Convex Relaxations for Manifold-Valued Markov Random Fields with Approximation Guarantees
Supplementary Materials

Robin Kenis, Emanuel Laude, and Panagiotis Patrinos

STADIUS ESAT, KU Leuven, Belgium
{robin.kenis,emanuel.laude,panos.patrinos}@kuleuven.be

A Additional Experiments

A.1 Denoised Images

In this subsection we include the denoised images (Fig. A.1) based on the chromaticity denoising experiment in Sec. 5.2. The chromaticity is multiplied by the intensity, which is denoised using a convex solver under TV-penalty. We record in Tab. A.1 the optimal cost P obtained when allowing our method to converge fully, as well as the relative gap G which shows that the results are numerically tight. Note that these costs are very nearly those obtained by our methods in Tab. 1 after 1000 seconds, showing that the results there are indeed nearly globally optimal. Finally, in Tab. A.2 we document the parameters used in the setup for the experiments. As in the cost function Eq. (15), L denotes the Lipschitz-scaling parameter. By σ we mean the standard deviation of the Gaussian noise added to corrupt the input data. The noise was generated using Matlab's `randn()`-function, with seed `12345`.

Table A.1: Primal energies, dual energies and relative gaps for the experiments in Fig. 2 (a)-(c) obtained by our relaxation upon convergence.

	Fig. 2a	Fig. 2b	Fig. 2c
P	3228.25	5564.62	5288.57
D	3228.17	5564.12	5288.52
G	2.479×10^{-5}	9.089×10^{-5}	9.586×10^{-6}

A.2 Convergence of the Optimal Transport Subproblem

In a first additional numerical experiment, we evaluate the modeling power of polynomials for solving the optimal transport problem, which is a subproblem

(a) (b) (c)

Fig. A.1: Denoised images obtained by separate ROF-type denoising chromaticity and intensity. Chromaticity is denoised using our own method (see Sec. 5.2).

Table A.2: Parameter values for experiments in 2.

	Fig. 2a	Fig. 2b	Fig. 2c	Fig. 2d
L	3×10^{-1}	3×10^{-1}	3×10^{-1}	1.5×10^{0}
σ	7.5×10^{1}	7.5×10^{1}	7.5×10^{1}	5×10^{-1}
size	300×200	300×200	300×200	60×40

of our formulation. Therefore we solve the semi-infinite problem

$$\begin{aligned} \max_{\varphi\in\mathbb{R}_n[x]} \quad & \langle\varphi,\mu_2-\mu_1\rangle \\ \text{subject to} \quad & L-\langle\xi,\Pi(x)\nabla\varphi(x)\rangle\geq 0, \quad \forall(x,\xi)\in\Omega\times\mathcal{S}^m \end{aligned} \tag{A.1}$$

for $\mu_1,\mu_2\in\mathcal{M}$ fixed. For this experiment we choose two diametrically opposite points x,y and let $\Omega=\mathcal{S}^1$, $\mu_1=\delta_x$ and $\mu_2=\delta_y$. Similarly to the $\mathcal{S}^1$-experiment in Sec. 5.2, diametrically opposite points form a worst case scenario for our relaxation. Consider optimizing the cost in (A.1) over $\mathrm{Lip}(\Omega,d)$ (the space of Lipschitz continuous functions w.r.t. the metric d on Ω) rather than the space $\mathbb{R}_n[x]$. It is straightforward to see that there exists a unique optimal function $\varphi\in\mathrm{Lip}(\Omega,d)$ with piecewise constant manifold gradient, which is constantly equal to 1 on half the circle, and -1 on the other half. The discontinuous gradient inhibits the convergence speed of the polynomial approximant. This is further exacerbated by the Lipschitz constraint, which requires that the norm of the manifold gradient of the polynomial never exceeds 1. We compute and plot for various degrees n the optimal polynomial multipliers φ_n restricted to $\mathcal{S}^1$ in terms of the angle θ, as well as their manifold gradients, in Fig. A.2. We also provide the multipliers ψ_n discussed in Theorem 6.

Furthermore, we display the gap between the polynomial approximations φ_n in terms of the degree n, and the actual geodesic distance between x and y ($d(x,y)=\pi$). We also plot the gap for the polynomial approximations ψ_n constructed in the proof of Theorem 6. These results are found in Fig. A.3, where for degrees n ranging from 1 to 10 the relative error between the geodesic distance and our polynomial approximation for various angles θ between x and y is also shown. Whenever x and y are not diametrically opposite, the approximation is

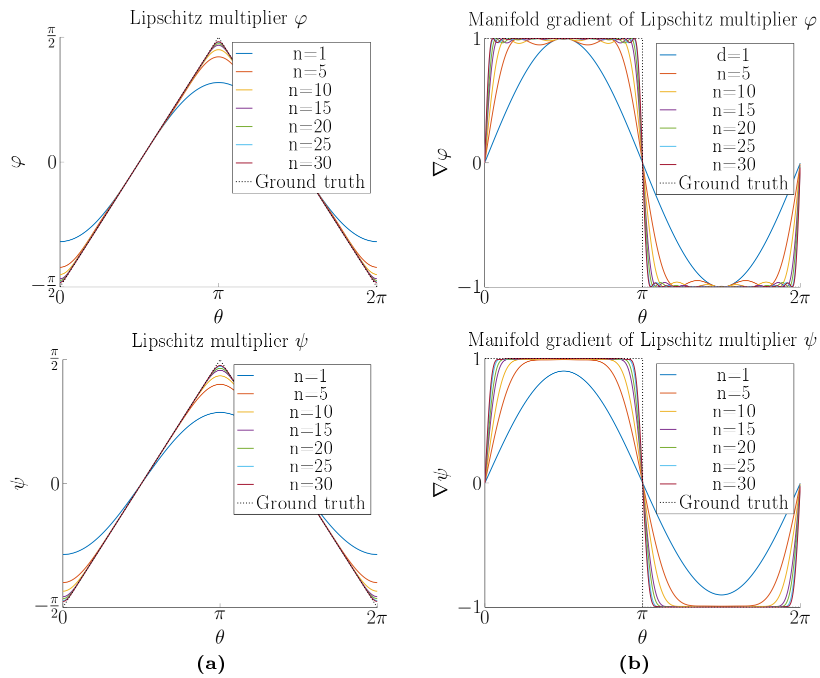


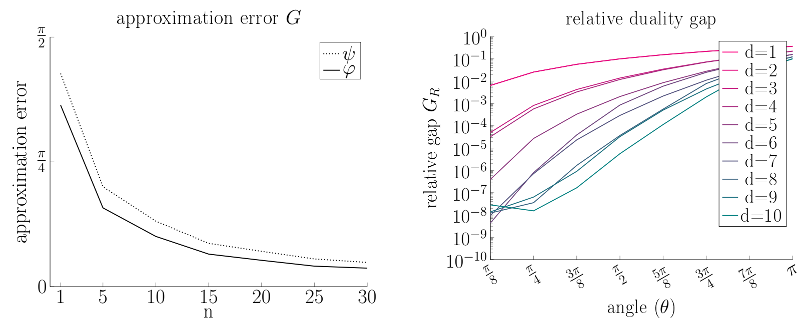
Fig. A.2: Lipschitz continuous Lagrange multiplier φ_n (optimal, obtained via our approach) and ψ_n (reference, as in Theorem 6), as well as their manifold gradients, for a connection between two vertices on opposing points on the unit circle for various degrees n.

always notably better than what is shown in the experiment; hence the particular focus on diametrically opposite points which actually pose a worst case problem for $\mathcal{S}^m$ more broadly.

A.3 Geodesic Denoising on $\mathcal{SO}(3)$

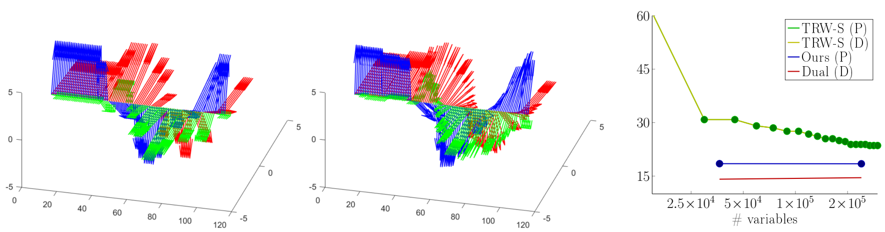
As an additional geodesic TV-denoising experiment, we implement our relaxation of (mR-D) for $\Omega = \mathcal{SO}(3)$, the manifold of 3-dimensional rotations. The results are displayed in Fig. A.4. We plot an element r of $\mathcal{SO}(3)$ by displaying how a standard orthonormal xyz-coordinate system is rotated by applying r from the left. The coordinate sytems have color-coded axes: green for the x-axis, red for the y-axis and blue for the z-axis. The ground truth and noisy input data are provided in Fig. A.5.

While we provide no guarantees for our metric relaxation beyond the sphere, we see that it performs very well in practice. We compare with the linear relaxation from 50 to 1000 labels, in steps of 50. We see that even at degree 2, our approach outperforms the reference approach with a maximum of 1000 samples. The relative primal-dual gap remains high, with $G = 0.2116$ at degree 4. Over



(a) (b)

Fig. A.3: (a) Error between polynomial approximation of the geodesic distance by φ_n (optimal, obtained via our approach) and ψ_n (reference, as in Theorem 6) and the actual geodesic distance ($d(x, y) = \pi$) for diametrically opposite points x and y in terms of degree n, (b) Relative duality gap for optimal multiplier φ_n for the optimal transport problem in function of degree n and angle θ between x and y.



(a) (b) (c)

Fig. A.4: Geodesic denoising of $\mathcal{SO}(3)$-valued data using (a) the reference approach and (b) our approach. (c) contains the cost obtained using the reference approach (*green line*) and the cost P after rounding using our own approach (*blue line*), as well as the dual cost D (*red dotted line*) expressed in the numbers of variables in the relaxation.

35% of this relative gap is due to the underestimation of the geodesic distance at the discontinuous jump in the dataset.

