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

Robin Kenis, Emanuel Laude, and Panagiotis Patrinos

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

Abstract. While neural network models have garnered significant attention in the imaging community, their application remains limited in important settings where optimality certificates are required or in the absence of extensive datasets. In such cases, classical models like (continuous) Markov Random Fields (MRFs) remain preferable. However, the associated optimization problem is nonconvex, and therefore very challenging to solve globally. This difficulty is further exacerbated in the case of nonconvex state spaces, such as the unit sphere. To address this, we propose a convex Semidefinite Programming (SDP) relaxation to provide lower bounds for these optimization challenges. Our relaxation provably approximates a certain infinite-dimensional convex lifting in measure spaces. Notably, our approach furnishes a certificate of (near) optimality when the relaxation (closely) approximates the unlifted problem. Our experiments show that our relaxation outperforms popular linear relaxations for many interesting problems.

Keywords: Global Optimization · Convex Relaxations · Polynomial Optimization · Continuous Markov Random Fields

1 Introduction

1.1 Motivation

In this paper we develop a convex relaxation approach for optimization problems with a graphical coupling structure whose variables are constrained to live on the Riemannian manifold Ω. Given a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ the problem takes the form

$$\text{minimize} \left\{ F(x) \equiv \sum_{u \in \mathcal{V}} f_u(x_u) + \sum_{(u,v) \in \mathcal{E}} f_{(u,v)}(x_u, x_v) : x \in \Omega^{\mathcal{V}} \right\}, \tag{P}$$

where we consider continuous data terms $f_u : \Omega \to \mathbb{R}$ and coupling terms $f_{(u,v)} : \Omega \times \Omega \to \mathbb{R}$. In the context of graphical models and exponential families this amounts to *maximum a posteriori* (MAP) inference in a continuous pairwise *Markov random field* (MRF) [38]. Optimization problems of this form appear in a variety of tasks such as language processing [32], image processing [4], bioinformatics [12] or graph-based *simultaneous localization and mapping* (SLAM) [14] to name a few. In image processing (P) can be used to formulate denoising or inpainting tasks where $\mathcal{G}$ represents the pixel grid, $\sum_{u\in\mathcal{V}} f_u(x_u)$ is a data fidelity term and $\sum_{(u,v)\in\mathcal{E}} f_{(u,v)}(x_u, x_v)$ is a prior such as the *total variation* (TV) that favors images x that are spatially smooth [34].

In many situations the state space is a manifold such as the unit shere $\mathcal{S}^m$ and the coupling terms $f_{(u,v)}(x_u, x_v) = d(x_u, x_v)$ amount to the (nonsmooth, nonconvex) geodesic distance. As a result the optimization problem (P) is nonconvex and nonsmooth and therefore local optimization techniques such as the subgradient method can get stuck in a poor local optimum. Global optimization techniques, such as measure-based lifting, are feasible for small-scale problems with few variables, but not when the optimization variable lies in a high-dimensional product space $(\mathcal{S}^m)^{\mathcal{V}}$ (or more generally, $\Omega^{\mathcal{V}}$).

Instead, we propose to reformulate the problem in terms of a more tractable linear program, via the *local marginal polytope* (LMP) relaxation [38]. The LMP relaxation is a convex lifting which in general only yields a lower bound to the so-called *global marginal polytope* (GMP) relaxation, which is tight but intractable. Despite this, the LMP solutions are often near globally optimal in practice. Many efficient methods exist to solve the LMP relaxation in the case of finite state spaces Ω (when (P) is discrete) [18, 19, 21, 36]. We consider continuous Ω on the other hand, in which case the LMP relaxation amounts to an infinite-dimensional linear program and is thus still intractable. As a remedy we consider a hierarchy of implementable dual programs obtained by polynomial subspace discretization and show that the duality gap between the dual approximation and the infinite-dimensional local marginal polytope relaxation vanishes if the degree goes to infinity. Our approach is related to optimization of (sparse) polynomials [22]. However, in contrast to [22] our theory covers cost functions f_u and $f_{(u,v)}$ that are in general not polynomial. Practically, we can implement pairwise couplings $f_{(u,v)} = d$, where d denotes the geodesic distance, which is not even smooth.

1.2 Related Work

Problem (P) can be viewed as MAP inference in MRFs with continuous state spaces [38]. The LMP relaxation is a popular linear programming relaxation to address these optimization problems. However, in the case of continuous state spaces it leads to an infinite-dimensional *linear program* (LP). In order to obtain a finite-dimensional relaxation, a piecewise polynomial approximation scheme for the dual problem was first proposed in [13] for the case where the state space is a compact subset of the reals. Then [2] built further upon the analysis of [13], and provided an implementable hierarchy for the unit interval. Our work is most closely related to that in [2], generalizing it to unit sphere state spaces. There are also strong relations to convex lifting for (continuous) variational problems [24, 28, 29, 31, 41].

Optimal transport with geodesic penalties on the circle was already considered in [10]. The results were based on formulae specific to the circle, and there is no clear generalization to higher-dimensional spheres. Tikhonov-type problems on the unit circle are considered in [11], and generalized to hyperspheres and rotation groups in [3]. The Tikhonov cost is exactly the squared Euclidean norm on the embedding space (which is polynomially representable) and falls within our general relaxation. Our polynomial approximation scheme builds heavily on existing literature on semidefinite programming relaxations for polynomial optimization [5, 15, 23, 30]. Finally, we remark that the LMP relaxation is tightly linked to approaches for exploiting structured sparsity in polynomial optimization [39, 40].

1.3 Contributions

In this work, we develop an implementable, convex lifting of the MAP inference problem (P) for continuous MRFs, without resorting to sampling, as in finite element-type discretizations (as in *e.g.* [25, 37]).

- In the fully polynomial setting (*i.e.* f_u and $f_{(u,v)}$ being piecewise polynomial, and Ω implementable with polynomial constraints), we provide an implementable hierarchy with convergence guarantees.
- In order to use polynomial optimization tools beyond the fully polynomial setting, we leverage optimal transport duality to provide an implementable hierarchy for (nonsmooth, nonvonvex) geodesic coupling terms, which we call the metric relaxation.
- In contrast to related existing approaches [25, 28, 37] that do not have a convergence theory, we prove convergence of the metric relaxation in the interesting case that $\Omega = \mathcal{S}^m$ (towards (R-P)).
- When the graph $\mathcal{G}$ is a tree, our relaxation (fully polynomial, or metric in the case $\Omega = \mathcal{S}^m$) converges to (P) for increasing degree. For some applications of interest when $\mathcal{G}$ is not a tree, we observe practically that the obtained duality gap is generally small, and may even vanish despite the lack of theoretical guarantees.
- We lower bound our relaxation hierarchy by studying the lowest stage, and noting its equivalence to a natural convex lower bound of (P).
- In our experiments we compare our approach to classical linear programming relaxations, and show that our relaxation returns better duality gaps for smaller relaxation size.

