quadratic – Schur complements allowed us to make it linear. [Similarly, we can bound the *lowest* eigenvalue: $\lambda_{\min}(A) \geq s \Leftrightarrow A \succeq sI$] □

- ***Example (portfolio optimization)***: Say we know $L_{ij} \leq \Sigma_{ij} \leq U_{ij}$. Given a portfolio $\boldsymbol{x}$, can maximize $\boldsymbol{x}^\top \Sigma \boldsymbol{x}$ s.t. that constraint and $\Sigma \succeq 0$ to get worst-case variance. We can add additional convex constraints
    - ***Known portfolio variances***: $\boldsymbol{u}_k^\top \Sigma \boldsymbol{u}_k = \sigma_k^2$
    - ***Estimation error***: If we estimate $\Sigma = \hat{\Sigma}$ but within an ellipsoidal confidence interval, we have $C\left(\Sigma - \hat{\Sigma}\right) \leq \alpha$, where $C(\cdot)$ is some positive definite quadratic form.
    - ***Factor models***: Say $p = Fz + d$, where z are random *factors* and d represents additional randomness. We then have $\Sigma = F\Sigma_{\text{factor}}F^\top + D$, and we can constraint each individually.
    - ***Correlation coefficients***: $\rho_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$. In a case where we know the volatilities exactly, constraints on ρ_{ij} are linear… □
- ***Example (expressing QCQP and SOCP as SDP)***: Using Schur complements, we can make these non-linear constraints linear

$$\tfrac{1}{2}\boldsymbol{x}^\top P\boldsymbol{x} + \boldsymbol{g}\cdot\boldsymbol{x} + \boldsymbol{h} \leq 0 \Leftrightarrow \begin{bmatrix} I & P^{1/2}\boldsymbol{x} \\ (P^{1/2}\boldsymbol{x})^\top & -\boldsymbol{g}\cdot\boldsymbol{x} - \boldsymbol{h} \end{bmatrix} \succeq 0$$

$$\left\| F\boldsymbol{x} + \boldsymbol{q} \right\| \leq \boldsymbol{g}\cdot\boldsymbol{x} + \boldsymbol{h} \Leftrightarrow \begin{bmatrix} (\boldsymbol{g}\cdot\boldsymbol{x} + \boldsymbol{h})I & F\boldsymbol{x} + \boldsymbol{q} \\ (F\boldsymbol{x} + \boldsymbol{q})^\top & \boldsymbol{g}\cdot\boldsymbol{x} + \boldsymbol{h} \end{bmatrix} \succeq 0$$

- *Geometric Programming*
    - A function $f(\boldsymbol{x}) = \sum_{k=1}^K c_k x_1^{a_{1k}} x_2^{a_{2k}} \cdots x_n^{a_{nk}}$, with $c_k > 0$ and $a_i \in \mathbb{R}$ is a ***posynomial*** (closed under $+, \times$). $K = 1$ gives a ***monomial*** (closed under $\times, \div$).
        - Posynomial $\times$ Monomial $=$ Posynomial
        - Posynomial $\div$ Monomial $=$ Posynomial
    - A ***geometric program*** is of the form

$$\min f_0(\boldsymbol{x}) \text{ s.t. } f_i(\boldsymbol{x}) \leq 1, h_i(\boldsymbol{x}) = 1, \boldsymbol{x} > \boldsymbol{0}$$

The dual function is $g(\boldsymbol{\lambda},\boldsymbol{\nu}) = \boldsymbol{\nu}\cdot(A\boldsymbol{x}-\boldsymbol{b}) + \min_s\left[(1-\boldsymbol{1}\cdot\boldsymbol{\lambda})s\right]$. This is only finite if $\boldsymbol{1}\cdot\boldsymbol{\lambda}=1$. So the dual is $d^* = \left(\max g(\boldsymbol{\lambda},\boldsymbol{\nu}) \text{ s.t. } \boldsymbol{1}\cdot\boldsymbol{\lambda}=1, \boldsymbol{\lambda}\ge\boldsymbol{0}\right)$. Provided strict feasibility holds, strong duality holds and $p^* = d^*$. So if the original system is infeasible ($p^* \geq 0$), then there exists a $g(\boldsymbol{\lambda},\boldsymbol{\nu}) \ge 0, \boldsymbol{\lambda} > \boldsymbol{0}$. Similarly, if there exists such a $(\boldsymbol{\lambda},\boldsymbol{\nu})$, then $p^* \ge 0$…