B Implementation Details

Following the rich literature on polynomial optimization via semidefinite programming spearheaded by Lasserre and Parillo [3, 5], we rewrite (R-D$^{\mathrm{n}}$) and (mR-D$^{\mathrm{n}}$) as a hierarchy of finite dimensional SDPs. We write down the final relaxation first, and set out the detail afterwards:

The fully polynomial case:

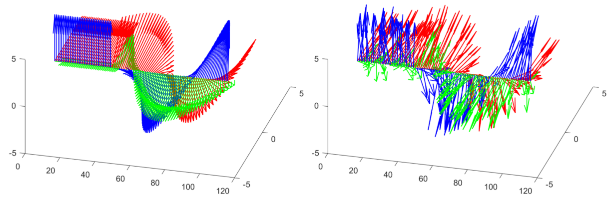


Fig. A.5: (a) ground truth and (b) noisy input data for the $\mathcal{SO}(3)$-valued experiment.

$$
\begin{aligned}
\min_{\boldsymbol{\mu}\in\mathbb{R}^{N\times\mathcal{V}}} \quad \max_{\substack{(\boldsymbol{\varphi},\boldsymbol{\psi})\in\left(\mathbb{R}^{P\times\mathcal{E}}\right)^2\\ \boldsymbol{S}\in(\mathbb{S}_+)^{\mathcal{E}}}} \quad & \sum_{u\in\mathcal{V}}\langle \boldsymbol{f}_u - A_u\boldsymbol{\varphi}, \boldsymbol{\mu}_u\rangle \\
\text{subject to} \quad & \langle \mathbf{1}, \boldsymbol{\mu}_u\rangle = 1, \quad F^*\boldsymbol{\mu}_u = 0, \quad \mathcal{B}_{\mathcal{V}}\boldsymbol{\mu}_u \succcurlyeq 0, \quad \forall\, u\in\mathcal{V} \\
& \boldsymbol{f}_{(\boldsymbol{u},\boldsymbol{v})} - \Pi_u^*\boldsymbol{\varphi}_u - \Pi_v^*\boldsymbol{\varphi}_v = \mathcal{B}_{\mathcal{E}}^*\boldsymbol{S}_{(u,v)}, \\
& \quad \boldsymbol{S}_{(u,v)} \succcurlyeq 0, \quad \forall\,(u,v)\in\mathcal{E}
\end{aligned}
$$

The geodesic case:

$$
\begin{aligned}
\min_{\boldsymbol{\mu}\in\mathbb{R}^{N\times\mathcal{V}}} \quad \max_{\substack{\boldsymbol{\varphi}\in\mathbb{R}^{N\times\mathcal{E}}\\ \boldsymbol{S}\in(\mathbb{S}_+)^{\mathcal{E}}}} \quad & \sum_{u\in\mathcal{V}}\langle \boldsymbol{f}_u - \mathbf{Div}\boldsymbol{\varphi}, \boldsymbol{\mu}_u\rangle \\
\text{subject to} \quad & \langle \mathbf{1}, \boldsymbol{\mu}_u\rangle = 1, \quad F^*\boldsymbol{\mu}_u = 0, \quad \mathcal{B}_{\mathcal{V}}\boldsymbol{\mu}_u \succcurlyeq 0 \quad \forall\, u\in\mathcal{V} \\
& \mathbf{1} - \langle \boldsymbol{\xi}, \Pi\boldsymbol{\nabla}\boldsymbol{\varphi}_{(u,v)}\rangle = \mathcal{B}_{\mathcal{E}}^*\boldsymbol{S}_{(u,v)} \\
& \quad \boldsymbol{S}_{(u,v)} \succcurlyeq 0 \quad \forall\,(u,v)\in\mathcal{E}
\end{aligned}
$$

$\mathbf{1}$ is the coefficient vector for the constant polynomial 1. $\boldsymbol{S}_{(u,v)}$ is a positive semidefinite matrix (denoted $\mathbb{S}^+$) of appropriate size. The linear operators $\boldsymbol{A}_u$ and $\mathbf{Div}$ are defined analogously to the operators (6) and (10), operating on coefficient vectors rather than functions.

We now discuss this relaxation, including the notation and operators therein in more detail.

B.1 Representation of Polynomials and Measures on Ω:

In order to implement our relaxation, we have to represent both polynomials and measures on Ω in a discrete way. We will represent polynomials by their coefficient vectors, and measures by a finite number of their moments.

Let $x, y \in \Omega$. For the vertices of $\mathcal{G}$, choose a basis of monomials for $\mathbb{R}_n[x]$, and collect them as the entries of a vector $\boldsymbol{b}_{\mathcal{V}}$. Similarly, for the edges of $\mathcal{G}$, choose a basis of monomials for $\mathbb{R}_n[x, y]$, collected in a vector $\boldsymbol{b}_{\mathcal{E}}$. The vectors $\boldsymbol{b}_u$ and $\boldsymbol{b}_{(u,v)}$ are often called *feature maps* in literature. For the standard choice of monomials in the univariate case, for example, we have:

$$
\boldsymbol{b}_{\mathcal{V}}(x) = \left[1, x, x^2, \ldots, x^n\right]^\top \qquad \text{and} \qquad \boldsymbol{b}_{\mathcal{E}}(x) = \left[1, x, y, x^2, xy, \ldots, y^n\right]^\top. \tag{B.1}
$$

From here on out we identify polynomials $\phi_u \in \mathbb{R}_n[x]$ and $\phi_{(u,v)} \in \mathbb{R}_n[x,y]$ with their coefficient vectors $\boldsymbol{\phi}_u$ and $\boldsymbol{\phi}_u$ in the monomial basis $\boldsymbol{b}_\mathcal{V}$ and $\boldsymbol{b}_\mathcal{E}$, respectively, such that

$$\phi_u(x_u) = \langle \boldsymbol{\phi}_u, \boldsymbol{b}_\mathcal{V}(x_u)\rangle \qquad \text{and} \qquad \phi_{(u,v)}(x_u, x_v) = \langle \boldsymbol{\phi}_{(u,v)}, \boldsymbol{b}_\mathcal{E}(x_u, x_v)\rangle. \quad \text{(B.2)}$$

Borel probability measures $\mu_u \in \mathcal{P}(\Omega)$ and $\mu_{(u,v)} \in \mathcal{P}(\Omega \times \Omega)$ are represented by vectors $\boldsymbol{\mu}_u$ and $\boldsymbol{\mu}_{(u,v)}$ containing their moments up to degree n w.r.t. $\boldsymbol{b}_\mathcal{V}$ and $\boldsymbol{b}_\mathcal{E}$ respectively:

$$\boldsymbol{\mu}_u = \int \boldsymbol{b}_\mathcal{V}(x_u)\, \mathrm{d}\mu_u(x_u) \qquad \text{and} \qquad \boldsymbol{\mu}_{(u,v)} = \int \boldsymbol{b}_\mathcal{E}(x_u, x_v)\, \mathrm{d}\mu_{(u,v)}(x_u, x_v). \quad \text{(B.3)}$$

When $\Omega \subset \mathrm{I\!R}^M$, let $N = \binom{M+n}{n}$ and $P = \binom{2M+n}{n}$, then $\boldsymbol{\phi}_u \in \mathrm{I\!R}^N$ and $\boldsymbol{\mu}_u \in \mathrm{I\!R}^N$ while $\boldsymbol{\phi}_{(u,v)} \in \mathrm{I\!R}^P$ and $\boldsymbol{\mu}_{(u,v)} \in \mathrm{I\!R}^P$. By linearity of integration, it is clear that

$$\int_\Omega \phi_u \,\mathrm{d}\mu_u = \langle \boldsymbol{\phi}_u, \boldsymbol{\mu}_u\rangle \qquad \text{and} \qquad \int_{\Omega\times\Omega} \phi_{(u,v)} \,\mathrm{d}\mu_{(u,v)} = \langle \boldsymbol{\phi}_{(u,v)}, \boldsymbol{\mu}_{(u,v)}\rangle \quad \text{(B.4)}$$

To assure that the measure μ_u is supported solely on Ω, we add the constraint $\langle \boldsymbol{f}, \boldsymbol{\mu}_u\rangle = 0$ for all polynomials $f \in \mathcal{I}(\Omega)$ up to degree n, where $\mathcal{I}(\Omega)$ is the set of polynomials which are identically zero on Ω ($\mathcal{I}(\Omega)$ is a polynomial ideal). We can construct a (finite) basis of coefficient vectors $\{\boldsymbol{f}_i\}$ such that for any $f \in \mathcal{I}(\Omega)$, $\boldsymbol{f}$ can be expressed as a linear combination of $\{\boldsymbol{f}_i\}$. Collecting the vectors $\{\boldsymbol{f}_i\}$ as columns in a matrix F, the constraint $\langle \boldsymbol{f}, \boldsymbol{\mu}_u\rangle = 0$ for all $f \in \mathcal{I}(\Omega)$ can be rephrased equivalently as a linear one: $F^* \boldsymbol{\mu}_u = 0$. A similar constraint holds for μ which is supported on the product manifold $\Omega \times \Omega$

B.2 Nonnegativity of polynomials and measures

Since polynomial nonnegativity is in general intractable to verify, it is a common idea in polynomial optimization literature to replace such conditions by an easy-to-verify nonnegativity certificate. There are a number of Positivstellensätze that have been considered for this goal [6, 8, 10, 15], but possibly the most popular option is Putinar's Positivstellensatz. For our problem, Putinar's certificate implies relaxing the polynomial nonnegativity constraint $\varphi \geq 0$ as $\varphi \in \Sigma$, where we have introduced the convex cone Σ as in (14):

$$\Sigma[x] = \left\{ q \in \mathbb{R}[x] \;\middle|\; q = \sum_i q_i^2 + \sum_j r_j k_j, \;\; k_j, q_i, r_j \in \mathbb{R}[x],\; k_j(x) = 0\; \forall x \in \Omega \right\}. \quad \text{(B.5)}$$

