2.4 (f, Γ)-Divergences on Polish Spaces

When working on a Polish space, S, and under further assumptions on f and Γ, we are able to show that D_f^Γ interpolates between the classical f-divergence, D_f, and the Γ-IPM, W^Γ. At various points, we will require f and Γ to have the following properties:

Definition 14 *We will call $f \in \mathcal{F}_1(a, b)$* **admissible** *if $\lim_{y \to -\infty} f^*(y) < \infty$ (note that this limit always exists by convexity) and $\{f^* < \infty\} = \mathbb{R}$. If f is also strictly convex at 1 then we will call f* **strictly admissible***. We will call $\Gamma \subset C_b(S)$* **admissible** *if $0 \in \Gamma$, Γ is convex, and Γ is closed in the weak topology generated by the maps τ_μ, $\mu \in M(S)$ (see Section 2.1). Γ will be called* **strictly admissible** *if it also satisfies the following property: There exists a $\mathcal{P}(S)$-determining set $\Psi \subset C_b(S)$ such that for all $\psi \in \Psi$ there exists $c \in \mathbb{R}$, $\epsilon > 0$ such that $c \pm \epsilon\psi \in \Gamma$.*

Our main result, Theorem 15, will require admissibility of both f and Γ. The functions f_{KL} and f_α, $\alpha > 1$, defined in Eq. (8), are strictly admissible but f_α, $\alpha \in (0, 1)$ is not admissible (however, Theorem 8 above does apply to f_α for $0 < \alpha < 1$). The admissibility requirements that Γ be convex and closed will let us express D_f^Γ as the infinite-dimensional convex conjugate of a convex and LSC functional. This will allow us to analyze D_f^Γ using tools from convex analysis. Strict admissibility will be key in proving the divergence property for both W^Γ and D_f^Γ.

Examples of strictly admissible Γ:

1. $\Gamma = C_b(S)$, which leads to the classical f-divergences.
2. $\Gamma = \mathrm{Lip}_b^1(S)$, i.e., all bounded 1-Lipschitz functions, which leads to generalizations of the Wasserstein metric.
3. $\Gamma = \{g \in C_b(S) : |g| \leq 1\}$, which leads to generalizations of the total variation metric.
4. $\Gamma = \{g \in \mathrm{Lip}_b^1(S) : |g| \leq 1\}$, which leads to generalizations of the Dudley metric.
5. $\Gamma = \{g \in X : \|g\|_X \leq 1\}$, the unit ball in a RKHS $X \subset C_b(S)$ (under appropriate assumptions given in Lemma 77). This yields a generalization of MMD and is also related to the recent KL-MMD interpolation method in Glaser et al. (2021); the latter employs a soft constraint rather than working on the RKHS unit ball and is based on the representation (10) instead of (11).

Note that the first two examples are shift invariant (hence Equation 18 is applicable) while the latter three are not.

We are now ready to present the second key theorem in this paper, where we derive the infimal convolution representation of D_f^Γ and provide alternative (to Theorem 8) conditions that ensure D_f^Γ possesses the divergence property. The proof can be found in Appendix C, Theorem 74.

Theorem 15 *Suppose f and Γ are admissible. For $Q, P \in \mathcal{P}(S)$ let $D_f^\Gamma(Q\|P)$ be defined by (15) and let $W^\Gamma(Q, P)$ be defined as in (16). These have the following properties:*