Tropical reproducing kernels and optimization

Pierre-Cyril Aubin,
joint work with Stéphane Gaubert, INRIA and CMAP

Seminar SIERRA, November 2022
Paper under review at *Integral Equations and Operator Theory* (available on arXiv)

Motivation for max-plus functional analysis: non-vector optimization

Max-plus analysis can be summarized as the field studying what happens when the operations $(+, \times)$ are replaced by $(\max, +)$.

This is important for convex analysis since convex functions are stable by $(\max, +)$ and not stable by taking the negative i.e. NO VECTOR SPACES HERE!

The prototypical problem is

$$\min_{f \in \mathcal{F}} \mathcal{L}((f(x_m))_{m \in \mathcal{I}}) \text{ e.g. } \sum |f(x_m) - y_m|^2$$

Can we still do some analysis when $\mathcal{F}$ is not a vector space, but the set of convex functions? (as in PEP at SIERRA: Adrien Taylor, Céline Moucer,...)

If $\mathcal{F}$ was a RKHS $\mathcal{H}_k$, then we would have representer theorems. Can we do the same if the set $\mathcal{F}_b$ depends on a kernel b?

Further motivation for max-plus: limit of quantum mechanics

Max-plus analysis is related to the passage from quantum to classical mechanics. Taken from [Litvinov, 2005]:
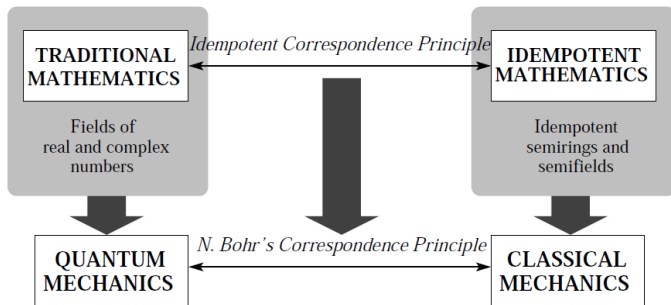
More precisely, set $u \oplus_\lambda v := \lambda \ln(e^{\frac{u}{\lambda}} + e^{\frac{v}{\lambda}}) \xrightarrow[\lambda \to 0]{} \max(u, v)$.

As in Laplace approximation of log-concave, $h(\lambda, x^*) \ln \int e^{\frac{u(x)}{\lambda}} f(x) dx \xrightarrow[\lambda \to 0]{} u(x^*) + \ln f(x^*)$ where $u(x^*) = \min u(x)$

Further motivation for max-plus functional analysis

Given a Lagrangian function $L : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, the action $J\left((t_0, r_0), (t_1, r_1), r(\cdot)\right)$ along an absolutely continuous trajectory $r(\cdot) : [t_0, t_1] \to \mathbb{R}^d$ going from (t_0, r_0) to (t_1, r_1) is defined as follows:

$$J\left((t_0, r_0), (t_1, r_1), r(\cdot)\right) := \int_{t_0}^{t_1} L(s, r(s), \dot{r}(s))ds \text{ with } r(t_0) = r_0, \ r(t_1) = r_1.$$

Given a set of absolutely continuous trajectories $S_{\text{traj}} \subset C^0(\mathbb{R}, \mathbb{R}^d)$ and a terminal cost, define the value function

$$\bar{V}(t_0, r_0) = \inf_{r_T \in \mathbb{R}^d} \inf_{\substack{r(\cdot) \in S_{\text{traj}} \\ r(t_0) = r_0, \ r(T) = r_T}} \int_{t_0}^{T} L(s, r(s), \dot{r}(s))ds + \psi_T(r_T).$$

The term in red depending on $(t_0, r_0), (T, r_T)$ is the fundamental solution of the HJB equation [Dower and McEneaney, 2015]. Why? Because it is the analogue of a Green kernel.

Reminder of linear PDEs

To solve Poisson's equation $\Delta u = f$ on a domain Ω, one just has to find the Green kernel $k(\cdot, x)$ s.t.

$$\Delta_1 k(\cdot, x) = \delta_x$$

then the solution is obtained through a kernel integral operator $u = Kf$, i.e.

$$u(y) = \int_x k(y, x) f(x) dx.$$

Indeed $\int (\Delta_1 k(y, x)) f(x) dx = f(y)$ (skipping boundary questions), so $\Delta u = f$. On the other hand we have

$$\bar{V}(t_0, r_0) = \inf_{t_1, r_1} b\left((t_0, r_0), (t_1, r_1)\right) + \psi_{t_1}(r_1) + \delta_T(t_1).$$

This is a fundamental solution of a nonlinear PDE. Can we find properties of tropical kernels that are analogue to Hilbertian kernels?

Your entry point: Fenchel's transform

Consider the inner product $(x, y)_2$ over $\mathbb{R}^d$. Given a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$, define its Fenchel transform

$$f^*(x) = \sup_{y \in \mathbb{R}^d} (x, y)_2 - f(y)$$

Then we know from Fenchel's theorem that

$$f^{**} = (f^*)^* = f \Leftrightarrow f \in \mathsf{CVEX}(\mathbb{R}^d) := \{ \sup_{y \in \mathbb{R}^d} (\cdot, y)_2 + a_y \mid a_y \in \mathbb{R} \cup \{-\infty\}\}$$

the space of lower-semicontinuous convex functions (Fenchel's theorem).

Your entry point: Fenchel's transform

Consider a kernel $b(x, y)$ over $\mathcal{X} \times \mathcal{X}$. Given a function $f : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$, define its B-transform

$$\bar{B}f(x) = \sup_{y \in \mathbb{R}^d} b(x, y) - f(y)$$

We will see later that

$$\bar{B}\bar{B}f = f \Leftrightarrow f \in \mathrm{Rg}(B) := \{\sup_{y \in \mathbb{R}^d} b(\cdot, y) + a_y \mid a_y \in \mathbb{R} \cup \{-\infty\}\}$$

Three examples of kernels (where d is some distance)

$$b_c(x, y) = (x, y)_2, \quad b_{sc}(x, y) = -\|x - y\|_2^2, \quad b_L(x, y) = -d(x, y). \tag{1}$$

Note: CVEX plays the role of L^2 in max-plus analysis.

Motivation
○○○○○●○
Kernels
○○○○○○○○
Properties
○○○○○○○○○
App: optimization
○○○○○
App: HJB
○○○○○○○○○○○

What we are going to see in this talk

- the characterization of tropical reproducing kernels is similar to the Hilbertian case

- it is possible to do optimization over tropical function spaces

- it is useful for value functions in control theory

Technical problems of max-plus analysis: infinities and signs

<div align="center">

WANTED! PUBLIC ENEMIES #1:

∞ and $-\infty$.