Verifying membership in the entire cone Σ is still not tractable. If we define Σ_n as

$$\Sigma_n[x] = \left\{ q \in \mathbb{R}[x] \;\middle|\; q = \sum_i q_i^2 + \sum_j r_j k_j, \;\; k_j, q_i, r_j \in \mathbb{R}_n[x],\; k_j(x) = 0\; \forall x \in \Omega \right\}. \quad \text{(B.6)}$$

we obtain a finite-dimensional cone condition $\varphi \in \Sigma_n$ as a relaxation of $\varphi \geq 0$. Checking the constraint $\varphi \in \Sigma_n$ can easily be realized as a semidefinite constraint w.r.t. a matrix constructed from the coefficient vector $\boldsymbol{\phi}$ [3]. For any monomial basis $\boldsymbol{b}_{\mathcal{V}}$, we introduce the operator $\mathcal{B}_{\mathcal{V}}$ such that $\phi \in \Sigma_n$ if and only if there exists a semidefinite matrix $\boldsymbol{S}_u \succcurlyeq 0$ such that $\boldsymbol{\phi}_u = \mathcal{B}_{\mathcal{V}} \boldsymbol{S}_u$. Similarly for the basis $\boldsymbol{b}_{\mathcal{E}}$, we introduce $\mathcal{B}_{\mathcal{E}}$ such that $\boldsymbol{\phi}_{(u,v)} = \mathcal{B}_{\mathcal{V}}^* \boldsymbol{S}_{(u,v)}$ for $\boldsymbol{S}_{(u,v)} \succcurlyeq 0$ is equivalent to $\phi_{(u,v)} \in \Sigma_n$ (where we mean $\Sigma_n[x,y]$). For more information about the operators $\mathcal{B}_{\mathcal{V}}$ and $\mathcal{B}_{\mathcal{E}}$, we refer to [3, 5]

Since any element of Σ_n is trivially nonnegative on Ω, Σ_n is an inner approximation of the cone of nonnegative polynomials. By definition, μ is an element of the dual cone Σ_n^* if and only if $\langle f, \mu \rangle \geq 0$ for all $f \in \Sigma_n$. Therefore, Σ_n^* can be thought of as an outer approximation to the cone of nonnegative measures on Ω. The condition $\mu \in \Sigma_n^*$ can be verified by a *linear matrix inequality* (LMI) constraint with weight vector $\boldsymbol{\mu}$. We have that $\mu_u \in \Sigma^*$ is verified by the LMI $\mathcal{B}_{\mathcal{V}} \boldsymbol{\mu}_u \succcurlyeq 0$. Similarly, $\mu_{(u,v)} \in \Sigma_n^*$ (meaning $\Sigma_n[x,y]^*$) is verified by the LMI $\mathcal{B}_{\mathcal{E}} \boldsymbol{\mu}_{(u,v)} \succcurlyeq 0$.

B.3 Marginalization and Lipschitz Constraints

In the general relaxation, we maximize over polynomials $\varphi \in (\mathbb{R}_n[x])^{\mathcal{E}}$ and $\psi \in (\mathbb{R}_n[y])^{\mathcal{E}}$, represented as coefficient vectors, such that

$$f_{(u,v)}(x,y) - \varphi_u(x) - \psi_v(y) \geq 0, \qquad \forall (x,y) \in \Omega \times \Omega, \tag{B.7}$$

which will be relaxed to $f_{(u,v)}(x,y) - \varphi_u(x) - \psi_v(y) \in \Sigma_n[x,y]$ in view of the previous subsection. The coefficient vectors $\boldsymbol{\varphi}_u$ and $\boldsymbol{\psi}_v$ are with respect to the monomials in $\boldsymbol{b}_{\mathcal{V}}$, whereas $\boldsymbol{f}_{(u,v)}$ is with respect to the monomials in $\boldsymbol{b}_{\mathcal{E}}$. Clearly, the monomials $\boldsymbol{b}_{\mathcal{E}}$ (as a basis of $\mathbb{R}_n[x,y]$) span those in $\boldsymbol{b}_{\mathcal{V}}$ (which is a basis of $\mathbb{R}_n[x]$ or equivalently $\mathbb{R}_n[y]$). Therefore, we can introduce the operator Π_u, which maps the coefficient vector $\boldsymbol{\varphi}_u$ expressing φ_u in the basis $\boldsymbol{b}_{\mathcal{V}}$ to $\Pi_u \boldsymbol{\varphi}_u$ expressing φ_u as a polynomial in the basis $\boldsymbol{b}_{\mathcal{E}}$ in the first variable ($\mathbb{R}_n[x] \to \mathbb{R}_n[x,y]$). Similarly we introduce Π_v such that $\Pi_v \boldsymbol{\psi}_v$ is the coefficient vector of ψ_v in the basis $\boldsymbol{b}_{\mathcal{E}}$, in the second variable ($\mathbb{R}_n[y] \to \mathbb{R}_n[x,y]$).

For the metric relaxation, on the other hand, it may not immediately be obvious how the $\langle \xi, \Pi \nabla \phi \rangle$-term in (13) is handled. Note however that the operators ∇ and $\Pi(x)$ from $\mathbb{R}[x]$ to $\mathbb{R}[x]$, as well as $\langle \xi, \cdot \rangle$ from $\mathbb{R}[x]$ to $\mathbb{R}[x, \xi]$ are all linear operators in terms of the coefficient vectors. Particularly, the composition $\langle \xi, \Pi \nabla \cdot \rangle$ is a linear map between coefficient vectors of polynomials from $\mathbb{R}[x]$ to coefficient vectors of polynomials in $\mathbb{R}[x, \xi]$. In line with our current notation, we will denote the corresponding linear operator that acts on the coefficient vectors as $\langle \boldsymbol{\xi}, \Pi \boldsymbol{\nabla} \cdot \rangle$.

B.4 Projection onto the Tangent Space of Ω

In Sec. 4.1 we consider the orthogonal projection matrix $\Pi(x) = I - xx^*$ which for any $x \in \mathcal{S}^m$ projects a vector $v \in \mathbb{R}^{m+1}$ onto the tangent space $\mathcal{T}_x \mathcal{S}^m$ at

x (that is: $\Pi(x)v \in \mathcal{T}_x\mathcal{S}^m$). In order to implement the metric relaxation for a more general manifold Ω, we require the existence of an orthogonal projection $\Pi(x)$ onto $\mathcal{T}_x\Omega$. Furthermore, we require the condition (13) to be polynomially implementable.

We assume that the Riemannian submanifold $\Omega \subset \mathbb{E}$ is defined in terms of polynomial equalities, meaning that there exists a finite set of polynomials $\{f_i\}_{i=1}^I$ such that Ω is exactly the set of points where $f_i = 0$ for every $i = 1 \dots I$. The existence of the orthogonal projector $\Pi(x)$ can be verified for such $\Omega \subset \mathbb{E}$, using standard results from differential and algebraic geometry.

It is known that the tangent space of Ω is exactly the orthogonal complement to $\operatorname{span}(J_f)$, where J_f denotes the Jacobian of the mapping $[f_1, f_2, \dots, f_I]^\top$ [4]. Since Ω is a Riemannian submanifold, the tangent space $\mathcal{T}_x\Omega$ at x has constant dimension m for any x. Therefore the columns of the Jacobian J_f are either always or never linearly independent. Assume for now that $m = I$ and the columns of J_f are linearly independent. In that case it is a standard result from linear algebra that the projection $\Pi(x)$ onto the complement of $\operatorname{rge}(J_f)$ is given as

$$\Pi(x) = I - J_f({J_f}^\top J_f)^{-1}{J_f}^\top = I - \tfrac{1}{\det({J_f}^\top J_f)} J_f \operatorname{Adj}({J_f}^\top J_f){J_f}^\top, \tag{B.8}$$

where $\operatorname{Adj}(A)$ is the adjugate matrix of A, which is specifically polynomial in the entries of A. We can then clearly see that (B.8) is in general a rational function with denominator $d(x) = \det(J_f(x)^\top J_f(x))$. By full column rank of J_f, the $d(x)$ does not vanish anywhere on Ω. The gradient inequality (13) can then be rewritten as a polynomial inequality as

$$p(x,\xi) = d(x)\big(L - \langle \xi, \Pi(x)\nabla\varphi(x)\rangle\big) \geq 0, \quad \forall (x,\xi) \in \Omega \times \mathcal{S}. \tag{B.9}$$

Here, $\mathcal{S}$ is the unit sphere in $\mathbb{E}^*$.

If the columns of J_f were not linearly independent, then we can construct a different set of polynomials $\{g_j\}_j^J$ such that the columns of J_g (defined analogously to J_f) are linearly independent. Let $\mathcal{I}(\Omega)$ be the largest ideal of polynomials that vanish on Ω (By Hilbert's Nullstellensatz, that is the radical ideal of Ω). Then Ω is indeed

$$\Omega = \{x \in \mathbb{E} \mid f(x) = 0,\ \forall\, g \in \mathcal{I}(\Omega)\}\,. \tag{B.10}$$

By Hilbert's basis theorem, the ideal $\mathcal{I}(\Omega)$ is generated by a finite basis of polynomials.

