

A convex approach to the Gilbert-Steiner problem

Mauro Bonafini, Edouard Oudet

▶ **To cite this version:**

Mauro Bonafini, Edouard Oudet. A convex approach to the Gilbert-Steiner problem. 2019. hal-02009425

HAL Id: hal-02009425

https://hal.univ-grenoble-alpes.fr/hal-02009425

Preprint submitted on 6 Feb 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A convex approach to the Gilbert–Steiner problem

M. Bonafini*, É. Oudet†

October 12, 2018

Abstract

We describe a convex relaxation for the Gilbert–Steiner problem both in $\mathbb{R}^d$ and on manifolds, extending the framework proposed in [9], and we discuss its sharpness by means of calibration type arguments. The minimization of the resulting problem is then tackled numerically and we present results for an extensive set of examples. In particular we are able to address the Steiner tree problem on surfaces.

1 Introduction

In the Steiner tree problem, at least in its classical Euclidean version, we are given N distinct points $P_1, \dots, P_N$ in $\mathbb{R}^d$ and we have to find the shortest connected graph containing the points P_i. From an abstract point of view this amounts to find a graph solving the variational problem

$$\text{(STP)} \qquad \inf\{\mathcal{H}^1(L),\ L \text{ connected},\ L \supset \{P_1, \dots, P_N\}\},$$

where $\mathcal{H}^1$ denotes the one dimensional Hausdorff measure in $\mathbb{R}^d$. An optimal (not necessarily unique) graph L always exists and, by minimality, L is indeed a tree. Every optimal tree can be described as a union of segments connecting the endpoints and possibly meeting at 120° in at most $N-2$ further branch points, called Steiner points.

On the other hand, the (single sink) Gilbert–Steiner problem [20] consists in finding a network L along which to flow unit masses located at the sources $P_1, \dots, P_{N-1}$ to the unique target point P_N. Such a network L can be viewed as $L = \cup_{i=1}^{N-1} \lambda_i$, with λ_i a path connecting P_i to P_N, corresponding to the trajectory of the particle located at P_i. To favour branching, one is led to optimize a cost which is a sublinear (concave) function of the mass density $\theta(x) = \sum_{i=1}^{N-1} \mathbf{1}_{\lambda_i}(x)$: i.e., for $0 \le \alpha \le 1$, find

$$(I_\alpha) \qquad \inf\left\{ E^\alpha(L) = \int_L |\theta(x)|^\alpha d\mathcal{H}^1(x) \right\}.$$

Problem (I_α) can be seen as a particular instance of an α-irrigation problem [8, 32] involving the irrigation of the atomic measures $\sum_{i=1}^{N-1} \delta_{P_i}$ and $(N-1)\delta_{P_N}$, and we notice

*Dipartimento di Matematica, Università di Trento, Italy, e-mail: mauro.bonafini@unitn.it

†Laboratoire Jean Kuntzmann, Université de Grenoble Alpes, France, e-mail: edouard.oudet@imag.fr

that (I_1) corresponds to the Monge optimal transport problem, while (I_0) corresponds to (STP) (the energy to be optimized reduces to the length of L). As for (STP) a solution to (I_α) is known to exist and any optimal network L turns out to be a tree [8].

The Steiner tree problem is known to be computationally hard (even NP complete in certain cases [21]), nonetheless in $\mathbb{R}^2$ and $\mathbb{R}^3$ we have efficient algorithms which allow us to obtain explicit solutions (see for instance [31, 19]), while a comprehensive survey on PTAS algorithms for (STP) can be found in [4, 5]. However, the general applicability of these schemes restricts somehow to the Steiner tree case. For this reason we stick here with a more abstract variational point of view, which allows us to treat in a unified way the Steiner and Gilbert–Steiner problems.

Many different variational approximations for (STP) and/or (I_α) have been proposed, starting form the simple situation where the points P_i lie on the boundary of a convex set: in this case (STP) is known to be an instance of an optimal partition problem [2, 3]. More recently several authors treated these problems in the spirit of Γ-convergence using approximating functionals modelled on Modica–Mortola or Ambrosio–Tortorelli type energies, initially focusing mainly on the two dimensional case [26, 11, 15], lately extending the same ideas also to higher dimensions [16, 10].

Within this sole we introduce in [9] a Γ-convergence type result in the planar case and at the same time we propose a convex framework for the Steiner and Gilbert–Steiner problem. The approach moves from the work of Marchese and Massaccesi [23] and considers ideas from [14] in order to obtain a convex relaxation of the energy we are dealing with. The aim of this paper is then to provide an extensive numerical investigation of the relaxation proposed in [9], adapting it to the treatment of more general Gilbert–Steiner problems (with multiple sources/sinks) and addressing its validity and applicability to problems defined on manifolds. In contrast to classical Γ-convergence type approaches, which may numerically end up in local minima (unless carefully taking initial guesses), this convex formulation is able to identify (in many cases) convex combinations of optimal networks, allowing us to have an idea of their structure. Furthermore, up to our knowledge, this is the very first formulation leading to a numerical approximation of the Steiner tree problem on manifolds.

The paper is organized as follows. In Section 2 we review the convex framework presented in [9] for the α-irrigation problem (I_α) and extend it to the treatment of more general situations with multiple sources/sinks, both in $\mathbb{R}^d$ and on manifolds. In Section 3 we see how the formulation simplifies for a network (STP) on graphs, with the relevant energy reducing to the norm introduced in [23]. We then proceed in Section 4 to describe our algorithmic scheme for the minimization of the proposed energy functional in the Euclidean setting and we present in Section 6 various results for (STP) and α-irrigation problems in two and three dimensions. In Section 7 we eventually detail our algorithmic approach on surfaces and present some results obtained on spheres, tori and other surfaces with boundaries.

2 Convex relaxation for irrigation type problems

In this section we first review the convex framework introduced in [9] for the α-irrigation problem (I_α) and then discuss how this same formulation can be extended to address more general Gilbert–Steiner problems with multiple sources/sinks in $\mathbb{R}^d$ or even on manifolds.

