THE TROPICAL ANALOGUES OF REPRODUCING KERNELS

PIERRE-CYRIL AUBIN-FRANKOWSKI AND STÉPHANE GAUBERT

Abstract. Hilbertian kernel methods and their positive semidefinite kernels have known an extensive use in various fields of applied mathematics and machine learning, owing to their several equivalent characterizations. We here unveil an analogy with concepts from tropical geometry, proving that tropical positive semidefinite kernels are also endowed with equivalent viewpoints, stemming from Fenchel-Moreau conjugations. We give a tropical analogue of Aronszajn theorem, showing that these kernels correspond to a feature map, define monotonous operators, and generate max-plus function spaces endowed with a reproducing property. They furthermore include all the Hilbertian kernels classically studied as well as Monge arrays. However, two relevant notions of tropical reproducing kernels must be distinguished, based either on linear or sesquilinear interpretations. The sesquilinear interpretation is the most expressive one, since reproducing spaces then encompass classical max-plus spaces, such as those of (semi)convex functions. In contrast, in the linear interpretation, the reproducing kernels are characterized by a restrictive condition, von Neumann regularity.

1. Introduction

Since the foundation of their theory (see Aronszajn, 1950, for a historical summary), reproducing kernel Hilbert spaces (RKHSs) have played an eminent role among linear function spaces (Saitoh and Sawano, 2016), all the more in machine learning (Steinwart and Christmann, 2008; Schölkopf and Smola, 2002). However optimization problems frequently involve some more intricate functions spaces requiring dedicated structures (Pallaschke and Rolewicz, 1997), such as the space of convex functions used in convex regression (e.g. Seijo and Sen, 2011). For instance, value functions, the solutions of the Hamilton-Jacobi-Bellman (HJB) equation in optimal control, are generically nonsmooth while still being semiconcave (Cannarsa and Sinestrari, 2004). Moreover, Lax-Oleinik semigroups, i.e., evolution semigroups of HJB equations with a Hamiltonian convex in the adjoint variable, are tropically linear. This fact spurred research in tropical functional analysis (Kolokoltsov and Maslov, 1997; McEneaney, 2006). At the price of a change to $(\max, +)$ operations, many concepts, e.g. operators, have been defined by analogy with the linear setting. This allowed to develop new classes of numerical methods for the HJB equation (Fleming and McEneaney, 2000; Akian et al., 2008; McEneaney, 2006,

Date: February 23, 2022.
2020 *Mathematics Subject Classification.* 46E22; 1410T; 52A01.
Key words and phrases. Reproducing kernels, Moreau conjugacies, Tropical geometry, Idempotent analysis, Positivity, Generalized convexity.

2007; Dower and McEneaney, 2015). Several research directions have been explored at the interface between tropical geometry, probablity theory and machine learning. These include studies of the tropicalization of stochastic processes (Akian et al., 1994) or of Gaussian measures (Tran, 2020), tropical support vector machines (Yoshida et al., 2021), tropical principal component analysis (Yoshida et al., 2019) inspired by phylogenetic studies, quantification of the expressivity of deep neural networks (Zhang et al., 2018; Montúfar et al., 2021) or their approximation (Calafiore et al., 2020) through tropical methods. A survey of some of these approaches can be found in Maragos et al. (2021). The proper tropical analogue of RKHSs still remained elusive nonetheless.

One key property of RKHSs is indeed the versatility entailed by the multiple entry points to their theory, either through a real-valued kernel, a nonlinear feature map, a function space or an integral operator. In this article, we uncover similar links between a tropical kernel, a.k.a. a coupling function, its factorization, a tropical function space and monotone operators. Our construction revolves around the famed Fenchel-Moreau conjugations (see e.g. Singer, 1997, Chapter 8) with a particular role played by the assumption of symmetry of the kernel. A conjugation such as the Legendre-Fenchel transform is at the core of convex optimization. Similarly, conjugations appear in the dual formulation of optimal transport (Santambrogio, 2017, Section 4.1). Conjugations have been known since Singer (1984) to correspond to a tropical kernel (see also Akian et al., 2005, for the extension to general Galois connections). Recent applications of conjugations were investigated for instance in Volle et al. (2013); Chancelier and Lara (2021). We here combine the relation between conjugations and kernels with insights from Pallaschke and Rolewicz (1997, Chapter 1) on monotonicity to provide a more complete picture.

More precisely, as shown by Aronszajn (1950), the key result in the Hilbertian case is that positive semidefinite kernels coincide with reproducing kernels, i.e.

Theorem 1 (Aronszajn (1950)). *Given a kernel $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, 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:* $\forall\, M \in \mathbb{N}^*$, $\forall\, (a_m, x_m) \in (\mathbb{R} \times X)^M$, $\sum_{n,m=1}^{M} a_n a_m k(x_n, x_m) \geqslant 0$;

ii) there exists a Hilbert space $(\mathcal{H}, (\cdot, \cdot)_{\mathcal{H}})$ *and a* feature map $\Phi : \mathcal{X} \to \mathcal{H}$ *such that*
   - $\forall\, x, y \in X,\ k(x, y) = (\Phi(x), \Phi(y))_{\mathcal{H}}$;

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

Our goal is to obtain the analogue of Theorem 1 in a max-plus context, and our first main result can be stated informally as follows:

Theorem 2 (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) k 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* $\{\sup_{n\in\{1,\dots,N\}} a_n + b(\cdot, x_n) \,|\, N \in \mathbb{N}^*, a_n \in \mathbb{R}_\perp, x_n \in \mathcal{X}\}$*, and* b *defines a tropical Cauchy-Schwarz inequality over* $\overline{\mathbb{R}}^{\mathcal{X}}$*.*

The precise statements are to be found in Proposition 2, Proposition 3, the definitions and notation being detailed below. While Aronszajn (1950) stresses the connection of reproducing kernels with inner products as in Theorem 1-ii), Schwartz (1964, Propositions 8 and 9) on the contrary approaches kernels through duality pairings. This duality viewpoint will prove important in the following and is summarized in a dictionary to be found in Table 1, translating the concepts from the Hilbertian to the tropical world.

Tropical positive semidefinite matrices have been investigated for discrete sets $\mathcal{X}$ (Yu, 2014; Cartwright and Chan, 2012). In particular, Yu defined the tropical positive matrices as the image by the nonarchimedean valuation of the set of positive semidefinite matrices over a real closed nonarchimedean field, and showed that they are characterized by the positivity of their 2×2 tropical minors. This result will appear as a special case of the tropical positivity of kernels. We also provide an extension to the infinite dimensional setting of a theorem of Cartwright and Chan (2012) showing that any tropical positive semidefinite matrices can be factorized as in Theorem 2-ii) above. The tropicalization of the subclass of totally positive matrices was also investigated in Gaubert and Niv (2018). While our analysis tackles general sets $\mathcal{X}$, we leverage the finite case to foster intuitions, and mostly for counter-examples.

As a matter of fact the tropical world can be equipped with two sets of operations, the "linear" $(\max, +)$ and the "sesquilinear" $(\max, -)$. We shall see, as hinted at by the sesquilinear Legendre-Fenchel transform, that the "linear" interpretation of the reproducing property is more restrictive than the "sesquilinear" version. Indeed, we show in Theorem 4 that the "linear" reproducing kernels are characterized by von Neumann regularity; in particular, their ranges are images of linear retractions (idempotent linear maps). von Neumann regularity is much more restrictive in the tropical setting than in the Hilbertian setting: whereas any closed subspace of a Hilbert space is the range of a linear retraction (the orthogonal projection over this space), ranges of tropically linear retractions are rare, and for instance they do not include spaces of convex functions. This comes as a negative result on the descriptive ability of "linear" approaches, studied by e.g. Litvinov (2011).

It has been often hinted that max-plus analysis can be interpreted as a limit case of log-sum-exp operations, a property known as Maslov's dequantization (Litvinov, 2005), akin to taking the limit in Planck's constant in passing from quantum to classical mechanics. Similar interpolating relations between optimal transport and RKHSs have been discussed recently in Feydy et al. (2019). Our efforts are thus directed toward bridging the gap between the max-plus and the Hilbertian worlds, defining the adequate tropical analogue to RKHSs. This first enquiry opens many theoretical questions, concerning for instance the analogue to the topological characterization of RKHSs or to the minimal factorization of the kernel (see Steinwart and Christmann, 2008, Theorem 4.21). It also opens up the question of applications of the theory and of the computational advantages of tropical positive semidefiniteness, since tropical kernels generally benefit from representer theorems (see Corollary 1 below).

The paper is structured as follows. Preliminaries on conjugations are summarized in Section 2. Section 3 introduces tropical positive definite kernels and their factorization. In Section 4, these kernels are identified with monotone operators and shown to lead to a tropical Cauchy-Schwarz inequality. In Section 5 we characterize the tropical ranges of symmetric kernels, which we define as tropical reproducing kernel spaces. The "linear" approach is investigated in Section 6. We conclude stating a representer theorem in Section 7.

2. Tropical functional analysis preliminaries

Notations: Let $\mathbb{N} = \{0, 1, \ldots\}$, $\mathbb{N}^* = \{1, 2, \ldots\}$, $\mathbb{R}_+$ and $\mathbb{R}_+^*$ denote the set of natural numbers, positive integers, non-negative and positive reals, respectively. We use the shorthand $[M] = \{1, \ldots, M\}$. The extended real line is denoted by $\overline{\mathbb{R}} = [-\infty, +\infty]$, we also use the notations $\mathbb{R}_\top = (-\infty, +\infty]$ and $\mathbb{R}_\perp = [-\infty, +\infty)$. These sets are equipped with the upper (resp. lower) addition and subtraction (Moreau, 1970, p.3), extending the usual operations by indicating through a dot whether $+\infty$ or $-\infty$ is absorbing, e.g. $\infty \dot{+} (-\infty) = -\infty$, $\infty \dot{-} \infty = +\infty$. When $\overline{\mathbb{R}}$ is equipped with the $(\max, +)$ (resp. $(\min, +)$) operations we denote it by $\overline{\mathbb{R}}_{\max}$ (resp. $\overline{\mathbb{R}}_{\min}$). Given any sets $\mathcal{X}$ and E, we denote by $E^{\mathcal{X}}$ the set of functions f from $\mathcal{X}$ to E. If E is equipped with $(\min, +)$ operations we denote generically its elements by $\hat{f}$. 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. When $\mathcal{X}$ is equipped with a topology, a function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is said to be lower semicontinuous (l.s.c.) if its epigraph is a closed subset of $\mathcal{X} \times \overline{\mathbb{R}}$.