C Proofs

C.1 Proof of Theorem 1 and Sharpness of Tree Assumption

The following theorem (well known for the case of discrete MRFs [16]) is a specialization of [7, Equation 3.5] noting that trees satisfy the *running intersection property* (RIP) [7, Equation 1.3]:

Proof (Theorem 1). Denote a minimizing sequence for (R-P) by $(\mu^k_\mathcal{V}, \mu^k_\mathcal{E}) \in \mathcal{P}(\Omega)^\mathcal{V} \times \mathcal{P}(\Omega \times \Omega)^\mathcal{E}$. Since the (RIP) holds for a tree, by [7, Lemma 6.4] there exists for any $k \in \mathbb{N}$ a measure $\mu^k \in \mathcal{P}(\Omega^\mathcal{V})$ such that

$$\pi_u \sharp \mu^k = \mu^k_{\mathcal{V},u} \quad \text{and} \quad \pi_{(u,v)} \sharp \mu^k = \mu^k_{\mathcal{E},(u,v)}.$$

Here $\pi_u : \Omega^\mathcal{V} \to \Omega$ denotes the canonical projection onto the u^{th} component, and $\pi_{(u,v)} : \Omega^\mathcal{V} \to \Omega^2$ denotes the canonical projection onto the u^{th} and v^{th} components. Then μ^k attains the same cost for (GMP) that $\mu^k_\mathcal{V}$ and $\mu^k_\mathcal{E}$ attain for (R-P):

$$\begin{aligned} \langle \mu^k, F \rangle &= \left\langle \mu^k, \sum_{u \in \mathcal{V}} f_u \circ \pi_u + \sum_{(u,v) \in \mathcal{E}} f_{(u,v)} \circ \pi_{(u,v)} \right\rangle \\ &= \sum_{u \in \mathcal{V}} \langle \pi_u \sharp \mu^k, f_u \rangle + \sum_{(u,v) \in \mathcal{E}} \langle \pi_{(u,v)} \sharp \mu^k, f_{(u,v)} \rangle \\ &= \sum_{u \in \mathcal{V}} \langle \mu^k_{\mathcal{V},u}, f_u \rangle + \sum_{(u,v) \in \mathcal{E}} \langle \mu^k_{\mathcal{E},(u,v)}, f_{(u,v)} \rangle. \end{aligned}$$

Together with (3) this implies that $\lim_{k\to\infty} \langle \mu^k, F \rangle = \inf_{\mu \in \mathcal{P}(\Omega^\mathcal{V})} \langle \mu, F \rangle$, and (GMP) = (R-P).

In [2] the authors prove tightness of the local marginal polytope relaxation for distance-like coupling terms on the unit interval by leveraging results about submodular functions from [1]. We construct a simple example that shows that for any graph containing at least one loop, the local marginal polytope relaxation is not tight for spheres under coupling terms penalizing geodesic distance.

Example 1 (Local marginal polytope relaxation is not tight). Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with K vertices, indexed by the integers $0 \dots K-1$. Assume that $\mathcal{E}$ contains all edges $(k, k+1)$ for $k = 0 \dots K-1$. Here, for notational convenience, we have indexed the 0^{th} vertex also by K so as to write the edge $(K-1, K)$ rather than $(K-1, 0)$. Furthermore, consider a set of $2K$ equally spaced points distributed around the unit circle, indexed by integers $0 \dots 2K-1$. These points are ordered around the circle in increasing order such that $d(y_k, y_{k+1}) = \frac{\pi}{K}$. Then $d(y_k, y_{k+K}) = \pi$ and $d(y_k, y_{k+K+1}) = \frac{(K-1)\pi}{K}$. Finally we state the optimization problem, where for every vertex k we optimize a variable $x_k \in \mathcal{S}^1$:

$$\min_{x \in (\mathcal{S}^1)^\mathcal{V}} F_K(x) = \min_{x \in (\mathcal{S}^1)^\mathcal{V}} \sum_{k=0}^{K-1} \iota_{\{y_k, y_{K+k}\}}(x_k) + \sum_{k=0}^{K-1} d(x_k, x_{k+1}). \tag{C.1}$$

Consider the pair (x_k, x_{k+1}). The k^{th} data term enforces that $x_k = y_k$ or $x_k = y_{k+K}$, and similarly for x_{k+1}. Notice that $d(x_k, x_{k+1}) = \frac{1}{K}\pi$ if $(x_k, x_{k+1}) = (y_k, y_{k+1})$ or $(x_k, x_{k+1}) = (y_{k+K}, y_{k+K+1})$, and $d(x_k, x_{k+1}) = \frac{K-1}{K}\pi$ otherwise. We prove that $x^* = (y_0, y_1, \dots, y_{K-1})$ is an optimizer to this problem.

To see this, note that there must be at least one edge (x_k, x_{k+1}) which achieves the cost $d(x_k, x_{k+1}) = \frac{K-1}{K}\pi$. Conversely, assume there exists a solution for which this is not true, and let $x_0 = y_l$ (where $l = 0$ or $l = K$). Since $d(x_0, x_1) = \frac{1}{K}\pi$, we must have $x_1 = y_{l+1}$. Repeating this argument for every edge, we obtain $x_K = y_{l+K}$, which is impossible since $x_0 = x_K$, and $y_l \neq y_{l+K}$. Hence at least one edge achieves cost $d(x_k, x_{k+1}) = \frac{K-1}{K}\pi$, and every other edge achieves a cost $d(x_k, x_{k+1}) \geq \frac{1}{K}\pi$, which means that $F_K(x) \geq 2\frac{K-1}{K}\pi$. This cost is achieved by x^* (as well as $2K$ equivalent solutions obtained by symmetry): $F(x^*) = 2\frac{K-1}{K}\pi$.

Now consider the local marginal polytope relaxation for this problem:

$$\min_{\mu \in \mathcal{M}(\mathcal{S}^1)^{\mathcal{V}}} \mathcal{F}(\mu) = \min_{\mu \in \mathcal{M}(\mathcal{S}^1)^{\mathcal{V}}} \sum_{k=0}^{K-1} \langle \iota_{\{y_k, y_{K+k}\}}, \mu_k \rangle + \sum_{k=0}^{K-1} \mathrm{OT}_d(\mu_k, \mu_{k+1}). \tag{C.2}$$

We can consider the solution μ^* with $\mu_k^* = (\frac{\delta_{y_k} + \delta_{y_{K+k}}}{2})$. This solution turns out to be optimal, but it suffices for this example that it is feasible and achieves lower cost than δ_{x^*}, the Dirac measure centered at x^*. To see this, note that μ_k^* is supported only on $\{y_k, y_{K+k}\}$, hence $\langle \iota_{\{y_k, y_{K+k}\}}, \mu_k^* \rangle = 0$. Furthermore, we have

$$\begin{aligned}
\mathrm{OT}_d(\mu_k^*, \mu_{k+1}^*) &= \mathrm{OT}_d(\tfrac{\delta_{y_k} + \delta_{y_{K+k}}}{2}, \tfrac{\delta_{y_{k+1}} + \delta_{y_{K+k+1}}}{2}) \\
&\leq \tfrac{1}{2}\langle d, \delta_{y_k} \times \delta_{y_{k+1}} + \delta_{y_{K+k}} \times \delta_{y_{K+k+1}} \rangle \\
&= \frac{d(y_k, y_{k+1}) + d(y_{K+k}, y_{K+k+1})}{2} = \tfrac{1}{K}\pi.
\end{aligned}$$

Since this inequality holds for every edge, we have that the cost of the local marginal polytope relaxation is upper bounded by $K \cdot \frac{1}{k}\pi = \pi$ (in fact $\mathcal{F}(\mu^*) = \pi$) which is strictly less than the minimal cost $2\frac{K-1}{K}\pi$ obtained for the unlifted problem. Therefore, the local marginal polytope relaxation is not tight.

It is interesting to note that it is not necessary to resort to extended real-valued functions for this result to hold. With some extra effort the same result can be shown for $f_k(x) = 2\pi|x_1(y_k)_2 - x_2(y_k)_1|$, rather than explicitly constraining x_k via the indicator functions $f_k(x) = \iota_{\{y_k, y_{K+k}\}}(x)$. The proof is obtained by showing that, for any fixed x_{k-1} and x_{k+1}, the optimal value for x_k is either y_k or y_{k+K}. The rest of the proof then follows the proof above. Finally, this example can be generalized to any arbitrary graph that contains at least one loop of size K, by simply setting $f_u = 0$ for all vertices u that are not part of the loop. All this shows the local marginal polytope relaxations is only tight for trees.

Remark 1. Note that the counterexample is essentially a specialization from a $\mathcal{S}^1$-valued MRF-problem problem to a specific discrete MRF-problem. It is well-known that the LMP-relaxation is not tight for discrete MRF-problems in general, and what we considered is one such example.

C.2 Proof of Theorem 2

Proof (Theorem 2). Define $F : \mathcal{C}(\Omega)^{\mathcal{V}} \to \overline{\mathbb{R}}$ and $G : (\mathcal{C}(\Omega)^{\mathcal{E}})^2 \to \overline{\mathbb{R}}$ as

$$F(\alpha) = \sum_{u \in \mathcal{V}} \sigma_{\mathcal{P}(\Omega)}(\alpha_u - f_u), \qquad \text{and} \qquad G(\varphi, \psi) = \sum_{e \in \mathcal{E}} \iota_{\mathcal{K}_e}(\varphi_e, \psi_e). \tag{C.3}$$

Then the problem (R-D) can be rewritten as

$$\inf_{(\varphi,\psi) \in (\mathcal{C}(\Omega)^{\mathcal{E}})^2} F(A(\varphi, \psi)) + G(\varphi, \psi).$$

By definition we have $(\iota_{\mathcal{P}(\Omega)} + \langle f_u, \cdot \rangle)^* = \sigma_{\mathcal{P}(\Omega)}(\cdot - f_u)$ which is proper, convex and lsc. In view of closedness of $\mathcal{P}(\Omega)$ in the weak* topology and [14, Lemma 1.6] we have that $\iota_{\mathcal{P}(\Omega)} + \langle f_u, \cdot \rangle$ is proper, convex and lsc as well implying that $(\sigma_{\mathcal{P}(\Omega)}(\cdot - f_u))^* = \iota_{\mathcal{P}(\Omega)} + \langle f_u, \cdot \rangle$. Therefore $F^* : \mathcal{M}(\Omega)^{\mathcal{V}} \to \overline{\mathbb{R}}$ with $F^*(\mu) = \sum_{u \in \mathcal{V}} \iota_{\mathcal{P}(\Omega)}(\mu) + \langle f_u, \mu_u \rangle$. We have $G^* : (\mathcal{M}(\Omega)^{\mathcal{E}})^2 \to \overline{\mathbb{R}}$ with

$$G^*(\mu, \nu) = \sum_{e \in \mathcal{E}} \sigma_{\mathcal{K}_e}(\mu, \nu) = \sup_{(\varphi,\psi) \in (\mathcal{C}(\Omega)^{\mathcal{E}})^2} \langle \varphi, \mu \rangle + \langle \psi, \nu \rangle - \sum_{e \in \mathcal{E}} \iota_{\mathcal{K}_e}(\varphi_e, \psi_e).$$

