$$\|\boldsymbol{x} - \boldsymbol{x}_0\|_2 \;\geq\; \langle \boldsymbol{x}_0, \boldsymbol{a}\rangle - h(\boldsymbol{a}),$$
where
$$h(\boldsymbol{a}) = \sup_{\boldsymbol{y}\in\mathcal{C}} \langle \boldsymbol{y}, \boldsymbol{a}\rangle.$$
This bound holds uniformly for feasible $\boldsymbol{x}$ and $\boldsymbol{a}$, so
$$\min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 \;\geq\; \max_{\|\boldsymbol{a}\|_2\leq 1} \langle \boldsymbol{x}_0, \boldsymbol{a}\rangle - h(\boldsymbol{a}).$$
We call the minimization program on the left the **primal** and the maximization program on the right the **dual**.

2. There is a pair of points where equality of the functionals above is achieved, so
$$\min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 = \max_{\|\boldsymbol{a}\|_2\leq 1} \langle \boldsymbol{x}_0, \boldsymbol{a}\rangle - h(\boldsymbol{a}).$$
Moreover, the minimum/maximum are achieved by unique $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{a}}$, and so
$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2 = \langle \boldsymbol{x}_0, \hat{\boldsymbol{a}}\rangle - h(\hat{\boldsymbol{a}}).$$

3. The dual optimal point (which is a normal vector for a hyperplane) is **aligned** with the error:
$$\hat{\boldsymbol{a}} = \frac{\boldsymbol{x}_0 - \hat{\boldsymbol{x}}}{\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2}.$$
This gives us the **optimality condition** for $\hat{\boldsymbol{x}}$:
$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}_0\|_2^2 = \langle \boldsymbol{x}_0, \boldsymbol{x}_0 - \hat{\boldsymbol{x}}\rangle - h(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}).$$

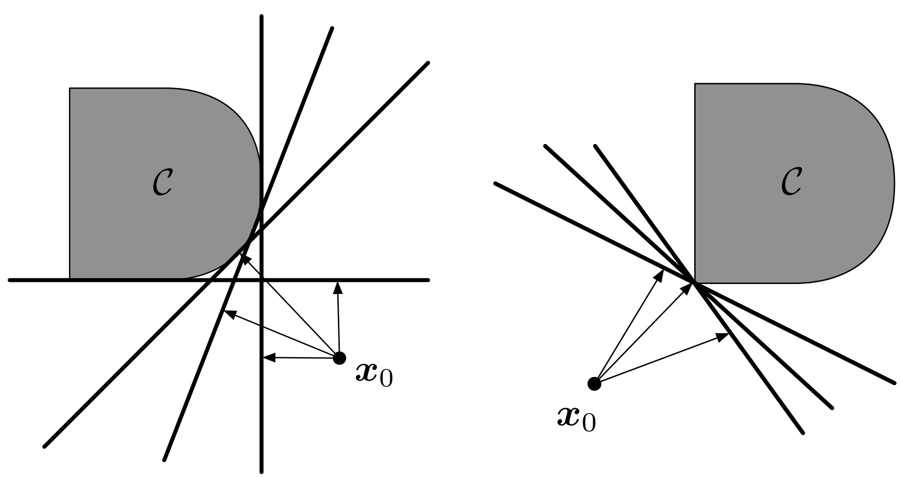
Let's take one more look at the dual problem,

$$\max_{\|\boldsymbol{a}\|_2 \leq 1} \langle \boldsymbol{x}_0, \boldsymbol{a} \rangle - h(\boldsymbol{a}),$$

and see how we can interpret it as finding the maximum distance from $\boldsymbol{x}_0$ to a hyperplane that separates $\boldsymbol{x}_0$ and $\mathcal{C}$. We know that the distance from $\boldsymbol{x}_0$ to a hyperplane $\{\boldsymbol{x} : \langle \boldsymbol{x}, \boldsymbol{a} \rangle = b\}$, when $\|\boldsymbol{a}\|_2 = 1$ is $|\langle \boldsymbol{x}_0, \boldsymbol{a} \rangle - b|$. Then quantity $h(\boldsymbol{a})$ is the "maximal" value of b that keeps such a hyperplane separating $\boldsymbol{x}_0$ and $\mathcal{C}$, so

$$\langle \boldsymbol{x}_0, \boldsymbol{a} \rangle - h(\boldsymbol{a})$$

is the distance of $\boldsymbol{x}_0$ to the furthest separating hyperplane that has $\boldsymbol{a}$ as a normal vector (again, with $\|\boldsymbol{a}\|_2 = 1$). Thus taking the maximum over all such normal vectors is a search for the separating hyperplane that is furthest from $\boldsymbol{x}_0$.



Problem: Let $\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^2 : \|\boldsymbol{x}\|_1 \leq 1\}$. Set $\boldsymbol{x}_0 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$. Find the closest point $\hat{\boldsymbol{x}} \in \mathcal{C}$ to $\boldsymbol{x}_0$ — you can do this by inspection after sketching $\mathcal{C}$ and $\boldsymbol{x}_0$. Calculate $\hat{\boldsymbol{a}}$ above, and verify that

$$\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2 = \langle \boldsymbol{x}_0, \hat{\boldsymbol{a}} \rangle - h(\hat{\boldsymbol{a}}).$$

Problem: Find $h(\boldsymbol{a}) = \sup_{\boldsymbol{x}\in\mathcal{C}} \langle \boldsymbol{x}, \boldsymbol{a} \rangle$ for the following sets:

1. $\mathcal{C} = \{\boldsymbol{x} \ : \ \|\boldsymbol{x}\|_2 \leq 1\}$

2. $\mathcal{C} = \{\boldsymbol{x} \ : \ \|\boldsymbol{x}\|_1 \leq 1\}$

3. $\mathcal{C} = \{\boldsymbol{x} \ : \ \|\boldsymbol{x}\|_\infty \leq 1\}$

Problem: Given $\boldsymbol{x}_0$, describe how to solve

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}_0 - \boldsymbol{x}\|_2 \quad \text{subject to} \quad \|\boldsymbol{x}\|_\infty \leq 1.$$

This can be solved using your intuition. After a moments thought, you know the solution $\hat{\boldsymbol{x}}$ must be given by

$$\hat{x}[n] = \left\{ \begin{array}{l} \\ \\ \end{array} \right.$$

Show that the optimality conditions hold for your solution:

$$\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2^2 = \langle \boldsymbol{x}_0, \boldsymbol{x}_0 - \hat{\boldsymbol{x}} \rangle - h(\boldsymbol{x}_0 - \hat{\boldsymbol{x}})$$

Problem: Given $\boldsymbol{x}_0$, describe how to solve

$$\min_{\boldsymbol{x}} \ \|\boldsymbol{x}_0 - \boldsymbol{x}\|_2 \quad \text{subject to} \quad \|\boldsymbol{x}\|_1 \leq 1.$$

Start by writing down the optimality conditions, and then thinking hard about what they mean.