In all that follows, given a set $\mathcal{X}$, a *kernel* b is a function from $\mathcal{X} \times \mathcal{X}$ to $\overline{\mathbb{R}}$. We shall be interested in at least three specific kernels over $\mathcal{X} = \mathbb{R}^N$ equipped with its Euclidean norm $\|\cdot\|_2$ and inner product $(\cdot, \cdot)_2$ or with a distance d:

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

These kernels are respectively related to the sets of proper convex l.s.c. functions, 1-semiconvex l.s.c. functions, and 1-Lipschitz functions w.r.t. the distance d (see p.13). 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

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

both with the convention that $-\infty$ is absorbing. For instance, for $b = b_c$, $\bar{B}$ is the Fenchel transform which is indeed sesquilinear as $\bar{B}(\min(f, g)) = \max(\bar{B}f, \bar{B}g)$ and $\bar{B}(f - \lambda) = \bar{B}f + \lambda$, effectively turning $(\min, -)$ into $(\max, +)$ operations. More formally:

Definition 1. 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 $B(f \dot{+} \lambda) = Bf \dot{+} \lambda$ (with $+\infty$ absorbing on both sides), for any finite index set I and $\lambda \in \overline{\mathbb{R}}$; we say in addition that B is continuous if $B(\sup\{f_i\}_{i \in I}) = \sup\{Bf_i\}_{i \in I}$ holds even for infinite families.

ii) $\overline{\mathbb{R}}_{\max}$-sesquilinear if $B(\inf\{f_i\}_{i\in I}) = \sup\{Bf_i\}_{i\in I}$ and $B(f \dot{+} \lambda) = Bf \dot{-} \lambda$ (with $+\infty$ absorbing on the l.h.s. and $-\infty$ absorbing on the r.h.s.), for any finite index set I and $\lambda \in \overline{\mathbb{R}}$; we say in addition that B is continuous if $B(\inf\{f_i\}_{i\in I}) = \sup\{Bf_i\}_{i\in I}$ holds even for infinite families.

The range $\mathrm{Rg}(B)$ of B is defined as the set of functions $g \in \overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ such that $g = Bf$ for some $f \in \overline{\mathbb{R}}_{\max}^{\mathcal{X}}$.

The continuity notion we introduced needs actually to be defined more generally for our analysis. Recall than an ordered set D is directed if for all $d, d' \in D$, there exists $d'' \in D$ such that $d \leqslant d''$ and $d' \leqslant d''$.

Definition 2. Let $\mathcal{G}$ and $\mathcal{H}$ be complete lattices. A map $B : \mathcal{G} \to \mathcal{H}$ will be said to be continuous, if for all directed subsets $D \subset \mathcal{G}$, $B(\sup D) = \sup B(D)$.

This is precisely the notion of continuity with respect to the *Scott topology* (Gierz et al., 2003), which, as noted in Akian (1999); Cohen et al. (2004), is a canonical one in idempotent analysis. In particular, a $\overline{\mathbb{R}}_{\max}$-linear map is an order preserving self-map of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$, and one can check it is continuous iff the relation $B(\sup\{f_i\}_{i\in I}) = \sup\{Bf_i\}_{i\in I}$ holds for all infinite families of functions $\{f_i\}_{i\in I} \subset \overline{\mathbb{R}}^{\mathcal{X}}$. Dually, we shall think of a $\overline{\mathbb{R}}_{\max}$-sesquilinear map B as being order preserving, from $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ equipped with the *opposite* to the standard order, to $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ equipped with the standard order. In this way a $\overline{\mathbb{R}}_{\max}$-sesquilinear B can be seen to be continuous iff the relation $B(\inf\{f_i\}_{i\in I}) = \sup\{Bf_i\}_{i\in I}$ holds for all infinite families of functions $\{f_i\}_{i\in I} \subset \overline{\mathbb{R}}^{\mathcal{X}}$, in which the infima and suprema refer to the standard order. Observe that no ambiguity will arise from our convention because a (Scott) continuous map is automatically order preserving.

We define the indicator functions $\delta_x^{\perp}, \delta_x^{\top} \in \overline{\mathbb{R}}^{\mathcal{X}}$ as follows[1]

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

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 1 (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 $\bar{B}f(x) = \sup_{y\in\mathcal{X}} b(x,y) \dot{-} f(y)$. Moreover in this case b is uniquely determined by $\bar{B}$ as $b(\cdot, x) = \bar{B}\delta_x^{\top}$.*

Remark 1. By a change of sign, this result also holds for $\overline{\mathbb{R}}_{\max}$-linear operators, $Bf(x) = \sup_{y\in\mathcal{X}} b(x,y) \dot{+} f(y)$. We refer to Akian et al. (2005, Theorem 2.1) for extensions of Proposition 1 which is the tropical analogue of Riesz representation theorem in Hilbert spaces (see also Martinez-Legaz and Singer, 1990). The $\overline{\mathbb{R}}_{\max}$-(sesqui)linear and continuous maps verify that $\mathrm{Rg}(B)$ is a max-plus completion in the sense that it is the smallest complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ containing

[1] Introducing two indicator functions is necessary because of change of signs and of the $\max - \min$ duality.

TABLE 1. Corresponding concepts between Hilbertian and tropical kernels

Concept	Hilbertian kernel	Tropical kernel	Reference
symmetry	$k(x,y)=k(y,x)$	$b(x,y)=b(y,x)$	Def. 4
positivity	$\sum_{i,j} a_i a_j k(x_i,x_j) \geqslant 0$	$b(x,x)+b(y,y) \geqslant b(x,y)+b(y,x)$	Def. 4
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	$K(\mu)(x)=\int_{\mathcal{X}} k(x,y)\mathrm{d}\mu(y)$	$\bar{B}(\hat{f})(x)=\sup_{y\in\mathcal{X}} b(x,y) \mathbin{\dot{-}} \hat{f}(y)$	Prop. 1
monotone operator	$\langle \mu, K(\mu)\rangle_{\mathbb{R}^{\mathcal{X},*}\times\mathbb{R}^{\mathcal{X}}} \geqslant 0$	$\langle \hat{f},\bar{B}\hat{f}\rangle \mathbin{\dot{+}} \langle \hat{g},\bar{B}\hat{g}\rangle \geqslant \langle \hat{f},\bar{B}\hat{g}\rangle \mathbin{\underset{\cdot}{+}} \langle \hat{g},\bar{B}\hat{f}\rangle$	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. 5

$\{\sup_{n\in\{1,\dots,N\}} a_n \mathbin{\underset{\cdot}{+}} b(\cdot,x_n) \mid N\in\mathbb{N}^*, a_n\in\mathbb{R}_\perp, x_n\in\mathcal{X}\}$ and

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

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 $\overline{\mathbb{R}}_{\min}^{\mathcal{X}}$,

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

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

Definition 3. 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).

3. TROPICAL POSITIVE SEMIDEFINITE KERNELS

Definition 4. 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: $\forall x,y\in\mathcal{X},\ b(x,x)+b(y,y)\geqslant b(x,y)+b(y,x)$.

Notice that all the three kernels of (1) are tropically positive semidefinite and finite-valued. Moreover Definition 4 implies that every Hilbertian positive semidefinite kernel is tropically positive semidefinite.[2] Any square Monge matrix[3] corresponds to a tpsd kernel, since the latter relaxes the Monge requirements on all

[2]This is also true for Hilbertian conditionally positive semidefinite (cpsd) kernels, those for which Theorem 1-i) only holds for $(a_m)_{m\in[M]}\in\mathbb{R}^M$ such that $\sum_{m=1}^M a_m=0$. Note in the tropical setting we only need to require the property to hold for $M=2$. For $M=1$, tropical positivity is always satisfied since $b(x,x)\geqslant-\infty$. For $M>2$ we refer to Proposition 3.

[3]A Monge matrix $B\in\overline{\mathbb{R}}^{M\times N}$ is one for which $b_{ij}+b_{mn}\leqslant b_{in}+b_{mj}$, for all $1\leqslant i<m\leqslant M$ and $1\leqslant j<n\leqslant N$. They correspond to submodular functions over discrete sets, we refer to

the 2-by-2 minors to only the principal minors. Similarly, the logarithm of any positive semidefinite kernel with nonnegative values is a tpsd kernel.[4] An immediate consequence of Definition 4 is that the set of tropically positive semidefinite kernels is stable for the sum, when adding a constant in $\mathbb{R}_\perp$, and for pointwise limits. Other examples of kernels include the opposite of powers of distances over any metric space, such as the Wasserstein distances in optimal transport, or the kernel $b(X,Y) = -\|\operatorname{Spec}\log(XY^{-1})\|_2^2$ over positive semidefinite matrices acting on $\mathbb{R}^N$, with Spec denoting the eigenspectrum.

We chose to consider $\mathbb{R}_\perp$-valued kernels even though most of our kernels of interest are $\mathbb{R}$-valued. Nevertheless, the Dirac mass $\delta_x^\perp$ is $\mathbb{R}_\perp$-valued and plays the role of a neutral element. Note that $\delta_x^\perp$ is the logarithm of the Hilbertian psd "0-1" kernel defined as $k(x,x) = 1$, $k(x,y) = 0$ if $x \neq y$. It is possible to extend our analysis to $\overline{\mathbb{R}}$-valued kernels, although it would be cumbersome to keep under check two infinite values in our computations, as stressed in Singer (1997, Remark 8.25b). We first show that all tpsd kernels are translations of symmetric diagonal-vanishing and nonpositive-valued kernels (which can be interpreted as costs in a game context). This property may remind of Berg et al. (1984, Lemma 2.1, Proposition 3.2) stated for negative definite kernels.

Lemma 1. *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 $b_0(x,y) \leqslant 0$ 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 $\operatorname{Rg}(B) = \phi + \operatorname{Rg}(B_0)$.

Proof. If b is tpsd, then we set $\phi(x) = b(x,x)/2$. If $(\phi(x), \phi(y)) \in \mathbb{R}^2$, we pose $b_0(x,y) := b(x,y) - \phi(x) - \phi(y)$, and set it to 0 otherwise. If $\phi(x) = -\infty$ for some $x \in \mathcal{X}$, then, by the tropical positivity of b, $b(x,y) = -\infty$ for all $x \in \mathcal{X}$, whence b_0 satisfies (6). Again the positivity of b yields that b_0 is symmetric, diagonal-vanishing and nonpositive-valued. Conversely, if b is defined by (6), it is trivially tropically positive since b_0 vanishes on the diagonal and is nonpositive-valued. Let $f \in \operatorname{Rg}(B)$, as per (4), we can find $(a_x)_{x\in\mathcal{X}} \in \mathbb{R}_\perp^{\mathcal{X}}$ such that, for any $y \in \mathcal{X}$,

$$f(y) = \sup_{x\in\mathcal{X}} a_x + b(y,x) = \phi(y) + \sup_{x\in\mathcal{X}} a_x + \phi(x) + b_0(y,x)$$

so $f \in \phi + \operatorname{Rg}(B_0)$, the converse is shown similarly. □

We now draw the connection with the characterization of Hilbertian positive semidefinite kernels k through feature maps, as in Theorem 1-ii). The result below shows the analogy with such a characterization for the tropical inner product over $\mathbb{R}_\perp^{\mathcal{Z}}$ defined by $(f,g)_{\sup} = \sup_{z\in\mathcal{Z}} f(z) + g(z)$.