1.4 Notation

Denote by (Ω, g) a compact, connected Riemannian manifold with metric g. We will focus on an embedded manifold in a Euclidean space, denoted $\mathbb{R}^m$ or $\mathbb{E}$ if the dimension is not important. In this case Ω is naturally endowed with a metric g by the embedding in $\mathbb{E}$, and we will often simply write Ω. For the m-dimensional (Euclidean) unit sphere, embedded as a Riemannian manifold in the natural way in $\mathbb{R}^{m+1}$, we write $\mathcal{S}^m$. Denote by $\mathcal{M}_+(\Omega)$ the convex cone of nonnegative Radon measures, and by $\mathcal{P}(\Omega)$ the space of Borel probability measures on Ω. Given measurable spaces (X_1, Σ_1) and (X_2, Σ_2), a nonnegative measure $\mu : \Sigma_1 \to [0, +\infty]$ and a measurable mapping $T : X_1 \to X_2$, $T \sharp \mu : \Sigma_2 \to$

$[0, +\infty]$ is the pushforward of μ under T defined by: $(T\sharp\mu)(A) = \mu(T^{-1}(A))$ for all $A \in \Sigma_2$. Let δ_x denote the Dirac measure centered at $x \in \Omega$. Furthermore let $\mathcal{C}(\Omega)$ be the space of continuous functions on Ω. We write lsc for lower semicontinuous. We will write $\langle \mu, f \rangle = \int_\Omega f(x)\,\mathrm{d}\mu(x)$, for $f : \Omega \to \mathbb{R}$, and f lsc, and $\mu \in \mathcal{M}(\Omega)$. The geodesic distance function on (Ω, g) is denoted d. We define the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. We say f admits the modulus of continuity ω if $|f(y) - f(x)| \le \omega(d(y,x))$ holds for the increasing extended real-valued function $\omega : [0, +\infty] \to [0, +\infty]$, with ω continuous at 0, and $\omega(0) = 0$. If $\omega : x \mapsto Lx$, we get $|f(y) - f(x)| \le Ld(y,x)$ and write that f is Lipschitz continuous with $\operatorname{Lip} f = L$, or in short L-Lipschitz continuous. The space of 1-Lipschitz continuous function on Ω under the distance function d is denoted by $\operatorname{Lip}(\Omega, d)$. For an extended real-valued function $f : \mathbb{E} \to \overline{\mathbb{R}}$ let $\operatorname{dom} f = \{x \in \mathbb{E} \mid f(x) < \infty\}$ be the domain of f, let $f^*(y) = \sup_{x \in \mathbb{E}} \langle x, y \rangle - f(x)$ denote the Fenchel conjugate of f at y and $f^{**} := (f^*)^*$ the Fenchel biconjugate. For a set $C \subset \mathbb{E}$ we denote by $\iota_C : \mathbb{E} \to \overline{\mathbb{R}}$ the indicator function with $\iota_C(x) = 0$ if $x \in C$ and $\iota_C(x) = \infty$ if $x \notin C$ and by $\sigma_C(x) = \sup_{y \in C} \langle x, y \rangle$ the support function of C at $x \in \mathbb{E}$. These notions are defined analogously for the topologically paired spaces $\mathcal{M}(\Omega)$ and $\mathcal{C}(\Omega)$. We denote by $\mathbb{R}[x]$ the ring of polynomials and by $\mathbb{R}_n[x] \subset \mathbb{R}[x]$ the finite-dimensional subspace of polynomials on $\mathbb{E}$ with degree at most n. As $\mathcal{S}^m$ is embedded in $\mathbb{R}^{m+1}$, both $\mathbb{R}[x]$ and $\mathbb{R}_n[x]$ induce function spaces on $\mathcal{S}^m$ by restriction to $\mathcal{S}^m$, which we will denote by the same notation.

2 Local Marginal Polytope Relaxation and Duality

2.1 Local vs. Global Marginal Polytope

Any nonconvex problem of the form (P) can be cast as a convex (linear) problem via a reformulation over the space of probability measures $\mathcal{P}(\Omega^{\mathcal{V}})$:

$$\min_{x \in \Omega^{\mathcal{V}}} F(x) = \min_{\mu \in \mathcal{P}(\Omega^{\mathcal{V}})} \langle \mu, F \rangle. \tag{GMP}$$

This is known in the context of Markov random fields as the global marginal polytope (GMP) relaxation. Even when the dimension of Ω is moderate, the GMP relaxation suffers from a curse of dimensionality for large $\mathcal{V}$. As a remedy, we consider the so-called local marginal polytope relaxation (LMP)

$$\inf_{\mu \in \mathcal{P}(\Omega)^{\mathcal{V}}} \left\{ \mathcal{F}(\mu) := \sum_{u \in \mathcal{V}} \langle \mu_u, f_u \rangle + \sum_{(u,v) \in \mathcal{E}} \mathrm{OT}_{f_{(u,v)}}(\mu_u, \mu_v) \right\}, \tag{R-P}$$

which lifts the optimization variable $x \in \Omega^{\mathcal{V}}$ to a product space of probability measures $\mu \in \mathcal{P}(\Omega)^{\mathcal{V}}$, rather than the space of measures on the product space $\mu \in \mathcal{P}(\Omega^{\mathcal{V}})$ as in (GMP). Here, OT is the optimal transportation [17] with marginals (μ_u, μ_v) and cost $f_{(u,v)}$ defined by

$$\mathrm{OT}_{f_{(u,v)}}(\mu_u, \mu_v) = \inf_{\mu_{(u,v)} \in \Pi(\mu_u, \mu_v)} \langle \mu_{(u,v)}, f_{(u,v)} \rangle. \tag{1}$$

The constraint set $\Pi(\mu_u, \mu_v)$ is the set of all Borel probability measures on $\Omega \times \Omega$ with prescribed marginals μ_u and μ_v:

$$\Pi(\mu_u, \mu_v) = \left\{\mu_{(u,v)} \in \mathcal{P}(\Omega \times \Omega) : \pi_u \sharp \mu_{(u,v)} = \mu_u,\ \pi_v \sharp \mu_{(u,v)} = \mu_v \right\}, \quad (2)$$

where $\pi_u : \Omega \times \Omega \to \Omega$ corresponds to the canonical projection onto the u^{th} component.

Since $\Pi(\delta_x, \delta_{x'}) = \{\delta_{(x,x')}\}$ restricting μ_u (and therefore $\mu_{(u,v)}$) to be Dirac probability measures, the formulation (R-P) is equivalent to the original problem (P). Thus optimizing over the larger set of all probability measures we obtain a lower bound to (GMP), meaning:

$$\text{(R-P)} \leq \text{(GMP)} = \text{(P)}. \quad (3)$$

As is well-known for the classical discrete MRF [38], the local marginal polytope relaxation is tight when $\mathcal{G}$ is a tree.

