Convex Optimization
Review Session 7 & 8

Question 1 (Theory of Banach Spaces)

In this "question", we'll be discussing Banach spaces.

Recall that in our discussions of Hilbert spaces in previous review sessions, many of the theorems we proved were adorned with a *. These theorems are identical in Banach space and will not be repeated here.

Solution

Having considered Hilbert spaces, we now move on to Banach spaces. Banach spaces are complete vector spaces together with a norm. In other words, the only thing that distinguishes Banach spaces from Hilbert spaces is that the latter are equipped with an inner product, whereas the former are only equipped with a norm.

Let's begin by listing some common Banach spaces...

> **Theorem 1. (Common Banach Spaces)**
>
> The following are Banach spaces
>
> - Any Hilbert space.
> - ℓ_p, the set of sequences $\boldsymbol{x}$ such that $\sum_i |x_i|^p < \infty$, with norm $\|\boldsymbol{x}\| = (\sum_i |x_i|^p)^{1/p}$.
> - L_p, the set of functions $\boldsymbol{x}$ such that $\int |x(t)|^p \mathrm{d}t < \infty$, with norm $\|\boldsymbol{x}\| = \left(\int |x(t)|^p\right)^{1/p}$
> - $C[a,b]$, the set of continuous functions on $[a,b]$, with norm $\|\boldsymbol{x}\| = \sup_{t\in[a,b]} |x(t)|$.
> - c, the set of sequences with a limit that exists, with norm $\|\boldsymbol{x}\| = \sup_k |x_k|$.
> - c_0, the set of sequences with limit 0, with the same norm as c.
> - $\mathrm{NBV}[a,b]$, the set of functions of bounded total variation on $[a,b]$, with value 0 at a, with norm equal to the total variation.
>
>   The *total variation* of a function $\boldsymbol{x}$ is given by
>
> $$\mathrm{TV}(\boldsymbol{x}) = \sup_{\substack{\text{paritions} \\ a=t_1\leq\cdots\leq t_n=b}} \sum_{t=1}^{n} |x(t_i) + x(t_{i-1})| = \int_a^b |\mathrm{d}x(t)|$$

As we will see, this seemingly small difference has an enormous impact. In particular, it means that the Riesz-Frechet Theorem no longer holds (because it make extensive use of inner products) – linear functionals in Banach spaces *cannot* be represented as an element in the space. This leads us to our first topic...

Dual Spaces

Definition 1. (Dual Space) The *dual space* of a Banach space X, denoted X^*, is the set of all bounded linear functionals on X.

As we discussed when we defined linear functionals, the form of any element $f \in X^*$ is

$$\|f\|^* = \sup_{\|\boldsymbol{x}\| \leq 1} |f(\boldsymbol{x})$$

It should be immediately apparent that the dual of a Hilbert space is the space itself, as a result of the Riesz-Frechet Theorem. Unfortunately, we no longer have recourse to this theorem in the more general case, and this is where complications begin.

Theorem 2. (Dual of Common Spaces) Given some p, let q be such that $\frac{1}{p} + \frac{1}{q} = 1$ (in other words, $q = p(p-1)$). Then

- The dual of ℓ_p, for $1 \leq p < \infty$, is ℓ_q. In other words, any bounded linear functional $f : \ell_p \to \mathbb{R}$ is representable in the form
$$f(\boldsymbol{x}) = \sum_{i=1}^{\infty} \eta_i x_i$$
where $\boldsymbol{\eta} \in \ell_q$. Furthermore, every element of ℓ_q defines a bounded linear functional on ℓ_p in this way, and we have
$$\|f\|_p^* = \|\boldsymbol{\eta}\|_q$$
(Note that the dual of ℓ_∞ is not ℓ_1).

- The dual of $L_p[0,1]$, for $1 \leq p < \infty$, is $L_q[0,1]$. In other words, any bounded linear functions $f : L_p \to \mathbb{R}$ is representable in the form
$$f(\boldsymbol{x}) = \int_0^1 x(t)y(t)\mathrm{d}t$$
where $\boldsymbol{y} \in L_q[0,1]$. Furthermore, every element of $L_q[0,1]$ defines a linear functional on $L_p[0,1]$ in this way, and
$$\|f\|_p^* = \|\boldsymbol{y}\|_q$$

(Again, the dual of L_∞ is not L_1).