Burkard et al. (1996) for a review of their properties and to Weiß et al. (2016) for their application to the assignment problem.

[4]This result is related to the relation between Hilbertian cpsd and indefinitely divisible psd kernels (see e.g. Berg et al., 1984). For positive semidefinite kernels, the Gram matrices $\begin{pmatrix} k(x,x) & k(x,y) \\ k(x,y) & k(y,y) \end{pmatrix}$ are positive semidefinite. Thus, when considering their determinant, we obtain that $k(x,x)k(y,y) \geqslant k(x,y)k(y,x)$. Taking the logarithm (since we allow tpsd kernels to have the $-\infty$ value), we obtain the tropical positivity of $\log(k)$.

Proposition 2. *Let $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_\perp$ be a kernel. The following properties are equivalent*

i) b is tpsd;

ii) there exists a set $\mathcal{Z}$ and a function $\psi : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}_\perp$ such that

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

Proof. i)⇒ii). Take $\mathcal{Z} = \mathcal{X} \times \mathcal{X}$, and consider the function ψ such that, for all $x, y \in \mathcal{X} \times \mathcal{X}$, $\psi(x,(x, y)) = b(x, x)/2$ and $\psi(x,(y, x)) = b(x, y) \underset{\cdot}{-} b(y, y)/2$, with $-\infty$ absorbing. We set to $-\infty$ all the other values of $\psi(x,(u, v))$, for which $x \notin \{u, v\}$, whence ψ takes its values in $\mathbb{R}_\perp$. Then, by definition of ψ, since $-\infty$ is absorbing, the only z for which the values $\psi(x, z)$ and $\psi(y, z)$ can be finite are $z \in \{(x, y),(y, x)\}$. As b is symmetric, we obtain that

$$\begin{aligned}\sup_{z\in\mathcal{Z}} \psi(x, z) \underset{\cdot}{+} \psi(y, z) &= \max\left(\psi(x,(x, y)) \underset{\cdot}{+} \psi(y,(x, y)), \psi(x,(y, x)) \underset{\cdot}{+} \psi(y,(y, x))\right) \\ &= \max\left(\frac{b(x, x)}{2} + b(y, x) \underset{\cdot}{-} \frac{b(x, x)}{2}, b(x, y) \underset{\cdot}{-} \frac{b(y, y)}{2} + \frac{b(y, y)}{2}\right) = b(y, x),\end{aligned}$$

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

ii)⇒i). The kernel b as defined by (7) is symmetric and we assumed it takes its values in $\mathbb{R}_\perp$. Then, for all $z \in \mathcal{Z}$,

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

taking the supremum over $z \in \mathcal{Z}$ allows us to conclude. □

Examples of factorizations $\psi(x, z)$ with $\mathcal{Z} = \mathcal{X}$: Note that if (7) holds for $\mathcal{Z} = \mathcal{X}$ and $\psi = b$, then $-b$ satisfies a triangular inequality. For instance, for the kernels defined in (1), it is true for b_L, but it does not hold for b_c or b_{sc}, for which the feature map cannot be as simple as $\psi = b$.

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. Indeed, the issue of the factorization of b has been studied in Cartwright and Chan (2012) when $\mathcal{X}$ is a finite set, of cardinality $|\mathcal{X}| = n$. Then, the minimal cardinality of Z such that there exists a factorization (7) (summarized by the notation $b = \psi\psi'$) is called the tropical symmetric Barvinok rank. It is shown in Cartwright and Chan (2012, Theorem 4), that, for every tpsd b, there exists a factorization (7) with $|\mathcal{Z}| \leqslant \max(n, \lfloor n^2/4\rfloor)$, and that this bound is tight. Hence there exists kernels b over finite sets $\mathcal{X}$ for which all $\mathcal{Z}$ verify $|\mathcal{Z}| \geqslant \lfloor n^2/4\rfloor) > n$. For instance for $n = 5$, defining the kernel through its Gram matrix, we set

$$B = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & -1 & -1 & 0 \end{pmatrix}.$$

Such B corresponds to the complete bipartite graph $K_{\lfloor \frac{n}{2} \rfloor, \lceil \frac{n}{2} \rceil}$, for which the smallest cardinality of $\mathcal{Z}$ as in (7) is the clique cover number, equal to $6 > n = 5$ (Cartwright and Chan, 2012, Remark 3.3).

Nevertheless the feature map $\Phi_b(x) := b(\cdot, x)$ remains of interest even if it does not correspond to a factorization (7). As a matter of fact, the notion of convexity w.r.t. the family of functions $\Phi_b(\mathcal{X}) = \{b(\cdot, x)\}_{x \in \mathcal{X}}$ has been extensively studied in Pallaschke and Rolewicz (1997) and Singer (1997, Chapter 9). In particular the tropical SDP inequality corresponds to the monotonicity of Φ_b in the sense that, for any $x, y \in \mathcal{X}$ (Pallaschke and Rolewicz, 1997, p.5, (1.1.8))

$$\Phi_b(x)(x) + \Phi_b(y)(y) - \Phi_b(x)(y) - \Phi_b(y)(x) = b(x,x) + b(y,y) - b(x,y) - b(y,x) \geqslant 0.$$

Monotonicity of Φ_b makes it a candidate to be the subgradient of a Φ_b-convex function. In Banach spaces, maximal monotone operators are obtained as subgradients of convex functions. In our tropical setting, while we actually do not have the desired maximality, we show in Section 4 that tropical positivity entails monotonicity and that we have an analogue of the Cauchy-Schwarz inequality. This will require to switch our view to the operator $\bar{B}$ corresponding to b.

4. Monotonicity of operators induced by tpsd kernels

The following result explains why, unlike for the Hilbertian kernels of Theorem 1-i), considering only pairs, rather than sequences of points, is sufficient in the tropical setting. Proposition 3 furthermore relates the tropical positive semidefiniteness of the kernel over elements of $\mathcal{X}$ to the (cyclic) monotonicity of the operator $\bar{B}$ over functions of $\overline{\mathbb{R}}^{\mathcal{X}}$.

Proposition 3. *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;

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

$$\sum_{m=1}^{M} b(x_m, x_m) \geqslant \sum_{m=1}^{M} b(x_m, x_{\sigma(m)}); \tag{8}$$

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

$$\forall\, \hat{f}, \hat{g} \in \overline{\mathbb{R}}^{\mathcal{X}},\ \langle \hat{f}, \bar{B}\hat{f} \rangle \dot{+} \langle \hat{g}, \bar{B}\hat{g} \rangle \geqslant \langle \hat{f}, \bar{B}\hat{g} \rangle \dot{+} \langle \hat{g}, \bar{B}\hat{f} \rangle; \tag{9}$$

iv) the operator $\bar{B}$ is $\overline{\mathbb{R}}_{\max}$-hermitian and

$$\forall\, \hat{f}, \hat{g} \in \overline{\mathbb{R}}^{\mathcal{X}},\ \max\left(\langle \hat{f}, \bar{B}\hat{f} \rangle, \langle \hat{g}, \bar{B}\hat{g} \rangle \right) \geqslant \langle \hat{f}, \bar{B}\hat{g} \rangle; \tag{10}$$

with, for any x, $b(x,x) = -\infty$ implying that for all $y \in \mathcal{X}$, $b(x,y) = -\infty$;

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

$$\forall\, M \in \mathbb{N}^*,\ \forall\, (\hat{f}_m)_{m\in[M]} \in \overline{\mathbb{R}}^{\mathcal{X},M} \text{ with the convention } f_{M+1} = f_1,$$

$$\sum_{m=1}^{M} \langle \hat{f}_m, \bar{B}\hat{f}_m\rangle \geqslant \sum_{m=1}^{M} \langle \hat{f}_m, \bar{B}\hat{f}_{m+1}\rangle; \tag{11}$$

vi) the operator $\bar{B}$ is $\overline{\mathbb{R}}_{\max}$-hermitian and

$$\forall\, M \in \mathbb{N}^*,\ \forall\, (\hat{f}_m)_{m\in[M]} \in \overline{\mathbb{R}}^{\mathcal{X},M} \text{ with the convention } f_{M+1} = f_1,$$

$$\max_{m=1,\dots M} \langle \hat{f}_m, \bar{B}\hat{f}_m\rangle \geqslant \max_{m=1,\dots M} \langle \hat{f}_m, \bar{B}\hat{f}_{m+1}\rangle, \tag{12}$$

with, for any x, $b(x,x) = -\infty$ implying that for all $y \in \mathcal{X}$, $b(x,y) = -\infty$.

Proof. ii)⇒i). This corresponds to the case $M = 2$ and the cycle $(\sigma(1), \sigma(2)) = (2,1)$.

i)⇒ii). Since every permutation can be decomposed as a product of cycles, we just have to prove the property for the latter. For $M = 1$, we have by definition that $b(x,x) \geqslant -\infty$ for all $x \in \mathcal{X}$. Fix $M \in \mathbb{N}^*$ with $M \geqslant 2$. Reindexing $(x_m)_{m\in[M]}$, we can assume w.l.o.g. that $\sigma(m) = m+1$ for $m \in [M-1]$ and $\sigma(M) = 1$. Hence, since the kernel is tpsd, with the convention that $x_{M+1} = x_1$

$$\begin{aligned}\sum_{m=1}^{M} b(x_m, x_m) &= \frac{1}{2}\sum_{m=1}^{M} b(x_m, x_m) + b(x_{m+1}, x_{m+1})\\ &\geqslant \sum_{m=1}^{M} b(x_m, x_{m+1}) = \sum_{m=1}^{M} b(x_m, x_{\sigma(m)}).\end{aligned}$$

i)⇔iii). Notice that for all $x, y \in \mathcal{X}$, $\langle \delta_x^\top, \bar{B}\delta_x^\top\rangle = b(x,y)$. Hence i) is just a specialization of iii). To prove i)⇒iii), note for every $\hat{f}, \hat{g} \in \overline{\mathbb{R}}_{\min}^{\mathcal{X}}$, according to (5) and since b is tpsd, we have that

$$\begin{aligned}\langle \hat{f}, \bar{B}\hat{f}\rangle \dotplus \langle \hat{g}, \bar{B}\hat{g}\rangle &= \sup_{x,u\in\mathcal{X}} b(x,u) \dot{-} \hat{f}(x) \dot{-} \hat{f}(u) \dotplus \sup_{y,v\in\mathcal{X}} b(y,v) \dot{-} \hat{g}(y) \dot{-} \hat{g}(v)\\ &\geqslant b(x,x) + b(y,y) \dot{-} 2\hat{f}(x) \dot{-} 2\hat{g}(y) \text{ setting } u = x, v = y\\ &\geqslant 2b(x,y) \dot{-} 2\hat{f}(x) \dot{-} 2\hat{g}(y).\end{aligned}$$