Theorem 1. *Let (Ω, d) be a metric space and assume that $(\mathcal{V}, \mathcal{E})$ is a tree. Then we have that* (R-P) = (GMP) = (P).

All proofs are deferred to the supplementary materials.

2.2 Duality

In this section, we derive a dual problem to (R-P), which can be tailored towards geodesic coupling terms by leveraging Kantorovich–Rubinstein duality. In light of theorem [35, Theorem 1.42], due to continuity of $f_{(u,v)}$ and compactness of Ω, the optimal transportation $\text{OT}_{f_{(u,v)}}(\mu_u, \mu_v)$ with cost $f_{(u,v)}$ between the marginals μ_u and μ_v admits a simple dual formulation which amounts to a supremum over continuous functions $(\varphi, \psi) \in \mathcal{C}(\Omega)$

$$\text{OT}_{f_{(u,v)}}(\mu_u, \mu_v) = \sup_{(\varphi,\psi) \in \mathcal{K}_{(u,v)}} \langle \mu_u, \varphi \rangle + \langle \mu_v, \psi \rangle \quad (4)$$

that satisfy the nonnegativity constraint

$$\mathcal{K}_{(u,v)} = \left\{ (\varphi, \psi) \in \mathcal{C}(\Omega)^2 \mid f_{(u,v)}(x,y) - \varphi(x) - \psi(y) \geq 0 \quad \forall x, y \in \Omega \right\}. \quad (5)$$

Replacing $\text{OT}_{f_{(u,v)}}$ in (R-P) with its dual formulation (4) an interchange of inf and sup results in

$$\sup_{(\varphi,\psi) \in (\mathcal{C}(\Omega)^{\mathcal{E}})^2} - \sum_{u \in \mathcal{V}} \sigma_{\mathcal{P}(\Omega)}(A_u(\varphi, \psi) - f_u) - \sum_{e \in \mathcal{E}} \iota_{\mathcal{K}_e}(\varphi_e, \psi_e), \quad \text{(R-D)}$$

where $A : (\mathcal{C}(\Omega)^{\mathcal{E}})^2 \to \mathcal{C}(\Omega)^{\mathcal{V}}$ is a linear mapping defined by

$$A_u(\varphi, \psi) = - \sum_{v:(u,v) \in \mathcal{E}} \varphi_{(u,v)} - \sum_{v:(v,u) \in \mathcal{E}} \psi_{(v,u)}. \quad (6)$$

Alternatively, the formulation is obtained by dualizing the marginalization constraints (2) for each edge $(u,v) \in \mathcal{E}$ with Lagrange multipliers $\varphi_{(u,v)}$ and $\psi_{(u,v)}$. Thanks to weak duality we have the relation

$$\text{(R-D)} \leq \text{(R-P)}. \tag{7}$$

Invoking the Fenchel–Rockafellar duality theorem in infinite dimensions [33] and optimal transport duality we show that strong duality holds and in particular (R-P) = (R-D):

Theorem 2. *Let (Ω, d) be a compact metric space and assume that $f_e : \Omega \times \Omega \to \mathbb{R}$ and $f_u : \Omega \to \mathbb{R}$ are lsc for all $e \in \mathcal{E}$ and $u \in \mathcal{V}$. Then* (R-P) = (R-D) *and a minimizer of* (R-P) *exists.*

If, in addition, f_e is continuous, following the proof of [35, Proposition 1.11], we also have existence of dual solutions: To this end we introduce the so-called c-transform

Definition 1 (c-transform and c-concavity). *Let $f : \Omega \to \overline{\mathbb{R}}$ and $c : \Omega \times \Omega \to \mathbb{R}$. Then we define the c-transform and $\bar{c}$-transform of f as*

$$f^c(y) := \inf_{x \in \Omega} \; c(x,y) - f(x), \quad f^{\bar{c}}(x) := \inf_{y \in \Omega} \; c(x,y) - f(y). \tag{8}$$

We say that f is c-concave, denoted $f \in \overline{\Gamma}_c(\Omega)$, if there exists some $g : \Omega \to \overline{\mathbb{R}}$ such that $f(x) = \inf_{y \in \Omega} \; c(x,y) - g(y)$ and we say that f is $\bar{c}$-concave, denoted $f \in \overline{\Gamma}_{\bar{c}}(\Omega)$, if there is some $g : \Omega \to \overline{\mathbb{R}}$ such that $f(y) = \inf_{x \in \Omega} \; c(x,y) - g(x)$.

If $c(x,y) = c(y,x)$, the $\bar{c}$-transform and c-transform of f are identical.

Theorem 3. *Let (Ω, d) be a compact metric space and assume that for all $e \in \mathcal{E}$ the pairwise term $f_e : \Omega \times \Omega \to \mathbb{R}$ is continuous with some modulus ω_e and for all $u \in \mathcal{V}$ the unary term $f_u : \Omega \to \mathbb{R}$ is lsc. Then there exists a maximizer $(\varphi, \psi) \in \mathcal{C}(\Omega)^{\mathcal{E}} \times \mathcal{C}(\Omega)^{\mathcal{E}}$ of* (R-D) *with the property $\varphi_e \in \overline{\Gamma}_{f_e}(\Omega)$ and $\psi_e = (\varphi_e)^{f_e}$ and thus φ_e, ψ_e inherit the modulus of continuity ω_e from f_e.*

In light of Theorem 3 the dual variables φ_e, ψ_e can be replaced by their respective f_e-transforms without changing the dual cost. For geodesic coupling terms, *i.e.* $f_{(u,v)} = d$ on Ω with d denoting the geodesic distance, thanks to [35, Proposition 3.1] $\varphi_e \in \overline{\Gamma}_{f_e}(\Omega)$ and $\psi_e = (\varphi_e)^{f_e}$ together imply that $\psi_e = -\varphi_e \in \mathrm{Lip}(\Omega, d)$ where

$$\mathrm{Lip}(\Omega, d) := \{\varphi : \Omega \to \mathbb{R} : |\varphi(x) - \varphi(y)| \leq d(x,y) \quad \forall x, y \in \Omega\}. \tag{9}$$

Invoking Theorem 2 this leads to the following reduced dual formulation of (R-P):

$$\sup_{\varphi \in \mathcal{C}(\Omega)^{\mathcal{E}}} \; -\sum_{u \in \mathcal{V}} \sigma_{\mathcal{P}(\Omega)}(\mathrm{Div}_u(\varphi) - f_u) - \sum_{e \in \mathcal{E}} \iota_{\mathrm{Lip}(\Omega,d)}(\varphi_e), \tag{mR-D}$$

where Div : $\mathcal{C}(\Omega)^{\mathcal{E}} \to \mathcal{C}(\Omega)^{\mathcal{V}}$ is a graph divergence operator defined by

$$\mathrm{Div}_u(\varphi) = -\sum_{v:(u,v)\in\mathcal{E}} \varphi_{(u,v)} + \sum_{v:(v,u)\in\mathcal{E}} \varphi_{(v,u)}. \tag{10}$$

We have the following result which is also known as Kantorovich–Rubinstein duality in the context of optimal transport.