</div>

Following J.-J. Moreau conventions, we have to introduce new operations: $\dot{+}, \dot{-}, \ddot{+}, \ddot{-}$. Very simply:

$$(\infty) \dot{+} (-\infty) = -\infty, \ (\infty) \dot{-} (\infty) = \infty.$$

Then we have to decide which analogue of $(+, \times)$ we take:

- linear, $(\sup, +)$ or $(\inf, +)$: $\int k(x, y) f(y) dy \to \inf_y b(x, y) \dot{+} f(y)$

- sesquilinear, $(\sup, -)$ or $(\inf, -)$: $\int k(x, y) f(y) dy \to \sup_y b(x, y) \dot{-} f(y)$

In all the proofs, we will have to be super careful with the infinite values! Same problem as in optimal transport theory (cf Villani, 2006, Chapters 5 and 10).

Reminders on reproducing kernels

A RKHS $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_{\mathcal{H}_K})$ is a Hilbert space of real-valued functions over a set $\mathcal{X}$ if

$\exists\, k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ s.t. $k_x(\cdot) = k(\cdot, x) \in \mathcal{H}_K$ and $f(x) = \langle f(\cdot), k_x(\cdot) \rangle_{\mathcal{H}_K}$ for all $x \in \mathcal{X}$ and $f \in \mathcal{H}_K$ (reproducing property)

or equivalently (Riesz's theorem for Hilbert spaces) if

the topology of $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_{\mathcal{H}_K})$ is stronger than pointwise convergence i.e. $\delta_x : f \in \mathcal{H}_K \mapsto F$ is continuous for all $x \in \mathcal{X}$.

A kernel is feature-map factorizable if there exists a Hilbert space $\mathcal{H}$ s.t.

k is s.t. $\exists\, \Phi : \mathcal{X} \to \mathcal{H}$ s.t. $k(t, s) = \langle \Phi(t), \Phi_k(s) \rangle_{\mathcal{H}}$

A kernel is positive semi-definite if

k is s.t. $\mathbf{G} = [k(x_i, x_j)]_{i,j=1}^n \succcurlyeq 0$.

Aronszajn's theorem

Theorem

Given a kernel $k : \mathfrak{X} \times \mathfrak{X} \to \mathbb{R}$, the three following properties are equivalent:

i) *k is a positive semidefinite kernel, i.e. a kernel being both:*

   - symmetric: $\forall x, y \in X,\ k(x, y) = k(y, x)$, and

   - positive: $\forall M \in \mathbb{N}^*,\ \forall (a_m, x_m) \in (\mathbb{R} \times X)^M,\ \sum_{n,m=1}^M a_n a_m k(x_n, x_m) \geq 0$;

ii) *there exists a Hilbert space $(\mathfrak{H}, (\cdot, \cdot)_{\mathfrak{H}})$ and a feature map $\Phi : \mathfrak{X} \to \mathfrak{H}$ such that*

   - $\forall x, y \in X,\ k(x, y) = (\Phi(x), \Phi(y))_{\mathfrak{H}}$;

iii) *k is the reproducing kernel of the Hilbert space (RKHS) of functions $\mathfrak{H}_k := \overline{\mathfrak{H}_{k,0}}$, the completion for the pre-scalar product $(k(\cdot, x), k(\cdot, y))_{k,0} = k(x, y)$ of the space $\mathfrak{H}_{k,0} := \mathsf{span}(\{k(\cdot, x)\}_{x \in \mathfrak{X}})$, in the sense that*

   - $\forall x \in X,\ k(\cdot, x) \in \mathfrak{H}_k$ and $\forall f \in \mathfrak{H},\ f(x) = (f, k(\cdot, x))_{\mathfrak{H}}$.

Main (informal) theorem: Aronszajn's analogue

Theorem (Tropical analogue of Aronszajn theorem)

Given a kernel $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$, the three following properties are equivalent

i) *b is a tropically positive semidefinite kernel;*

ii) *there exists a factorization of b by a feature map $\psi : \mathcal{X} \to \overline{\mathbb{R}}_{\max}^{\mathcal{Z}}$ for some set $\mathcal{Z}$;*

iii) *b is the sesquilinear reproducing kernel of a max-plus space of functions $\mathrm{Rg}(B)$, the max-plus completion of $\{\sup_{n \in \{1,\dots,N\}} a_n + b(\cdot, x_n) \mid N \in \mathbb{N}^*, a_n \in \mathbb{R}_\perp, x_n \in \mathcal{X}\}$, and b defines a tropical Cauchy-Schwarz inequality over $\overline{\mathbb{R}}^{\mathcal{X}}$.*

Defining properly the tropical concepts: (sesqui)linear continuous operator

The extended real line is denoted by $\overline{\mathbb{R}} = [-\infty, +\infty]$, $\mathbb{R}_\top = (-\infty, +\infty]$ and $\mathbb{R}_\bot = [-\infty, +\infty)$
In all that follows, given a set $\mathcal{X}$, a *kernel* b is a function from $\mathcal{X} \times \mathcal{X}$ to $\overline{\mathbb{R}}$. Given a kernel b,
we also consider the max-plus linear B and sesquilinear $\bar{B}$ operators, defined over $\overline{\mathbb{R}}^{\mathcal{X}}$ as

$$Bf(x) = \sup_{y \in \mathcal{X}} b(x, y) \dotplus f(y), \quad \bar{B}f(x) = \sup_{y \in \mathcal{X}} b(x, y) \dotdiv f(y), \, \forall x \in \mathcal{X}, \, f \in \overline{\mathbb{R}}^{\mathcal{X}}. \tag{2}$$

Definition

A map $B : \overline{\mathbb{R}}^{\mathcal{X}} \to \overline{\mathbb{R}}^{\mathcal{X}}$ is said to be

  i) $\overline{\mathbb{R}}_{\max}$-linear if $B(\sup\{f_i\}_{i \in I}) = \sup\{Bf_i\}_{i \in I}$ and $B(f \dotplus \lambda) = Bf \dotplus \lambda$ (with $+\infty$ absorbing on both sides), for any finite index set I and $\lambda \in \overline{\mathbb{R}}$; B is *continuous* if I can be taken infinite.

  ii) $\overline{\mathbb{R}}_{\max}$-sesquilinear if $B(\inf\{f_i\}_{i \in I}) = \sup\{Bf_i\}_{i \in I}$ and $B(f \dotplus \lambda) = Bf \dotdiv \lambda$ (with $+\infty$ absorbing on the l.h.s. and $-\infty$ absorbing on the r.h.s.), for any finite index set I and $\lambda \in \overline{\mathbb{R}}$; B is *continuous* if I can be taken infinite.