$$\boldsymbol{f}(\boldsymbol{x}) < \boldsymbol{0}, A\boldsymbol{x} = \boldsymbol{b} \ \ \frac{\text{feasible}}{\text{infeasible}} \Leftrightarrow g(\boldsymbol{\lambda},\boldsymbol{\nu}) \ge 0, \boldsymbol{\lambda} > \boldsymbol{0} \ \ \frac{\text{infeasible}}{\text{feasible}}$$

- ***Non-strict inequalities***: Consider $\boldsymbol{f}(\boldsymbol{x}) \le \boldsymbol{0}, A\boldsymbol{x} = \boldsymbol{b}$ the program is the same as above, but we need the optimum to be attained so that $p^* > 0$ if the system is infeasible. In that case, $\boldsymbol{\lambda} \ge \boldsymbol{0}, g(\boldsymbol{\lambda},\boldsymbol{\nu}) > 0$ is clearly feasible.

- ***Example***: Consider $A\boldsymbol{x} \le \boldsymbol{b}$. Then $g(\boldsymbol{\lambda}) = -\boldsymbol{\lambda}\cdot\boldsymbol{b}$ if $A^\top\boldsymbol{\lambda} = \boldsymbol{0}$ and $-\infty$ o.w. The strong system of alternative inequalities is $\boldsymbol{\lambda} \ge \boldsymbol{0}, A^\top\boldsymbol{\lambda} = 0, \boldsymbol{\lambda}\cdot\boldsymbol{b} < 0$. □
- ***Example***: Take m ellipsoids $\mathcal{E}_i = \left\{\boldsymbol{x} : f_i(\boldsymbol{x}) = \boldsymbol{x}^\top A_i\boldsymbol{x} + 2\boldsymbol{b}^i\cdot\boldsymbol{x} + \boldsymbol{c}^i \le 0\right\}, A_i \in \mathbb{S}^n_{++}$. We ask if the intersection has a non-empty interior. This is equivalent to solving the system $\boldsymbol{f}(\boldsymbol{x}) < \boldsymbol{0}$. Here, $g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}} \boldsymbol{x}^\top\left(\sum\lambda_i A_i\right)\boldsymbol{x} + 2\left(\sum\lambda_i\boldsymbol{b}^i\right)\cdot\boldsymbol{x} + \left(\sum\lambda_i\boldsymbol{c}^i\right)$. Differentiating, setting to 0 and using obvious notation, $g(\boldsymbol{\lambda}) = -\boldsymbol{b}_\lambda^\top A_\lambda^{-1}\boldsymbol{b}_\lambda + \boldsymbol{c}_\lambda$. As such, the alternative system is $\boldsymbol{\lambda} > \boldsymbol{0}, -\boldsymbol{b}_\lambda^\top A_\lambda^{-1}\boldsymbol{b}_\lambda + \boldsymbol{c}_\lambda \ge \boldsymbol{0}$.

  To explain geometrically, consider that the ellipsoid with $f(\boldsymbol{x}) = \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x})$ contains the intersection of all the ellipsoids above, because if $f(\boldsymbol{x}) \le \boldsymbol{0}$, then clearly a positive linear combination of them is also ≤ 0. This ellipsoid is empty if and only if the alternative is satisfied [prove by finding $\inf f(\boldsymbol{x})$]. □
- ***Example***: Farkas' Lemma: the following two systems are strong alternatives

  $$A\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \ge \boldsymbol{0} \qquad\qquad A^\top\boldsymbol{y} \ge 0, \boldsymbol{y}\cdot\boldsymbol{b} < 0$$
- ………………..