2.1 The Euclidean Gilbert–Steiner problem

Fix a set of N distinct points $A = \{P_1, \dots, P_N\} \subset \mathbb{R}^d$, $d \geq 2$. A candidate minimizer for (I_α) is given as a family of simple rectifiable curves $(\gamma_i)_{i=1}^{N-1}$, each one connecting P_i to P_N. For optimality reasons we can choose these curves so that the resulting network $L = \cup_i \lambda_i$ contains no cycles (see Lemma 2.1 in [23]), restricting this way the set of possible minimizers to the set of (connected) *acyclic graphs* L that can be described as

$$L = \bigcup_{i=1}^{N-1} \lambda_i, \quad \text{s.t.} \quad \begin{array}{l} \cdot\ \lambda_i \text{ is a simple rectifiable curve connecting } P_i \text{ to } P_N, \\ \cdot\ \text{each } \lambda_i \text{ is oriented by an } \mathcal{H}^1\text{-measurable unit vector field } \tau_i, \\ \cdot\ \tau_i(x) = \tau_j(x) \text{ for } \mathcal{H}^1\text{-a.e. } x \in \lambda_i \cap \lambda_j, \end{array}$$

where the last condition requires the $N-1$ pieces composing L to share the same orientation on intersections. Let us call $\mathcal{G}(A)$ the set of acyclic graphs L having such a representation. Hence, we can reduce ourself to consider

$$\inf \left\{ \int_L |\theta(x)|^\alpha d\mathcal{H}^1, \quad L \in \mathcal{G}(A), \ \ \theta(x) = \sum_{i=1}^{N-1} \mathbf{1}_{\lambda_i}(x) \right\}.$$

To each $L \in \mathcal{G}(A)$ we now associate a measure taking values in $\mathbb{R}^{d\times(N-1)}$ as follows: identify the curves λ_i with the vector measures $\Lambda_i = \tau_i \cdot \mathcal{H}^1 \llcorner \lambda_i$, and consider the rank one tensor valued measure $\Lambda = (\Lambda_1, \dots, \Lambda_{N-1})$, which can be written as $\Lambda = \tau \otimes g \cdot \mathcal{H}^1 \llcorner L$, with

- $\tau \colon \mathbb{R}^d \to \mathbb{R}^d$ a unit vector field providing a global orientation for L, satisfying $\operatorname{spt} \tau = L$ and $\tau = \tau_i$ $\mathcal{H}^1$-a.e. on λ_i,
- $g \colon \mathbb{R}^d \to \mathbb{R}^{N-1}$ a multiplicity function whose entries satisfy $g_i \cdot \mathcal{H}^1 \llcorner L = \mathcal{H}^1 \llcorner \lambda_i$.

Observe that $g_i \in \{0, 1\}$ a.e. for any $1 \leq i \leq N-1$ (in particular $g_i(x) = 1$ if $x \in \lambda_i$), and by construction the measures Λ_i verify

$$\operatorname{div} \Lambda_i = \delta_{P_i} - \delta_{P_N}. \tag{2.1}$$

Definition 2.1. *Given any graph $L \in \mathcal{G}(A)$, we call the above constructed measure $\Lambda = \tau \otimes g \cdot \mathcal{H}^1 \llcorner L$ the canonical (rank one) tensor valued measure representation of the acyclic graph L and denote the set of such measures as $\mathcal{L}(A)$.*

Let us define on the space of matrix valued Radon measures $\mathcal{M}(\mathbb{R}^d;\mathbb{R}^{d\times(N-1)})$ the functional

$$\mathcal{F}^\alpha(\Lambda)=\begin{cases}\displaystyle\int_{\mathbb{R}^d}||g||_{1/\alpha}\,d\mathcal{H}^1\llcorner L & \text{if } \Lambda=\tau\otimes g\cdot\mathcal{H}^1\llcorner L\in\mathcal{L}(A)\\ +\infty & \text{otherwise}\end{cases}$$

where we assume $1/0=\infty$. When $\Lambda=\tau\otimes g\cdot\mathcal{H}^1\llcorner L\in\mathcal{L}(A)$, since by construction $g_i\in\{0,1\}$ on L and $g_i(x)=1$ whenever $x\in\lambda_i$, one immediately gets

$$\mathcal{F}^\alpha(\Lambda)=\int_L\left(\sum_{i=1}^{N-1}g_i(x)^{1/\alpha}\right)^\alpha d\mathcal{H}^1=\int_L\left(\sum_{i=1}^{N-1}g_i(x)\right)^\alpha d\mathcal{H}^1=\int_L\left(\sum_{i=1}^{N-1}\mathbf{1}_{\lambda_i}(x)\right)^\alpha d\mathcal{H}^1,$$

which is exactly the cost E^α associated to L in (I_α). We recognize that minimizing $\mathcal{F}^\alpha$ among measures $\Lambda\in\mathcal{L}(A)$ corresponds to minimize E^α among graphs $L\in\mathcal{G}(A)$, and thus solves (I_α) in $\mathbb{R}^d$.