Motivation
0000000

Kernels
00000●000

Properties
000000000

App: optimization
00000

App: HJB
00000000000

Defining properly the tropical concepts: ranges and kernels

The range of B is $\mathrm{Rg}(B) := \{g \in \overline{\mathbb{R}}_{\max}^{\mathcal{X}} \mid \exists f \in \overline{\mathbb{R}}_{\max}^{\mathcal{X}}, g = Bf\}$. The indicator functions

$$\delta_x^\perp(y) := \begin{cases} 0 & \text{if } y = x, \\ -\infty & \text{otherwise,} \end{cases} \quad \delta_x^\top(y) := \begin{cases} 0 & \text{if } y = x, \\ +\infty & \text{otherwise.} \end{cases} \tag{3}$$

The $\overline{\mathbb{R}}_{\max}$-sesquilinear and continuous maps, a.k.a. *(Fenchel-Moreau) conjugations*, have been characterized as the ones having a kernel as in (2):

Proposition (Theorem 3.1, [Singer, 1984])

A map $\bar{B} : \overline{\mathbb{R}}^{\mathcal{X}} \to \overline{\mathbb{R}}^{\mathcal{X}}$ is $\overline{\mathbb{R}}_{\max}$-sesquilinear and continuous if and only if there exists a kernel $b : \mathcal{X} \times \mathcal{X} \to \overline{\mathbb{R}}$ such that $\bar{B}f(x) = \sup_{y \in \mathcal{X}} b(x, y) \dotplus f(y)$. Moreover in this case b is uniquely determined by $\bar{B}$ as $b(\cdot, x) = \bar{B}\delta_x^\top$ (same as Hilbertian $k(\cdot, x) = K\delta_x$).

We refer to [Akian et al., 2005, Theorem 2.1] for extensions of this tropical analogue of Riesz representation theorem in Hilbert spaces [Martinez-Legaz and Singer, 1990].

$$\mathrm{Rg}(B) = \{\sup_{x \in \mathcal{X}} a_x \dotplus b(\cdot, x) \mid a_x \in \mathbb{R}_\perp\}. \tag{4}$$

Defining properly the tropical concepts: tropical duality

In the following, we also extensively use the following duality product over $\overline{\mathbb{R}}_{\min}^{\mathcal{X}} \times \overline{\mathbb{R}}_{\max}^{\mathcal{X}}$, denoting by $\hat{g}$ the elements of $\overline{\mathbb{R}}_{\min}^{\mathcal{X}}$,

$$\langle \hat{g}, f \rangle := \sup_{x \in \mathcal{X}} f(x) \dotminus \hat{g}(x) \quad \forall (\hat{g}, f) \in \overline{\mathbb{R}}_{\min}^{\mathcal{X}} \times \overline{\mathbb{R}}_{\max}^{\mathcal{X}}. \tag{5}$$

This duality product allows to define the adjoint of an operator $B : \overline{\mathbb{R}}^{\mathcal{X}} \to \overline{\mathbb{R}}^{\mathcal{X}}$.

Definition

If it exists, the adjoint map $\bar{B}'$ of a $\overline{\mathbb{R}}_{\max}$-sesquilinear map $\bar{B} : \overline{\mathbb{R}}^{\mathcal{X}} \to \overline{\mathbb{R}}^{\mathcal{X}}$ is defined as the one such that

$$\langle \hat{g}, \bar{B}\hat{f} \rangle = \langle \hat{f}, \bar{B}'\hat{g} \rangle, \quad \forall (\hat{g}, \hat{f}) \in \overline{\mathbb{R}}_{\min}^{\mathcal{X}} \times \overline{\mathbb{R}}_{\min}^{\mathcal{X}}$$

If $\bar{B}' = \bar{B}$, then $\bar{B}$ is said to be $\overline{\mathbb{R}}_{\max}$-hermitian. If $\bar{B}$ is continuous with kernel $b(x, y)$, then $\bar{B}'$ exists and corresponds to $b(y, x)$ [Singer, 1997, Theorem 8.4].

Defining properly the tropical concepts: tpsd kernel

Definition

We say that a kernel $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_\perp$ is a *tropical positive semidefinite* (tpsd) kernel if it is

i) symmetric: $\forall x, y \in \mathcal{X},\ b(x, y) = b(y, x)$, and

ii) tropically positive: $\forall x, y \in \mathcal{X},\ b(x, x) + b(y, y) \geq b(x, y) + b(y, x)$.

Notice that all the three kernels

$$b_c(x, y) = (x, y)_2, \quad b_{sc}(x, y) = -\|x - y\|_2^2, \quad b_L(x, y) = -d(x, y).$$

are tropically positive semidefinite and finite-valued. Moreover every Hilbertian (conditionally) positive semidefinite kernel is tropically positive semidefinite.

Note: a submodular function over a discrete set defines a tpsd matrix.

Full analogy between Hilbertian and tropical kernels

Dedicated to kernel lovers:

Concept	Hilbertian kernel	Tropical kernel	Reference
symmetry	$k(x, y) = k(y, x)$	$b(x, y) = b(y, x)$	Def. 5
positivity	$\sum_{i,j} a_i a_j k(x_i, x_j) \geq 0$	$b(x, x) + b(y, y) \geq b(x, y) + b(y, x)$	Def. 5
feature map	$k(x, y) = (\Phi(x), \Phi(y))_{\mathcal{H}}$	$b(x, y) = \sup_{z \in \mathcal{Z}} \psi(x, z) + \psi(y, z)$	Prop. 2
duality bracket	$\langle \mu, f \rangle_{\mathbb{R}^{\mathcal{X},*} \times \mathbb{R}^{\mathcal{X}}} = \int_{\mathcal{X}} f(y) \mathrm{d}\mu(y)$	$\langle \hat{g}, f \rangle = \sup_{x \in \mathcal{X}} f(x) - \hat{g}(x)$	(5)
kernel operator	$K(\mu)(x) = \int_{\mathcal{X}} k(x, y) \mathrm{d}\mu(y)$	$\bar{B}(\hat{f})(x) = \sup_{y \in \mathcal{X}} b(x, y) \dot{-} \hat{f}(y)$	Prop. 1
monotone operator	$\langle \mu, K(\mu) \rangle_{\mathbb{R}^{\mathcal{X},*} \times \mathbb{R}^{\mathcal{X}}} \geq 0$	$\langle \hat{f}, \bar{B}\hat{f} \rangle \dot{+} \langle \hat{g}, \bar{B}\hat{g} \rangle \geq \langle \hat{f}, \bar{B}\hat{g} \rangle + \langle \hat{g}, \bar{B}\hat{f} \rangle$	Prop. 3
function space	$\mathcal{H}_k = \overline{\mathrm{span}(\{k(\cdot, x)\}_{x \in \mathcal{X}})}$	$\mathrm{Rg}(B) = \{ \sup_{x \in \mathcal{X}} [a_x + b(\cdot, x)] \mid a_x \in \mathbb{R}_\perp \}$	Prop. 1+(4)
reproducing property	$f(x) = (k(\cdot, x), f(\cdot))_{\mathcal{H}_k}$	$\hat{g}(x) = \langle \bar{B}\hat{g}, \bar{B}\delta_x^\top \rangle = (\bar{B}\bar{B}\hat{g})(x)$	Def. 8