- ***Duality & Decentralization***
  - Consider $\min\sum_{i=1}^k f_i(\boldsymbol{x}^i)$ s.t. $\sum_{i=1}^k \boldsymbol{g}^i(\boldsymbol{x}^i) \le \boldsymbol{0}, \boldsymbol{x}^i \in \Omega_i$ [note: the vector $\boldsymbol{g}$ represents a number of inequality constraints]. The Lagrangian is $\mathcal{L}(\boldsymbol{x},\boldsymbol{\mu}) = \sum_{i=1}^k f_i(\boldsymbol{x}^i) + \boldsymbol{\mu}\cdot\sum_{i=1}^k \boldsymbol{g}^i(\boldsymbol{x}^i)$. The dual is $g(\boldsymbol{\mu}) = \sum_{i=1}^k g_i(\boldsymbol{\mu})$ s.t. $\boldsymbol{\mu} \ge \boldsymbol{0}$ where $g_i(\boldsymbol{\mu}) = \inf_{\boldsymbol{x}^i\in\Omega_i} f_i(\boldsymbol{x}^i) + \boldsymbol{\mu}\cdot\boldsymbol{g}^i(\boldsymbol{x}^i)$

- If bounded, $\left|\varphi(\boldsymbol{z})\right| \le M\left\|\boldsymbol{z}\right\| \ \forall \boldsymbol{z}$ and so $\left|\varphi(\boldsymbol{z})\right| \le \varepsilon \ \forall \boldsymbol{z} : \left\|\boldsymbol{z}\right\| \le \frac{\varepsilon}{M}$. So continuous at 0, and therefore everywhere.
- ***Example of a non-bounded linear functional***: Let V be the space of all sequences with finitely many non-zero elements, with norm $\left\|\boldsymbol{x}\right\| = \max_k \left|x_k\right|$. Then $\varphi(\boldsymbol{x}) = \max_k \left|k x_k\right|$ is unbounded because we can push the non-zero elements of $\boldsymbol{x}$ to infinity without changing the norm but making the functional grow to infinity. □

- ***Theorem (Riesz-Frechet)***: If $\varphi(\boldsymbol{x})$ is a continuous linear functional, then there exists a $\boldsymbol{z} \in H$ such that $\varphi(\boldsymbol{x}) = \left\langle \boldsymbol{x}, \boldsymbol{z} \right\rangle$

  Proof: Let $M = \left\{\boldsymbol{y} : \varphi(\boldsymbol{y}) = 0\right\}$. Since the functional is continuous, M is closed. If $M = H$, set $\boldsymbol{z} = \boldsymbol{0}$. Else, choose $\boldsymbol{\gamma} \in M^{\perp}$.

$$\begin{aligned}\varphi\left(\boldsymbol{x} - \tfrac{\varphi(\boldsymbol{x})}{\varphi(\boldsymbol{\gamma})}\boldsymbol{\gamma}\right) = \varphi(\boldsymbol{x}) - \varphi(\boldsymbol{x}) = 0 \Rightarrow \boldsymbol{x} - \tfrac{\varphi(\boldsymbol{x})}{\varphi(\boldsymbol{\gamma})}\boldsymbol{\gamma} &\in M \\ \Rightarrow 0 = \left\langle \boldsymbol{x} - \tfrac{\varphi(\boldsymbol{x})}{\varphi(\boldsymbol{\gamma})}\boldsymbol{\gamma}, \boldsymbol{\gamma} \right\rangle &= \left\langle \boldsymbol{x}, \boldsymbol{\gamma} \right\rangle - \tfrac{\varphi(\boldsymbol{x})}{\varphi(\boldsymbol{\gamma})}\left\langle \boldsymbol{\gamma}, \boldsymbol{\gamma} \right\rangle \\ \Rightarrow \varphi(\boldsymbol{x}) = \left\langle \boldsymbol{x}, \tfrac{\overline{\varphi(\boldsymbol{\gamma})}}{\left\|\boldsymbol{\gamma}\right\|^2}\boldsymbol{\gamma} \right\rangle &= \left\langle \boldsymbol{x}, \boldsymbol{z} \right\rangle\end{aligned}$$

  Note also that by Cauchy-Schwarz, $\left|\varphi(\boldsymbol{x})\right| \le \left\|\boldsymbol{z}\right\|\left\|\boldsymbol{x}\right\| \Rightarrow \left\|\varphi\right\| = \left\|\boldsymbol{z}\right\|$. ■

- This means that Hilbert spaces are self-dual (see later), and that we can write $\varphi(\boldsymbol{x}) = \left\langle \boldsymbol{x}, \varphi \right\rangle$.