Then the dual problem amounts to

$$\sup_{\mu \in \mathcal{P}(\Omega)^{\mathcal{V}}} -G^*(-A^*(\mu)) - \sum_{u \in \mathcal{V}} \langle f_u, \mu_u \rangle, \tag{C.4}$$

for

$$\begin{aligned}
G^*(-A^*(\mu)) &= \sup_{\varphi,\psi \in \mathcal{C}(\Omega)^{\mathcal{E}}} -\langle (\varphi, \psi), A^*(\mu) \rangle - \sum_{e \in \mathcal{E}} \iota_{\mathcal{K}_e}(\varphi_e, \psi_e) \\
&= \sup_{\varphi,\psi \in \mathcal{C}(\Omega)^{\mathcal{E}}} -\langle A(\varphi, \psi), \mu \rangle - \sum_{e \in \mathcal{E}} \iota_{\mathcal{K}_e}(\varphi_e, \psi_e) \\
&= \sup_{\varphi,\psi \in \mathcal{C}(\Omega)^{\mathcal{E}}} \sum_{u \in \mathcal{V}} \langle -A_u(\varphi, \psi), \mu_u \rangle - \sum_{e \in \mathcal{E}} \iota_{\mathcal{K}_e}(\varphi_e, \psi_e) \\
&= \sup_{\varphi,\psi \in \mathcal{C}(\Omega)^{\mathcal{E}}} \sum_{u \in \mathcal{V}} \sum_{v:(u,v) \in \mathcal{E}} \langle \varphi_{(u,v)}, \mu_u \rangle + \sum_{v \in \mathcal{V}} \sum_{u:(u,v) \in \mathcal{E}} \langle \psi_{(u,v)}, \mu_v \rangle - \sum_{e \in \mathcal{E}} \iota_{\mathcal{K}_e}(\varphi_e, \psi_e) \\
&= \sup_{\varphi,\psi \in \mathcal{C}(\Omega)^{\mathcal{E}}} \sum_{(u,v) \in \mathcal{E}} \langle \varphi_{(u,v)}, \mu_u \rangle + \langle \psi_{(u,v)}, \mu_v \rangle - \iota_{\mathcal{K}_{(u,v)}}(\varphi_{(u,v)}, \psi_{(u,v)}) \\
&= \sum_{(u,v) \in \mathcal{E}} \sup_{\varphi,\psi \in \mathcal{C}(\Omega)} \langle \varphi, \mu_u \rangle + \langle \psi, \mu_v \rangle - \iota_{\mathcal{K}_{(u,v)}}(\varphi, \psi) \\
&= \sum_{(u,v) \in \mathcal{E}} \mathrm{OT}_{f_{(u,v)}}(\mu_u, \mu_v),
\end{aligned}$$

where the last equality follows from [14, Theorem 1.42] and the fact that Ω is compact and $f_{(u,v)}$ lsc.

Next we are going to verify that $\inf_{\varphi,\psi\in\mathcal{C}(\Omega)^{\mathcal{E}}} F(A(\varphi,\psi)) + G(\varphi,\psi)$ satisfies the conditions in [12, Theorem 1]: We have already seen that F is proper, convex and lsc since it is the convex conjugate of a proper, lsc convex function.

Consider a sequence $(\varphi_e^k, \psi_e^k) \in \mathcal{K}_e$ that converges uniformly to (φ_e, ψ_e), i.e., $(\varphi_e^k, \psi_e^k) \to (\varphi_e, \psi_e)$ for $k \to \infty$. Then we have

$$-\varphi_e^k(x) - \psi_e^k(y) + f_e(x,y) \geq 0 \quad \forall x, y \in \Omega \times \Omega,$$

for any $k \in \mathbb{N}$. Let $x, y \in \Omega \times \Omega$ arbitrary but fixed. Adding $\varphi_e(x) + \psi_e(y)$ to both sides of the inequality this implies that

$$\varphi_e(x) - \varphi_e^k(x) + \psi_e(y) - \psi_e^k(y) + f_e(x,y) \geq \varphi_e(x) + \psi_e(y).$$

Passing $k \to \infty$ we obtain since uniform convergence implies pointwise convergence that $\varphi_e(x) - \varphi_e^k(x) \to 0$ and $\psi_e(y) - \psi_e^k(y) \to 0$. This implies that

$$f_e(x,y) \geq \varphi_e(x) + \psi_e(y). \tag{C.5}$$

Since $x, y \in \Omega$ are arbitrary this implies that $(\varphi_e, \psi_e) \in \mathcal{K}_e$ and thus $\mathcal{K}_e$ is closed implying that G is lsc. Clearly G is also convex and proper.

Note that A is linear and continuous as a weighted sum of $(\varphi, \psi) \in \mathcal{C}(\Omega)^{\mathcal{E}} \times \mathcal{C}(\Omega)^{\mathcal{E}}$. Without loss of generality assume that $f_e(x,y) \geq 0$. Choose $\varphi_e = \psi_e \equiv 0$. Then $\varphi_e(x) + \psi_e(y) = 0 \leq f_e(x,y)$ for all $e \in \mathcal{E}$, $(x,y) \in \Omega^2$ and $G(\varphi,\psi) < \infty$. In addition we have that $A_u(\varphi,\psi) \equiv 0$. Note that $F(0) = -\sum_{u\in\mathcal{V}} \inf_{\mu\in\mathcal{P}(\Omega)} \int_\Omega f_u \,\mathrm{d}\mu = -\sum_{u\in\mathcal{V}} \min_{x\in\Omega} f_u(x) < \infty$. Clearly, $\sigma_{\mathcal{P}(\Omega)}(\cdot - f_u)$ is continuous in the sup-norm topology at 0 for any $u \in \mathcal{V}$. Since F is a separable finite sum it is continuous at 0 as well.

By [12, Theorem 1] we obtain that $\inf_{(\varphi,\psi)\in(\mathcal{C}(\Omega)^{\mathcal{E}})^2} F(A(\varphi,\psi)) + G(\varphi,\psi)$ is stably set. Thus we can invoke [12, Theorem 3] and obtain that

$$\inf_{(\varphi,\psi)\in(\mathcal{C}(\Omega)^{\mathcal{E}})^2} F(A(\varphi,\psi)) + G(\varphi,\psi) = \sup_{\mu\in\mathcal{P}(\Omega)^{\mathcal{V}}} -G^*(-A^*(\mu)) - \sum_{u\in\mathcal{V}} \langle f_u, \mu_u \rangle$$

and there exists $\nu \in \mathcal{P}(\Omega)^{\mathcal{V}}$ such that

$$\sup_{\mu\in\mathcal{P}(\Omega)^{\mathcal{V}}} -G^*(-A^*(\mu)) - \sum_{u\in\mathcal{V}} \langle f_u, \mu_u \rangle = -G^*(-A^*(\nu)) - \sum_{u\in\mathcal{V}} \langle f_u, \nu_u \rangle.$$

C.3 Proof of Theorem 3

Proof (Theorem 3). Assume that $f_e : \Omega \times \Omega \to \mathbb{R}$ are continuous with modulus ω_e for all $e \in \mathcal{E}$. Let $(\varphi, \psi) \in \mathcal{C}(\Omega)^{\mathcal{E}} \times \mathcal{C}(\Omega)^{\mathcal{E}}$ satisfying $(\varphi_e, \psi_e) \in \mathcal{K}_e$. Consider the score

$$-F(A(\varphi,\psi)) = \min_{\mu\in\mathcal{P}(\Omega)^{\mathcal{V}}} \sum_{u\in\mathcal{V}} \int_\Omega (f_u - A_u(\varphi,\psi)) \,\mathrm{d}\mu_u \tag{C.6}$$

$$= \min_{\mu\in\mathcal{P}(\Omega)^{\mathcal{V}}} \sum_{u\in\mathcal{V}} \langle f_u, \mu_u \rangle + \sum_{(u,v)\in\mathcal{E}} \langle \varphi_{(u,v)}, \mu_u \rangle + \langle \psi_{(u,v)}, \mu_v \rangle. \tag{C.7}$$

Since $(\varphi_e, \psi_e) \in \mathcal{K}_e$ we have that $\psi_e(y) \le f_e(x,y) - \varphi_e(x)$ for all $(x,y) \in \Omega^2$ implying that

$$\psi_e(y) \le \inf_{x \in \Omega} f_e(x,y) - \varphi_e(x) = (\varphi_e)^{f_e}(y), \tag{C.8}$$

and we have $\varphi_e(x) + (\varphi_e)^{f_e}(y) \le f_e(x,y)$ for all $(x,y) \in \Omega^2$. Similarly we obtain that

$$\varphi_e(x) \le \inf_{y \in \Omega} f_e(x,y) - (\varphi_e)^{f_e}(y) = (\varphi_e)^{f_e \bar{f}_e}(x), \tag{C.9}$$

and

$$(\varphi_e)^{f_e \bar{f}_e}(x) + (\varphi_e)^{f_e}(y) \le f_e(x,y).$$

Since by assumption f_e is continuous with modulus ω_e, both $(\varphi_e)^{f_e \bar{f}_e}, (\varphi_e)^{f_e}$ inherit the same modulus [14, Box 1.8] and are in particular continuous. Overall this means that $((\varphi_e)^{f_e \bar{f}_e}, (\varphi_e)^{f_e}) \in \mathcal{K}_e$ is still feasible. In addition the inequalities (C.9) and (C.8) imply that for every $\mu \in \mathcal{P}(\Omega)$ we have that $\int_\Omega \varphi_e \,\mathrm{d}\mu \le \int_\Omega (\varphi_e)^{f_e \bar{f}_e} \,\mathrm{d}\mu$ and $\int_\Omega \psi_e \,\mathrm{d}\mu \le \int_\Omega (\varphi_e)^{f_e} \,\mathrm{d}\mu$. Since $-A_u(\varphi, \psi)$ is a conic combination of multipliers φ_e and ψ_e the infimum in $-F(A(\varphi,\psi))$ increases if anything when φ_e is replaced with $(\varphi_e)^{f_e \bar{f}_e}$ and ψ_e is replaced with $(\varphi_e)^{f_e}$.