Factorization of tropical kernels

Lemma

A kernel $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_\perp$ is tpsd if and only if there exists a function $\phi : \mathcal{X} \to \mathbb{R}_\perp$ and a symmetric kernel $b_0 : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_\perp$, with $b_0(x,x) = 0$ and $b_0(x,y) \le 0$ for all $x, y \in \mathcal{X}$, such that

$$b(x,y) = \phi(x) + b_0(x,y) + \phi(y). \tag{6}$$

Moreover, given $b(\cdot, \cdot)$, we have that $\phi(x) = b(x,x)/2$ and $\mathrm{Rg}(B) = \phi + \mathrm{Rg}(B_0)$.

Proposition

A kernel $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_\perp$ is tpsd iff there exists a set $\mathcal{Z}$ and a function $\psi : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}_\perp$

$$b(x,y) = \sup_{z \in \mathcal{Z}} \psi(x,z) + \psi(y,z). \tag{7}$$

Proof of factorization

$\Rightarrow$. Take $\mathcal{Z} = \mathcal{X} \times \mathcal{X}$, and consider the function ψ such that, for all $x, y \in \mathcal{X} \times \mathcal{X}$,

$$\psi(x, (x, y)) = b(x, x)/2 \text{ and } \psi(x, (y, x)) = b(x, y) \dot{-} b(y, y)/2$$
$$\psi(x, (u, v)) = -\infty \text{ if } x \notin \{u, v\}$$

Thus $\psi(x, z)$ and $\psi(y, z)$ can be finite iff $z \in \{(x, y), (y, x)\}$. As b is symmetric, we obtain that

$$\sup_{z \in \mathcal{Z}} \psi(x, z) + \psi(y, z) = \max\left(\psi(x, (x, y)) + \psi(y, (x, y)), \psi(x, (y, x)) + \psi(y, (y, x))\right)$$

$$= \max\left(\frac{b(x, x)}{2} + b(y, x) \dot{-} \frac{b(x, x)}{2}, b(x, y) \dot{-} \frac{b(y, y)}{2} + \frac{b(y, y)}{2}\right) = b(y, x),$$

the equality holding for $b(x, x) = -\infty$ or $b(y, y) = -\infty$, since $b(y, x) < \infty$ by assumption.

$\Leftarrow$. The kernel $b(x, y) = \sup_{z \in \mathcal{Z}} \psi(x, z) + \psi(y, z)$ is symmetric and we assumed it takes its values in $\mathbb{R}_\perp$.

$$2(\psi(x, z) + \psi(y, z)) \leq 2 \sup_{z \in \mathcal{Z}} \psi(x, z) + 2 \sup_{z \in \mathcal{Z}} \psi(y, z) = b(x, x) + b(y, y), \forall z \in \mathcal{Z}.$$

Examples of factorization

Examples of factorizations $\psi(x, z)$ **with** $\mathcal{Z} = \mathcal{X}$:

i) For $\mathcal{X} = \mathbb{R}^N$ and $b_c(x, y) = (x, y)_2$, we can choose $\psi(x, z) = \frac{1}{2}\|x\|_2^2 - \|x - z\|_2^2$.

ii) For $\mathcal{X} = \mathbb{R}^N$ and $b_{sc}(x, y) = -\|x - y\|_2^2$, we can choose $\psi(x, z) = -2\|x - z\|_2^2$.

iii) For $(\mathcal{X}, d)$ a metric space, $b(x, y) = -d(x, y)^p$ with $p \in (0, 1]$, we can choose $\psi = b$ as a consequence of the subadditivity of $t \in \mathbb{R}_+ \mapsto t^p$.

However we cannot always take $\mathcal{Z} = \mathcal{X}$, even for finite sets (Barvinok rank problem, clique cover number of bipartite graph), counter-example with $\#\mathcal{X} = 5$.

Note: for Hilbertian kernels, we always had the choice $\Phi(x) = k(\cdot, x)$ for
$k(x, y) = (\Phi(x), \Phi(y))_{\mathcal{H}_k} \rightarrow$ we do not have the same here

Proposition ((Cyclical) monotony)

Given a kernel $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_\perp$, set $\bar{B}$ as in (2). Then the following statements are equivalent:

i) *the kernel b is tpsd, i.e. symmetric and $b(x,x) + b(y,y) \geq b(x,y) + b(y,x)$;*

ii) *for all $M \in \mathbb{N}^*$, $(x_m)_{m \in [M]} \in \mathcal{X}^M$ and permutations $\sigma : [M] \to [M]$, the Gram matrices*
   $\mathbf{G} := [b(x_n, x_m)]_{n,m \in [M]}$ are symmetric and satisfy $\sum_{m=1}^{M} b(x_m, x_m) \geq \sum_{m=1}^{M} b(x_m, x_{\sigma(m)})$

iii) *the operator $\bar{B}$ is $\overline{\mathbb{R}}_{\max}$-hermitian and monotone for the duality pairing in the sense that*

$$\forall \hat{f}, \hat{g} \in \overline{\mathbb{R}}^{\mathcal{X}}, \langle \hat{f}, \bar{B}\hat{f} \rangle \dotplus \langle \hat{g}, \bar{B}\hat{g} \rangle \geq \langle \hat{f}, \bar{B}\hat{g} \rangle \dotplus \langle \hat{g}, \bar{B}\hat{f} \rangle; \tag{8}$$

iv) *the operator $\bar{B}$ is $\overline{\mathbb{R}}_{\max}$-hermitian and cyclic monotone for the duality pairing in the sense that $\forall M \in \mathbb{N}^*$, $\forall (\hat{f}_m)_{m \in [M]} \in \overline{\mathbb{R}}^{\mathcal{X},M}$ with the convention $f_{M+1} = f_1$,*

$$\sum_{m=1}^{M} \langle \hat{f}_m, \bar{B}\hat{f}_m \rangle \geq \sum_{m=1}^{M} \langle \hat{f}_m, \bar{B}\hat{f}_{m+1} \rangle; \textit{ (}\Sigma \textit{ can be replaced with } \max\textit{)} \tag{9}$$

Characterizing tropical spaces

Can we give any characterization of the spaces that are Rg(B)?