This reformulation of (I_α) involves the minimization of a convex energy, namely $\mathcal{F}^\alpha$, but the problem is still non convex due to the non convexity of $\mathcal{L}(A)$ (the domain of definition of $\mathcal{F}^\alpha$). In view of a convex formulation the optimal choice would be to consider the convex envelope $(\mathcal{F}^\alpha)^{**}$ of the energy, but such an object (up to our knowledge) has no explicit representation. Hence, following [14], we instead look for a "local" convex envelope of the form

$$\mathcal{R}^\alpha(\Lambda)=\int_{\mathbb{R}^d}\Psi_\alpha(\Lambda)\tag{2.2}$$

with $\Psi_\alpha\colon\mathbb{R}^{d\times(N-1)}\to[0,+\infty)$ a 1-homogeneous, convex, continuous function such that $\mathcal{R}^\alpha(\Lambda)=\mathcal{F}^\alpha(\Lambda)$ whenever $\Lambda\in\mathcal{L}(A)$. The integral in (2.2), as outlined in [12], can be defined as

$$\begin{aligned}&\int_{\mathbb{R}^d}\Psi_\alpha(\Lambda)=\int_{\mathbb{R}^d}\Psi_\alpha\left(\frac{d\Lambda_a}{d\mathcal{L}^d}\right)dx+\int_{\mathbb{R}^d}\Psi_\alpha\left(\frac{d\Lambda_s}{d|\Lambda_s|}\right)d|\Lambda_s|\\&=\sup_{\varphi\in C_c^\infty\left(\mathbb{R}^d;\mathbb{R}^{d\times(N-1)}\right)}\left\{\sum_{i=1}^{N-1}\int_{\mathbb{R}^d}\varphi_i\,d\Lambda_i-\int_{\mathbb{R}^d}\Psi_\alpha^*(\varphi)\,dx,\quad\Psi_\alpha^*(\varphi)\in L^1(\mathbb{R}^d)\right\}\end{aligned}\tag{2.3}$$

where $\Lambda=\Lambda_a+\Lambda_s$ is the Lebesgue decomposition of Λ w.r.t. the d-dimensional Lebesgue measure $\mathcal{L}^d$, $|\Lambda_s|$ is the total variation of Λ_s, φ_i are the columns of the function $\varphi(x)=(\varphi_1(x),\dots,\varphi_{N-1}(x))$ and Ψ_α^* is the Legendre-Fenchel conjugate of Ψ_α on $\mathbb{R}^{d\times(N-1)}$: for $p=(p_1,\dots,p_{N-1})\in\mathbb{R}^{d\times(N-1)}$ and $q=(q_1,\dots,q_{N-1})\in\mathbb{R}^{d\times(N-1)}$ we have

$$\Psi_\alpha^*(q)=\sup_p\left[\langle q,p\rangle-\Psi_\alpha(p)\right]=\sup_p\left[\sum_{i=1}^{N-1}q_i\cdot p_i-\Psi_\alpha(p)\right].$$

We immediately see that the evaluation of $\mathcal{R}^\alpha$ on any $\Lambda\in\mathcal{L}(A)$, i.e. $\Lambda=\tau\otimes g\cdot\mathcal{H}^1\llcorner L$ with $||\tau||_2=1$ and $g_i\in\{0,1\}$, only involves the singular part of the decomposition, so that

$$\mathcal{R}^\alpha(\Lambda)=\int_{\mathbb{R}^d}\Psi_\alpha(\tau\otimes g)d\mathcal{H}^1\llcorner L.$$

Since we require $\mathcal{R}^\alpha(\Lambda) = \mathcal{F}^\alpha(\Lambda)$ on these measures, we then look for a 1-homogeneous, convex, continuous function Ψ_α such that

$$\Psi_\alpha(p) = ||g||_{1/\alpha} \quad \text{whenever } p \in K_\alpha = \{\tau \otimes g, \quad ||\tau||_2 = 1,\ g_i \in \{0,1\}\}.$$

The maximal function satisfying this condition can be computed as the 1-homogeneous convex envelope of the function

$$\Phi_\alpha(p) = \begin{cases} ||g||_{1/\alpha} & \text{if } p \in K_\alpha \\ +\infty & \text{otherwise} \end{cases}$$

and, as show in [9], it turns out to be $\Phi_\alpha^{**}(p) = \sup_{q \in K^\alpha} \langle p, q \rangle$, which is to say the support function of the set

$$K^\alpha = \left\{ p \in \mathbb{R}^{d\times(N-1)}, \quad \left\| \sum_{j\in J} p_j \right\|_2 \leq |J|^\alpha \quad \forall\, J \subset \{1, ..., N-1\} \right\},$$

with $|J|$ the cardinality of the set J. Thanks to (2.3), setting $\Psi_\alpha = \Phi_\alpha^{**}$, we can finally define

$$\mathcal{R}^\alpha(\Lambda) = \sup \left\{ \sum_{i=1}^{N-1} \int_{\mathbb{R}^d} \varphi_i \, d\Lambda_i, \quad \varphi \in C_c^\infty(\mathbb{R}^d; K^\alpha) \right\},$$

and consider the relaxed problem

$$\inf \{\mathcal{R}^\alpha(\Lambda), \quad \operatorname{div} \Lambda_i = \delta_{P_i} - \delta_{P_N} \text{ for all } i = 1, \ldots, N-1\}. \tag{2.4}$$

This formulation provides the convex framework we were looking for: the problem is now defined on the whole space of matrix valued Radon measures and the energy is convex as it is a supremum of linear functionals.

However the functional $\mathcal{R}^\alpha$ is obtained only as a "local" convex envelope of $\mathcal{F}^\alpha$ and as such it is not expected to always coincide with the true convex envelope, as we will see in Example 2.2. Thus, given a minimizer $\bar{\Lambda}$ of (2.4) we can end up in three different situations:

1. $\bar{\Lambda} \in \mathcal{L}(A)$, then $\bar{\Lambda}$ is also a minimizer of $\mathcal{F}^\alpha$ and we have solved our original problem;
2. $\mathcal{R}^\alpha(\bar{\Lambda}) = \inf_\Lambda \mathcal{F}^\alpha(\Lambda)$, then $\bar{\Lambda}$ is a convex combination of minimizers of $\mathcal{F}^\alpha$;
3. $\mathcal{R}^\alpha(\bar{\Lambda}) < \inf_\Lambda \mathcal{F}^\alpha(\Lambda)$, which means that the relaxation is not tight and generally speaking minima of $\mathcal{R}^\alpha$ have no relation with minima of $\mathcal{F}^\alpha$.