- The dual of $C[a,b]$, the space of continuous functions on $[a,b]$, is $\mathrm{NBV}[a,b]$.

  In other words, any bounded linear functional $f : C[a,b] \to \mathbb{R}$ is representable in the form

  $$f(\boldsymbol{x}) = \int_a^b x(t)\, \mathrm{d}v(t)$$

  such that $\boldsymbol{v} \in \mathrm{NBV}[a,b]$. Furthermore, every element of $\mathrm{NBV}[a,b]$ defines a bounded linear functional in this way and

  $$\|\boldsymbol{f}\|^*_{C[a,b]} = \|\boldsymbol{v}\|_{\mathrm{NBV}[a,b]}$$

(We will consider the duals of c_0 and c in exercises.)

Proof. We will carry out the proof for ℓ_p – the proof for L_p is similar. We'll show this in two steps

$\boxed{\ell_q \subseteq \ell_p^*}$ Consider any element $\boldsymbol{\eta} \in \ell_q$, and consider the functional

$$f(\boldsymbol{x}) = \sum_{i=1}^{\infty} \eta_i x_i$$

Clearly, the resulting linear is functional. Furthermore, by Holder's Inequality,

$$|f(\boldsymbol{x})| \leq \sum_{i=1}^{\infty} |\eta_i x_i| \leq \|\boldsymbol{x}\|_p \|\boldsymbol{\eta}\|_q$$

and this inequality is tight. Thus,

$$\|f\|_p^* = \|\boldsymbol{\eta}\|_q$$

$\boxed{\ell_p^* \subseteq \ell_q}$ Consider some linear functional $f \in \ell_p^*$, and consider some $\boldsymbol{x} \in \ell_p$. We then carry out the following steps

Truncate $\boldsymbol{x}$ First, define 'basis functions' $\boldsymbol{e}^{(i)} \in \ell_p$, consisting of sequences which are identically 0 except for the i^{th} component, which is equal to 1. We then define the approximation

$$\boldsymbol{x}_{(N)} = \sum_{i=1}^{n} x_i \boldsymbol{e}^{(i)} \to \boldsymbol{x}$$

Apply f to the truncated vectors Let $\eta_i = f(\boldsymbol{e}^{(i)}) \in \mathbb{C}$.

Then, by linearity of f, we have that

$$f(\boldsymbol{x}_{(N)}) = \sum_{i=1}^{N} \eta_i x_i \to \sum_{i=1}^{\infty} y_i x_i$$

By *continuity* of f, we have that

$$f(\boldsymbol{x}_N) = \sum_{i=1}^{\infty} \eta_i x_i$$

Show that $\boldsymbol{\eta}$ is a vector in ℓ_q Define $\eta_{(N)}$ by

$$(\eta_{(N)})_i = \begin{cases} |\eta_i|^{q/p}\text{sign}(\eta_i) & i \leq N \\ 0 & i > N \end{cases}$$

(This vectors is both in ℓ_p and ℓ_q). Now, consider that (bearing in mind that $\frac{1}{p} + \frac{1}{q} = 1$),

$$\|\boldsymbol{\eta}_{(N)}\|_p = \left(\sum_{i=1}^{N} |\eta_i|^q \right)^{1/p}$$

$$f(\boldsymbol{\eta}_{(N)}) = \sum_{i=1}^{n} |\eta_i|^{q/p}\text{sign}(\eta_i)\eta_i = \sum_{i=1}^{n} |\eta_i|^q$$

Finally, consider that by the definition of the dual norm, $|f(\boldsymbol{\eta}_{(N)})| \leq \|\boldsymbol{f}\|_p^* \|\boldsymbol{\eta}_{(N)}\|_p$, and so

$$\frac{|f(\boldsymbol{\eta}_{(N)})|}{\|\boldsymbol{\eta}_{(N)}\|} \leq \|\boldsymbol{f}\|_p^* < \infty \qquad \forall N$$

Combining these three results, we find that

$$\|\boldsymbol{\eta}\|_q < \infty \qquad \forall N$$

As such, the sequence $\{\boldsymbol{\eta}_{(i)}\}$ is in ℓ_q. And since the space is complete and $\boldsymbol{\eta}_{(i)} \to \boldsymbol{\eta}$, we have that $\boldsymbol{\eta} \in \ell_q$.