A set $\mathcal{G}$ is a complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ if it is stable under arbitrary sups and addition of constants.

Theorem

Let $\mathcal{G}$ be a complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$. Then the following statements are equivalent:

i) there exists a symmetrical kernel $b : \mathcal{X} \times \mathcal{X} \to \overline{\mathbb{R}}$ such that $\mathcal{G} = \text{Rg}(B)$;

ii) there exists a $\overline{\mathbb{R}}_{\max}$-sesquilinear map $\bar{F} : \mathcal{G} \to \mathcal{G}$ such that $\bar{F}\bar{F} = \text{Id}_{\mathcal{G}}$, i.e. $\bar{F}$ is an anti-involution over $\mathcal{G}$.

If these properties hold, then $\bar{F}$ can be taken as the restriction of $\bar{B}$ to Rg(B).

Defining tropical reproducing spaces

We now leverage the characterization, $\bar{B}\bar{B}\bar{B}\hat{f} = \bar{B}\hat{f}$, expressed for symmetric kernels in Theorem 7, to define tropical sesquilinear reproducing kernel spaces. As shown before,

$$\text{Rg}(B) = \{\sup_{x \in \mathcal{X}} a_x + b(\cdot, x) \mid a_x \in \mathbb{R}_\perp\},$$

so using spaces $\text{Rg}(B)$ to define RKMSs provides a direct analogy with the fact that a RKHS $\mathcal{H}_k$ is the completion for its norm $\|\cdot\|_k$ of $\text{span}(\{k(\cdot, x)\}_{x \in \mathcal{X}})$.

Definition

We call *reproducing kernel Moreau spaces* (RKMS) the complete submodules $\text{Rg}(B)$ of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ where $\bar{B}$ is a $\overline{\mathbb{R}}_{\max}$-sesquilinear continuous and hermitian operator associated with the symmetric kernel b. For all $\hat{g} \in \text{Rg}(B)$ and $x \in \mathcal{X}$, we say that they satisfy a *sesquilinear reproducing property*

$$\hat{g}(x) = (\bar{B}\bar{B}\hat{g})(x) = \langle \bar{B}\hat{g}, \bar{B}\delta_x^\top \rangle = \sup_{z \in \mathcal{X}} b(z, x) \dotminus [\sup_{y \in \mathcal{X}} b(z, y) \dotminus \hat{g}(y)]. \tag{10}$$

Around the reproducing property

$$\hat{g}(x) = (\bar{B}\bar{B}\hat{g})(x) = \langle \bar{B}\hat{g}, \bar{B}\delta_x^\top \rangle = \sup_{z \in \mathcal{X}} b(z, x) \dotdiv [\sup_{y \in \mathcal{X}} b(z, y) \dotdiv \hat{g}(y)].$$

The sesquilinear reproducing property is not an empty statement. It characterizes the elements of $\mathcal{G} = \mathrm{Rg}(B)$ through an immediate lemma, proved again using the identity $\bar{B}\bar{B}\bar{B} = \bar{B}$.

Lemma ([Singer, 1997] Corollary 8.5)

Let $\bar{B}$ be a $\overline{\mathbb{R}}_{\max}$-sesquilinear and continuous operator. Then for any $g \in \overline{\mathbb{R}}^{\mathcal{X}}$, $\hat{g} = \bar{B}\bar{B}\hat{g}$ holds if and only if $g \in \mathrm{Rg}(B)$.

For $b(x, y) = (x, y)_2$, $\bar{B}$ is the Fenchel conjugate operator, whence (10) is equivalent to Fenchel's theorem stating that convex l.s.c. functions are the only fixed points of the Fenchel biconjugate. Note that the difficult part in Fenchel's theorem is to prove that all convex l.s.c. functions are outputs of the Fenchel transform, i.e. hard to identify $\mathrm{Rg}(B)$.

Examples of spaces

We choose to interpret $\hat{g} = \bar{B}\bar{B}\hat{g}$ as a reproducing property, however $\bar{B}\bar{B}$ is not a $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-(sesqui)linear operator. There is a more direct interpretation with a $\overline{\mathbb{R}}_{\min}^{\mathcal{X}}$-linear operator.

Lemma

For all $\hat{g} \in \mathrm{Rg}(B)$ and $x \in \mathcal{X}$, (10) is equivalent to

$$\hat{g}(x) = \inf_{y \in \mathcal{X}} \hat{g}(y) \dotplus \sup_{z \in \mathcal{X}}[b(z,x) \dotdiv b(z,y)] =: \inf_{y \in \mathcal{X}} \hat{g}(y) \dotplus c(x,y) =: (C^{op}\hat{g})(x) \qquad (11)$$

where $c(x,y) := \sup_{z \in \mathcal{X}}[b(z,x) \dotdiv b(z,y)]$ is the Funk distance between $b(\cdot,x)$ and $b(\cdot,y)$, and $C^{op} : \overline{\mathbb{R}}_{\min}^{\mathcal{X}} \to \overline{\mathbb{R}}_{\min}^{\mathcal{X}}$ the related $\overline{\mathbb{R}}_{\min}$-linear operator.

Unfortunately c does not correspond to a unique b, nor to a unique $\mathrm{Rg}(B)$. More in the article about this and the difference with max-plus linear operators

Examples of spaces (cont.)

Examples of tpsd $b(x, y)$, $c(x, y)$ and $\mathrm{Rg}(B)$ [Singer, 1997]:

i) For $\mathcal{X} = \mathbb{R}^N$, $b(x, y) = (x, y)_2$ gives $c(x, y) = \delta_x^\top(y)$ whereas $\mathrm{Rg}(B)$ is the set of proper convex l.s.c. functions adding the constant functions $\pm\infty$.

ii) For $\mathcal{X} = \mathbb{R}^N$, $b(x, y) = -\|x - y\|^2$ gives $c(x, y) = \delta_x^\top(y)$ whereas $\mathrm{Rg}(B)$ is the set of proper 1-semiconvex l.s.c. functions adding the constant functions $\pm\infty$, i.e. $f + \|\cdot\|^2$ is convex.

iii) For any $\mathcal{X}$ and $\alpha \geq 0$, $b(x, y) = \begin{cases} 0 & \text{if } y = x, \\ -\alpha & \text{otherwise,} \end{cases}$ gives $c(x, y) = -b(x, y)$ whereas $\mathrm{Rg}(B)$ is the set of functions f which difference $f(x) - f(y)$ is smaller than α. For $\alpha = +\infty$, $b(x, y) = \delta_x^\perp(y)$, $\mathrm{Rg}(B)$ corresponds to the whole $\overline{\mathbb{R}}^{\mathcal{X}}$.

iv) For $(\mathcal{X}, d)$ a metric space, $b(x, y) = -d(x, y)^p$ gives $c(x, y) = d(x, y)^p$ whereas $\mathrm{Rg}(B)$ is the set of $(1, p)$-Hölder continuous functions w.r.t. the distance d (i.e. $|f(x) - f(y)| \leq 1 \cdot d(x, y)^p$), when adding the constant functions $\pm\infty$.