- ***Theorem (Special case of the Hahn-Banach Theorem)***: Let $M \subseteq H$ be a closed subspace and φ_M be a continuous linear functional on M. Then there exists a continuous linear functional φ on H such that $\varphi(\boldsymbol{x}) = \varphi_M(\boldsymbol{x}) \ \forall \boldsymbol{x} \in M$ and $\left\|\varphi\right\| = \left\|\varphi_M\right\|$.

  Proof: Easy in the case of a Hilbert space. Since M is closed, it is also a Hilbert space, and so $\exists \boldsymbol{m} \in M$ such that $\varphi_M(\boldsymbol{x}) = \left\langle \boldsymbol{x}, \boldsymbol{m} \right\rangle$. Then define $\varphi(\boldsymbol{x}) = \left\langle \boldsymbol{x}, \boldsymbol{m} \right\rangle$ for $\boldsymbol{x} \in H$. By the CS inequality, $\left\|\varphi_M\right\| = \left\|\varphi\right\| = \left\|\boldsymbol{m}\right\|$. ■

Banach Spaces & Their Duals

- A Banach space is a normed, complete vector space with *no* inner product.
  - $C[0,1]$ is the space of continuous function son $[0,1]$, with

$$\left\|\boldsymbol{f}\right\| = \max_{0 \le t \le 1}\left|\boldsymbol{f}(t)\right|$$

[As we showed above, the choice of this norm ensures completeness]. An example of a linear functional on this space is

$$\varphi(\boldsymbol{f}) = \int_0^1 \boldsymbol{x}(t)\ \mathrm{d}v(t) \le \left\|\boldsymbol{x}\right\| \int_0^1 \mathrm{d}v(t) \le \left\|\boldsymbol{x}\right\| \mathrm{TV}(v)$$

Provided the total variation of v, $\mathrm{TV}(v) < \infty$, where

$$\mathrm{TV}(v) = \sup_{\text{All partitions } 0 = t_1 < t_2 < \cdots < t_n = 1} \sum_{i=1}^n \left|v(t_i) - v(t_{i-1})\right|$$

- $\ell_p = \left\{\boldsymbol{x} \in \mathbb{R}^\infty : \left\|\boldsymbol{x}\right\|_p < \infty\right\}$, where

$$\left\|\boldsymbol{x}\right\|_p = \left(\sum_{i=1}^\infty \left|x_i\right|^p\right)^{1/p} \qquad \text{or} \quad \left\|\boldsymbol{x}\right\|_p = \sup_i \left|x_i\right| \text{ if } p = \infty$$

- $\mathcal{L}_p[0,1] = \left\{\boldsymbol{x} : \int_0^1 \left|\boldsymbol{x}(t)\right|^p \mathrm{d}t < \infty\right\}$, with

$$\left\|\boldsymbol{f}\right\|_p = \left(\int_0^1 \left|\boldsymbol{f}(t)\right|^p \mathrm{d}t\right)^{1/p} \qquad 1 \le p < \infty \qquad \square$$

- ***Definition***: We say $V^* = \left\{\varphi : \varphi \text{ is continuous linear functional on } V\right\}$ is the *dual space* of V, with norm $\left\|\varphi\right\|_* = \sup\left\{\left|\varphi(\boldsymbol{x})\right| : \left\|\boldsymbol{x}\right\| \le 1\right\}$. $\left(V^*, \left\|\ \right\|_*\right)$ is always a Banach space.