For a given set of terminal points $A = \{P_1, \ldots, P_N\}$ we will then call the relaxation (2.4) to be tight (or sharp) whenever one of its minimizers satisfies 1. or 2., i.e. whenever its minimizers are related to the actual minimizers of $\mathcal{F}^\alpha$ as it is the case with real convex envelopes. Unfortunately, as the following counterexample shows, the relaxation is not always sharp.

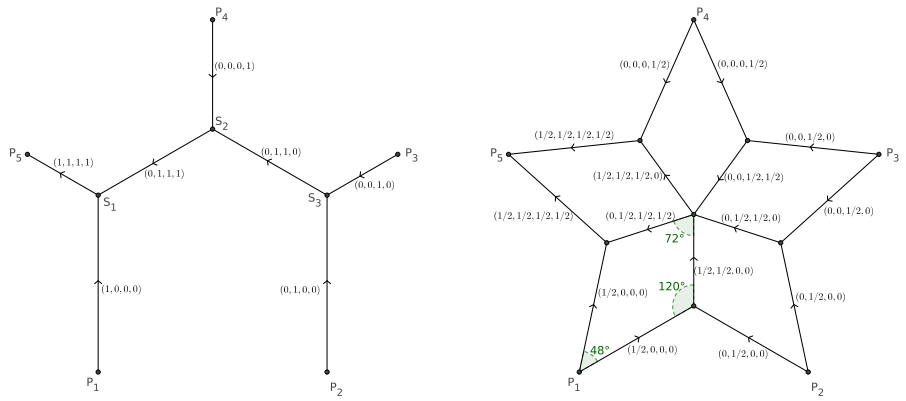


Figure 1: Left: an optimal Steiner tree viewed as its corresponding measure Λ. Right: a rank one tenor valued measure $\Sigma = \tau \otimes g \cdot \mathcal{H}^1 \llcorner L$, with L the graph itself, τ and g as displayed.

Example 2.2. [Non sharpness for pentagon configurations] Consider as terminal points the five vertices of a regular pentagon of side $\ell > 0$ and let $\beta = \frac{3}{10}\pi$. In this situation (STP) has 5 minimizers which are the one in the left picture of figure 1 and its 4 rotations. The energy $\mathcal{R}^0$ of a Steiner tree, which corresponds by construction to its length, is equal to $\ell \tan\beta\,(1 + \sin\beta + \sqrt{3}\cos\beta) \approx 3.8911 \cdot \ell$. However none of the optimal Steiner trees is a minimizer for (2.4). Indeed we can exhibit an admissible tensor valued measure Σ with an energy strictly less than the energy of a Steiner tree: consider for example the rank one tensor valued measure Σ constructed in the right picture of figure 1. Such a measure satisfies the divergence constraints and its energy, which amounts to $1/2$ the length of its support, is equal to $\frac{5}{4}\ell(\sqrt{3} + \tan\beta) \approx 3.8855 \cdot \ell$. Hence we are in the third case of the previous list: the relaxation is not tight and as we already said there is in general no way of reconstructing an optimum for (STP) staring from a minimizer of $\mathcal{R}^0$ (in this case our numerical results suggest Σ as the actual minimizer of $\mathcal{R}^0$). Another example of non-sharpness can be obtained considering as terminal points the vertices of the pentagon plus the center: also in this case Σ has less energy than any optimal Steiner tree.

Despite the previous example, the proposed relaxation can be proved to be sharp in many situations. Indeed, thanks to the duality nature of $\mathcal{R}^\alpha$, we can prove minimality of certain given measures by means of calibration type arguments. This implies that whenever we are able to find a calibration for a given $\bar{\Lambda} \in \arg\min_\Lambda \mathcal{F}^\alpha(\Lambda)$ then the relaxation is sharp because $\bar{\Lambda}$ will also be a minimizer for $\mathcal{R}^\alpha$. A calibration, at least in the simple case of $\mathbb{R}^2$, can be defined as follows

Definition 2.3. *Fix a matrix valued Radon measure $\Lambda = (\Lambda_1, \dots, \Lambda_{N-1})$ and $\varphi \in C_c^\infty(\mathbb{R}^2; K^\alpha)$. We say that φ is a calibration for Λ if $\nabla \times \varphi_i = 0$ for all $i = 1, \dots, N-1$, and φ realizes the supremum in the definition of $\mathcal{R}^\alpha$, i.e.*

$$\sum_{i=1}^{N-1} \int_{\mathbb{R}^2} \varphi_i \, d\Lambda_i = \mathcal{R}^\alpha(\Lambda).$$

The only existence of such an object certifies the optimality of Λ in (2.4). Indeed, let $\Sigma = (\Sigma_1, \dots, \Sigma_{N-1})$ be another competitor, with $\mathcal{R}^\alpha(\Sigma) < \infty$ and $\operatorname{div} \Sigma_i = \delta_{P_i} - \delta_{P_N}$ for each $i = 1, \dots, N-1$. Hence $\operatorname{div}(\Lambda_i - \Sigma_i) = 0$ and we have[1]

$$\int_{\mathbb{R}^2} \varphi_i \, d(\Lambda_i - \Sigma_i) = 0, \tag{2.5}$$

so that

$$\begin{aligned} \mathcal{R}^\alpha(\Lambda) &= \sum_{i=1}^{N-1} \int_{\mathbb{R}^2} \varphi_i \, d\Lambda_i = \sum_{i=1}^{N-1} \left(\int_{\mathbb{R}^2} \varphi_i \, d(\Lambda_i - \Sigma_i) + \int_{\mathbb{R}^2} \varphi_i \, d\Sigma_i \right) \\ &\leq 0 + \mathcal{R}^\alpha(\Sigma) = \mathcal{R}^\alpha(\Sigma). \end{aligned}$$

In $\mathbb{R}^d$ with $d > 2$, the definition of a calibration extends as it is, where now $\nabla \times \varphi_i$ stands for the exterior derivative of the 1-form associated to the vector field φ_i. Also (2.5) generalizes and the proof carries over directly.