Corollary 1. *Let (Ω, d) be a compact metric space and assume that $f_{(u,v)} = d$ and $f_u : \Omega \to \mathbb{R}$ is lsc. Then* (R-P) = (mR-D) *and there exists a minimizer of* (R-P) *and a maximizer of* (mR-D).

As we will see (mR-D) can be used to derive a tailored implementation for geodesic coupling terms leveraging a simple characterization of Lipschitz continuity on manifolds whereas (R-D) is applicable for polynomial couplings only.

3 Polynomial Approximation and Moment Relaxation

In this section we consider polynomial subspace approximations of (R-D) and (mR-D). Although both problems are equivalent in infinite dimensions they lead to different hierachies once we restrict the maximization to the finite-dimensional subspace of polynomials. This leads us to examine the convergence properties of both hierachies in parallel restricting Ω to a unit sphere.

3.1 General Dual Problem

Firstly, we consider the general dual problem (R-D). We restrict the dual variables φ_e, ψ_e in (R-D) to be polynomials with degree n on the embedding space $\mathbb{E} \supset \Omega$. Then the discretized dual problem amounts to

$$\sup_{\varphi,\psi\in(\mathbb{R}_n[x])^{\mathcal{E}}} \quad -\sum_{u\in\mathcal{V}} \sigma_{\mathcal{P}(\Omega)}(A_u(\varphi,\psi) - f_u) - \sum_{e\in\mathcal{E}} \iota_{\mathcal{K}_e}(\varphi_e, \psi_e). \tag{R-D$^{\mathrm{n}}$}$$

Next we will show that the duality gap (R-P) − (R-D$^{\mathrm{n}}$) vanishes as $n \to \infty$, roughly following the proof of [13, Theorem 2].

Theorem 4. *Let (Ω, d) be a compact metric space and let $f_u : \Omega \to \mathbb{R}$ be lsc and $f_{(u,v)} : \Omega \times \Omega \to \mathbb{R}$ be continuous. Then we have that the duality gap vanishes,* i.e., (R-P) − (R-D$^{\mathrm{n}}$) $\searrow 0$ *for $n \to \infty$. If, in addition, $(\mathcal{V}, \mathcal{E})$ is a tree we have in particular* (P) − (R-D$^{\mathrm{n}}$) $\searrow 0$ *for $n \to \infty$.*

Further assuming that $f_{(u,v)} = d$, with d denoting the geodesic distance, we have the following result:

Corollary 2. *Let (Ω, g) be the m-dimensional unit sphere $\mathcal{S}^m$ with induced Riemannian metric g and let $f_u : \Omega \to \mathbb{R}$ be lsc and $f_{(u,v)} = d$, where d is the geodesic distance function. Then for any $m \in \mathbb{N}$ there exists a constant C_m such that the duality gap vanishes with rate* (R-P) − (R-D$^{\mathrm{n}}$) $\leq 4C_m|\mathcal{E}|\frac{1}{n}$. *If, in addition, $(\mathcal{V}, \mathcal{E})$ is a tree we have in particular* (P) − (R-D$^{\mathrm{n}}$) $\leq 4C_m|\mathcal{E}|\frac{1}{n}$.

This corollary proves convergence of the general relaxation for geodesic penalty terms, which is equivalent to the metric problem (mR-D) in light of Corollary 1. Unfortunately, (R-D$^{\mathrm{n}}$) with geodesic coupling terms is not readily implementable. In contrast, the metric problem is structurally simpler, involving only a Lipschitz constraint on the dual variable φ, and we will consider it next.

3.2 Metric Dual Problem

In this section we consider a hierarchy of polynomial discretizations for the metric dual formulation (mR-D) which amounts to

$$\sup_{\varphi \in (\mathbb{R}_n[x])^{\mathcal{E}}} \quad -\sum_{u \in \mathcal{V}} \sigma_{\mathcal{P}(\Omega)}(\mathrm{Div}_u(\varphi) - f_u) - \sum_{e \in \mathcal{E}} \iota_{\mathrm{Lip}(\Omega,d)}(\varphi_e). \tag{mR-D$^{\mathrm{n}}$}$$

Next we state the main result of this section which shows that the duality gap vanishes at rate $\mathcal{O}(1/n)$ where n is the degree of the dual variable:

Theorem 5. *Let (Ω, g) be the m-dimensional unit sphere $\mathcal{S}^m$ with induced Riemannian metric g and let $f_u : \Omega \to \mathbb{R}$ be lsc and $f_{(u,v)} = d$, where d is the geodesic distance function. For any $m \in \mathbb{N}$ there exists a constant C_m such that the duality gap vanishes sublinearly at rate* (R-P) $-$ (mR-D$^{\mathrm{n}}$) $\leq 2|\mathcal{E}|C_m \frac{1}{n}$. *If, in addition, $(\mathcal{V}, \mathcal{E})$ is a tree we have in particular* (P) $-$ (mR-D$^{\mathrm{n}}$) $\leq 2|\mathcal{E}|C_m \frac{1}{n}$.

In contrast to the approximation result for general MRFs, this theorem does not follow from applying standard results from approximation theory. In particular, we require the approximants to exhibit the same Lipschitz constant as the optimal dual multiplier. For that purpose, we prove for any L-Lipschitz continuous function φ on $\mathcal{S}^m$ the existence of a sequence of L-Lipschitz continuous polynomials $\{\varphi_n\}_{n=0}^{\infty}$ on $\mathcal{S}^m$ which uniformly approximate φ as $n \to \infty$:

Theorem 6. *Let $m \in \mathbb{N}$. Then there exists a constant C_m such that for every L-Lipschitz continuous function $\varphi : \mathcal{S}^m \to \mathbb{R}$ and any $n \in \mathbb{N}$ we can construct an L-Lipschitz continuous $\varphi_n \in \mathbb{R}_n[x]$ such that $\|\varphi - \varphi_n\|_\infty \leq C_m \frac{L}{n}$.*

3.3 Degree 1-Case

Given that our interest is in a tractable relaxation, we are mainly interested in the quality of our relaxation using dual polynomial approximants of low degree. In this section we show a lower bound by relating the lowest stage of our hierarchy to a natural convexification of (P). The first part of the following theorem is a specialization of [13, Theorem 1]:

Theorem 7 (degree 1 relaxation). *Let the degree $n = 1$. Then* (R-D^1) *is equivalent to*

$$\inf_{x \in (\mathbb{E})^{\mathcal{V}}} \Big\{ \sum_{u \in \mathcal{V}} f_u^{**}(x_u) + \sum_{(u,v) \in \mathcal{E}} f_{(u,v)}^{**}(x_u, x_v) \Big\}, \tag{11}$$

where f_u *and* $f_{(u,v)}$ *are to be understood as extended real-valued functions on the embedding space,* i.e. $f_u : \mathbb{E} \to \overline{\mathbb{R}}$, *with* $f_u(x) = +\infty$, $x \notin \Omega$ *and finite otherwise, and similarly for* $f_{(u,v)} : \mathbb{E} \times \mathbb{E} \to \overline{\mathbb{R}}$. *In particular,* $\mathrm{dom}(f_u^{**}) = \mathrm{con}(\Omega)$. *Furthermore, when* $\Omega = \mathcal{S}^m$, (mR-D^1) *is equivalent to*

$$\inf_{x \in (\mathbb{E})^{\mathcal{V}}} \Big\{ \sum_{u \in \mathcal{V}} f_u^{**}(x_u) + \sum_{(u,v) \in \mathcal{E}} \|x_u - x_v\| \Big\}, \tag{12}$$

where $\mathrm{dom}(f_u^{**}) = \mathcal{B}^m$, *the* $m+1$*-dimensional unit ball with boundary* $\mathcal{S}^m$.

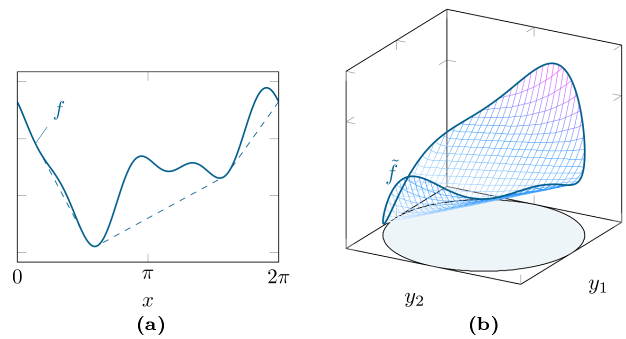

Fig. 1: (a) Naive convexification via a Euclidean parametrization vs. (b) convex hulls obtained by convexification over the manifold. In (a), the convex hull (*dotted line*) of a 2π-periodic function f (*full line*) loses the overall shape of f. Looking at (b), the periodic function f is naturally interpreted as a function $\tilde{f}$ (*blue line*) on the circle, where our proposed convexification returns a convex function on the unit ball, but notably retains the cost $\tilde{f}$ on the boundary of its domain (the unit circle).

This clearly identifies the first level of our proposed hierarchy as a simple but 'natural' convexification. We argue that this simple convexification is preferable over the naive convexification that one may obtain via a Euclidean parametrization of the manifold Ω. In [2, Definition 4.8] the concept of an extremal moment curve is defined. Taking the mapping φ in this definition to be the identity map, the unit circle (more generally the unit sphere $\mathcal{S}^m$) is an extremal moment curve. Then [2, Theorem 4.11] proves that any lower semi-continuous cost on Ω is preserved under our convexification. This is depicted in Fig. 1.

4 Implementation of the Semidefinite Relaxation

In this section we give a high level overview of an implementable algorithm to optimize manifold-constrained MRF problems. We defer the details of implementation to the supplementary materials. We assume that Ω is a Riemannian submanifold of $\mathbb{E}$ that can be encoded using polynomial equality constraints, which is in particular true for the unit sphere $\mathcal{S}^m$. Furthmore, we assume that the unary terms are polynomial and that the coupling terms are either polynomial (for (R-D$^{\mathrm{n}}$)) or geodesic penalties (for (mR-D$^{\mathrm{n}}$)). For the case of (mR-D$^{\mathrm{n}}$), we rewrite the Lipschitz continuity condition as a box constraint with respect to the manifold gradient. Then we use tools from polynomial optimization literature to rephrase the problem as an implementable semidefinite problem.

4.1 Implementation of the Manifold Gradient Condition

The implementation of the metric relaxation (mR-D$^{\mathrm{n}}$), requires a certificate of Lipschitz continuity for a polynomial. Similar to the Euclidean case, it is known [7, Proposition 10.43] that a function φ on Ω is Lipschitz continuous with $\operatorname{Lip}\varphi = L$ if and only if $\|\nabla_\Omega \varphi(x)\| \leq L$ for all $x \in \Omega$. Here, ∇_Ω denotes the manifold gradient on Ω. One can calculate the manifold gradient $\nabla_\Omega \varphi$ at x via a projection of the (Euclidean) gradient $\nabla \varphi$ onto the tangent space of Ω.

For $\Omega = \mathcal{S}^m$, the matrix $\Pi(x) = I - xx^\top$ is the projector onto the tangent space at $x \in \mathcal{S}^m$ such that $\nabla_\Omega \varphi(x) = \Pi(x)\nabla\varphi(x)$. Via dualization of the norm, we can then assure Lipschitz continuity as $\|\nabla_\Omega \varphi(x)\| = \sup_{\xi \in \mathcal{S}^m} \langle \xi, \Pi(x)\nabla\varphi(x)\rangle \leq L$. Since this inequality holds at the supremum, it holds for every $\xi \in \mathcal{S}^m$ and we obtain that the gradient condition is met if and only if

$$p(x,\xi) = L - \langle \xi, \Pi(x)\nabla\varphi(x)\rangle \geq 0, \quad \forall (x,\xi) \in \mathcal{S}^m \times \mathcal{S}^m, \tag{13}$$

which is a polynomial inequality in the variables (x, ξ). This procedure generalizes straightforwardly to a more general manifold Ω. Some details on the existence of $\Pi(x)$ can be found in the supplementary materials.

4.2 Semidefinite Problem Formulation

The core principle of the semidefinite hierarchies is to verify the semi-infinite polynomial nonnegativity constraints (5) and (13) by *sum-of-squares* (SOS) nonnegativity certificates [5,15]. That is, a polynomial nonnegativity constraint $p \geq 0$ is replaced by the convex cone constraint $p \in \Sigma$, where

$$\Sigma = \left\{ q \in \mathbb{R}[x] \;\middle|\; q = \sum_i q_i^2 + \sum_j r_j k_j, \;\; k_j, q_i, r_j \in \mathbb{R}[x], \; k_j(x) = 0 \; \forall x \in \Omega \right\}. \tag{14}$$

An element of Σ is clearly nonnegative for any $x \in \Omega$, and therefore Σ is an inner approximation to the cone of nonnegative polynomials. When representing polynomials by their coefficient vectors, $p \in \Sigma$ can be ensured by a semidefinite constraint [5]. Dually, the probability measures are represented by a truncated moment sequence, and the dual cone Σ^* to Σ gives rise to an outer approximation to the cone of nonnegative measures [15]. Minimization over moment sequences in Σ^* allows us to evaluate the support functions in (R-D$^{\mathrm{n}}$) and (mR-D$^{\mathrm{n}}$). Using these well-established ideas, we construct our semidefinite relaxation following exactly the procedure in the references [5,15]. A detailed description of our relaxation is found in the supplementary materials.