Taking the supremum over $x, y \in \mathcal{X}$ yields the result.

iii)⇔v). This equivalence follows from the same computations as for the proof of i)⇔ii) Cyclic monotonicity is precisely the σ-cycle considered in the proof above.

iii)⇒iv). This stems from the fact that, applying (9) and the hermitianity of $\bar{B}$,

$$2\max\left(\langle \hat{f}, \bar{B}\hat{f}\rangle, \langle \hat{g}, \bar{B}\hat{g}\rangle\right) \geqslant \langle \hat{f}, \bar{B}\hat{f}\rangle \dotplus \langle \hat{g}, \bar{B}\hat{g}\rangle \geqslant \langle \hat{f}, \bar{B}\hat{g}\rangle \dotplus \langle \hat{g}, \bar{B}\hat{f}\rangle = 2\langle \hat{f}, \bar{B}\hat{g}\rangle.$$

iv)⇒i). Since $\bar{B}$ is $\overline{\mathbb{R}}_{\max}$-hermitian, b is symmetric. Fix $x, y \in \mathcal{X}$. If $b(x,x) + b(y,y) \neq -\infty$, then we can define the functions $\hat{f}(\cdot) = \delta_x^\top(\cdot) - \frac{b(x,x)}{2}$ and $\hat{g}(\cdot) = \delta_y^\top(\cdot) - \frac{b(y,y)}{2}$ and apply (10), whence

$$0 = \max\left(\langle \hat{f}, \bar{B}\hat{f}\rangle, \langle \hat{g}, \bar{B}\hat{g}\rangle\right) \geqslant \langle \hat{f}, \bar{B}\hat{g}\rangle = b(x,y) - \frac{b(x,x)}{2} - \frac{b(y,y)}{2}.$$

If $b(x,x)+b(y,y)=-\infty$, then, by iv), $b(x,y)=-\infty$. In both cases, we have shown that b is tropically positive.

iv)⇔vi) We just have to consider $M=2$ for vi)⇒iv), while iv)⇒vi) is a consequence of the transitivity of the maximum. □

The two formulas (9)-(10) can be seen as two different ways of defining a tropical analogue of the monotonicity of a nonlinear operator over a Hilbert space, informally written as "$\langle \hat{f}-\hat{g}, K\hat{f}-K\hat{g}\rangle \geqslant 0$". We can moreover interpret (10) as a tropical Cauchy-Schwarz inequality, informally written as "$\langle \hat{f}, K\hat{f}\rangle^{\frac{1}{2}}\langle \hat{g}, K\hat{g}\rangle^{\frac{1}{2}} \geqslant \langle \hat{f}, K\hat{g}\rangle$" From the notion of monotonicity (9), we can define a form of tropical "quadratic" discrepancy $d_B : \overline{\mathbb{R}}_{\min}^{\mathcal{X}} \times \overline{\mathbb{R}}_{\min}^{\mathcal{X}} \to \overline{\mathbb{R}}$:

$$d_B(\hat{f},\hat{g}) := \frac{1}{2}\left[\langle \hat{f}, \bar{B}\hat{f}\rangle \dot{+} \langle \hat{g}, \bar{B}\hat{g}\rangle \dot{-} \langle \hat{f}, \bar{B}\hat{g}\rangle \dot{-} \langle \hat{g}, \bar{B}\hat{f}\rangle\right]. \tag{13}$$

Examples of d_B: For a metric space $(\mathcal{X},d)$ and $b(x,y)=-d(x,y)$, we shall see below (p.13) that $\mathrm{Rg}(B)$ is the set of 1-Lipschitz functions (including the constant functions in $\overline{\mathbb{R}}$) and that $\bar{B}\hat{f}=-\hat{f}$ for $\hat{f}\in \mathrm{Rg}(B)$. Hence, for any $\hat{f},\hat{g}\in\mathrm{Rg}(B)$, we have that $\langle \hat{g},\bar{B}\hat{f}\rangle = -\inf_{\mathcal{X}}(\hat{f}+\hat{g})$ and $2d_B(\hat{f},\hat{g}) = \inf_{\mathcal{X}}(\hat{f}+\hat{g}) - \inf_{\mathcal{X}}\hat{f} - \inf_{\mathcal{X}}\hat{g}$. So $d_B(\hat{f},\hat{g})=0$ indicates that $\hat{f}$ and $\hat{g}$ have a common infimum. A similar computation for $b(x,y)=\delta_x^{\perp}(y)$ gives the same d_B, but defined over $\overline{\mathbb{R}}_{\min}^{\mathcal{X}}$. Nevertheless, for non-compact $\mathcal{X}$, even with smooth kernels, d_B can bee infinite. Indeed for $b(x,y)=(x,y)_2$, $\langle b(\cdot,w),\bar{B}b(\cdot,w)\rangle = \sup_{x,y\in\mathbb{R}^d}(x,y)_2 \dot{-} (w,x+y)_2 = \infty$. Note that Proposition 3 could be rewritten in a $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-linear rather than $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-sesquilinear setting, replacing the duality product by a sup-inner product and changing signs, $\langle \hat{f},\bar{B}\hat{g}\rangle = \sup_{x,y\in\mathcal{X}} b(x,u) \dot{-} \hat{f}(x) \dot{-} \hat{g}(y) =: (-\hat{f}, B(-\hat{g}))_{\sup}$.

5. A sesquilinear theory: defining Reproducing Kernel Moreau Spaces based on symmetry

We shall be interested in spaces that are ranges of tropical operators. We start with a characterization of such ranges when the kernels are symmetric.

Theorem 3. *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 $\mathcal{G}=\mathrm{Rg}(B)$;

ii) there exists a $\overline{\mathbb{R}}_{\max}$-sesquilinear map $\bar{F}:\mathcal{G}\to\mathcal{G}$ such that $\bar{F}\bar{F}=\mathrm{Id}_{\mathcal{G}}$, 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 $\mathrm{Rg}(B)$.

Proof. i)⇒ii). We use the classical property that $\bar{B}\bar{B}\bar{B}=\bar{B}$, see e.g. (Akian et al., 2005, p.3), so $\bar{B}$ is an anti-involution over $\mathrm{Rg}(B)$.

ii)⇒i). We first show that $\bar{F}$ is necessarily continuous in the sense of Definition 2 and that it can be extended to a $\overline{\mathbb{R}}_{\max}$-hermitian and continuous map $\tilde{F}:\overline{\mathbb{R}}^{\mathcal{X}}\to\overline{\mathbb{R}}^{\mathcal{X}}$.

Given an arbitrary index set $\mathcal{A}$, fix a family $(g_\alpha)_{\alpha\in\mathcal{A}}\in\mathcal{G}^{\mathcal{A}}$. We define its infimum relatively to $\mathcal{G}$ as

$$\inf\nolimits_\alpha^{\mathcal{G}} g_\alpha := \max\{g\in\mathcal{G}\,|\,\forall\alpha\in\mathcal{A},\, g\leqslant g_\alpha\},$$

noting that the latter set does admit a greatest element, as $\mathcal{G}$ is a complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$. In particular, if the family $(g_\alpha)_{\alpha\in\mathcal{A}}$ consists of a single element g, then,

$$\inf^{\mathcal{G}} g = \max\{h \in \mathcal{G} \,|\, h \leqslant g\}.$$

The operation of relative supremum with respect to $\mathcal{G}$ is defined in a dual manner, however, since $\mathcal{G}$ is stable by arbitrary suprema, it merely coincides with the ordinary supremum of functions. As $\bar{F}$ is $\overline{\mathbb{R}}_{\max}$-sesquilinear, it satisfies that

$$\forall f, g \in \mathcal{G} \text{ s.t. } f \leqslant g, \bar{F}(f) = \bar{F}(\min(f,g)) = \max(\bar{F}(f), \bar{F}(g)) \geqslant \bar{F}(g),$$

in other words, $\bar{F}$ is antitone. Hence, as $\inf^{\mathcal{G}}_{\alpha\in\mathcal{A}} g_\alpha \leqslant g_\beta$ for any $\beta \in \mathcal{A}$,

$$\bar{F}(\inf^{\mathcal{G}}_{\alpha\in\mathcal{A}} g_\alpha) \geqslant \sup_{\alpha\in\mathcal{A}} \bar{F}(g_\alpha) \geqslant \bar{F}(g_\beta).$$

Since $\mathcal{G}$ is a complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$, $\sup_{\alpha\in\mathcal{A}} \bar{F}(g_\alpha) \in \mathcal{G}$, so composing by $\bar{F}$ and using that $\bar{F}$ is antitone, we derive that

$$\bar{F}\bar{F}(\inf^{\mathcal{G}}_{\alpha\in\mathcal{A}} g_\alpha) \leqslant \bar{F}(\sup_{\alpha\in\mathcal{A}} \bar{F}(g_\alpha)) \leqslant \bar{F}\bar{F}(g_\beta).$$

As $\bar{F}$ is an anti-involution, taking the infimum in $\mathcal{G}$ over β on the r.h.s., yields

$$\inf^{\mathcal{G}}_{\alpha\in\mathcal{A}} g_\alpha = \bar{F}\bar{F}(\inf_{\alpha\in\mathcal{A}} g_\alpha) \leqslant \bar{F}(\sup_{\alpha\in\mathcal{A}} \bar{F}(g_\alpha)) \leqslant \inf^{\mathcal{G}}_{\beta\in\mathcal{A}} g_\beta \ .$$

So $\inf^{\mathcal{G}}_{\alpha\in\mathcal{A}} g_\alpha = \bar{F}\left(\sup_{\alpha\in\mathcal{A}} \bar{F}(g_\alpha)\right)$, and, composing again by $\bar{F}$, we deduce that $\bar{F}$ is continuous on its domain, meaning that $\bar{F}$ sends relative infima with respect to $\mathcal{G}$ to relative suprema (which coincide with ordinary suprema).