(For $p = 1$ and $q = \infty$, the proof is similar but slightly different – see Luenberger pp 108). □

Proof. The proof for $C[a, b]$ requires the Hahn-Banach Theorem. Feel free to skip it and to return to it later. As ever, let's show this in two steps:

$\boxed{\text{NBV}[a, b] \subseteq C[a, b]^*}$ Consider any element $\boldsymbol{v} \in \text{NBV}[a, b]$. Clearly, any functional f defined as in the question is linear. Furthermore, it is bounded, because

$$\begin{aligned} f(\boldsymbol{x}) &= \int_a^b x(t)\ \mathrm{d}v(t) \\ &\leq \max_{t \in [a,b]} |x(t)| \text{TV}(\boldsymbol{v}) \\ &\leq \|\boldsymbol{x}\| \text{TV}(\boldsymbol{v}) \end{aligned}$$

Choosing a constant $\boldsymbol{x}$ reveals that this inequality is tight, and so we do indeed find that

$$\|\boldsymbol{f}\|^* = \|\boldsymbol{v}\|_{\mathrm{NBV}[a,b]}$$

$\boxed{C[a,b]^* \subseteq \mathrm{NBV}[a,b]}$ Consider some bounded linear functional $\boldsymbol{f} \in C[a,b]^*$. Since $C[a,b]$ is a subspace of $B[a,b]$ (the space of *bounded* functions in $[0,1]$), the Hahn-Banach Theorem implies that there is some functional $\boldsymbol{F} \in B[a,b]^*$ such that $\|\boldsymbol{F}\|^* = \|\boldsymbol{f}\|^*$. Now, carry out the following steps

Approximate $\boldsymbol{x}$ by discretizing it. Define a set of step functions $\boldsymbol{u}^{(s)}(t) = \mathbb{I}_{t \leq s} \in B[0,1]$. Then, note that

$$x(\tau) \approx z^{\pi}(\tau) = \sum_{i=1}^{n} x(t_i) \left(u^{(t_i)}(\tau) - y^{(t_{i-1})}(\tau) \right) \in B[0,1]$$

where π is some partition of $[a,b]$.

Find the image of each approximand . It will now become apparent why we needed to us the HB Theorem to find an extension of $\boldsymbol{f}$. Let $v(s) = F(\boldsymbol{u}^{(s)})$ be the image of these 'basis functions'. Since $\boldsymbol{F}$ is linear, we can write

$$F(\boldsymbol{z}^{\pi}) = \sum_{i=1}^{n} x(t_i) \left(v(t_i) - v(t_{i-1}) \right) \to \int_0^1 x(t) \, \mathrm{d}v(t)$$

Bridge the gap between B and C . By uniform continuity of $\boldsymbol{x}$, the approximation $\boldsymbol{z}^{\pi}$ becomes arbitrarily good (using the max norm). Since $\boldsymbol{F}$ is continuous, this means that

$$F(\boldsymbol{z}^{\pi}) \to F(\boldsymbol{x}) = f(\boldsymbol{x})$$

and so

$$f(\boldsymbol{x}) = \int_a^b x(t) \, \mathrm{d}v(t)$$

Show that $\boldsymbol{v}$ has bounded TV . Consider that (we let $\epsilon_i = \pm 1$, to take care of the absolute value)

$$\begin{aligned}
\mathrm{TV}(\boldsymbol{v}) &= \sum_{i=1}^{n} |v(t_i) - v(t_{i-1})| \\
&= \boldsymbol{F}\left(\sum_{i=1}^{n} \epsilon_i [\boldsymbol{u}^{t_i} - \boldsymbol{u}^{t_{i-1}}] \right) \\
&\leq \|\boldsymbol{F}\|^* \left\| \sum_{i=1}^{n} \epsilon_i [\boldsymbol{u}^{t_i} - \boldsymbol{u}^{t_{i-1}}] \right\| \\
&= \|\boldsymbol{F}\|^* = \|\boldsymbol{f}\|^* < \infty
\end{aligned}$$

(where the last step follows because $\boldsymbol{u}$ are step functions and so their maximum norms will be 1.)