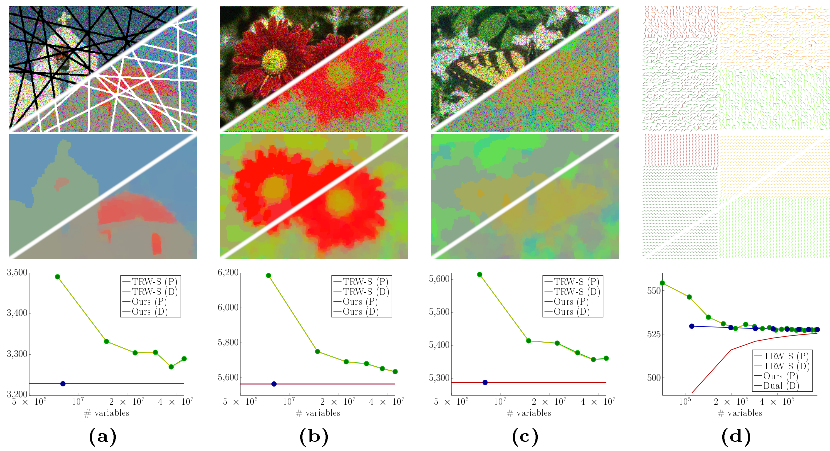

Fig. 2: Geodesic denoising results for (a) chromaticity denoising with inpainting, (b) and (c) pure chromaticity denoising and (d) a $\mathcal{S}^1$-valued toy dataset. The noisy datasets are shown in the first row, where for the images the noisy image is shown (*top left*) as well as the noisy chromaticity (*bottom right*). In the second row, part of the denoised results by the TRW-S baseline (*top left*) and our approach (*bottom right*) is displayed. The third row contains the primal (*green line*) and dual (*yellow line*) costs per experiment obtained using TRW-S, as well as the primal cost P after rounding our approach (*blue line*) and dual cost D (*red line*) without rounding, expressed in the numbers of variables in the relaxation. For the imaging experiments, P and D are indistinguishable. The primal and dual cost for TRW-S coincide, indicating it has converged.

5 Experimental Evaluation

5.1 Setup and Rounding Scheme

In this section, we use our semidefinite relaxation to solve a collection of geodesic denoising and pose graph optimization problems. Our approach is implemented[1] using MOSEK's interior point solver [1]. The implementation of a specialized solver which leverages parallelizability is left for future work. As we have lifted the problem over the space of measures, the optimizers obtained by our relaxation are moment vectors. The degree 1 moments are extracted and projected onto Ω to obtain a feasible point $x^* \in \Omega^{\mathcal{V}}$ for the primal problem. We denote by P the cost obtained by x^* for the unlifted problem, and by D the optimal lower bound of our relaxation. The relative duality gap G is defined as $\frac{P-D}{P}$.

5.2 $\mathcal{S}^m$-Denoising Problem

In the following experiments, we consider the geodesic denoising problem [16]:

$$T_{\text{denoised}} = \underset{Y \in (\Omega)^{\mathcal{V}}}{\arg\min} \sum_{u \in \mathcal{V}} \frac{(T_u - Y_u)^2}{2} + L \sum_{(u,v) \in \mathcal{E}} d(Y_u, Y_v) \tag{15}$$

[1] https://github.com/Robin-Kenis/ManifoldMRFs.git

Table 1: Primal and dual energies for the experiments in Fig. 2 (a)-(c) with a fixed runtime budget of 1000 seconds. The number in square brackets denotes the degree of the dual multipliers in our relaxation, or the number of labels used in the LP-discretization.

Time = 1000 s		Ours [2]	TRW-S [150]	TRW-S [200]	LR-Solver [150]	LR-Solver [200]
Rooftop	P	3228.25	3288.05	3266.50	3287.44	3402.78
Fig. 2a	D	3228.17	3287.37	3256.48	3287.44	2782.90
Flowers	P	5565.06	5635.22	5650.13	5634.27	5815.74
Fig. 2b	D	5565.82	5634.19	5623.32	5634.24	5104.79
Butterfly	P	5288.99	5362.19	5367.40	5360.76	5670.43
Fig. 2c	D	5289.92	5360.68	5338.28	5360.75	3935.67

where T is used to denote the given dataset. This is a generalization of TV-a.k.a. ROF-denoising [34] to manifold-valued signals. In [16] the model (15) was proposed in the context of chromaticity denoising for images, where chromaticity is defined as the projection of the RGB-value of a pixel onto $\mathcal{S}^2$. In Fig. 2 we consider this problem for three images from [27]. For Fig. 2a some pixels are corrupted. Here we set the data cost term for these pixels to zero, and perform denoising and inpainting simultaneously. Finally, we consider an $\mathcal{S}^1$-valued denoising toy problem. We compare our approach to a standard linear relaxation based on discretization: the manifold is sampled in order to obtain a discrete MRF problem, for which linear relaxations are well known. We solve the resulting problem using TRW-S [19] and the method presented in [20]. We use a number of labels, denoted l, in the set $\{25, 50, \ldots, 150\}$, and compare the methods in space (number of variables) and time efficiency. For the number of variables, we use the formula $l \times (|\mathcal{V}| + 2 \times |\mathcal{E}|)$ from the section on efficient implementation in [19]. For our own approach, we sum the number of moments for each vertex and edge in the purely primal formulation of our problem (as it is implemented in MOSEK).