Proof: Want to show that $\forall \left\{\boldsymbol{x}_n^*\right\} \subseteq V^*$ with $\left\|\boldsymbol{x}_n^* - \boldsymbol{x}_m^*\right\|_* \le \varepsilon\ \forall n, m \ge M_\varepsilon$ converges to a point $\boldsymbol{x}^* = \lim_{n\to\infty} \boldsymbol{x}_n^* \in V^*$. First fix $\boldsymbol{x} \in V$ and note that

$$\left|\boldsymbol{x}_n^*(\boldsymbol{x}) - \boldsymbol{x}_m^*(\boldsymbol{x})\right| = \left|(\boldsymbol{x}_n^* - \boldsymbol{x}_m^*)(\boldsymbol{x})\right| \le \left\|\boldsymbol{x}_n^* - \boldsymbol{x}_m^*\right\|_* \left\|\boldsymbol{x}\right\|$$

As such, $\left\{\boldsymbol{x}_n^*(\boldsymbol{x})\right\}$ is a Cauchy sequence in $\mathbb{R}$. Since $\mathbb{R}$ is complete, $\boldsymbol{x}^*(\boldsymbol{x}) = \lim_{n\to\infty} \boldsymbol{x}_n^*(\boldsymbol{x})$ exists. Define $\boldsymbol{x}^*$ pointwise using this limit. Now

- ***Linearity***: By linearity of expectations, $\boldsymbol{x}^*$ is linear.
- ***Continuity/boundedness***: Fix m_0 such that $\left\|\boldsymbol{x}_n^* - \boldsymbol{x}_m^*\right\| \le \varepsilon\ \forall n, m \ge m_0$. Then by the definition of $\boldsymbol{x}^*(\boldsymbol{x})$, $\left|\boldsymbol{x}^*(\boldsymbol{x}) - \boldsymbol{x}_m^*(\boldsymbol{x})\right| \le \varepsilon \left\|\boldsymbol{x}\right\|$, and

$$\left|\boldsymbol{x}^*(\boldsymbol{x})\right| \le \left|\boldsymbol{x}^*(\boldsymbol{x}) - \boldsymbol{x}_{m_0}^*(\boldsymbol{x})\right| + \left|\boldsymbol{x}_{m_0}^*(\boldsymbol{x})\right| \le \left(\varepsilon + \left\|\boldsymbol{x}_{m_0}^*\right\|_*\right)\left\|\boldsymbol{x}\right\| \Rightarrow \text{bounded} \qquad \blacksquare$$

Examples

- We have already shown (Riesz-Frechet Theorem) that Hilbert spaces are self-dual.

- **_Theorem_**: If $M \subseteq X$, then ${}^{\perp}(M^{\perp}) = M$.

  Proof: Clearly, $M \subseteq {}^{\perp}(M^{\perp})$. To show the converse, we'll show that $\boldsymbol{x} \notin M \Rightarrow \boldsymbol{x} \notin {}^{\perp}(M^{\perp})$. Define a linear functional $\boldsymbol{f}$ on the space spanned by M and $\boldsymbol{x}$ which vanishes on M so that $\boldsymbol{f}(\boldsymbol{m} + \alpha\boldsymbol{x}) = \alpha$. It can be shown that $\|\boldsymbol{f}\| < \infty$, and so by the HB Theorem, we can extend it to some $\boldsymbol{F}$ which also vanishes on M. As such, $\boldsymbol{F} \in M^{\perp}$. However, $\boldsymbol{F}(\boldsymbol{x}) = \langle \boldsymbol{F}, \boldsymbol{x} \rangle = 1 \neq 0$, and so $\boldsymbol{x} \notin {}^{\perp}(M^{\perp})$. ■

Minimum Norm Problems

- Let us consider a vector $\boldsymbol{x} \in X$. There are clearly two ways to take the norm of that vector – as an element of X or as an element of X^{**} (a functional on X^*).

$$\|\boldsymbol{x}\| \qquad \text{or} \qquad \max_{\|\boldsymbol{x}^*\|=1} \langle \boldsymbol{x}, \boldsymbol{x}^* \rangle$$