For each $(\varphi_e, \psi_e) \in \mathcal{K}_e$ we can add and subtract a constant c_e such that $(\varphi_e - c_e, \psi_e + c_e)$ achieves the same score (C.7). Choose $c_e = \min_{x \in \Omega} \varphi_e(x)$ then $\varphi_e(x) - c_e \in [0, \omega_e(\operatorname{diam} \Omega)]$ and $\psi_e(x) + c_e \in [\min f_e - \omega_e(\operatorname{diam} \Omega), \max f_e]$, i.e., $\varphi_e - c_e, \psi_e + c_e$ are both uniformly bounded.

Now consider a maximizing sequence (φ^k, ψ^k) of (R-D). In particular this means $(\varphi_e^k, \psi_e^k) \in \mathcal{K}_e$. Then (φ_e^k, ψ_e^k) can be replaced by $((\varphi_e^k)^{f_e \bar{f}_e} - c_e^k, (\varphi_e^k)^{f_e} + c_e^k)$ for $c_e^k := \min_{x \in \Omega} (\varphi_e^k)^{f_e \bar{f}_e}(x)$ which is still a maximizing sequence whose elements are uniformly bounded and equicontinuous. By the Arzelà-Ascoli Theorem [13, Theorem 11.28] there exists a uniformly converging subsequence. Since the objective (and the constraints) are upper semicontinuous the limit must be a maximizer.

C.4 Proof of Theorem 4

Proof (Theorem 4). For proving the claim we resort to an unconstrained version of (R-D$^{\mathrm{n}}$) obtained via the change of variable $\varphi_e = \gamma_e + \xi_e$ for $\xi_e \in \mathbb{R}$ for each edge $e \in \mathcal{E}$. Then

$$\begin{aligned} -A_u(\varphi, \psi) &= \sum_{v:(u,v) \in \mathcal{E}} \gamma_{(u,v)} + \sum_{v:(v,u) \in \mathcal{E}} \psi_{(v,u)} + \sum_{v:(u,v) \in \mathcal{E}} \xi_{(u,v)} \\ &= -A_u(\gamma, \psi) + \sum_{v:(u,v) \in \mathcal{E}} \xi_{(u,v)} \end{aligned}$$

and (R-D$^{\mathrm{n}}$) becomes

$$\begin{aligned}\sup_{\substack{\gamma,\psi\in\mathcal{C}(\Omega)^{\mathcal{E}}\\ \xi\in\mathbb{R}^{\mathcal{E}}}} \quad & \sum_{u\in\mathcal{V}} \min_{x\in\Omega} f_u(x) - A_u(\gamma,\psi)(x) + \sum_{e\in\mathcal{E}} \xi_e \\ \text{subject to} \quad & f_e(x,y) - \gamma_e(x) - \psi_e(y) \geq \xi_e \quad \forall x,y\in\Omega.\end{aligned}$$

Interpreting ξ_e in terms of a slack variable the problem can be written equivalently in terms of an unconstrained one

$$\sup_{\gamma,\psi\in\mathcal{C}(\Omega)^{\mathcal{E}}} \sum_{u\in\mathcal{V}} \min_{x\in\Omega} f_u(x) - A_u(\gamma,\psi)(x) + \sum_{e\in\mathcal{E}} \min_{x,y\in\Omega} f_e(x,y) - \gamma_e(x) - \psi_e(y), \tag{C.10}$$

and in particular (R-D) = (C.10). Restricting $\gamma_e, \psi_e \in \mathbb{R}_n[x]$ to the subspace of polynomials with degree n we consider

$$\sup_{\gamma,\psi\in(\mathbb{R}_n[x])^{\mathcal{E}}} \sum_{u\in\mathcal{V}} \min_{x\in\Omega} f_u(x) - A_u(\gamma,\psi)(x) + \sum_{e\in\mathcal{E}} \min_{x,y\in\Omega} f_e(x,y) - \gamma_e(x) - \psi_e(y). \tag{C.11}$$

First note that (C.10) − (C.11) ≥ 0. Let (α,β) be the maximizer of (C.10) which exists thanks to Theorem 3 and the equality (R-D) = (C.10). Let $(\gamma,\psi) \in \mathcal{C}(\Omega)^{\mathcal{E}} \times \mathcal{C}(\Omega)^{\mathcal{E}}$. Then we have for every $x \in \Omega$:

$$\begin{aligned} f_u(x) - A_u(\alpha,\beta)(x) &= f_u(x) - A_u(\gamma,\psi)(x) + A_u(\gamma,\psi)(x) - A_u(\alpha,\beta)(x) \\ &\leq f_u(x) - A_u(\gamma,\psi)(x) + \|A_u(\gamma,\psi) - A_u(\alpha,\beta)\|_\infty. \end{aligned}$$

Minimizing both sides wrt $x \in \Omega$ and summing over $u \in \mathcal{V}$ this implies that

$$\begin{aligned} \sum_{u\in\mathcal{V}} \min_{x\in\Omega}(f_u - A_u(\alpha,\beta))(x) - \min_{x\in\Omega}(f_u - A_u(\gamma,\psi))(x) \\ \leq \sum_{u\in\mathcal{V}} \|A_u(\gamma,\psi) - A_u(\alpha,\beta)\|_\infty. \end{aligned} \tag{C.12}$$

Furthermore we have

$$\begin{aligned} &\sum_{u\in\mathcal{V}} \|A_u(\gamma,\psi) - A_u(\alpha,\beta)\|_\infty \\ &\quad = \sum_{u\in\mathcal{V}} \max_{x\in\Omega} \left| \sum_{v:(u,v)\in\mathcal{E}} (\alpha_{(u,v)} - \gamma_{(u,v)})(x) + (\beta_{(v,u)} - \psi_{(v,u)})(x) \right| \\ &\quad \leq \sum_{u\in\mathcal{V}} \max_{x\in\Omega} \sum_{v:(u,v)\in\mathcal{E}} |(\alpha_{(u,v)} - \gamma_{(u,v)})(x)| \\ &\qquad\quad + \sum_{v:(u,v)\in\mathcal{E}} |(\beta_{(v,u)} - \psi_{(v,u)})(x)| \\ &\quad \leq \sum_{e\in\mathcal{E}} \|\alpha_e - \gamma_e\|_\infty + \|\beta_e - \psi_e\|_\infty. \end{aligned}$$

We also have for every $x, y \in \Omega$

$$\begin{aligned} &f_e(x,y) - \alpha_e(x) - \beta_e(y) \\ &\qquad = f_e(x,y) - \gamma_e(x) - \psi_e(y) + \gamma_e(x) - \alpha_e(x) + \psi_e(y) - \beta_e(y) \\ &\qquad \le f_e(x,y) - \gamma_e(x) - \psi_e(y) + \|\gamma_e - \alpha_e\|_\infty + \|\psi_e - \beta_e\|_\infty. \end{aligned}$$

Minimizing both sides wrt $x, y \in \Omega$ and summing over $e \in \mathcal{E}$ this implies that

$$\begin{aligned} \sum_{e\in\mathcal{E}} \min_{x,y\in\Omega} f_e(x,y) - \alpha_e(x) - \beta_e(y) - \min_{x,y\in\Omega} f_e(x,y) - \gamma_e(x) - \psi_e(y) \\ \le \sum_{e\in\mathcal{E}} \|\gamma_e - \alpha_e\|_\infty + \|\psi_e - \beta_e\|_\infty. \end{aligned} \tag{C.13}$$

Now choose $\varepsilon > 0$. Since Ω is compact and $\mathbb{R}[x]$ is a subalgebra of $\mathcal{C}(\Omega)$ by the Stone–Weierstrass theorem we can find polynomials $\gamma_e, \psi_e \in \mathbb{R}[x]$ such that $\|\gamma_e - \alpha_e\|_\infty \le \frac{\varepsilon}{4|\mathcal{E}|}$ and $\|\psi_e - \beta_e\|_\infty \le \frac{\varepsilon}{4|\mathcal{E}|}$. Let $n = \max_{e\in\mathcal{E}} \max\{\deg(\gamma_e), \deg(\psi_e)\}$. Summing (C.12) and (C.13) since (α, β) is a maximizer of (C.10) we have

$$\begin{aligned} (\text{C.10}) - (\text{C.11}) &\le 2 \sum_{e\in\mathcal{E}} \|\gamma_e - \alpha_e\|_\infty + \|\psi_e - \beta_e\|_\infty \\ &\le 2 \sum_{e\in\mathcal{E}} \frac{\varepsilon}{2|\mathcal{E}|} = \varepsilon. \end{aligned}$$

Thanks to Theorem 2 we have (C.10) = (R-D) = (R-P). Since (C.11) = (R-D$^{\text{n}}$) this means that for any $\varepsilon > 0$ we can find $n \in \mathbb{N}$ sufficiently large so that

$$0 \le (\text{R-P}) - (\text{R-D}^{\text{n}}) \le \varepsilon.$$

If, in addition, $(\mathcal{V}, \mathcal{E})$ is a tree we have in light of Theorem 1 that (R-P) = (P).

C.5 Proof of Corollary 2

Proof (Corollary 2). In view of the proof of the previous theorem we consider the equivalent unconstrained problem (C.10) = (R-D). Let (α, β) be the maximizer of (C.10). As shown in the proof of the previous theorem the duality gap can be bounded as

$$(\text{C.10}) - (\text{C.11}) \le 2 \sum_{e\in\mathcal{E}} \|\gamma_e - \alpha_e\|_\infty + \|\psi_e - \beta_e\|_\infty.$$

Thanks to Theorem 3 $\alpha_e, \beta_e \in \mathcal{C}(\Omega)$ inherit the modulus of continuity ω_e. Now fix $n \in \mathbb{N}$. Thanks to [9, Theorem 2] there exist polynomials ρ_e and λ_e with maximum degree n such that $\|\rho_e - \alpha_e\|_\infty \le C_m \omega_e(1/n)$ and $\|\lambda_e - \beta_e\|_\infty \le C_m \omega_e(1/n)$. Thus

$$(\text{C.10}) - (\text{C.11}) \le 4C_m \sum_{e\in\mathcal{E}} \omega_e(1/n).$$

If, in addition, $(\mathcal{V}, \mathcal{E})$ is a tree we have in light of Theorem 1 that (R-P) = (P).