For the case $\alpha = 0$, which corresponds to (STP), we can take advantage of calibration arguments of [23] to justify sharpness of (2.4) for some classical choices of $\{P_1, \dots, P_N\}$. Indeed, as we will see in the next section, whenever Λ is a rank one tensor valued measure, for instance whenever it concentrates on a graph and has real-valued weights, $\mathcal{R}^\alpha$ coincides with the norm introduced in [23] to study (STP) as a mass-minimization problem for 1-dimensional currents with coefficients in a suitable normed group. Thus, every calibrated example in that context turns out to be a calibrated configuration in our framework, i.e. a situation where $\mathcal{R}^0$ is sharp (see [23, 24]).

2.2 Extensions: generic Gilbert–Steiner problems and manifolds

The same ideas developed in the previous paragraph can be extended beyond the (single sink) Gilbert–Steiner problem (I_α) in order to address problems with possibly multiple sources/sinks in an Euclidean setting or even problems formulated within manifolds.

Following the strategy introduced in [22] the energy $\mathcal{F}^\alpha$ can also be used to address the general (oriented version of) "who goes where" problem. In this context we do not have to move all the mass to a single sink but instead we are given a family of source/sink couples and we have to move a unit mass from each source to each given destination. Thus, letting $\{S_1, \dots, S_m\} \subset \mathbb{R}^d$ be the set of (unit) sources and $\{T_1, \dots, T_m\} \subset \mathbb{R}^d$ the corresponding set of (unit) sinks, we optimize the same energy E^α involved in the definition of (I_α) but this time among oriented networks of the form $L = \cup_{i=1}^m \lambda_i$, with λ_i a simple rectifiable curve connecting S_i to T_i. The same derivation as above can then be repeated, leading us to the relaxed formulation

$$\inf\{\mathcal{R}^\alpha(\Lambda), \quad \Lambda = (\Lambda_1, \dots, \Lambda_m), \operatorname{div} \Lambda_i = \delta_{S_i} - \delta_{T_i} \text{ for all } i = 1, \dots, m\}. \tag{2.6}$$

We remark that in the previous who goes where problem, differently to what happens in [8], we do not allow two paths λ_i, λ_j to have opposite orientation on intersections, i.e. particles have to go the same way when flowing in the same region.

[1] This generalizes the "smooth" case: thinking to Λ_i and Σ_i as "regular" vector fields we have that $\Lambda_i - \Sigma_i$ is a gradient, whence integrating by parts and using that φ_i is curl-free we get zero.

The previous approach to the "who goes where" problem can now be used within the formulation of more general branched transportation problems, where we are just required to move mass from a set of (unit) sources $\{S_1, \dots, S_m\} \subset \mathbb{R}^d$ to a set of (unit) sinks $\{T_1, \dots, T_m\} \subset \mathbb{R}^d$, without prescribing the final destination of each particle. In this context the problem can be tackled as follows: for every possible coupling between sources and sinks, i.e. among all permutations $\sigma \in \mathcal{S}_m$, solve the corresponding "who goes where" problem with pairs $(S_i, T_{\sigma(i)})_{i=1}^m$, and then take the coupling realizing the minimal energy. Each "who goes where" can be relaxed as done in (2.6), providing this way a relaxed formulation also for the case of generic multiple sources/sinks.

We point out how the extension of the previous discussion to a manifold framework is direct: the derivation that led us to the energy $\mathcal{R}^\alpha$, together with problems (2.4) and (2.6), is still valid on surfaces embedded in the three dimensional space, with the only difference that divergence constraints have to be intended as involving the tangential divergence operator on the given surface.

3 A first simple approximation on graphs

In this section we first see how the previous formulation simplifies when we consider the Steiner tree problem in the context of graphs, in which case the energy reduces to the norm introduced in [23]. Then, once we are able to address (STP) on networks, we try to approximate the Euclidean (STP) by means of a discretization of the domain through an augmented graph.

3.1 The Steiner tree problem on graphs

Consider a connected graph $G = (V, E)$ in $\mathbb{R}^d$, where $V = \{v_1, \dots, v_n\} \subset \mathbb{R}^d$ and $E = \{e_1, \dots, e_m\}$ is a set of m segments. Each $e_j = [v_j^1, v_j^2]$ connects two vertices v_j^1, v_j^2, has length $\ell(e_j) = ||v_j^2 - v_j^1||_2$ and is oriented by $\tau_{e_j} = (v_j^2 - v_j^1)/|v_j^2 - v_j^1|$. Furthermore, we can assume without loss of generality that edges intersect each other in at most 1 point. The Steiner Tree Problem within G can be formulated in the same fashion as its Euclidean counterpart: given a set of terminal points $A = \{P_1, \dots, P_N\} \subset V$ find the shortest connected sub-graph spanning A. As in the Euclidean case a solution always exists and optimal sub-graphs are indeed sub-trees (they contain no cycles).

Following what we did above in the Euclidean case, we can decompose any candidate sub-graph $L \subset G$ into the superposition of $N - 1$ paths λ_i within the graph, each one connecting P_i to P_N. Each path is identified as the support of a flow $V_i \colon E \to \{-1, 0, 1\}$ flowing a unit mass from P_i to P_N: we set $V_i(e) = 1$ if edge e is travelled in its own direction within path λ_i, -1 if it is travelled in the opposite way and 0 otherwise. By construction we satisfy the discrete version of (2.1), i.e. the classical Kirchhoff conditions: for all "interior" vertices $v \in V \setminus \{P_i, P_N\}$ we have

$$\sum_{e \in \delta^+(v)} V_i(e) - \sum_{e \in \delta^-(v)} V_i(e) = 0, \tag{3.1a}$$

with $\delta^{\pm}(v)$ the set of outgoing/incoming edges at vertex v, and (P_i, P_N) is the source/sink couple, meaning

$$\sum_{e\in\delta^+(P_i)} V_i(e) - \sum_{e\in\delta^-(P_i)} V_i(e) = 1, \qquad \sum_{e\in\delta^+(P_N)} V_i(e) - \sum_{e\in\delta^-(P_N)} V_i(e) = -1. \tag{3.1b}$$