Optimization on tropical function spaces

Dedicated to PEP people:

$$\min_{f \in \mathcal{F}} \mathcal{L}((f(x_m))_{m \in \mathcal{I}}) \text{ e.g. } \sum |f(x_m) - y_m)|^2$$

Given two sets $\mathcal{X}$ and $\mathcal{X}'$, a kernel $b : \mathcal{X} \times \mathcal{X}' \to \mathbb{R}_\perp$, and a subset $\hat{\mathcal{X}} \subset \mathcal{X}$, define

$$\mathsf{Rg}_{\partial\text{-}\hat{\mathcal{X}}}(B) := \left\{ f \in \mathsf{Rg}(B) \,|\, \forall x \in \hat{\mathcal{X}}, \, \exists p_x \in \mathcal{X}', \, f(x) = b(x, p_x) \dot{-} \bar{B}f(p_x) = \sup_{p \in \mathcal{X}'} b(x, p) \dot{-} \bar{B}f(p) \right\}$$

This set can be understood as the subset of functions of $\mathsf{Rg}(B)$ for which there exists a B-subdifferential at every point of $\hat{\mathcal{X}}$.

For instance, for $b_c(x, y) = (x, y)_2$ and $\mathcal{X} = \mathcal{X}' = \mathbb{R}^N$, $\mathsf{Rg}_{\partial\text{-}\mathcal{X}}(B)$ contains the continuous convex functions, but it is strictly smaller than $\mathsf{Rg}(B)$ since convex l.s.c. functions may have an empty subdifferential at points that do not lie in the relative interior of their domain.

Optimization on tropical function spaces: interpolation theorem

Proposition (Tropical interpolation)

Let $\mathcal{I}$ be a nonempty index set, given $(x_m, y_m)_{m \in \mathcal{I}} \in (\mathcal{X} \times \mathbb{R})^{\mathcal{I}}$, setting $\hat{\mathcal{X}} = \{x_m\}_{m \in \mathcal{I}}$, the three following statements are equivalent:

i) there exists $f \in \mathrm{Rg}_{\partial\text{-}\hat{\mathcal{X}}}(B)$ such that $y_m = f(x_m)$ for all $m \in \mathcal{I}$;

ii) there exists $(p_m)_{m \in \mathcal{I}} \in (\mathcal{X}')^{\mathcal{I}}$ such that $y_m = f^0(x_m)$ for all $m \in \mathcal{I}$, for

$$f^0(\cdot) := \bar{B}(\inf_{m \in \mathcal{I}}[\delta_{p_m}^\top(\cdot) \dotplus b(x_m, p_m) - y_m]) = \max_{m \in \mathcal{I}} b(\cdot, p_m) \dotdiv b(x_m, p_m) + y_m;$$

iii) there exists $(p_m)_{m \in \mathcal{I}} \in (\mathcal{X}')^{\mathcal{I}}$ such that $y_n - y_m \geq b(x_n, p_m) \dotdiv b(x_m, p_m)$ for all $n, m \in \mathcal{I}$.

Proof of interpolation theorem

i)$\Rightarrow$ii). Set p_m for each x_m s.t. $f(x_m) = b(x_m, p_m) \dot{-} \bar{B}f(p_m)$, then, for all $n \in \mathcal{I}$,

$$y_n = b(x_n, p_n) \dot{-} \bar{B}f(p_n) \leq \max_{m \in \mathcal{I}} b(x_n, p_m) \dot{-} \underbrace{(b(x_m, p_m) - y_m)}_{=\bar{B}f(p_m)} = f^0(x_n)$$

$$\leq \sup_{p \in \mathcal{X}'} b(x_n, p) \dot{-} \bar{B}f(p) = \bar{B}\bar{B}f(x_n) = f(x_n) = y_n,$$

so $f^0(x_n) = y_n$.

ii)$\Rightarrow$i). We directly have that $f^0 \in \text{Rg}(B)$ and that p_m is a subdifferential at each x_m, whence $f^0 \in \text{Rg}_{\partial \cdot \hat{\mathcal{X}}}(B)$.

ii)$\Leftrightarrow$iii). This follows from the definition of f^0.

Optimization on tropical function spaces: representer theorem

---

Corollary (Representer theorem)

Given points $(x_m)_{m \in \mathcal{I}} \in \mathcal{X}^{\mathcal{I}}$ and a function $\mathcal{L} : \overline{\mathbb{R}}^{\mathcal{I}} \to \overline{\mathbb{R}}$, fix $\hat{\mathcal{X}} = \{x_m\}_{m \in \mathcal{I}}$. Then, if the problem

$$\min_{f \in \mathrm{Rg}(B)} \mathcal{L}((f(x_m))_{m \in \mathcal{I}}) \tag{12}$$

has a solution $\bar{f} \in \mathrm{Rg}_{\partial \cdot \hat{\mathcal{X}}}(B)$ with finite values $(f(x_m))_{m \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$, it also has a solution f^0 as in Proposition 4-ii) which can be obtained solving

$$\min_{(p_m, y_m)_{m \in \mathcal{I}} \in (\mathcal{X}' \times \mathbb{R})^M} \mathcal{L}((y_m))_{m \in \mathcal{I}}) \tag{13}$$

$$s.t. \ y_n - y_m \geq b(x_n, p_m) \ \dot{-} \ b(x_m, p_m), \ \forall \, n, m \in \mathcal{I}.$$

Conversely, if (13) has a solution, then it is also a solution in $\mathrm{Rg}_{\partial \cdot \hat{\mathcal{X}}}(B)$ of (12).

WE DO NOT NEED ANY PROPERTY OF THE KERNEL b!