C.6 Proof of Theorem 6

For any $m \in \mathbb{N}$, there exists a measure μ on $\mathcal{S}^m$ that is invariant to the left group action of $\mathcal{SO}(m+1)$, where $\mathcal{SO}(m+1)$ is the special orthogonal group in dimension $m+1$ which can be identified with the space of rotation matrices in $\mathbb{R}^{(m+1)\times(m+1)}$. This defining property of the $\mathcal{SO}(m+1)$-invariant measure μ is written mathematically as follows:

$$\mu(B) = \mu(r^{-1}B), \quad \forall B \in \mathcal{B}(\mathcal{S}^m),\ r \in \mathcal{SO}(m+1).$$

In [9] it is proved that for all $n \in \mathbb{N}$: there exists a $\kappa_n \in \mathbb{R}_n[x]$ such that $\kappa_n(\tau)$ is nonnegative for $\tau \in [-1,+1]$, and for any $x \in \mathcal{S}^m$,

$$\int_{\mathcal{S}^m} \kappa_n(\langle x, y\rangle)\, \mathrm{d}\mu(y) = 1. \tag{C.14}$$

Then an approximant φ_n of φ can be computed as

$$\varphi_n(x) = [K_n\varphi](x) := \int_{\mathcal{S}^m} \varphi(y)\kappa_n(\langle x, y\rangle)\, \mathrm{d}\mu(y). \tag{C.15}$$

It can be shown [11, Theorem 3.3] that for all $m \in \mathbb{N}$ there exists a $C_m \in \mathbb{R}$ such that

$$\|\varphi - K_n\varphi\|_\infty \leq C_m \tfrac{L}{n}. \tag{C.16}$$

Before proving that $K_n\varphi$ is Lipschitz continuous with $\operatorname{Lip}\varphi = L$, we require a lemma:

Lemma 1. *For any $x, x' \in \mathcal{S}^m$ there exists $r \in \mathcal{SO}(m+1)$ such that $x' = rx$ and $\min_{y\in\mathcal{S}^m} \langle y, ry\rangle = \langle x, x'\rangle$.*

Proof. We will identify any element of $y \in \mathcal{S}^m$ with points in $y \in \mathbb{R}^{m+1}$ such that $\|y\| = 1$, and $r \in \mathcal{SO}(m+1)$ with a rotation matrix $R \in \mathbb{R}^{(m+1)\times(m+1)}$. Assume x and x' are not linearly dependent, and let $U_S = [x, \frac{1}{\sqrt{1-\langle x,x'\rangle^2}}x' - \frac{\langle x,x'\rangle}{\sqrt{1-\langle x,x'\rangle^2}}x]$, which has orthonormal columns that span $S = \operatorname{rge}([x, x'])$. Furthermore, let $U_{S^\perp}$ be any matrix such that its columns are an orthonormal basis for $S^\perp$. One can define

$$R = \begin{bmatrix}U_{S^\perp} & U_S\end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & \langle x, x'\rangle & -\sqrt{1-\langle x, x'\rangle} \\ 0 & \sqrt{1-\langle x, x'\rangle} & \langle x, x'\rangle \end{bmatrix} \begin{bmatrix}U_{S^\perp} & U_S\end{bmatrix}^\mathsf{T} =: U\tilde{R}U^\mathsf{T}$$

and readily check that $Rx = x'$. Furthermore, R is unitarily similar to the rotation matrix $\tilde{R}$, hence R is itself a rotation matrix. Constructing $\frac{R+R^\mathsf{T}}{2}$, we have

$$\tfrac{R+R^\mathsf{T}}{2} = \begin{bmatrix}U_{S^\perp} & U_S\end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & \langle x, x'\rangle & 0 \\ 0 & 0 & \langle x, x'\rangle \end{bmatrix} \begin{bmatrix}U_{S^\perp} & U_S\end{bmatrix}^\mathsf{T}$$

which is an eigendecomposition of $\frac{R+R^\mathsf{T}}{2}$. Since for all $x, x' \in \mathcal{S}^m$ we have $\langle x, x' \rangle \leq 1$, the smallest eigenvalue of $\frac{R+R^\mathsf{T}}{2}$ is $\lambda_{\min}(\frac{R+R^\mathsf{T}}{2}) = \langle x, x' \rangle$. Finally, we have the Rayleigh quotient

$$\min_{y \in \mathcal{S}^m} \langle y, Ry \rangle = \min_{y \in \mathcal{S}^m} \langle y, \tfrac{R+R^\mathsf{T}}{2} y \rangle = \lambda_{\min}(\tfrac{R+R^\mathsf{T}}{2}) = \langle x, x' \rangle.$$

If x and x' are linearly dependent, the argument holds via a limiting process by replacing x' by $\sqrt{1-\varepsilon^2}x' + \varepsilon r$ where $r \in \mathcal{S}^m$ can be chosen random, when $\varepsilon \to 0$. In particular, when $x = x'$, we get $R = I$ and we have $\langle y, Ry \rangle = \|y\|^2 = 1 = \langle x, x' \rangle$ for all $y \in \mathcal{SO}(m)$. If $x = -x'$, then R can be any rotation such that $Rx = x'$ (depending on the choice of r), but since $Ry \in \mathcal{S}^m$ for any $y \in \mathcal{S}^m$, we have by Cauchy-Schwartz $\langle y, Ry \rangle \geq -\|y\|\|Ry\| \geq -1 = \langle x, x' \rangle$ with equality for $y = x$.

We prove that this approximant is Lipschitz continuous on $\mathcal{S}^m$:

Proof (Theorem 6). Denoting the action of $r \in \mathcal{SO}(m+1)$ on a function $\varphi : \mathcal{S}^m \to \mathbb{R}$ by $r \cdot \varphi(x) := \varphi(r^{-1}x)$, [11, Proposition 3.1] shows that K_n is an $\mathcal{SO}(m+1)$-invariant operator:

$$r \cdot K_n \varphi = K_n(r \cdot \varphi), \quad \forall r \in \mathcal{SO}(m+1).$$

For any $x, x' \in \mathcal{S}^m$, choose $r \in \mathcal{SO}(m+1)$ as in Lemma 1, for which we have $rx = x'$ and that

$$\min_{z \in \mathcal{S}^m} z^\mathsf{T} r z = \langle x, x' \rangle. \tag{C.17}$$

Then

$$\varphi_n(x') = [K_n \varphi](x') = [K_n \varphi](r^{-1}x) = r \cdot [K_n \varphi](x) = [K_n(r \cdot \varphi)](x),$$

and $\varphi_n(x) = [K_n \varphi](x)$. By linearity of K_n, we then obtain

$$\varphi_n(x') - \varphi_n(x) = [K_n(r \cdot \varphi - \varphi)](x).$$

Writing K_n out in full, we obtain

$$\begin{aligned}
|\varphi_n(x') - \varphi_n(x)| &= \Big| \int_{\mathcal{S}^m} \varphi(r^{-1}y) - \varphi(y) \kappa_n(\langle x, y \rangle) \,\mathrm{d}\mu(y) \Big| \\
&\leq \int_{\mathcal{S}^m} |\varphi(r^{-1}y) - \varphi(y)| \kappa_n(\langle x, y \rangle) \,\mathrm{d}\mu(y) \\
&\leq L \int_{\mathcal{S}^m} d(r^{-1}y, y) \kappa_n(\langle x, y \rangle) \,\mathrm{d}\mu(y) \\
&\leq L \|d(r^{-1}\cdot, \cdot)\|_\infty \int_{\mathcal{S}^m} \kappa_n(\langle x, y \rangle) \,\mathrm{d}\mu(y) \\
&= L \|d(r^{-1}\cdot, \cdot)\|_\infty.
\end{aligned} \tag{C.18}$$

The first inequality follows since $\kappa_n(\tau) \geq 0$, for any $\tau \in [-1, +1]$, the second inequality follows from the Lipschitz continuity of φ. The final equality follows from (C.14).

The result is proved if $\|d(r^{-1}\cdot,\cdot)\|_\infty = d(x,x')$. For any m, we have that $d(x,x') = \arccos(\langle x,x'\rangle)$ for $x,x' \in \mathcal{S}^m$. Hence $d(r^{-1}y,y) = \arccos(y^\mathsf{T} r^{-\mathsf{T}} y)$. Since arccos is a monotonically decreasing function, we have

$$\max_{y\in\mathcal{S}^m} \left\{ \arccos(y^\mathsf{T} r^{-\mathsf{T}} y)\right\} = \arccos\Big(\min_{y\in\mathcal{S}^m} \left\{y^\mathsf{T} r^{-\mathsf{T}} y\right\}\Big) = \arccos\Big(\min_{z\in\mathcal{S}^m} \left\{z^\mathsf{T} r z\right\}\Big), \tag{C.19}$$

where the last equality follows from the change of variables $y = rz$. We may write min and max as we are optimizing a bounded function on a compact set. By (C.17) together with (C.19) this yields

$$\|d(r^{-1}\cdot,\cdot)\|_\infty = \max_{y\in\mathcal{S}^m} \left\{ \arccos(y^\mathsf{T} r^{-\mathsf{T}} y)\right\} = \arccos\big(\langle x,x'\rangle\big) = d(x,x').$$

and using (C.18) we get

$$|\varphi_n(x') - \varphi_n(x)| \le L d(x,x')$$

C.7 Proof of Theorem 5

Proof (Theorem 5). There exists a maximizer of (mR-D) by Corollary 1 which we denote by $\psi \in \mathrm{Lip}_d(\Omega)^{\mathcal{E}}$. Then for any φ, we have

$$\begin{aligned}
-\sigma_{\mathcal{P}(\Omega)}\big[\mathrm{Div}_u(\psi) - f_u\big] &= \min_{x\in\Omega} \big[f_u - \mathrm{Div}_u(\psi)\big](x)\\
&= \min_{x\in\Omega} f_u(x) - \mathrm{Div}_u\,\varphi(x) - \mathrm{Div}_u\,\psi(x) + \mathrm{Div}_u\,\varphi(x)\\
&\le \min_{x\in\Omega} f_u(x) - \mathrm{Div}_u\,\varphi(x) + \|\mathrm{Div}_u(\varphi-\psi)\|_\infty\\
&\le \sigma_{\mathcal{P}(\Omega)}\big[\mathrm{Div}_u(\varphi) - f_u\big] + \|\mathrm{Div}_u(\varphi-\psi)\|_\infty
\end{aligned}$$

Since we are maximizing over a strictly smaller set in (mR-D$^\mathrm{n}$) as compared to (mR-D), clearly (mR-D) $\ge$ (mR-D$^\mathrm{n}$). On the other hand we can now bound

$$\begin{aligned}
(\text{mR-D}) - (\text{mR-D}^\mathrm{n}) &\le \sum_{u\in\mathcal{V}} \|\mathrm{Div}_u(\varphi-\psi)\|_\infty\\
&\le \sum_{u\in\mathcal{V}} d_u \sup_{e\in\mathcal{E}} \|\varphi_e - \psi_e\|_\infty\\
&\le 2|\mathcal{E}| \sup_{e\in\mathcal{E}} \|\varphi_e - \psi_e\|_\infty
\end{aligned} \tag{C.20}$$

with d_u denoting the degree of the vertex u.