Define now $\tilde{F} : \overline{\mathbb{R}}^{\mathcal{X}} \to \overline{\mathbb{R}}^{\mathcal{X}}$ by, for any $f \in \overline{\mathbb{R}}^{\mathcal{X}}$, $\tilde{F}(f) := \bar{F}(\inf^{\mathcal{G}} f) \in \mathcal{G}$. The function $\tilde{F}$ is automatically an extension of $\bar{F}$. We now show that $\tilde{F}$ is $\overline{\mathbb{R}}_{\max}$-hermitian. Fix $f \in \overline{\mathbb{R}}^{\mathcal{X}}$ and $\lambda \in \mathbb{R}$, then, as $\bar{F}$ is $\overline{\mathbb{R}}_{\max}$-sesquilinear, $\tilde{F}(f \dot{+} \lambda) = \bar{F}(\lambda \dot{+} \inf^{\mathcal{G}} f) = \tilde{F}(f) \dot{-} \lambda$. Let $(f_\alpha)_{\alpha\in\mathcal{A}} \in (\overline{\mathbb{R}}^{\mathcal{X}})^{\mathcal{A}}$. By definition, $\inf^{\mathcal{G}}(\inf_{\alpha\in\mathcal{A}} f_\alpha) = \inf^{\mathcal{G}}_\alpha(\inf^{\mathcal{G}} f_\alpha)$. As $\bar{F}$ is continuous,

$$\tilde{F}(\inf_{\alpha\in\mathcal{A}} f_\alpha) = \bar{F}(\inf^{\mathcal{G}}_\alpha(\inf^{\mathcal{G}} f_\alpha)) = \sup_\alpha \bar{F}(\inf^{\mathcal{G}} f_\alpha) = \sup_\alpha \tilde{F}(f_\alpha).$$

Hence $\tilde{F}$ is $\overline{\mathbb{R}}_{\max}$-hermitian and continuous. We can now apply Proposition 1 to derive a kernel b associated with $\tilde{F}$. By construction, $\mathrm{Rg}(\tilde{F}) \subset \mathcal{G}$ and since $\bar{F}$ is an anti-involution, $\mathcal{G} = \mathrm{Rg}(\bar{F}) \subset \mathrm{Rg}(\tilde{F})$, so $\mathcal{G} = \mathrm{Rg}(\tilde{F})$. As $\bar{F}$ is an anti-involution,

$$\tilde{F}\tilde{F}(f) = \bar{F}\bar{F}(\inf^{\mathcal{G}} f) = \inf^{\mathcal{G}} f \leqslant f.$$

We have thus recovered one of the characterizations of dual Galois connections over $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$, as given in (Akian et al., 2005, p.3, Eq.(2a)), so $\tilde{F}$ is equal to its dual connection. Akian et al. (2005, Theorem 2.1, Proposition 2.3) then allows to conclude that, since the dual connection is simply the transpose $\tilde{F}^\top$ (see also Singer, 1997, Theorem 8.4), the kernel b is symmetric. □

Remark 2. Theorem 3 should be compared with (Develin and Sturmfels, 2004, Th. 23) and (Cohen et al., 2004, Th. 42), which entail that the row and column spaces of a tropical matrix are anti-isomorphic lattices. By specializing this result, we deduce that the row space of a symmetric matrix is anti-isomorphic to itself. Theorem 3 refines this result (and also extends it to the infinite dimensional setting), showing that a version of the latter anti-isomorphism property characterizes the ranges of symmetric operators.

We now leverage the characterization, $\bar{B}\bar{B}\bar{B}\hat{f} = \bar{B}\hat{f}$, expressed for symmetric kernels in Theorem 3, to define tropical sesquilinear reproducing kernel spaces. As shown in (4), using spaces $\mathrm{Rg}(B)$ 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 $\mathrm{span}(\{k(\cdot,x)\}_{x\in\mathcal{X}})$.

Definition 5. We call *reproducing kernel Moreau spaces* (RKMS) the complete submodules $\mathrm{Rg}(B)$ 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 $\hat{g} \in \mathrm{Rg}(B)$ and $x \in \mathcal{X}$, we say that they satisfy a *sesquilinear reproducing property*

$$\hat{g}(x) = (\bar{B}\bar{B}\hat{g})(x) = \langle \bar{B}\hat{g}, \bar{B}\delta_x^\top \rangle = \sup_{z\in\mathcal{X}} b(z,x) \dot{-} [\sup_{y\in\mathcal{X}} b(z,y) \dot{-} \hat{g}(y)]. \tag{14}$$

The sesquilinear reproducing property as defined in (14) 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 2. *(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 $b = b_c$, $\bar{B}$ is the Fenchel conjugate operator, whence (14) 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 identify $\mathrm{Rg}(B)$, i.e. proving that all convex l.s.c. functions are outputs of the Fenchel transform.

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. One may thus wonder whether there is a more direct interpretation. Indeed we can relate (14) to a $\overline{\mathbb{R}}_{\min}^{\mathcal{X}}$-linear operator, but it unfortunately cannot characterize $\mathrm{Rg}(B)$. This will be further emphasized in Section 6.

Lemma 3. *For all $\hat{g} \in \mathrm{Rg}(B)$ and $x \in \mathcal{X}$, (14) is equivalent to*

$$\hat{g}(x) = \inf_{y\in\mathcal{X}} \hat{g}(y) \dot{+} \sup_{z\in\mathcal{X}} [b(z,x) - b(z,y)] =: \inf_{y\in\mathcal{X}} \hat{g}(y) \dot{+} c(x,y) =: (C^{op}\hat{g})(x) \tag{15}$$

where $c(x,y) := \sup_{z\in\mathcal{X}}[b(z,x) \dot{-} b(z,y)]$ 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.

Proof. The term $c(x,y)$ appears by permuting the 'sup inf' into 'inf sup' and using (14) since

$$\begin{aligned}\hat{g}(x) &= \sup_{z\in\mathcal{X}} b(z,x) \dot{-} [\sup_{y\in\mathcal{X}} b(z,y) \dot{-} \hat{g}(y)] = \sup_{z\in\mathcal{X}} \inf_{y\in\mathcal{X}} b(z,x) \dot{-} b(z,y) \dot{+} \hat{g}(y) \\ &\leqslant \inf_{y\in\mathcal{X}} \hat{g}(y) \dot{+} \sup_{z\in\mathcal{X}} [b(z,x) \dot{-} b(z,y)] = \inf_{y\in\mathcal{X}} \hat{g}(y) \dot{+} c(x,y) \overset{y=x}{\leqslant} \hat{g}(x).\end{aligned}$$

□

Unfortunately c does not correspond to a unique b, nor to a unique $\mathrm{Rg}(B)$, as shown in the examples below:

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

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 $\pm\infty$ (Singer, 1997, Theorem 3.7).

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 $\pm\infty$ (Singer, 1997, Theorem 3.16).

iii) For any $\mathcal{X}$ and $\alpha \geqslant 0$, $b(x, y) = \begin{cases} 0 & \text{if } y = x, \\ -\alpha & \text{otherwise,} \end{cases}$ 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 $\overline{\mathbb{R}}^{\mathcal{X}}$ (Singer, 1997, Remark 3.2).

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. $|f(x) - f(y)| \leqslant 1 \cdot d(x, y)^p$), when adding the constant functions $\pm\infty$ (Singer, 1997, Theorem 3.14).

Informal analogy between RKMS and RKHS: We can draw an informal analogy between (15) and the reproducing property of functions f of a RKHS $\mathcal{H}_k$ defined through a differential operator D (and its adjoint D^*) over a bounded open set $\mathcal{X}$, i.e.

$$f(x) = (f(\cdot), k(\cdot, x))_{\mathcal{H}_k} \overset{(*)}{=} \int_{y \in \mathcal{X}} f(y) D_y^* D_y k(x, y) \mathrm{d}y \tag{16}$$

where $(*)$ holds for Green kernels with null integral on the boundaries, as per Saitoh and Sawano (2016, Section 1.7). Equation (16) underlines the fact that the integral was formally replaced with an inf, the product with a sum and the differential operator D^*D by the sup of a difference, all this in a classical tropical fashion. Moreover for the RKHSs defined through differential operators, (16) expresses the fact the functions $f \in \mathcal{H}_k$ are fixed points for a kernel integral operator, but whose kernel is not simply $k(x, y)$.

Reproducing but not positive: By Theorem 3, any kernel $b(x, y)$ that is symmetric, even non-tpsd, gives $\bar{B}\bar{B}\bar{B} = \bar{B}$, so the reproducing property (14) can be defined for spaces and kernels that do have not have a notion of positivity. This is true also for Hilbertian kernels, since reproducing properties can be defined for indefinite kernels and their related Krein spaces (see e.g. Ong et al., 2004, and references within). Even the symmetry requirement could be relaxed as was done for tropical operators by Akian et al. (2005); Singer (1997) and for RKHSs by Mary (2005).

Relation between $\bar{B}$ and C^{op}: The tropical positive operators $\bar{B}$ are $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-sesquilinear over $\overline{\mathbb{R}}^{\mathcal{X}}$ whereas C^{op} is $\overline{\mathbb{R}}_{\min}^{\mathcal{X}}$-linear. Based on the non-symmetric $c(x, y)$ one can define the set of c-Lipschitz functions.

$$\mathrm{Lip}_{\mathrm{c}}(\mathcal{X}, \overline{\mathbb{R}}) := \{g \in \mathbb{R}^{\pm, \mathcal{X}} \mid g(x) \leqslant g(y) \dotplus c(x, y)\}.$$

For finite-valued kernels b, since $c(x, x) = 0$, $\mathrm{Lip}_{\mathrm{c}}(\mathcal{X}, \overline{\mathbb{R}})$ corresponds to the fixed points $(g = C^{op}g)$ of C^{op}. By (14),

$$\mathrm{Rg}^{\max}(B) \subseteq \mathrm{Lip}_{\mathrm{c}}(\mathcal{X}, \overline{\mathbb{R}}) \subseteq \mathrm{Rg}^{\min}(C^{op}) := \{\inf_{x \in \mathcal{X}} a_x + c(\cdot, x) \mid a_x \in \mathbb{R}_\top\}$$

with equality when $c(x,y) = -b(x,y)$ (i.e. $-b$ is idempotent). Notice that $\mathrm{Rg}^{\max}(B)$ is sup-stable, $\mathrm{Lip}_c(\mathcal{X}, \overline{\mathbb{R}})$ is sup-stable and inf-stable while $\mathrm{Rg}^{\min}(C^{op})$ is inf-stable.

6. A linear theory: only idempotent operators and spaces of Lipschitz functions

We have seen in Section 5 that one could introduce some tropically linear, rather than sesquilinear, operators. This choice can be related to the identification of kernel integral operators in the Hilbertian context with $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-linear operators in the tropical setting as was done by Litvinov (Section 7 2011). We prove below that discarding sesquilinearity imposes considerable restrictions on the types of spaces that one can consider. These are either ranges of idempotent kernels or lattices of Lipschitz functions. This result further emphasizes why we need to consider $\overline{\mathbb{R}}_{\max}$-sesquilinear operators to properly define the max-plus analogue of RKHSs. Recall for instance that convex l.s.c. functions are defined based on the $\overline{\mathbb{R}}_{\max}$-sesquilinear Fenchel transform, not a linear version of it.