Setting $V = (V_1, \dots, V_{N-1})$ and $L = \operatorname{supp} V = \cup\{e \in E : V(e) \neq 0\}$, we have

$$\mathcal{H}^1(L) = \sum_{e\in E} \ell(e) \cdot ||V(e)||_\infty =: \mathcal{F}(V),$$

and as before a solution to the network (STP) can be found minimizing $\mathcal{F}$ among vector valued flows $V\colon E \to \{-1, 0, 1\}^{N-1}$ satisfying the above flux conditions (3.1). Let us identify each family V with a tensor valued measure $\Lambda = (\Lambda_1, \dots, \Lambda_{N-1})$ defined on the whole $\mathbb{R}^d$ by setting

$$\Lambda_i = \sum_{e\in E} V_i(e)\, \tau_e \cdot \mathcal{H}^1 \llcorner e, \quad \text{i.e.} \quad \Lambda = \sum_{e\in E} \tau_e \otimes V(e) \cdot \mathcal{H}^1 \llcorner e. \tag{3.2}$$

The idea is now to drop the integer constraint $\{-1, 0, 1\}$ on each V_i and optimize the previously defined energy $\mathcal{R}^0$ among tensor valued measures of the form (3.2), obtaining the relaxed energy

$$\mathcal{R}(V) = \mathcal{R}^0(\Lambda) = \sup_{\varphi\in C_c^\infty(\mathbb{R}^d;K^0)} \sum_{i=1}^{N-1} \int_{\mathbb{R}^d} \varphi_i \, d\Lambda_i = \sup_{\varphi\in C_c^\infty(\mathbb{R}^d;K^0)} \sum_{i=1}^{N-1} \sum_{e\in E} \left(V_i(e) \int_e \varphi_i \, ds \right).$$

Since edges intersect in at most 1 point it is possible to interpret the last supremum as a supremum over test functions entirely supported on the graph and of the form $\varphi = \sum_e \tau_e \otimes W(e)$ with $W\colon E \to \mathbb{R}^{N-1}$. By assumption, for almost every point x on the graph (except at intersections) there exists only one edge e containing x; hence, the pointwise constraint $\varphi(x) \in K^0$ translates into $\varphi \llcorner e \in K^0$ for all edges $e \in E$, i.e.

$$\left\| \sum_{j\in J} W_j(e)\tau_e \right\|_2 = \left| \sum_{j\in J} W_j(e) \right| \leq 1 \quad \forall\, J \subset \{1, ..., N-1\}.$$

These new constraints involve only vectors $W(e)$ and are equivalent to the unique constraint

$$||W(e)||_* = \left[\sum_{j=1}^{N-1} (W_j(e) \vee 0) \right] \vee \left[-\sum_{j=1}^{N-1} (W_j(e) \wedge 0) \right] \leq 1,$$

which amounts to require that the maximum between the ℓ^1 norm of the positive part and the ℓ^1 norm of the negative part of $W(e)$ has to be less or equal to 1. The energy can be finally rewritten as

$$\begin{aligned} \mathcal{R}(V) &= \sup \left\{ \sum_{e\in E} \ell(e)\, V(e) \cdot W(e), \quad ||W(e)||_* \leq 1 \ \forall e \in E \right\} \\ &= \sum_{e\in E} \ell(e) \left(\sup_i [V_i(e) \vee 0] - \inf_i [V_i(e) \wedge 0] \right) = \sum_{e\in E} \ell(e) ||V(e)||. \end{aligned}$$

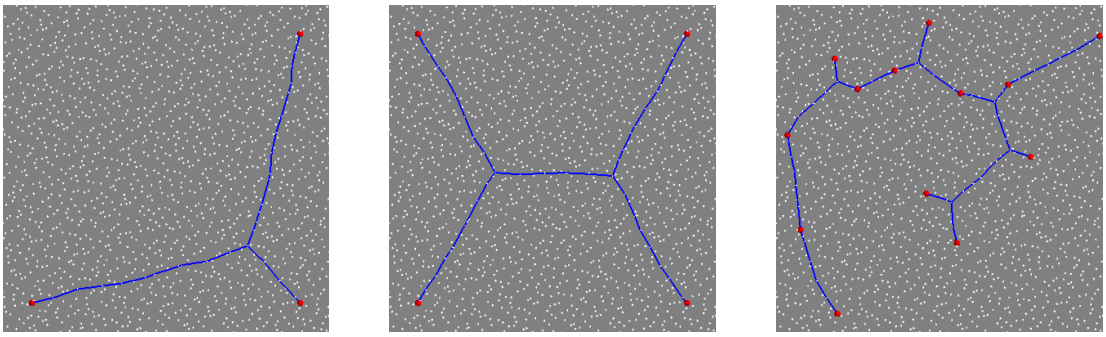


Figure 2: Approximations of (STP) for 3, 4 and 13 terminal points (red) using the augmented graph idea, $K = 1681$, $M = 30$. Edges carrying a non-zero flux are displayed.

The norm $||\cdot||$ is exactly the norm used in [23] to study (STP) using currents with coefficients in normed groups and hence we can take advantage of calibration arguments of [23] to justify the sharpness of the relaxation for calibrated configurations of terminal points. Of course the counterexample 2.2 still applies to this discrete version of the problem using as graph G the union of the two graphs of picture 1: the minimizer concentrates on the star and not on the Steiner structure.