For the imaging problems, our approach is numerically tight, achieving the optimality certificate of near-zero primal-dual gaps ($G = (2.479, 9.089, 0.959) \times 10^{-5}$ for (a), (b) and (c) respectively). On the other hand, Fig. 2c shows that the sampling approach converges only slowly to the optimum, and is far from optimal even for as many as 150 labels. Additionally, we only sample the positive octant of the sphere, which is a great benefit for the linear hierarchy. In the second row of Fig. 2 we plot the denoised result using 50 labels, where the sampling bias of the linear relaxation is clear from to the nonsmooth piecewise constant results. Our own approach does not suffer from this drawback. In Tab. 1, we record the results of each approach after a runtime of 1000 seconds using all three approaches. Despite not having converged to a high-accuracy solution within the given time (as evidenced by a slightly negative primal-dual gap G), the rounded solutions are near-globally optimal. The LP-solvers achieve either a significantly

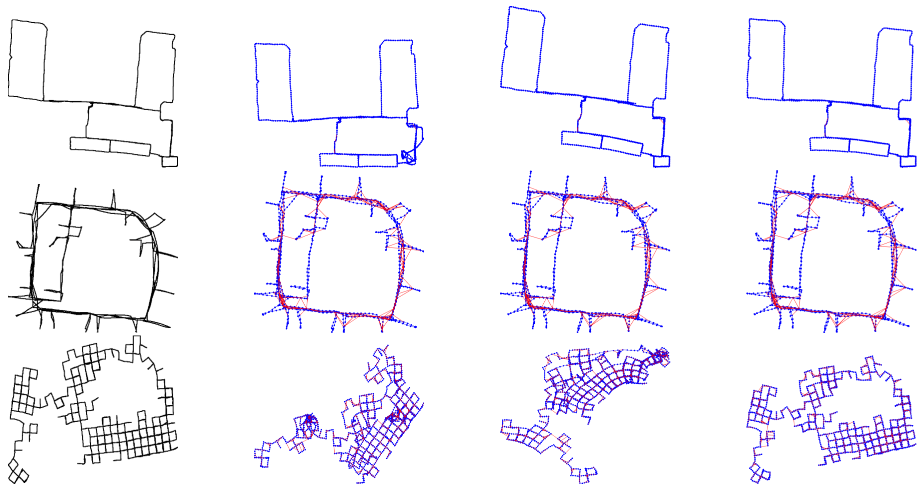

Fig. 3: Results for the pose graph optimization problem: (a) ground truth, (b) local method, (c) ours with least squares rounding, (d) ours after local optimization with ManOpt [6]. The datasets are INTEL, MIT and M3500b, taken from [9].

worse cost (for 150 labels), or do not achieve a small primal-dual gap in time (200 labels).

For the $\mathcal{S}^1$-valued dataset (Fig. 2d) we consider an ‘adversarial’ example for our own approach. It is solved with degrees ranging from 2 to 16 for our approach, and labels in the set $\{10, 15, \dots, 100\}$ for the discrete relaxation. As the numbers of labels is increased the reference approach reaches similar or better costs for the same memory complexity (number of variables). Furthermore, the primal-dual gap for our approach converges relatively slowly. This discrepancy can be explained by the edge in the dataset where the optimal denoised result requires a $180°$ jump. Some details on why this poses a worst-case scenario for our approach are laid out in the supplementary materials. Nonetheless, our approach is almost as good as the reference approach which has access to exact values of the geodesic penalty. Furthermore, our approach achieves better results in the first stage of our hierarchy as compared to the first stages of the linear relaxation. Rounding our approach with a local method would straightforwardly alleviate this drawback.

5.3 Pose Graph Optimization

In robotics and computer vision, *pose graph optimization* (PGO) consists of refining estimates of the positions and orientations (poses) of an agent or camera system based on noisy sensor measurements. The aim is to find the most likely configuration of poses that is consistent with the observed data while also accounting for loop closures, which occur when the system revisits a previously visited location. The PGO problem amounts to minimizing the following cost [14]:

$$\min_{(y,q)\in(\mathcal{SE}(2))^{\mathcal{V}}} \sum_{(u,v)\in\mathcal{E}} \tfrac{\kappa_e}{2}\left\|y_v - y_u - q_u x_{(u,v)}\right\|_2^2 + \tfrac{\lambda_e}{2}\left\|q_v - q_u r_{(u,v)}\right\|_2^2 \tag{16}$$

where $(x,r) \in (\mathcal{SE}(2))^{\mathcal{E}}$ contains the noisy relative measurements between positions, including those estimated for the loop closure edges. Here, $\mathcal{SE}(2)$ is the special Euclidean group of order 2. The solution $(y,q) \in (\mathcal{SE}(2))^{\mathcal{V}}$ contains the absolute poses.

Semidefinite relaxations for the PGO problem have been studied widely, including from the viewpoint of sparse polynomial programming [8, 26, 42]. We solve the PGO problem using our relaxation, using polynomials of degree 2 as dual variables, and either use a least squares optimization step to update the positions while keeping the angles fixed (LSQ rounding), or initialize a local Riemannian optimization method with our solution (Manopt [6] rounding). The results are found in Fig. 3. While a local method is required to refine the solution for the PGO problem, we remark that for the rotation estimation problem (which is obtained by fixing the poses y and optimizing only over the rotations q in (16)) our relaxation is empirically tight in low-noise regimes even at degree 2. The appeal of our hierarchy for these problems is that it leads to smaller optimization problems than existing semidefinite relaxations.

6 Conclusion

In this paper we have considered an efficient moment relaxation for manifold-valued optimization problems with graphical structure. Leveraging optimal transport duality we derive a tractable dual subspace approximation of the infinite-dimensional problem which allows us to tackle possibly nonpolynomial optimization problems with manifold constraints and geodesic coupling terms. In contrast to similar existing approaches that do not provide approximation guarantees, we prove convergence of our relaxation for fully polynomial problems, as well as geodesic penalties on the sphere. We showed that the duality gap vanishes in the limit by proving that a Lipschitz continuous dual multiplier on a unit sphere can be approximated as closely as desired in terms of a Lipschitz continuous polynomial. The formulation is applied to manifold-valued imaging problems with total variation regularization and graph-based SLAM. We compare our SDP relaxation to a standard LP relaxation for imaging problems, and a local method for SLAM. In imaging tasks our approach achieves almost-zero duality gaps for a moderate degree. In contrast the LP relaxation suffers from a 'labeling bias' and thus requires a large amount of samples for a faithful approximation of the unlifted problem. In graph-based SLAM our approach often finds solutions which after refinement with a local method are near the ground truth solution. While we implement our approach in an off-the-shelf SDP solver (MOSEK), it is hugely parallelizable and could benefit greatly from a GPU-based solver. This will allow us to apply our method to even larger problems. Implementing a specialized solver is left for future work.

Acknowledgments

Work supported by the Research Foundation Flanders (FWO) research projects G081222N, G033822N, and G0A0920N; European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 953348;Research Council KU Leuven C1 project No. C14/24/103; KU Leuven Internal Funds postdoctoral mandate PDMt1/22/023.