  It is clear these two should be equal, because $\langle \boldsymbol{x}, \boldsymbol{x}^* \rangle \leq \|\boldsymbol{x}\| \|\boldsymbol{x}^*\|$ (or, more intuitively, because the second norm finds the most $\boldsymbol{x}$ can yield under a functional of norm 1 – clearly, the answer is its norm). Let us now restrict ourselves to a subspace M of X. We can, again, define two norms

$$\|\boldsymbol{x}\|_M = \inf_{\boldsymbol{m} \in M} \|\boldsymbol{x} - \boldsymbol{m}\| \qquad \text{or} \qquad \|\boldsymbol{x}\|_{M^{\perp}} = \sup_{\substack{\|\boldsymbol{x}^*\|=1 \\ \boldsymbol{x}^* \in M^{\perp}}} \langle \boldsymbol{x}, \boldsymbol{x}^* \rangle$$

  The first simply consists of the minimum distance between $\boldsymbol{x}$ and M (as opposed to between $\boldsymbol{x}$ and 0). The second is the most $\boldsymbol{x}$ can yield under a functional of norm 1 that annihilates any element of M. Intuitively, the "remaining bit" that's "not annihilated" is $\boldsymbol{x} - \boldsymbol{m}$; this is maximized when it is aligned with $\boldsymbol{x}^*$ – at $\boldsymbol{m}_0$. So it makes sense that the two should be equal.

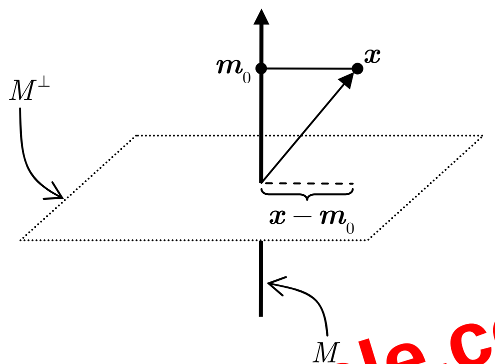
- **_Theorem_**: Consider a normed linear space X and a subspace M therein. Let $\boldsymbol{x} \in X$. Then

$$d = \inf_{\boldsymbol{m} \in M} \|\boldsymbol{x} - \boldsymbol{m}\| = \max_{\substack{\|\boldsymbol{x}^*\|_* \leq 1 \\ \boldsymbol{x}^* \in M^{\perp}}} \langle \boldsymbol{x}, \boldsymbol{x}^* \rangle \qquad \left(= \max_{\substack{\|\boldsymbol{x}^*\|_* = 1 \\ \boldsymbol{x}^* \in M^{\perp}}} \langle \boldsymbol{x} - \boldsymbol{m}_0, \boldsymbol{x}^* \rangle \right).$$

  Or, in our terminology above, $\|\boldsymbol{x}\|_M = \|\boldsymbol{x}\|_{M^{\perp}}$. The maximum on the right is achieved for some $\boldsymbol{x}_0^* \in M^{\perp}$; if the infemum on the left is achieved for some $\boldsymbol{m}_0 \in M$, then $\boldsymbol{x} - \boldsymbol{m}_0$ is aligned with $\boldsymbol{x}_0^*$.

Intuitively, this is because at the optimal $\boldsymbol{m}$, the residual $\boldsymbol{x} - \boldsymbol{m}_0$ is aligned to some vector in $M^\perp$. As such, for that vector, $\left\langle \boldsymbol{x} - \boldsymbol{m}_0, \boldsymbol{x}^* \right\rangle = \left\| \boldsymbol{x} - \boldsymbol{m}_0 \right\|$. For every other $\boldsymbol{x}^*$, it'll be smaller than that.

Pictorially, looking for the point on M that minimizes the norm is equivalent to looking for a point on $M^\perp$ that is aligned with $\boldsymbol{x} - \boldsymbol{m}_0$.



This also implies that a vector $\boldsymbol{m}_0$ is the minimum-norm projection if and only if there is a non-zero vector $\boldsymbol{x}^* \in M^\perp$ aligned with $\boldsymbol{x} - \boldsymbol{m}_0$.