Optimization on tropical function spaces: example of convex regression

When b is the standard scalar product and $\mathcal{I}$ is finite, each p_m can be interpreted as a subgradient at x_m, and Theorem 11 recovers a well-known property in convex regression, [Boyd and Vandenberghe, 2004][Section 6.5.5]

$$\min_{f \in \text{CVEX}} \sum |f(x_m) - \bar{y}_m|^2 \Leftrightarrow \min_{\substack{(p_m, y_m)_{m \in \mathcal{I}} \in (\mathbb{R}^d \times \mathbb{R})^M, \\ y_n - y_m \geq (x_n, p_m)_2 - (x_m, p_m)_2}} \sum |y_m - \bar{y}_m|^2.$$

We have thus shown that this result also holds for very general kernels and uncountable set $\mathcal{I}$, not even assuming symmetry or tropical positivity of b. Consequently Proposition 4 should be related to interpolation theorems such as [Taylor et al., 2016, Theorem 4].

For instance for μ-strongly convex functions with L-bounded gradient (which corresponds to $\mathcal{X} = \mathbb{R}^d$, $\mathcal{X}' = \{p \in \mathbb{R}^d, \|p\|_2 \leq L\}$ and $b(x, p) = (x, p)_2 + \mu \|x\|_2^2$).

Fundamental spacetime solution of HJB

Take $\mathcal{X} \subset \mathbb{R} \times \mathbb{R}^d$ the spacetime vector space, i.e. $x = (t, r)$. Given a Lagrangian function $L : \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_\top$, the action $J((t_0, r_0), (t_1, r_1), r(\cdot))$ along an absolutely continuous trajectory $r(\cdot) : [t_0, t_1] \to \mathbb{R}^d$ going from (t_0, r_0) to (t_1, r_1) is defined as follows:

$$J((t_0, r_0), (t_1, r_1), r(\cdot)) := \int_{t_0}^{t_1} L(s, r(s), \dot{r}(s)) ds \text{ with } r(t_0) = r_0, \ r(t_1) = r_1. \tag{14}$$

The optimal stopping time problem, where $w(t_1, r_1)$ is a final cost obtained for choosing to leave the game at (t_1, r_1) [Bensoussan and Lions, 1982, Barles and Perthame, 1987].

$$\forall x_0 = (t_0, r_0) \in \mathcal{X}, \ f(x_0) = - \inf_{\substack{(t_1, r_1) \in \mathcal{X}, \ t_1 \geq t_0, \\ r(\cdot) \in S_{\text{traj}}, \ r(t_0) = r_0 \\ r(t_1) = r_1}} \int_{t_0}^{t_1} L(s, r(s), \dot{r}(s)) ds \dotplus w(t_1, r_1).$$

The player has then to determine, given w and starting from (t_0, r_0), the corresponding final and optimal (t_1, r_1).

Definition (Maupertuis kernel)

Given a set of absolutely continuous trajectories $S_{\text{traj}} \subset C^0(\mathbb{R}, \mathbb{R}^d)$, we define the Maupertuis kernel $b_{\text{Maup}} : \mathcal{X} \times \mathcal{X} \to \overline{\mathbb{R}}$, and the asymmetrical b_{Maup}^{asym}, between $x_0 = (t_0, r_0)$ and $x_1 = (t_1, r_1)$

$$b_{\text{Maup}}(x_0, x_1) := -\operatorname{sign}(t_1 - t_0) \inf_{\substack{r(\cdot) \in S_{\text{traj}} \\ r(t_0) = r_0, r(t_1) = r_1}} \int_{t_0}^{t_1} L(s, r(s), \dot{r}(s)) ds, \tag{15}$$

$$b_{\text{Maup}}^{asym}(x_0, x_1) := (1 + \delta_{t_1 \geq t_0}^\top) b_{\text{Maup}}(x_0, x_1), \tag{16}$$

with $\delta_{t_1 \geq t_0}^\top = 0$ if $t_1 \geq t_0$ and $+\infty$ otherwise.

Lemma (Tropical positivity for nonnegative Lagrangian)

If $L(\cdot) \geq 0$, then b_{Maup} is tpsd.

Evaluating the Maupertuis kernel

The evaluation of a Maupertuis kernel amounts to solving an optimal control problem [Kolokoltsov and Maslov, 1997, McEneaney, 2006]. In general, computing b_{Maup} is as hard as solving a HJB PDE.

However there are cases where the value of b_{Maup} is known, for instance through Lax-Hopf formulas [Cannarsa and Sinestrari, 2004][Theorem 1.3.1]. This corresponds to the case where $L(s, r, v) = L(v)$, L is convex, and $S_{\text{traj}} = W^{1,1}(\mathbb{R}, \mathbb{R}^d)$, in which case (15) writes simply as a perspective function for $t_1 \neq t_0$

$$b_{\text{Maup}}^{Hopf}(x_0, x_1) := -|t_1 - t_0| L\left(\frac{r_1 - r_0}{t_1 - t_0}\right) \qquad (17)$$

and 0 if $x_0 = x_1$, $-\infty$ if $r_1 \neq r_0$ and $t_0 = t_1$.

Unlike the off-the-shelf Hilbertian kernels used typically in machine learning, the kernel b_{Maup} is canonically defined by the triplet $(\mathcal{X}, S_{\text{traj}}, L)$.

Identifying $\mathrm{Rg}(B^{\mathsf{asym}}_{\mathsf{Maup}})$

For every function $f \in \mathrm{Rg}(b^{\mathsf{asym}}_{\mathsf{Maup}})$, there exists $w : \mathfrak{X} \mapsto \mathbb{R}_\top$ such that

$$\forall\, x_0 = (t_0, r_0) \in \mathfrak{X}, \ f(x_0) = \sup_{x_1 \in \mathfrak{X}} b^{\mathsf{asym}}_{\mathsf{Maup}}(x_0, x_1) \dotminus w(x_1)$$

$$= - \inf_{\substack{(t_1, r_1) \in \mathfrak{X},\, t_1 \geq t_0, \\ r(\cdot) \in S_{\mathsf{traj}},\, r(t_0) = r_0 \\ r(t_1) = r_1}} \int_{t_0}^{t_1} L(s, r(s), \dot{r}(s)) ds \dotplus w(t_1, r_1). \qquad (18)$$

This is precisely the optimal stopping time problem!

Fundamental spacetime solution of HJB: example 1

Inverse optimal control problem with stopping time

If the terminal cost w in the optimal stopping problem (18) is unknown, but the Lagrangian L is known. How to infer w or the value function?