□

The Hahn-Banach Theorem

It turns out that the Hahn-Banach Theorem, which we proved for Hilbert spaces, also holds for Banach spaces, in both its forms. We re-state the theorem here, for convenience.

> **Theorem 3. (Hahn-Banach)**
>
> Let $M \subseteq H$ be a closed subspace of a Banach space X, and let p be a seminorm on X.
>
> Let ϕ be a continuous linear functional on M satisfying $\phi(\boldsymbol{m}) \leq p(\boldsymbol{m})$ for all $\boldsymbol{m} \in M$. Then there exists a continuous linear extension of ϕ on X, Φ, such that $\Phi(\boldsymbol{x}) \leq p(\boldsymbol{x})$ for all $\boldsymbol{x} \in X$.

Proof. See Luenberger pp. 111. □

Alignment & Orthogonality

In a nutshell, the reminder of this section will be devoted to getting around the difficulties caused by a lack of inner product, and to derive a set of results analogous to those we derived for Hilbert space.

We begin by introducing some rather clever notation that will obscure the lack of an inner product in Banach space. Consider a Banach space X, an element $\boldsymbol{x} \in X$, and a linear functional $\boldsymbol{f} \in X^*$. We will write

$$f(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{f} \rangle$$

In some sense, this is nonsense, because Banach spaces are not equipped with inner products. However, as a point of notation, this will turn out to be extremely useful. This is particularly true in light of the following theorem

> **Theorem 4.** Given a Banach space X and for a fixed $\boldsymbol{x} \in X$, the quantity $\langle \boldsymbol{x}, \boldsymbol{f} \rangle$ as a function of $\boldsymbol{f} \in X^*$ defines a linear functional on $\boldsymbol{X}^*$.
>
> Furthermore,
> $$\langle \boldsymbol{x}, \boldsymbol{f} \rangle \leq \|\boldsymbol{x}\|_X \|\boldsymbol{f}\|_X^*$$

and there is some functional $\boldsymbol{f} \in X^*$ (not necessarily unique) such that this holds with equality. Any such vector is said to be *aligned* with $\boldsymbol{x}$.

Proof. First, consider $\boldsymbol{f}_1 \in X^*$ and $\boldsymbol{f}_2 \in X^*$. We have that

$$\langle \boldsymbol{x}, \alpha \boldsymbol{f}_1 + \beta \boldsymbol{f}_2 \rangle = \alpha f(\boldsymbol{x}) + \beta f_2(\boldsymbol{x}) = \alpha \langle \boldsymbol{x}, \boldsymbol{f}_1 \rangle + \beta \langle \boldsymbol{x}, \boldsymbol{f}_2 \rangle$$

So the functional is linear.

Furthermore,

$$\langle \boldsymbol{x}, \boldsymbol{f} \rangle = f(\boldsymbol{x}) \leq \|\boldsymbol{x}\|_X \|\boldsymbol{f}\|_X^*$$

so the linear functional is bounded.

Finally, to show that there is a $\boldsymbol{f} \in X^*$ such that this holds in equality, consider the linear functional $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_X$ defined over the subspace $\{\boldsymbol{y} : \boldsymbol{y} = \alpha \boldsymbol{x}\}$. This is clearly bounded with norm unity. By the Hahn-Banach Theorem (with seminorm $p(\boldsymbol{x}) = \|\boldsymbol{x}\|$), we can find an extension of this functional that also has norm unity. □

Note that the discussion above implies that $X \subseteq X^{**}$. For certain spaces, $X = X^{**}$ – these spaces are called *reflexive*.

We now take the 'inner product' analogy even further

Definition 2. (Alignment & Orthogonality) Two vectors $\boldsymbol{x} \in X$ and $\boldsymbol{x}^* \in X^*$ are said to be *orthogonal* if $\langle \boldsymbol{x}, \boldsymbol{x}^* \rangle = 0$.

The vectors are said to be *aligned* if $\langle \boldsymbol{x}, \boldsymbol{x}^* \rangle = \|\boldsymbol{x}\| \|\boldsymbol{x}\|^*$.