For RKHSs $(\mathcal{H}_k, (\cdot,\cdot)_{\mathcal{H}_k})$, the reproducing property writes as follows

$$\forall\, f \in \mathcal{H}_k,\, \forall\, x \in \mathcal{X},\; f(x) = \langle \delta_x^{lin}, f(\cdot)\rangle_{\mathbb{R}^{\mathcal{X},*}\times\mathbb{R}^{\mathcal{X}}} = (k(\cdot,x), f(\cdot))_{\mathcal{H}_k}, \tag{17}$$

where $k(\cdot,x) \in \mathcal{H}_k$ and the space of finite measures $\mathbb{R}^{\mathcal{X},*}$ is the dual of $\mathbb{R}^{\mathcal{X}}$ for the pointwise convergence. The latter is generated by the linear Dirac masses $\delta_x^{lin} : f \in \mathbb{R}^{\mathcal{X}} \mapsto f(x)$.

Consider a set $\mathcal{G} \subset \mathbb{R}_{\top}^{\mathcal{X}}$. A $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-linear analogy to (17) would be the existence of a kernel $c : \mathcal{X}\times\mathcal{X} \to \overline{\mathbb{R}}$ satisfying

$$\forall\, g \in \mathcal{G},\, \forall\, x \in \mathcal{X},\; g(x) = \sup_{y\in\mathcal{X}} \delta_x^{\perp}(y) \dotplus g(y) = \sup_{y\in\mathcal{X}} c(x,y) \dotplus g(y). \tag{18}$$

Although it is always possible to consider $c(x,y) = \delta_x^{\perp}(y)$ in (18), since $\mathcal{G} \subset \mathbb{R}_{\top}^{\mathcal{X}}$, $\delta_x^{\perp}(\cdot) \notin \mathcal{G}$. Nonetheless among the possible c, there is one that is maximal:

Proposition 4. *For any set $\mathcal{G} \subset \mathbb{R}_{\top}^{\mathcal{X}}$, there is a kernel $c_{\mathcal{G}}$ that is maximal over the kernels c satisfying* (18)*, for the partial order inherited from $\overline{\mathbb{R}}$. Moreover $c_{\mathcal{G}} : \mathcal{X}\times\mathcal{X} \to \overline{\mathbb{R}}$ is given by*

$$c_{\mathcal{G}}(x,y) := \inf_{g\in\mathcal{G}} g(x) \dot{-} g(y). \tag{19}$$

If, for some $x \in \mathcal{X}$, there exists a function $g \in \mathcal{G}$ such that $g(x) < \infty$, i.e. $\mathcal{G}$ is proper *at x, then $c_{\mathcal{G}}(x,x) = 0$ (otherwise $c_{\mathcal{G}}(x,x) = +\infty$). Moreover $c_{\mathcal{G}}$ is idempotent in the* $(\max,+)$ *sense*

$$c_{\mathcal{G}}(x,y) = \sup_{z\in\mathcal{Z}} c_{\mathcal{G}}(x,z) \dotplus c_{\mathcal{G}}(z,y). \tag{20}$$

Proof. We first show that $c_{\mathcal{G}}$ satisfies (18). Fix any $x \in \mathcal{X}$ and $g \in \mathcal{G}$, and define

$$\tilde{g}(x) = \sup_{y\in\mathcal{X}} g(y) \dotplus c_{\mathcal{G}}(x,y) = \sup_{y\in\mathcal{X}} g(y) \dotplus \inf_{h\in\mathcal{G}} h(x) \dot{-} h(y) \overset{h=g}{\leqslant} g(x). \tag{21}$$

If $h(x) = +\infty$ for all $h \in \mathcal{G}$, since $+\infty$ is absorbing, $c_{\mathcal{G}}(x,y) = +\infty$ for all $y \in \mathcal{X}$. So $\tilde{g}(x) \geqslant g(x) + c_{\mathcal{G}}(x,x) = \infty + \infty = \infty = g(x)$. Assume now that $\{g \in \mathcal{G}, g(x) < \infty\} \neq \emptyset$. Since $+\infty$ is absorbing and $\mathcal{G} \subset \mathbb{R}_{\top}^{\mathcal{X}}$, $c_{\mathcal{G}}(x,x) = \inf_{g\in\mathcal{G}, g(x)<\infty} g(x) \dot{-} g(x) = 0$. Hence, by (21), $g(x) \geqslant \tilde{g}(x) \geqslant g(x) \dotplus c_{\mathcal{G}}(x,x) = g(x)$, whence $g = \tilde{g}$. By definition of $\tilde{g}$ in (21), we deduce that $c_{\mathcal{G}}$ satisfies (18). Consider now any kernel c satisfying (18), whence, for all $g \in \mathcal{G}$ and all $x, y \in \mathcal{X}$,

$$g(x) \geqslant g(y) \dotplus c(x, y).$$

Taking the infimum over $g \in \mathcal{G}$, by definition of $c_{\mathcal{G}}$ (19), $c_{\mathcal{G}}$ is indeed larger than any c.

Let $x \in \mathcal{X}$. If $h(x) = +\infty$ for all $h \in \mathcal{G}$, then $c_{\mathcal{G}}(x, y) = +\infty$ for all $y \in \mathcal{X}$, so (20) holds. Otherwise, we can fix $f \in \mathcal{G}$ such that $f(x) < \infty$. Let $z \in \mathcal{X}$. If $f(z) < \infty$, then

$$c_{\mathcal{G}}(x, z) \dotplus c_{\mathcal{G}}(z, y) = \inf_{g \in \mathcal{G}} g(x) \dot{-} g(z) \dotplus \inf_{h \in \mathcal{G}} h(z) \dot{-} h(y) \overset{g=h=f}{\leqslant} f(x) \dot{-} f(y). \tag{22}$$

Assume now that $f(z) = \infty$. As $f(x) < +\infty$, $c_{\mathcal{G}}(x, z) = -\infty$ which is absorbing for the l.h.s. of (22). Hence $c_{\mathcal{G}}(x, z) \dotplus c_{\mathcal{G}}(z, y) = -\infty \leqslant f(x) \dot{-} f(y)$. We are thus allowed to take the infimum of the r.h.s. of (22) over all $f \in \mathcal{G}$, which yields $c_{\mathcal{G}}(x, z) \dotplus c_{\mathcal{G}}(z, y) \leqslant c_{\mathcal{G}}(x, y)$. Since, for $\mathcal{G}$ proper at x, we also have

$$\sup_{z \in \mathcal{Z}} c_{\mathcal{G}}(x, z) \dotplus c_{\mathcal{G}}(z, y) \overset{z=x}{\geqslant} c_{\mathcal{G}}(x, y),$$

we deduce that (20) holds, which concludes the proof. □

Alike C^{op} in Section 5, the main limitation of $c_{\mathcal{G}}$ is that it cannot characterize uniquely each $\mathcal{G} \subset \mathbb{R}_{\top}^{\mathcal{X}}$. Indeed we show in Proposition 5 below that $\operatorname{Rg}(C_{\mathcal{G}}) := \{\sup_{x \in \mathbb{R}^N} [c_{\mathcal{G}}(\cdot, x) \dotplus a_x] \mid a_x \in \mathbb{R}_{\perp}, x \in \mathcal{X}\}$ is the "closure" in a max-min sense of $\mathcal{G}$ and is hence inf and sup-stable.

Proposition 5. *Let $\mathcal{G} \subset \mathbb{R}_{\top}^{\mathcal{X}}$ be a set of $\overline{\mathbb{R}}$-proper functions, and $c_{\mathcal{G}}$ as in* (19). *Consider $C_{\mathcal{G}} : \overline{\mathbb{R}}_{\max}^{\mathcal{X}} \to \overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ as in* (2). *Then $C_{\mathcal{G}} \geqslant \operatorname{Id}_{\overline{\mathbb{R}}^{\mathcal{X}}}$, in the sense that $C_{\mathcal{G}}(f)(x) \geqslant f(x)$ for all $f \in \overline{\mathbb{R}}^{\mathcal{X}}$, and* $\operatorname{Rg}(C_{\mathcal{G}})$ *is the smallest, in the inclusion sense, inf-stable complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ containing $\mathcal{G}$. It corresponds to the set of (-$c'_{\mathcal{G}}$)-Lipschitz functions,* $\operatorname{Rg}(C_{\mathcal{G}}) = \{f \in \overline{\mathbb{R}}^{\mathcal{X}} \mid \forall\, x, y \in \mathcal{X},\ f(x) \leqslant f(y) \dot{-} c_{\mathcal{G}}(y, x)\}$ *where $c'_{\mathcal{G}}$ is the transpose of $c_{\mathcal{G}}$.*

Proof. Let $\mathcal{F}$ be an inf-stable complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ containing $\mathcal{G}$. Hence $c_{\mathcal{G}}(\cdot, y) = \inf_{g \in \mathcal{G}} g(\cdot) \dot{-} g(y) \in \mathcal{F}$ since $g(\cdot) \dot{-} g(y) \in \mathcal{F}$. The set $\mathcal{F}$ is also sup-stable, so $\operatorname{Rg}(C_{\mathcal{G}}) \subset \mathcal{F}$. Since $c_{\mathcal{G}}(x, x) \in \{0, \infty\}$ by Proposition 4, $C_{\mathcal{G}}(f)(x) = \sup_{y \in \mathcal{X}} c_{\mathcal{G}}(x, y) \dotplus f(y) \overset{y=x}{\geqslant} f(x)$ for all $f \in \overline{\mathbb{R}}^{\mathcal{X}}$.

We now prove that $\operatorname{Rg}(C_{\mathcal{G}})$ is an inf-stable complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$. The set $\operatorname{Rg}(C_{\mathcal{G}})$ is by definition sup-stable. Let us show that it is also inf-stable. Take any family $(f_\alpha)_{\alpha \in \mathcal{A}}$ with $f_\alpha \in \overline{\mathbb{R}}^{\mathcal{X}}$ and $\mathcal{A}$ an arbitrary index set. Using successively that $C_{\mathcal{G}} = C_{\mathcal{G}} C_{\mathcal{G}}$ as shown in Proposition 4 (as $\mathcal{G}$ is a set of proper functions), then a permutation of inf and sup, and that $C_{\mathcal{G}} \geqslant \operatorname{Id}_{\overline{\mathbb{R}}_{\max}^{\mathcal{X}}}$, we derive that

$$\begin{aligned}\inf_{\alpha \in \mathcal{A}} C_{\mathcal{G}} f_\alpha = \inf_{\alpha \in \mathcal{A}} C_{\mathcal{G}} C_{\mathcal{G}} f_\alpha &= \inf_{\alpha \in \mathcal{A}} \sup_{y \in \mathcal{X}} C_{\mathcal{G}} f_\alpha(y) \dotplus c_{\mathcal{G}}(\cdot, y) \\ &\geqslant \sup_{y \in \mathcal{X}} \inf_{\alpha \in \mathcal{A}} C_{\mathcal{G}} f_\alpha(y) \dotplus c_{\mathcal{G}}(\cdot, y) = C_{\mathcal{G}}(\inf_{\alpha \in \mathcal{A}} C_{\mathcal{G}} f_\alpha) \geqslant \inf_{\alpha \in \mathcal{A}} C_{\mathcal{G}} f_\alpha,\end{aligned}$$

so $\inf_{\alpha \in \mathcal{A}} C_{\mathcal{G}} f_\alpha = C_{\mathcal{G}}(\inf_{\alpha \in \mathcal{A}} C_{\mathcal{G}} f_\alpha) \in \operatorname{Rg}(C_{\mathcal{G}})$ which is thus inf-stable. Finally, as $c_{\mathcal{G}}$ satisfies (18) by Proposition 4, for any $g \in \mathcal{G}$, $g = C_{\mathcal{G}} g \in \operatorname{Rg}(C_{\mathcal{G}})$, so $\mathcal{G} \subset \operatorname{Rg}(C_{\mathcal{G}})$.