Assume we know $\bar{y}_m \simeq f(x_m)$, $x_m = (t_m, r_m) \in \mathfrak{X}$. Cast this as an inverse problem by considering a loss function

$$\mathcal{L}((f(x_m))_{m \in \mathcal{I}}) = \sum_{m \in \mathcal{I}} |\bar{y}_m - f(x_m)|$$

Then, by Theorem 11, reconstructing the unknown stopping cost amounts to finding solutions $(p_m, y_m)_{m \in \mathcal{I}} \in (\mathfrak{X} \times \mathbb{R})^{\mathcal{I}}$ minimizing $\mathcal{L}((y_m)_{m \in \mathcal{I}})$ under the constraints

$$\forall\, n, m \in \mathcal{I},\ y_n - y_m \geq b_{\mathrm{Maup}}^{\mathrm{asym}}(x_n, p_m) \dot{-} b_{\mathrm{Maup}}^{\mathrm{asym}}(x_m, p_m).$$

Using that b_{Maup} is idempotent, we have that $b_{\mathrm{Maup}}^{\mathrm{asym}}(x_n, p_m) \dot{-} b_{\mathrm{Maup}}^{\mathrm{asym}}(x_m, p_m) \geq b_{\mathrm{Maup}}^{\mathrm{asym}}(x_n, x_m)$, achieved for $p_m = x_m$, whence the problem reduces to minimizing the function $\mathcal{L}$ over $(y_m)_{m \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$ such that $\forall\, n, m \in \mathcal{I},\ y_n - y_m \geq b_{\mathrm{Maup}}^{\mathrm{asym}}(x_n, x_m)$. If $(y_m^*)_{m \in \mathcal{I}}$ is an optimal solution of the above problem, an admissible stopping cost is simply $w = \inf_{m \in \mathcal{I}}[\delta_{x_m}^{\top}(\cdot) - y_m^*]$.

Fundamental spacetime solution of HJB: example 2

Space-restrictions of b_{Maup}

With fixed $t_0, t_1 \in \mathbb{R}$, one can also consider a kernel $b_{\text{Maup}}^{[t_0, t_1]}(r_0, r_1) := b_{\text{Maup}}(x_0, x_1)$.
[Kolokoltsov and Maslov, 1997, McEneaney, 2006, Dower and Zhang, 2015]

Interpolation of the value function with a fixed final time

Assume t_0 and T are fixed, and that the Lagrangian L is known. Our aim is to recover an unknown terminal cost ψ_T, and thus, the value function $\bar{V}$ everywhere, given measurements $(\bar{y}_m)_{m \in \mathcal{I}} \in \mathbb{R}^{\mathcal{I}}$ at sample points. Consider the exact interpolation problem, with $\bar{y}_m = -\bar{V}(t_0, r_m)$ known at some point r_m. By Proposition 4, the interpolation problem corresponds to finding solutions $(p_m)_{m \in \mathcal{I}} \in (\mathbb{R}^d)^{\mathcal{I}}$ of

$$\forall\, n, m \in \mathcal{I},\ \bar{y}_n - \bar{y}_m \geq b_{\text{Maup}}^{[t_0, T]}(r_n, p_m) \dot{-} b_{\text{Maup}}^{[t_0, T]}(r_m, p_m).$$

Here each p_m can be interpreted as a point to reach at time T starting from (t_0, r_m). Given such p_m, an admissible terminal cost is $\psi_T(r) = \inf_{m \in \mathcal{I}}[\delta_{p_m}^\top(r) \dot{+} b(x_m, p_m) - \bar{y}_m]$.

Conclusion

- the characterization of tropical reproducing kernels is similar to the Hilbertian case

- it is possible to do optimization over tropical function spaces

- it is useful for value functions in control theory

Open problems:

- link max-plus space with log-sum-exp applications in ML

- link tpsd kernels with cyclic monotonicity in optimal transport

- why is tpsd not necessary for optimization? Find something specific to tpsd

- do some learning applications, e.g. apply the kernels to do systematic PEP

Motivation
0000000

Kernels
00000000

Properties
000000000

App: optimization
00000

App: HJB
0000000●0000

Conclusion

- the characterization of tropical reproducing kernels is similar to the Hilbertian case

- it is possible to do optimization over tropical function spaces

- it is useful for value functions in control theory

Open problems:
- link max-plus space with log-sum-exp applications in ML

- link tpsd kernels with cyclic monotonicity in optimal transport

- why is tpsd not necessary for optimization? Find something specific to tpsd

- do some learning applications, e.g. apply the kernels to do systematic PEP

Thank you for your attention!
Future interests: majorization-minimization, backward SDEs, diffusion models

References I

📄 Akian, M., Gaubert, S., and Kolokoltsov, V. N. (2005).

Set coverings and invertibility of functional Galois connections.

In Litvinov, G. L. and Maslov, V. P., editors, *Idempotent Mathematics and Mathematical Physics*, Contemporary Mathematics, pages 19–51. American Mathematical Society.

📄 Barles, G. and Perthame, B. (1987).

Discontinuous solutions of deterministic optimal stopping time problems.

ESAIM: Mathematical Modelling and Numerical Analysis, 21(4):557–579.

📄 Bensoussan, A. and Lions, J.-L. (1982).

Applications of variational inequalities in stochastic control.

Studies in mathematics and its applications. Elsevier Science.

📄 Boyd, S. and Vandenberghe, L. (2004).

Convex Optimization.

Cambridge University Press.

References II

📄 Cannarsa, P. and Sinestrari, C. (2004).
Semiconcave Functions, Hamilton—Jacobi Equations, and Optimal Control.
Birkhäuser Boston.

📄 Dower, P. M. and McEneaney, W. M. (2015).
A max-plus dual space fundamental solution for a class of operator differential Riccati equations.
SIAM Journal on Control and Optimization, 53(2):969–1002.

📄 Dower, P. M. and Zhang, H. (2015).
A new fundamental solution for differential riccati equations arising in l2-gain analysis.
In *2015 5th Australian Control Conference (AUCC)*, pages 65–68.

📄 Kolokoltsov, V. N. and Maslov, V. P. (1997).
Idempotent analysis and its applications, volume 401 of *Mathematics and its Applications*.
Kluwer Academic Publishers Group.

References III

📄 Litvinov, G. L. (2005).

Maslov dequantization, idempotent and tropical mathematics: A brief introduction.

Journal of Mathematical Sciences, 140:426–444.

📄 Martinez-Legaz, J. and Singer, I. (1990).

Dualities between complete lattices.

Optimization, 21(4):481–508.

📄 McEneaney, W. M. (2006).

Max-Plus Methods for Nonlinear Control and Estimation.

Birkhäuser-Verlag.

📄 Singer, I. (1984).

Conjugation operators.

In *Lecture Notes in Economics and Mathematical Systems*, pages 80–97. Springer Berlin Heidelberg.

References IV

Singer, I. (1997).
Abstract Convex Analysis.
Wiley-Interscience and Canadian Mathematics Series of Monographs and Texts. Wiley-Interscience, 1 edition.

Taylor, A. B., Hendrickx, J. M., and Glineur, F. (2016).
Smooth strongly convex interpolation and exact worst-case performance of first-order methods.
Mathematical Programming, 161(1-2):307–345.