Example

Let $\boldsymbol{x} \in X = C[a, b]$, and let Γ be the following set of points

$$\Gamma = \left\{ t : |\boldsymbol{x}(t)| = \|\boldsymbol{x}\| = \sup_{t \in [a,b]} |x(t)| \right\}$$

Now, consider a bounded linear functional $\boldsymbol{x}^* \in C[a, b]^*$. We proved above that it can be expressed as

$$\langle \boldsymbol{x}^*, \boldsymbol{x} \rangle = x^*(\boldsymbol{x}) = \int_a^b x(t) \, \mathrm{d}v(t)$$

$\boldsymbol{x}$ and $\boldsymbol{x}^*$ are aligned if and only if

$$\langle \boldsymbol{x}^*, \boldsymbol{x} \rangle = \|\boldsymbol{x}\| \|\boldsymbol{x}^*\|^*$$

or, more specifically,

$$\int_a^b x(t)\ \mathrm{d}v(t) = \left[\sup_{t\in[a,b]} |x(t)|\right] \mathrm{TV}(v(t))$$

This makes it pretty clear that $\boldsymbol{x}^*$ is aligned with $\boldsymbol{x}$ only if

- It only varies on Γ.
- It is non-decreasing at t if $x(t) > 0$.
- It is non-increasing at t if $x(t) < 0$.

Minimum-norm problems

Consider a vector $\boldsymbol{x}$ in some normed linear space X. Because $X \subseteq X^{**}$ (see above), there are two ways to take the norm of that vector; either by considering it as an element of X, or by considering it as an element of X^{**}. Furthermore, by Theorem 4, both norms gives the same result. In other words,

$$\|\boldsymbol{x}\| = \max_{\|\boldsymbol{f}\|^* \leq 1} \langle \boldsymbol{x}, \boldsymbol{f} \rangle$$

The following theorem is a very similar result, but instead of finding the *norm* of $\boldsymbol{x}$ (ie: its distance from 0), it concerns the distance of $\boldsymbol{x}$ from a certain subspace M.

Theorem 5. (Duality)

Let $\boldsymbol{x}$ be an element in a real normed linear space X and d denote its distance from the subspace M. Then

$$d = \inf_{\boldsymbol{m}\in M} \|\boldsymbol{x} - \boldsymbol{m}\| = \max_{\substack{\|\boldsymbol{f}\|^* \leq 1 \\ \boldsymbol{f}\in M^{\perp}}} \langle \boldsymbol{x}, \boldsymbol{f} \rangle$$

where the maximum on the right is achieved for some $\boldsymbol{f}^{(0)} \in M^{\perp}$. If the infemum on the left is achieved, for some $\boldsymbol{m}^{(0)} \in M$, then $\boldsymbol{f}^{(0)}$ is aligned with $\boldsymbol{x} - \boldsymbol{m}^{(0)}$.

Proof. See Luenberger pp. □

We'll (try) to understand this problem in intuitively on the chalkboard in our review session.

As a companion to this theorem, we have

Theorem 6. Let M be a subspace in a real normed space X. Consider some $\boldsymbol{f} \in X^*$. Then

$$d = \min_{\boldsymbol{\phi} \in M^\perp} \|\boldsymbol{f} - \boldsymbol{\phi}\|^* = \sup_{\substack{\boldsymbol{x} \in M \\ \|\boldsymbol{x}\| \leq 1}} \langle \boldsymbol{x}, \boldsymbol{f} \rangle$$

where the minimum on the left is achieved for some $\boldsymbol{\phi}^{(0)} \in M^\perp$. If the supremum on the right is achieved for some $\boldsymbol{x}^{(0)} \in M$, then $\boldsymbol{f} - \boldsymbol{\phi}^{(0)}$ is aligned with $\boldsymbol{x}^{(0)}$.

This theorem guarantees the existence of a solution to the minimum norm problem *if* the problem is formulated in the dual of a normed space. This simply reflects the fact that Hahn-Banach Theorem establishes the existence of certain linear functionals, not of certain vectors.

This establishes a general rule – which we will be using in applications – that minimum norm problems must be formulated in a dual space if one is to guarantee the existence of solutions.