Optimization of $\mathcal{R}$ under the (linear) flux constraints (3.1) can then be performed solving a linear program: in order to linearize the objective we introduce two sets of variables $\{s_e\}_{e\in E}$, $\{i_e\}_{e\in E}$, and for each $e \in E$ we require $i_e \leq 0$, $s_e \geq 0$ and

$$i_e \leq V_i(e) \leq s_e \quad \text{for all } i = 1, \dots, N-1,$$

so that the objective reduces to $\sum_e \ell(e)(s_e - i_e)$. Whenever the size of the resulting linear program is too big to be treated by standard interior point solvers we can alternatively apply the cheaper first order scheme proposed in [27] (see Section 4 for details).

3.2 Graphs and the Euclidean (STP)

Once we have a method to approximate (STP) on networks we can try to address the Euclidean (STP) through the use of an augmented graph. The core idea is the following: let $\{x_1, \dots, x_K\}$, $K \in \mathbb{N}$, be a set of scattered points that uniformly covers an open convex domain Ω such that $\{P_1, \dots, P_N\} \subset \Omega$ and let $V = \{x_1, \dots, x_K\} \cup \{P_1, \dots, P_N\}$. Fix $M \in \mathbb{N}$ and construct the graph $G = (V, E)$ where each $v \in V$ is connected through segments to its M closest neighbours. For M sufficiently large the network G is connected and solving (STP) within G provides an approximation of the underlying Euclidean Steiner tree.

We see in figure 2 two examples with $K = 1681$ and $M = 30$. In both cases results are very close to the optimal Steiner tree and for obtaining them we simply solve a medium scale linear program. However the use of a fixed underlying graph has some drawbacks. For example we cannot expect edges meeting at triple points to satisfy the 120° condition and what should be a straight piece in the optimal tree is only approximated by a

sequence of (non-aligned) edges. A possible remedy for obtaining "straighter" solutions is to increase M, allowing this way longer edges, but this would increase the size of the problem. Furthermore obtaining convex combinations of minimizers is almost impossible because the underlying graph is not regular and having two sub-graphs with the exact same energy is very rare. On the other hand taking regularly distributed points generates many equivalent solutions even when there should be only one.

We also observe that this simplified framework is specific to the Euclidean Steiner tree case: the corresponding graph framework for (I_α) does not end up in a linear program and no direct extension to the manifold case is possible. This lack of generality, together with the intrinsic low precision of the approach as a consequence of working on a graph, leads us to switch our focus on the direct minimization of $\mathcal{R}^\alpha$ on the whole of $\mathbb{R}^2/\mathbb{R}^3$.

4 Generic Euclidean setting, the algorithmic approach

Motivated by the shortcomings of the previous simplified framework, we present in this section our approach for solving (2.4) in $\mathbb{R}^2$ (the same ideas extends to the three dimensional setting). Our resolution is based on a staggered grid for the discretization of the unknowns coupled with a conic solver (or a primal-dual scheme) for the optimization of the resulting finite dimensional problem.

4.1 Spatial discretization

Assume without loss of generality that $P_1, \dots, P_N$ are contained in the interior of $\Omega = [0,1] \times [0,1]$, which will be our computational domain. From a discrete standpoint we view the unknown vector measures $(\Lambda_1, \dots, \Lambda_{N-1})$ in (2.4) as a family $V = (V_1, \dots, V_{N-1})$ of vector fields in Ω and, due to the divergence constraints that we need to satisfy, we discretize these unknown fields on a staggered grid (this way our degrees of freedom are directly related to the flux of each vector field through the given grid interface). Fix then a regular Cartesian grid of size $M \times M$ over Ω and let $h = 1/M$. The first component $V_{i,1}$ of each vector field is placed on the midpoints of the vertical cells interfaces whereas the second components $V_{i,2}$ on the horizontal ones, so that to have on each element (k, ℓ)

$$V_i|_{(k,\ell)} = \begin{pmatrix} (V_{i,1}^{k+1,\ell} - V_{i,1}^{k,\ell})(x - (k-1)h)/h + V_{i,1}^{k,\ell} \\ (V_{i,2}^{k,\ell+1} - V_{i,2}^{k,\ell})(y - (\ell-1)h)/h + V_{i,2}^{k,\ell} \end{pmatrix}.$$

The component $V_{i,1}$ is described by $(M+1) \times M$ unknowns whereas $V_{i,2}$ is described by $M \times (M+1)$ parameters. Regarding the test functions $\varphi = (\varphi_1, \dots, \varphi_{N-1})$ we define them to be piecewise constant on each element of the grid, i.e. for any cell (k, ℓ) we have $\varphi_i^{k,\ell} = (\varphi_{i,1}^{k,\ell}, \varphi_{i,2}^{k,\ell}) \in \mathbb{R}^2$.

Within this setting the optimization of the energy $\mathcal{R}^\alpha$ translates into

$$\min_{(V_{i,d}^{k,\ell})} \sup_{(\varphi_{i,d}^{k,\ell}) \in K^\alpha} \sum_{k,\ell} \sum_{i=1}^{N-1} h^2 \left[\frac{V_{i,1}^{k,\ell} + V_{i,1}^{k+1,\ell}}{2} \varphi_{i,1}^{k,\ell} + \frac{V_{i,2}^{k,\ell} + V_{i,2}^{k,\ell+1}}{2} \varphi_{i,2}^{k,\ell} \right] \tag{4.1}$$

under the condition $\operatorname{div} V_i = \delta_{P_i} - \delta_{P_N}$ for all $i = 1, \dots, N-1$. Since the flux of each V_i over the generic cell (k, ℓ) is given by

$$F_i^{k,\ell} = h(V_{i,1}^{k+1,\ell} - V_{i,1}^{k,\ell}) + h(V_{i,2}^{k,\ell+1} - V_{i,2}^{k,\ell}),$$

the divergence constraints translate, at a discrete level, into

$$\begin{cases} F_i^{k,\ell} = 0 & \text{whenever cell } (k,\ell) \text{ does not contain } P_i \text{ or } P_N, \\ F_i^{k,\ell} = 1 & \text{if cell } (k,\ell) \text{ contains } P_i, \\ F_i^{k,\ell} = -1 & \text{if cell } (k,\ell) \text{ contains } P_N, \end{cases} \tag{4.2}$$

complemented with a "zero flux" condition at the boundary, i.e. we set $V_{i,d}^{k,\ell} = 0$ whenever it refers to a boundary interface. We finally observe that, by construction, $\varphi \in K^\alpha$ if for each cell (k, ℓ) in the grid the matrix $\varphi^{k,\ell} = (\varphi_1^{k,\ell}, \dots, \varphi_{N-1}^{k,\ell})$ satisfies

$$\left\| \sum_{j \in J} \varphi_j^{k,\ell} \right\|_2 \leq |J|^\alpha \text{ for all } J \subset \{1, \dots, N-1\}.$$

For the resolution of this finite dimensional optimization problem we then propose two different and somehow complementary approaches.