Finally we show that $\mathrm{Rg}(C_{\mathcal{G}})$ is the set of (-$c'_{\mathcal{G}}$)-Lipschitz functions. Let f be a (-$c'_{\mathcal{G}}$)-Lipschitz function. By definition, for any $x, y \in \mathcal{X}$, $f(x) \dot{+} c_{\mathcal{G}}(y,x) \leqslant f(y)$, with equality when $x = y$, so $f(y) = \sup_x c_{\mathcal{G}}(y,x) \dot{+} f(x)$, and $f \in \mathrm{Rg}(C_{\mathcal{G}})$. Take now $f = C_{\mathcal{G}} g \in \mathrm{Rg}(C_{\mathcal{G}})$, thus we obtain that

$$\begin{aligned} C_{\mathcal{G}} g(x) - C_{\mathcal{G}} g(y) &= \sup_z c_{\mathcal{G}}(x,z) \dot{+} g(z) - \sup_w c_{\mathcal{G}}(y,w) \dot{+} g(w) \\ &\overset{w=z}{\leqslant} \sup_z c_{\mathcal{G}}(x,z) - c_{\mathcal{G}}(y,z) \overset{(20)}{\leqslant} \sup_z c_{\mathcal{G}}(x,z) - c_{\mathcal{G}}(y,x) - c_{\mathcal{G}}(x,z) = -c_{\mathcal{G}}(y,x), \end{aligned}$$

which concludes the proof.

□

What if we had considered a weighted inner product in (18)? Then we could indeed obtain a formula akin to a reproducing property but only in a very limited context of ranges of idempotent operators. Notice also that this approach does not require the symmetry of the kernels involved.

Theorem 4. *Let $\mathcal{G}$ be a complete submodule of $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$. The following properties are equivalent:*

i) there exists two kernels $a, b : \mathcal{X} \times \mathcal{X} \to \overline{\mathbb{R}}$ such that, when considering the weighted sup-inner product $(f,g)_{\sup,a} = \sup_{y,z\in\mathcal{X}} f(y) + g(z) + a(y,z)$, then, for all $x \in \mathcal{X}$ and $f \in \mathcal{G}$,

$$f(x) = (f, b(x,\cdot))_{\sup,a} \text{ and } b(\cdot,x) \in \mathcal{G}; \tag{23}$$

ii) $\mathcal{G} = \mathrm{Rg}(P)$ for some linear continuous idempotent operator $P : \overline{\mathbb{R}}_{\max}^{\mathcal{X}} \to \overline{\mathbb{R}}_{\max}^{\mathcal{X}}$, i.e. $P = P^2$;

iii) $\mathcal{G} = \mathrm{Rg}(B)$ for some Von Neumann regular operator $B : \overline{\mathbb{R}}_{\max}^{\mathcal{X}} \to \overline{\mathbb{R}}_{\max}^{\mathcal{X}}$, i.e. a linear and continuous operator such that there exists a linear continuous $A : \overline{\mathbb{R}}_{\max}^{\mathcal{X}} \to \overline{\mathbb{R}}_{\max}^{\mathcal{X}}$ satisfying $B = BAB$.

Proof. i)⇒iii). Fix a, b as in i), then, for all $f \in \mathcal{G}$,

$$f(\cdot) = \sup_{y,z\in\mathcal{X}} b(\cdot,y) + f(z) + a(y,z),$$

so f is an arbitrary supremum of elements of $\mathrm{Rg}(B)$, hence $f \in \mathrm{Rg}(B)$. Conversely, since $b(\cdot,x) \in \mathcal{G}$ for all $x \in \mathcal{X}$, $\mathrm{Rg}(B) \subset \mathcal{G}$, so $\mathrm{Rg}(B) = \mathcal{G}$. Hence (23) is equivalent to say that $f = BAf$ for all $f \in \mathrm{Rg}(B)$, so $B = BAB$ and B is Von Neumann regular.

iii)⇒i). Since $\mathcal{G} = \mathrm{Rg}(B)$, and $B = BAB$, for all $x \in \mathcal{X}$ and $f \in \mathcal{G}$, (23) is satisfied when applying Proposition 1 to the $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-linear and continuous operators A, B to derive the corresponding kernels a, b (see Remark 1).

ii)⇒iii). Setting $B = P$ and $A = \mathrm{Id}_{\overline{\mathbb{R}}^{\mathcal{X}}}$ yields the result. In other words, every $(\max,+)$ idempotent operator is Von Neumann regular.

iii)⇒ii). Since $BA = BABA$, setting $P = BA$ proves that P idempotent. Since $B = BAB = PB$, we have that $\mathrm{Rg}(B) \subset \mathrm{Rg}(P)$ and, as $P = BA$, $\mathrm{Rg}(P) \subset \mathrm{Rg}(B)$. Hence $\mathrm{Rg}(P) = \mathrm{Rg}(B) = \mathcal{G}$. □

However the $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-linear formula Theorem 4-(23) does not cover cases of interest which the $\overline{\mathbb{R}}_{\max}^{\mathcal{X}}$-sesquilinear Definition 5-(14) can tackle. Take for instance $\mathcal{X} =$

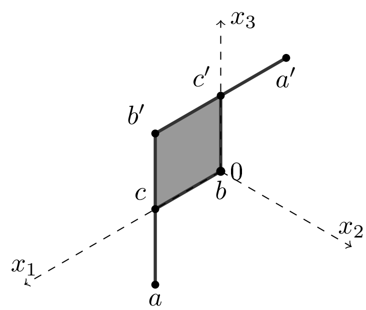

FIGURE 1. A representation of the range of a tropically PSD matrix that is $\overline{\mathbb{R}}_{\max}$-sesquilinear reproducing (Definition 5), but not $\overline{\mathbb{R}}_{\max}$-linear reproducing (Theorem 4), since it is not the range of an idempotent matrix.

$\{-1, 0, 1\}$, and $b(x, y) = (x, y)_2$, then $\mathrm{Rg}(B)$ is not the range of an idempotent $\overline{\mathbb{R}}^{\mathcal{X}}_{\max}$-linear operator. Indeed, the latter are exactly tropical eigenspaces which have been characterized in the setting of tropical spectral theory (Baccelli et al., 1992, Th. 3.100). The only full-dimensional tropical eigenspaces are indeed *alcoved polyhedra*, i.e. sets defined by collections of inequalities $x_i - x_j \geqslant M_{ij}$ for some matrix $M \cup \{-\infty\}$ (Lam and Postnikov, 2007). Here, the range of B, whose cross section by the hyperplane orthogonal to the main diagonal is shown on Figure 1, is the union of such a three dimensional alcoved polyhedron (the two dimensional cell of the cross section), and of two cells of lower dimension 2 (edges of the cross section). The points a, b, a', which correspond to the columns of the matrix B, are the tropical generators of the range of B. The action of the involution $\bar{F}$ (see Theorem 3) is illustrated on the figure, it acts as a central symmetry on the parallelogram (c, b, c', b'), whereas it sends the segment $[c, a]$ to $[c', a']$.

7. APPLICATION TO OPTIMIZATION PROBLEMS: A REPRESENTER THEOREM

One important perk of Hilbertian kernel methods stems from the fact that some infinite-dimensional optimization problems can be solved through equivalent finite-dimensional problems. This behavior is expressed through "representer theorems", which ensure that the solutions of an optimization problem live in a finite-dimensional subspace of the RKHS and consequently enjoy a finite representation. These "representer theorems" can be informally summarized as: "a finite number of evaluations implies a finite number of coefficients" or "all the information is contained in the samples". Representer theorems can be found as early as in Kimeldorf and Wahba (1971, Theorems 3.1 and 5.1) for quadratic norm-penalties and exact or (quadratic) approximate interpolation (see Schölkopf et al., 2001, for an extension to more general objectives). We show below that tropical kernels enjoy a similar property.

Given a finite-valued kernel $b : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, and a subset $\hat{\mathcal{X}} \subset \mathcal{X}$, define

$$\operatorname{Rg}_{\partial\hat{\mathcal{X}}}(B) := \Big\{ f \in \operatorname{Rg}(B) \,|\, \forall x \in \hat{\mathcal{X}},\ \exists p_x \in \mathcal{X},\ f(x) = b(x, p_x) - \bar{B}f(p_x) = \sup_{p \in \mathcal{X}} b(x,p) - \bar{B}f(p) \Big\}. \tag{24}$$

This set can be understood as the subset of functions of $\operatorname{Rg}(B)$ for which there exists a B-subdifferential at every point of $\hat{\mathcal{X}}$.[5] For instance, for $b_c(x, y) = (x, y)_2$ and $\mathcal{X} = \mathbb{R}^N$, it is well-known that $\operatorname{Rg}_{\partial\mathcal{X}}(B)$ contains the continuous convex functions, but that is strictly smaller than $\operatorname{Rg}(B)$ since convex l.s.c. functions may have an empty subdifferential at points that do not lie in the relative interior of their domain. For functions belonging to $\operatorname{Rg}_{\partial\hat{\mathcal{X}}}(B)$, Proposition 6, below, states that there always exists another function with a finite representation that interpolates a finite number of their values.[6] This entails straightforwardly a representer theorem, Corollary 1.

Proposition 6 (Finite interpolation). *Given $(x_m, y_m)_{m\in[M]} \in (\mathcal{X} \times \mathbb{R})^M$, setting $\hat{\mathcal{X}} = \{x_m\}_{m\in[M]}$, the three following statements are equivalent:*

i) there exists $f \in \operatorname{Rg}_{\partial\hat{\mathcal{X}}}(B)$ such that $y_m = f(x_m)$ for all $m \in [M]$;

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

$$f^0(\cdot) := \bar{B}\big(\inf_{m\in[M]} [\delta^{\top}_{p_m}(\cdot) + b(x_m, p_m) - y_m]\big);$$

iii) there exists $(p_m)_{m\in[M]} \in \mathcal{X}^M$ such that $y_n - y_m \geqslant b(x_n, p_m) - b(x_m, p_m)$ for all $n, m \in [M]$.