Like in Hilbert space, many optimization problems in Banach spaces are not concerned with finding the minimum distance between a point and a subspace, but instead seek the point of minimum norm in an *affine set*. The following theorem concerns such problems

Theorem 7. Consider a set of vectors $\{\boldsymbol{y}^{(1)}, \cdots, \boldsymbol{y}^{(N)}\}$ and a set of constants $\{c_1, \cdots, c_N\}$ (collected together into a vector $\boldsymbol{c}$). Then

$$\min_{\boldsymbol{f} \in X^* : \langle \boldsymbol{y}^{(i)}, \boldsymbol{f} \rangle = c_i} \|\boldsymbol{f}\|^* = \sup_{\boldsymbol{a} : \| \sum a_i \boldsymbol{y}^{(i)} \| \leq 1} \boldsymbol{c} \cdot \boldsymbol{a}$$

Furthermore, the optimal $\boldsymbol{f}$ on the LHS is aligned with the optimal $\sum a_i \boldsymbol{y}^{(i)}$ on the RHS.

Proof. As a first step, let's convert this problem to a standard minimum-norm problem. Let M denote the space generated by the $\boldsymbol{y}^{(i)}$, and let $\bar{\boldsymbol{f}}$ be *some* vector in the affine space (ie: satisfying all the constraints). Then

$$\begin{aligned} \min_{\boldsymbol{f} \in X^* : \langle \boldsymbol{y}^{(i)}, \boldsymbol{f} \rangle = c_i} \|\boldsymbol{f}\|^* &= \min_{\bar{\boldsymbol{\phi}} \in (\bar{\boldsymbol{f}} + M^\perp)} \|\bar{\boldsymbol{\phi}}\|^* \\ &= \min_{\boldsymbol{\phi} \in M^\perp} \|\bar{\boldsymbol{f}} - \boldsymbol{\phi}\|^* \end{aligned}$$

Using Theorem 6, we find that this is equivalent to

$$\sup_{\substack{\boldsymbol{x} \in M \\ \|\boldsymbol{x}\| \leq 1}} \langle \bar{\boldsymbol{f}}, \boldsymbol{x} \rangle$$

Since M is the subspace generated by the $\boldsymbol{y}^{(i)}$, we can write $\boldsymbol{x} = \sum a_i \boldsymbol{y}^{(i)}$, and this becomes

$$\sup_{\|\sum a_i \boldsymbol{y}^{(i)}\| \leq 1} \langle \bar{\boldsymbol{f}}, \sum a_i \boldsymbol{y}^{(i)} \rangle$$

And finally, using linearity and since by definition of $\bar{\boldsymbol{f}}$ we have that $\langle \bar{\boldsymbol{f}}, \boldsymbol{y}^{(i)} = c_i$ for all i, this becomes

$$\sup_{\|\sum a_i \boldsymbol{y}^{(i)}\| \leq 1} \boldsymbol{c} \cdot \boldsymbol{a}$$

As required. □

Hyperplanes & The Geometric Hahn-Banach Theorem

We are now ready to look at the geometric form of the Hahn-Banach Theorem.

> **Theorem 8. (Geometric Hahn-Banach*)**
>
> Let $\mathscr{K}$ be a convex set with non-empty interior in a Hilbert space H. Suppose $\mathscr{V}$ is an affine set in X (which could be a single point $\boldsymbol{x}_0$) that contains no interior points of $\mathscr{K}$. Then there is a closed hyperplane in H containing $\mathscr{V}$ but containing no interior point of $\mathscr{K}$.
>
> In other words, there exists an element $\boldsymbol{h}^* \in H$ such that
>
> $$\langle \boldsymbol{v}, \boldsymbol{h}^* \rangle = c \qquad \forall \boldsymbol{v} \in \mathscr{V}$$
>
> $$\langle \boldsymbol{k}, \boldsymbol{h}^* \rangle < c \qquad \forall \boldsymbol{k} \in \text{int}(\mathscr{K})$$

Proof. Given the extension form of the Hahn-Banach Theorem, the proof is the same as it was for Hilbert spaces. □

■ □ ■

Question 2 (A Control Problem)

Consider an electric motor governed by the equation

$$\ddot{\theta}(t) + \dot{\theta}(t) = u(t)$$

were $u(t)$ is a driving current. Suppose that at $t = 0$, the motor starts at rest, and at $\theta = 0$, and that we want to apply some driving current so that at $t = 1$, the motor ends up at rest at $\theta = 1$.