- ***Theorem***: Let M be a subspace in a real normed space X. Let $\boldsymbol{x}^* \in X^*$. Then

$$d = \min_{\boldsymbol{m}^* \in M^\perp} \left\| \boldsymbol{x}^* - \boldsymbol{m}^* \right\| = \sup_{\boldsymbol{x} \in M, \|\boldsymbol{x}\| \leq 1} \left\langle \boldsymbol{x}, \boldsymbol{x}^* \right\rangle$$

  where the minimum on the left is achieved for some $\boldsymbol{m}_0^* \in M^\perp$. If the supremum is achieved for some $\boldsymbol{x}_0 \in M$, then $\boldsymbol{x}^* - \boldsymbol{m}_0^*$ is aligned with $\boldsymbol{x}_0$.

  Because the minimum on the left is always achieved, it is *always* more desirable to express optimization problems in a dual space.

- In many optimization problems, we seek to minimize a norm over an affine subset of a dual space rather than subspace. More specifically, subject to a set of linear constraints of the form $\left\langle \boldsymbol{y}_i, \boldsymbol{x}^* \right\rangle = c_i$. In that case, if $\bar{\boldsymbol{x}}^*$ is some vector that satisfies these constraints,

$$\begin{aligned}
x(T) &= \int_0^T \int_0^t u(s)\ \mathrm{d}s\ \mathrm{d}t - \frac{T^2}{2} \\
x(T) &= \left[t\int_0^t u(s)\ \mathrm{d}s\right]_0^T - \int_0^T tu(t)\ \mathrm{d}t - \frac{T^2}{2} \\
x(T) &= T\int_0^T u(s)\ \mathrm{d}s - \int_0^T tu(t)\ \mathrm{d}t - \frac{T^2}{2} \\
\int_0^T (T-t)u(s)\ \mathrm{d}s &= x(T) + \frac{T^2}{2}
\end{aligned}$$

$$\boxed{\left\langle T-t, v\right\rangle = x(T) + \frac{T^2}{2}}$$

Where v is the function in $NBV[0,1]$ associated with u, as described above.

- Our problem is then a minimum norm problem subject to a single linear constraint. We want $x(T) = 1$, and using our theorem, as we did above

$$\min_{\left\langle T-t,v\right\rangle = 1+\frac{1}{2}T^2} \left\|v\right\| = \max_{\left\|(T-t)a\right\|\le 1}\left[a\left(1+\tfrac{1}{2}T^2\right)\right]$$

This is a one-dimensional problem. The norm is in $C[0,1]$, the space to which NBV is dual. As such $\left\|(T-t)a\right\| = \max_{t\in[0,1]}\left|(T-t)a\right| = T\left|a\right|$, and the optimum occurs at $a = 1/T$. We then have $\min_{\left\langle T-t,v\right\rangle = 1+\frac{1}{2}T^2}\left\|v\right\| = \frac{1}{T} + \frac{1}{2}T$. Differentiating this with respect to T, we find that the minimum fuel expenditure of $\sqrt{2}$ is achieved at $T = \sqrt{2}$.

- To find the optimal u, note that the optimal v must be aligned to $(T-t)a$. As we discussed above when characterizing alignment of C and NBV, this means that v must be a step function at $t = 0$, rising to $\sqrt{2}$ at $t = 0$, and as such, u must be an impulse (delta function) at $t = 0$.

- ***Hyperplanes & the Geometric Hahn-Banach Theorem***
  - ***Definition***: A *hyperplane* H of a normed linear space X is a maximal proper affine set. ie: if $H \subseteq A$ and A is affine, then either $A = H$ or $A = X$.
  - ***Theorem***: A set H is a hyperplane if and only if it is of the form $\left\{\boldsymbol{x}\in X : f(\boldsymbol{x}) = c\right\}$ where f is a non-zero linear functional, and c is a scalar.

    Proof:
    - ***If***: Let $H = \boldsymbol{x}_0 + M$, where M is a linear subspace