By Theorem 6 and (C.16) there exists a sequence of Lipschitz continuous polynomials φ_n indexed by degree n such that $\|\varphi - \varphi_n\|_\infty \le C_m \frac{L}{n}$, with $\mathrm{Lip}\,\varphi_n = \mathrm{Lip}\,\varphi$, for all $n \in \mathbb{N}$. Now to obtain a lower bound of (mR-D$^\mathrm{n}$), choose for every edge $e \in \mathcal{E}$ that φ_e is such a polynomial $\varphi_e = (\psi_e)_n$ in the sequence approximating ψ_e. We obtain that

$$\|\varphi_e - \psi_e\|_\infty \le C_m \tfrac{L}{n},$$

which together with (C.20) yields

$$(\text{mR-D}) - (\text{mR-D}^\mathrm{n}) \le 2|\mathcal{E}| C_m \tfrac{L}{n}.$$

C.8 Proof of Theorem 7

Proof (Theorem 7). It can easily be verified that both (R-D$^{\mathrm{n}}$) and (mR-D$^{\mathrm{n}}$) are invariant to the constant part of the dual multiplier γ and ψ (of φ), so it is natural to further restrict the optimization to linear functionals; i.e. when the constant part(s) of γ and ψ (or φ) equal(s) zero. Hence we restrict the optimization to linear functionals $\gamma, \psi \in \mathbb{E}^*$, define F, G as

$$F(x) = \sum_{u \in \mathcal{V}} f_u(x_u) \quad \text{and} \quad G(x, y) = \sum_{e \in \mathcal{E}} f_e(x_e, y_e)$$

and let A be defined as in (6). We have that

$$-F^*\big(A(\gamma, \psi)\big) = \sum_{u \in \mathcal{V}} -f_u^*\big(A_u(\gamma, \psi)\big) = \sum_{u \in \mathcal{V}} \min_{x \in \mathbb{E}^*} f_u(x) - \langle A_u(\gamma, \psi), x \rangle.$$

The first equality follows from the properties of the Fenchel conjugate for separable F. Similarly for G^*, we obtain

$$-G^*\big(\gamma, \psi\big) = \sum_{e \in \mathcal{E}} -f_e^*(\gamma_e, \psi_e) = \sum_{e \in \mathcal{E}} \min_{x, y \in \mathbb{E}^*} f_e(x, y) - \langle \gamma_e, x \rangle - \langle \psi_e, y \rangle$$

As in the proof for Theorem 4, we may rewrite (R-D$^{\mathrm{n}}$) as (C.11), where, in light of the above derivation, restriction to linear functionals yields

$$\begin{aligned}(\text{R-D}^1) &= \sup_{\gamma, \psi \in (\mathbb{E}^*)^{\mathcal{E}}} F^*\big(A(\gamma, \psi)\big) + G^*\big(\gamma, \psi\big) = \inf_{x \in (\mathbb{E}^*)^{\mathcal{V}}} F^{**}(x) + G^{**}(-A^* x) \\ &= \inf_{x \in (\mathbb{E}^*)^{\mathcal{V}}} \sum_{u \in \mathcal{V}} f_u^{**}(x) + \sum_{(u,v) \in \mathcal{E}} f_{(u,v)}^{**}(x_u, x_v)\end{aligned}$$

where the first equality is an application of Fenchel–Rockafellar duality, and the second is a straightforward rewriting using the definition of the operator A.

In the case of (mR-D$^{\mathrm{n}}$) when $\Omega = \mathcal{S}^m$, we have that any linear functional $\varphi \in \mathbb{E}^*$ which is Lipschitz continuous with $\mathrm{Lip}_{(\mathbb{E}, d)}\, \varphi = 1$ on the embedding space is necessarily Lipschitz continuous with $\mathrm{Lip}_{(\mathcal{S}^m, g)}\, \varphi = 1$ on $\mathcal{S}^m$, since $1 \geq \|\nabla \varphi\| \geq \|\nabla_\Omega \varphi\|$. This can be easily verified by noting that $\nabla_\Omega \varphi$ is an orthogonal projection of $\nabla \varphi$, and an orthogonal projection is a contraction mapping. On the other hand, for any linear functional φ, we have that for all $x \in \mathcal{S}^m$ such that $\langle \varphi, x \rangle = 0$ the Euclidean and Riemannian gradients agree: $\nabla_\Omega \varphi(x) = \nabla \varphi(x)$. Therefore, for any linear functional $\varphi \in \mathbb{E}^*$ we have $\iota_{\mathrm{Lip}(\mathcal{S}^m, g)}(\varphi_e) = \iota_{\mathrm{Lip}(\mathbb{E}, \|\cdot\|)}(\varphi_e)$. Defining F as before, we have

$$-F^*(\mathrm{Div}\, \varphi) = -\sum_{u \in \mathcal{V}} f_u^*(\mathrm{Div}_u\, \varphi) = \sum_{u \in \mathcal{V}} -\sigma_{\mathcal{P}(\Omega)}(\mathrm{Div}_u\, \varphi - f_u)$$

For the metric case, we have that

$$-G^*(\varphi) = \sum_{e \in \mathcal{E}} \iota_{\mathrm{Lip}(\mathcal{S}^m, g)}(\varphi_e) = \sum_{e \in \mathcal{E}} \iota_{\mathrm{Lip}(\mathbb{E}, \|\cdot\|)}(\varphi_e)$$

where the last equality holds only in the of linear dual variables, when $\varphi_e \in \mathbb{E}^*$. Then

$$(\text{mR-D}^1) = \sup_{\varphi \in (\mathbb{E}^*)^{\mathcal{E}}} F^*\big(\operatorname{Div}\varphi\big) + G^*\big(\varphi\big) = \inf_{x \in (\mathbb{E}^*)^{\mathcal{V}}} F^{**}(x) + G^{**}(-\operatorname{Div}^* x)$$
$$= \inf_{x \in (\mathbb{E}^*)^{\mathcal{V}}} \sum_{u \in \mathcal{V}} f_u^{**}(x) + \sum_{(u,v) \in \mathcal{E}} \|x_u - x_v\|.$$

where the first equality is an application of Fenchel–Rockafellar duality, and the second is a straightforward rewriting using the definition of the graph divergence operator Div and that for $y \in \mathbb{E}^*$ one has $\iota^*_{\mathrm{Lip}(\mathbb{E},\|\cdot\|)}(y) = \iota^*_{\mathcal{B}^{M-1}}(y) = \|y\|$.

References

1. Bach, F.: Submodular functions: from discrete to continous domains. Mathematical Programming **175** (10 2015)
2. Bauermeister, H., Laude, E., Möllenhoff, T., Moeller, M., Cremers, D.: Lifting the convex conjugate in Lagrangian relaxations: A tractable approach for continuous markov random fields. SIAM Journal on Imaging Sciences **15**(3), 1253–1281 (2022)
3. Blekherman, G., Parrilo, P.A., Thomas, R.R.: Semidefinite Optimization and Convex Algebraic Geometry. Society for Industrial and Applied Mathematics, Philadelphia, PA (2012)
4. Boumal, N.: An introduction to optimization on smooth manifolds. Cambridge University Press (2023)
5. Henrion, D., Korda, M., Lasserre, J.B.: The Moment-SOS Hierarchy, Optimization and its Applications, vol. 4. World Scientific Publishing Europe Ltd. (Dec 2020)
6. Krivine, J.L.: Anneaux préordonnés. HAL **1964**(0) (1964)
7. Lasserre, J.B.: Convergent SDP-relaxations in polynomial optimization with sparsity. SIAM Journal on Optimization **17**(3), 822–843 (2006)
8. Lasserre, J., Toh, K.C., Yang, S.: A bounded degree SOS hierarchy for polynomial optimization. EURO Journal on Computational Optimization **5**(1), 87–117 (2017)
9. Newman, D.J., Shapiro, H.S.: Jackson's Theorem in Higher Dimensions, pp. 208–219. Springer Basel, Basel (1964)
10. Putinar, M.: Positive polynomials on compact semi-algebraic sets. Indiana University Mathematics Journal **42**(3), 969–984 (1993)
11. Ragozin, D.L.: Constructive polynomial approximation on spheres and projective spaces. Transactions of the American Mathematical Society **162**, 157–170 (1971)
12. Rockafellar, R.: Duality and stability in extremum problems involving convex functions. Pacific Journal of Mathematics **21**(1), 167–187 (1967)
13. Rudin, W.: Real and Complex Analysis. McGraw-Hill (1987)
14. Santambrogio, F.: Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling, vol. 87. Birkhäuser (2015)
15. Stengle, G.: A nullstellensatz and a positivstellensatz in semialgebraic geometry. Mathematische Annalen **207**(2), 87–97 (Jun 1974)
16. Wainwright, M.J., Jordan, M.I.: Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning **1**(1–2), 1–305 (2008)