Show that the driving function $u(t)$ that minimizes the maximum current applied must be 'bang-bang' – ie: show that the function only takes values $\pm M$ for some M, and changes signs at most once.

Solution

We first need to choose a space in which to carry out this optimization problem. We might be tempted to choose $C[0,1]$, which has the correct norm. Unfortunately, this wouldn't guarantee a solution, because $C[0,1]$ is not the dual of any other space. Thus, we choose $L_\infty[0,1]$ (with norm $\|\boldsymbol{x}\| = \max_{t\in[0,1]} x(t)$), which is the dual of $L_1[0,1]$ (with norm $\|\boldsymbol{x}\| = \int_0^1 |x(t)| \, \mathrm{d}t$). This space still has the correct norm, but also has the advantage of being a dual space.

Having said that, let's massage our differential equation so as to express our constraints in a more palatable way. First, multiply throughout by the integrating factor e^t

$$\ddot{\theta}(t)e^t + \dot{\theta}(t)e^t = e^t u(t)$$

Simplifying

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^t\dot{\theta}(t)\right) = e^t u(t)$$

Integrating with respect to t

$$\left[e^t\dot{\theta}(t)\right]_0^1 = \int_0^1 e^t u(t) \, \mathrm{d}t$$

Simplifying, and recalling that $\dot{\theta}(0) = 0$, we find that

$$\dot{\theta}(1) = \int_0^1 e^{t-1} u(t) \, \mathrm{d}t$$

Consider, however, that since $e^{t-1} \in L_1[0,1]$, the dual space of $L_\infty[0,1]$ in which u resides, the integral above defines a linear functional on $L_\infty[0,1]$. Thus, we can use inner product notation to write

$$\boxed{\dot{\theta}(1) = \langle e^{t-1}, u\rangle}$$

We can also integrate the governing equation directly to get

$$[\dot{\theta}(t)]_0^1 + [\theta(t)]_0^1 = \int_0^1 u(t) \, \mathrm{d}t$$

Using the initial conditions, we find that

$$\theta(1) = \int_0^1 u(t) \, \mathrm{d}t - \dot{\theta}(1)$$

Finally, using our boxed result above (and similar logic to deal with the integral), we find that

$$\boxed{\theta(1) = \langle 1 - e^{t-1}, u \rangle}$$

Thus, our boundary conditions become

$$\langle e^{t-1}, u \rangle = 0$$

$$\langle 1 - e^{t-1}, u \rangle = 1$$

and we need to minimize the norm of u subject to those constraints. This problem is precisely in the form we discussed in theorem 7 – finding the vector of minimum norm in an affine space – and using the results there, its optimal solution is equal to the optimal solution of

$$\max_{\|a_1 e^{t-1} + a_2(1-e^{t-1})\| \leq 1} a_2$$

in other words, it is equal to the largest constant a_2 that allows

$$\int_0^1 \left|(a_1 - a_2)e^{t-1} + a_2\right| \ \mathrm{d}t \leq 1$$

for any a_1.

Assuming we have found such a constant a_2, it remains to characterize our optimal solution $\boldsymbol{u}$. We do this by using the second part of theorem 7, which states that the optimal $\boldsymbol{u}$ is aligned with the optimal $(a_1 - a_2)e^{t-1} + a_2$ (we will denote this function by $\boldsymbol{x}$). Thus, we need

$$\langle \boldsymbol{x}, \boldsymbol{u} \rangle = \|\boldsymbol{x}\| \|\boldsymbol{u}\|^*$$

Feeding in the definition of the inner product for this primal-dual pair (ie: the form of a linear functional), as well as the two norms, we find that we need

$$\int_0^1 x(t)u(t) \ \mathrm{d}t = \max_{t \in [0,1]} |u(t)| \int_0^1 |x(t)| \ \mathrm{d}t$$

It is clear that for this to be true, $u(t)$ *must* be equal to its maximum absolute value everywhere *and* its sign must agree with the sign of $x(t)$ everywhere.

But since $x(t)$ is the sum of an exponential and a constant term, it can only change sign once. Thus, $u(t)$ can only change sign once, and must be 'bang-bang'.

■ □ ■