Proof. i)⇒ii). Set p_m for each x_m as in (24), then, for all $n \in [M]$,

$$\begin{aligned} y_n &\leqslant \max_{m\in[M]} b(x_n, p_m) - b(x_m, p_m) + y_m = f^0(x_n) \\ &\leqslant \sup_{p\in\mathcal{X}} b(x_n, p) - \bar{B}f(p) = \bar{B}\bar{B}f(x_n) = f(x_n) = y_n, \end{aligned}$$

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

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

ii)⇔iii). This follows from the definition of f^0. □

Corollary 1 (Representer theorem). *Given points $(x_m)_{m\in[M]} \in \mathcal{X}^M$ and a function $L : \overline{\mathbb{R}}^M \to \overline{\mathbb{R}}$, fix $\hat{\mathcal{X}} = \{x_m\}_{m\in[M]}$. Then, if the problem*

$$\min_{f\in\operatorname{Rg}(B)} L(f(x_1), \dots, f(x_m))$$

[5] We refer to Martinez-Legaz and Singer (1995, Section 2) and Akian et al. (2005, Section 3) for more comments on this notion and on its relation with the continuity and coercivity of b.

[6] This may be thought of as an analogue of Carathéodory theorem, since, as soon as they are satisfiable by an element $f \in \operatorname{Rg}_{\partial\hat{\mathcal{X}}}(B)$, the M constraints $y_m = f(x_m)$ can be also satisfied by an element f written as a supremum of at most M generators $b(\cdot, p_m)$. We refer the reader to Develin and Sturmfels (2004); Gaubert and Katz (2007); Butkovič et al. (2007) for background on the discrete tropical analogue of Carathéodory theorem.

has a solution $\bar{f} \in \mathrm{Rg}_{\partial\hat{\mathcal{X}}}(B)$ *with finite values* $(f(x_m))_{m\in[M]} \in \mathbb{R}^M$*, it also has a solution* f^0 *as in Proposition 6-ii) which can be obtained solving*

$$\begin{aligned} &\min_{(p_m,y_m)_{m\in[M]}\in(\mathcal{X}\times\mathbb{R})^M} L(y_1,\ldots,y_M) \\ &s.t.\ y_n - y_m \geqslant b(x_n,p_m) - b(x_m,p_m),\ \forall n,m \in [M]. \end{aligned} \tag{25}$$

When b is the standard scalar product, Corollary 1 recovers a well-known property in convex regression (e.g. Boyd and Vandenberghe, 2004, Section 6.5.5), where (25) is then a convex problem provided L is convex. We have shown that this result thus also holds for very general kernels, not even assuming symmetry or tropical positivity of b. This opens the question of the computational advantages of tropical positive semidefiniteness, both for finite and uncountable sets $\mathcal{X}$.

References

Akian, M. (1999). Densities of idempotent measures and large deviations. *Transactions of the American Mathematical Society*, 351(11):4515–4543.

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.

Akian, M., Gaubert, S., and Lakhoua, A. (2008). The max-plus finite element method for solving deterministic optimal control problems: basic properties and convergence analysis. *SIAM J. Control Optim.*, 47(2):817–848.

Akian, M., Quadrat, J., and Viot, M. (1994). Bellman processes. In *11th International Conference on Analysis and Optimization of Systems : Discrete Event Systems*, volume 199 of *Lecture notes in control and information sciences*. Springer Verlag.

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404.

Baccelli, F., Cohen, G., Olsder, G., and Quadrat, J. (1992). *Synchronization and Linearity*. Wiley.

Berg, C., Christensen, J. P. R., and Ressel, P. (1984). *Harmonic Analysis on Semigroups*. Springer New York.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.

Burkard, R. E., Klinz, B., and Rudolf, R. (1996). Perspectives of Monge properties in optimization. *Discrete Applied Mathematics*, 70(2):95–161.

Butkovič, P., Schneider, H., and Sergeev, S. (2007). Generators, extremals and bases of max cones. *Linear Algebra Appl.*, 421(2-3):394–406.

Calafiore, G. C., Gaubert, S., and Possieri, C. (2020). A universal approximation result for difference of log-sum-exp neural networks. *IEEE Trans. Neural Networks Learn. Syst.*, 31(12):5603–5612.

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

Cartwright, D. and Chan, M. (2012). Three notions of tropical rank for symmetric matrices. *Combinatorica*, 32(1):55–84.

Chancelier, J.-P. and Lara, M. D. (2021). Capra-convexity, convex factorization and variational formulations for the ℓ_0 pseudonorm. *Set-Valued and Variational Analysis*. on line.

Cohen, G., Gaubert, S., and Quadrat, J.-P. (2004). Duality and separation theorems in idempotent semimodules. *Linear Algebra and its Applications*, 379:395–422.

Develin, M. and Sturmfels, B. (2004). Tropical convexity. *Doc. Math.*, 9:1–27. (Erratum pp. 205–206).

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.

Feydy, J., Séjourné, T., Vialard, F.-X., Amari, S.-i., Trouve, A., and Peyré, G. (2019). Interpolating between optimal transport and MMD using Sinkhorn divergences. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89, pages 2681–2690.

Fleming, W. H. and McEneaney, W. M. (2000). A max-plus-based algorithm for a Hamilton-Jacobi-Bellman equation of nonlinear filtering. *SIAM J. Control Optim.*, 38(3):683–710.

Gaubert, S. and Katz, R. (2007). The Minkowski theorem for max-plus convex sets. *Linear Algebra and Appl.*, 421:356–369.

Gaubert, S. and Niv, A. (2018). Tropical totally positive matrices. *Journal of Algebra*, 515:511 – 544.

Gierz, G., Hofmann, K., Keimel, K., Lawson, J., Mislove, M., and Scott, D. (2003). *Continuous Lattices and Domains*. Encyclopedia of Mathematics and its Applications. Cambridge University Press.

Kimeldorf, G. and Wahba, G. (1971). Some results on Tchebycheffian spline functions. *Journal of Mathematical Analysis and Applications*, 33(1):82–95.

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

Lam, T. and Postnikov, A. (2007). Alcoved polytopes. I. *Discrete Comput. Geom.*, 38(3):453–478.

Litvinov, G. (2011). Tropical mathematics, idempotent analysis, classical mechanics and geometry. In *Spectral Theory and Geometric Analysis*, pages 159–186. American Mathematical Society.

Litvinov, G. L. (2005). Maslov dequantization, idempotent and tropical mathematics: A brief introduction. *Journal of Mathematical Sciences*, 140:426–444.

Maragos, P., Charisopoulos, V., and Theodosis, E. (2021). Tropical geometry and machine learning. *Proceedings of the IEEE*, 109(5):728–755.

Martinez-Legaz, J. and Singer, I. (1990). Dualities between complete lattices. *Optimization*, 21(4):481–508.

Martinez-Legaz, J.-E. and Singer, I. (1995). Subdifferentials with respect to dualities. *ZOR Zeitschrift für Operations Research Methods and Models of Operations Research*, 42(1):109–125.

Mary, X. (2005). Theory of subdualities. *Journal d'Analyse Mathématique*, 97(1):203–241.

McEneaney, W. M. (2006). *Max-Plus Methods for Nonlinear Control and Estimation*. Birkhäuser-Verlag.

McEneaney, W. M. (2007). A curse-of-dimensionality-free numerical method for solution of certain HJB PDEs. *SIAM J. Control Optim.*, 46(4):1239–1276.

Montúfar, G., Ren, Y., and Zhang, L. (2021). Sharp bounds for the number of regions of maxout networks and vertices of Minkowski sums. arXiv:2104.08135.

Moreau, J. J. (1970). Inf-convolution, sous-additivité, convexité des fonctions numériques. *Journal de Mathématiques Pures et Appliquées*, pages 109–154.

Ong, C. S., Mary, X., Canu, S., and Smola, A. J. (2004). Learning with non-positive kernels. In *International Conference on Machine Learning (ICML)*, pages 81–88. ACM Press.

Pallaschke, D. and Rolewicz, S. (1997). *Foundations of Mathematical Optimization*. Springer Netherlands.

Saitoh, S. and Sawano, Y. (2016). *Theory of Reproducing Kernels and Applications*. Springer Singapore.

Santambrogio, F. (2017). {Euclidean, metric, and Wasserstein} gradient flows: an overview. *Bulletin of Mathematical Sciences*, 7(1):87–154.

Schölkopf, B., Herbrich, R., and Smola, A. J. (2001). A generalized representer theorem. In *Computational Learning Theory (COLT)*, pages 416–426. Springer Berlin Heidelberg.

Schölkopf, B. and Smola, A. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.

Schwartz, L. (1964). Sous-espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants). *Journal d'Analyse Mathématique*, 13:115–256.

Seijo, E. and Sen, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. *The Annals of Statistics*, 39(3):1633–1657.

Singer, I. (1984). Conjugation operators. In *Lecture Notes in Economics and Mathematical Systems*, pages 80–97. Springer Berlin Heidelberg.

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

Steinwart, I. and Christmann, A. (2008). *Support Vector Machines*. Springer.

Tran, N. M. (2020). Tropical Gaussians: a brief survey. *Algebraic Statistics*, 11(2):155–168.

Volle, M., Martínez-Legaz, J. E., and Vicente-Pérez, J. (2013). Duality for closed convex functions and evenly convex functions. *Journal of Optimization Theory and Applications*, 167(3):985–997.

Weiß, C., Knust, S., Shakhlevich, N., and Waldherr, S. (2016). The assignment problem with nearly Monge arrays and incompatible partner indices. *Discrete Applied Mathematics*, 211:183–203.

Yoshida, R., Takamori, M., Matsumoto, H., and Miura, K. (2021). Tropical support vector machines: Evaluations and extension to function spaces. (`https://arxiv.org/abs/2101.11531`).

Yoshida, R., Zhang, L., and Zhang, X. (2019). Tropical principal component analysis and its application to phylogenetics. *Bulletin of Mathematical Biology*, 81:568–597.

Yu, J. (2014). Tropicalizing the positive semidefinite cone. *Proceedings of the American Mathematical Society*, 143(5):1891–1895.

Zhang, L., Naitzat, G., and Lim, L.-H. (2018). Tropical geometry of deep neural networks. In *International Conference on Machine Learning (ICML)*, volume 80, pages 5824–5832.

INRIA and Département d'Informatique, École Normale Supérieure, PSL Research University

INRIA and CMAP, École polytechnique, IP Paris, CNRS
Email address: pierre-cyril.aubin@inria.fr
Email address: stephane.gaubert@inria.fr