References

1. ApS, M.: The MOSEK optimization toolbox for MATLAB manual. Version 10.1. (2024), `http://docs.mosek.com/latest/toolbox/index.html`
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. Beinert, R., Bresch, J., Steidl, G.: Denoising of sphere- and SO(3)-valued data by relaxed tikhonov regularization. CoRR **abs/2307.10980** (2023). `https://doi.org/10.48550/ARXIV.2307.10980`, `https://doi.org/10.48550/arXiv.2307.10980`
4. Blake, A., Kohli, P., Rother, C.: Markov Random Fields for Vision and Image Processing. The MIT Press (2011)
5. Blekherman, G., Parrilo, P.A., Thomas, R.R.: Semidefinite Optimization and Convex Algebraic Geometry. Society for Industrial and Applied Mathematics, Philadelphia, PA (2012)
6. Boumal, N., Mishra, B., Absil, P.A., Sepulchre, R.: Manopt, a Matlab toolbox for optimization on manifolds. Journal of Machine Learning Research **15**(42), 1455–1459 (2014), `https://www.manopt.org`
7. Boumal, N.: An introduction to optimization on smooth manifolds. Cambridge University Press (2023)
8. Carlone, L., Calafiore, G.C.: Convex relaxations for pose graph optimization with outliers. IEEE Robotics and Automation Letters **3**(2), 1160–1167 (2018)
9. Carlone, L., Censi, A.: From angular manifolds to the integer lattice: Guaranteed orientation estimation with application to pose graph optimization. IEEE Transactions on Robotics **30**(2), 475–492 (2014)
10. Condat, L.: Atomic norm minimization for decomposition into complex exponentials and optimal transport in Fourier domain. Journal of Approximation Theory **258**, 105456 (2020)
11. Condat, L.: Tikhonov regularization of circle-valued signals. IEEE Transactions on Signal Processing **70**, 2775–2782 (2022). `https://doi.org/10.1109/TSP.2022.3179816`
12. Durbin, R., Eddy, S.R., Krogh, A., Mitchison, G.: Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press (1998)
13. Fix, A., Agarwal, S.: Duality and the continuous graphical model. In: Computer Vision – ECCV 2014. pp. 266–281. Springer International Publishing, Cham (2014)
14. Grisetti, G., Kümmerle, R., Stachniss, C., Burgard, W.: A tutorial on graph-based SLAM. IEEE Intelligent Transportation Systems Magazine **2**(4), 31–43 (2010)
15. Henrion, D., Korda, M., Lasserre, J.B.: The Moment-SOS Hierarchy, Optimization and its Applications, vol. 4. World Scientific Publishing Europe Ltd. (Dec 2020)

16. Kang, S.H., March, R.: Variational models for image colorization via chromaticity and brightness decomposition. IEEE Transactions on Image Processing **16**(9), 2251–2261 (2007)
17. Kantorovich, L.V.: Mathematical methods of organizing and planning production. Management Science **6**, 366–422 (1960)
18. Kappes, J.H., Andres, B., Hamprecht, F.A., Schnörr, C., Nowozin, S., Batra, D., Kim, S., Kausler, B.X., Kröger, T., Lellmann, J., Komodakis, N., Savchynskyy, B., Rother, C.: A Comparative Study of Modern Inference Techniques for Structured Discrete Energy Minimization Problems. International Journal of Computer Vision **115**(2), 155–184 (Nov 2015). https://doi.org/10.1007/s11263-015-0809-x, https://doi.org/10.1007/s11263-015-0809-x
19. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**, 1568–1583 (2006), https://api.semanticscholar.org/CorpusID:8616813
20. Kolmogorov, V.: Solving relaxations of MAP-MRF problems: Combinatorial in-face Frank-Wolfe directions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11980–11989 (2023)
21. Komodakis, N., Paragios, N., Tziritas, G.: MRF optimization via dual decomposition: Message-passing revisited. In: 2007 IEEE 11th International Conference on Computer Vision. pp. 1–8 (2007)
22. Lasserre, J.B.: Convergent SDP-relaxations in polynomial optimization with sparsity. SIAM Journal on Optimization **17**(3), 822–843 (2006)
23. Lasserre, J.B.: Global optimization with polynomials and the problem of moments. SIAM Journal on Optimization **11** (09 2004)
24. Laude, E., Möllenhoff, T., Möller, M., Lellmann, J., Cremers, D.: Sublabel-accurate convex relaxation of vectorial multilabel energies. In: European Conference on Computer Vision (2016), https://api.semanticscholar.org/CorpusID:6156695
25. Lellmann, J., Strekalovskiy, E., Koetter, S., Cremers, D.: Total variation regularization for functions with values in a manifold. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2944–2951 (2013)
26. Mangelson, J.G., Liu, J., Eustice, R.M., Vasudevan, R.: Guaranteed globally optimal planar pose graph and landmark SLAM via sparse-bounded sums-of-squares programming. In: 2019 International Conference on Robotics and Automation (ICRA). pp. 9306–9312. IEEE (2019)
27. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision. vol. 2, pp. 416–423 (July 2001)
28. Mollenhoff, T., Cremers, D.: Sublabel-accurate discretization of nonconvex free-discontinuity problems. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 1192–1200. IEEE Computer Society, Los Alamitos, CA, USA (oct 2017)
29. Mollenhoff, T., Cremers, D.: Lifting vectorial variational problems: a natural formulation based on geometric measure theory and discrete exterior calculus. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11117–11126 (2019)
30. Parrilo, P.A.: Semidefinite programming relaxations for semialgebraic problems. Mathematical programming **96**, 293–320 (2003)
31. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. SIAM Journal on Imaging Sciences **3**(4), 1122–1145 (2010)

32. Rabiner, Lawrence R.m Juang, B.H.: Fundamentals of speech recognition. Englewood Cliffs (1993)
33. Rockafellar, R.: Duality and stability in extremum problems involving convex functions. Pacific Journal of Mathematics **21**(1), 167–187 (1967)
34. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(1), 259–268 (1992)
35. Santambrogio, F.: Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling, vol. 87. Birkhäuser (2015)
36. Savchynskyy, B.: Discrete graphical models — an optimization perspective. Foundations and Trends® in Computer Graphics and Vision **11**(3-4), 160–429 (2019). https://doi.org/10.1561/0600000084, http://dx.doi.org/10.1561/0600000084
37. Vogt, T., Strekalovskiy, E., Cremers, D., Lellmann, J.: Lifting methods for manifold-valued variational problems, pp. 95–119. Springer International Publishing (Apr 2020)
38. 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)
39. Waki, H., Kim, S., Kojima, M., Muramatsu, M.: Sums of squares and semidefinite program relaxations for polynomial optimization problems with structured sparsity. SIAM Journal on Optimization **17**(1), 218–242 (2006)
40. Weisser, T., Lasserre, J.B., Toh, K.C.: Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. Mathematical Programming Computation **10**(1), 1–32 (Mar 2018)
41. Windheuser, T., Cremers, D.: A convex solution to spatially-regularized correspondence problems. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. pp. 853–868. Springer (2016)
42. Yang, H., Carlone, L.: One ring to rule them all: Certifiably robust geometric perception with outliers. Advances in neural information processing systems **33**, 18846–18859 (2020)