4.2 Optimization via conic duality

The inf-sup problem (4.1) can be written, thanks to conic duality (see e.g. Lecture 2 of [7]), as a pure minimization problem involving the degrees of freedom $(V_{i,d}^{k,\ell})$ and a set of dual variables $(\psi_{J,d}^{k,\ell})$ indexed over subsets $J \subset \{1, \dots, N-1\}$. Indeed, for fixed $1 \leq k, \ell \leq M$ and $J \subset \{1, \dots, N-1\}$, one has

$$\inf_{\psi_J^{k,\ell} \in \mathbb{R}^2} \left(|J|^\alpha ||\psi_J^{k,\ell}||_2 - \langle\, \psi_J^{k,\ell}, \sum_{j \in J} \varphi_j^{k,\ell} \,\rangle \right) = \begin{cases} 0 & \text{if } \left\| \sum_{j \in J} \varphi_j^{k,\ell} \right\|_2 \leq |J|^\alpha \\ -\infty & \text{otherwise,} \end{cases}$$

so that, if we denote $\tilde{V}_i^{k,\ell} = ((V_{i,1}^{k,\ell} + V_{i,1}^{k+1,\ell})/2, (V_{i,2}^{k,\ell} + V_{i,2}^{k,\ell+1})/2) \in \mathbb{R}^2$, (4.1) is equivalent to

$$\min_{(V_{i,d}^{k,\ell})} \sup_{(\varphi_{i,d}^{k,\ell})} \left[\sum_{k,\ell} \sum_{i=1}^{N-1} h^2 \langle \tilde{V}_i^{k,\ell}, \varphi_i^{k,\ell} \rangle + \inf_{(\psi_{J,d}^{k,\ell})} \sum_{k,\ell} \sum_{J} h^2 \left(|J|^\alpha ||\psi_J^{k,\ell}||_2 - \langle\, \psi_J^{k,\ell}, \sum_{j \in J} \varphi_j^{k,\ell} \,\rangle \right) \right].$$

Switching the sup over $(\varphi_{i,d}^{k,\ell})$ and the inf over $(\psi_{J,d}^{k,\ell})$ we obtain

$$\min_{(V_{i,d}^{k,\ell})} \inf_{(\psi_{J,d}^{k,\ell})} \left[\sum_{k,\ell} \sum_{J} h^2 |J|^\alpha ||\psi_J^{k,\ell}||_2 + h^2 \sup_{(\varphi_{i,d}^{k,\ell})} \sum_{k,\ell} \left(\sum_{i=1}^{N-1} \langle \tilde{V}_i^{k,\ell}, \varphi_i^{k,\ell} \rangle - \langle\, \psi_J^{k,\ell}, \sum_{j \in J} \varphi_j^{k,\ell} \,\rangle \right) \right].$$

Since the inner sup is either 0 if $\tilde{V}_i^{k,\ell} = \sum_{J \ni i} \psi_J^{k,\ell}$ for all $1 \le k, \ell \le M$ and $1 \le i \le N-1$ or $+\infty$ otherwise, the previous problem eventually leads to

$$\min_{(V_{i,d}^{k,\ell}),(\psi_{J,d}^{k,\ell})} \sum_{k,\ell} \sum_{J} h^2 |J|^\alpha \, \|\psi_J^{k,\ell}\|_2 \tag{4.3}$$

where each V_i satisfies the same flux constraints (4.2) and for all cells (k, ℓ) and all $i = 1, \dots, N-1$ we must satisfy

$$\frac{V_{i,1}^{k,\ell} + V_{i,1}^{k+1,\ell}}{2} = \sum_{J \ni i} \psi_{J,1}^{k,\ell} \quad \text{and} \quad \frac{V_{i,2}^{k,\ell} + V_{i,2}^{k,\ell+1}}{2} = \sum_{J \ni i} \psi_{J,2}^{k,\ell}. \tag{4.4}$$

Problem (4.3) under the set of linear constraints (4.2) and (4.4) can now be solved invoking the conic solver of the library MOSEK [25] within the framework provided by [17].

4.3 Optimization via primal-dual schemes

Collect all the $(V_{i,d}^{k,\ell})$ into a vector $\mathbf{v} \in \mathbb{R}^{n_v}$, $n_v = (N-1)(2M^2 + 2M)$, and all the $(\varphi_{i,d}^{k,\ell})$ into $\boldsymbol{\varphi} \in \mathbb{R}^{n_\varphi}$, $n_\varphi = (N-1)2M^2$. Moving the constraints on φ into the objective via the convex indicator function, the discrete energy (4.1) can be written down as

$$\min_{\mathbf{v}} \max_{\boldsymbol{\varphi}} \langle \boldsymbol{\varphi}, B\mathbf{v} \rangle - \chi_{K^\alpha}(\boldsymbol{\varphi})$$

for a suitable (sparse) matrix B of size $n_\varphi \times n_v$, while the divergence constraints reduce to $A\mathbf{v} = \mathbf{b}$ for a suitable (sparse) matrix A of size $n_\lambda \times n_v$ and a vector $\mathbf{b} \in \mathbb{R}^{n_\lambda}$. To the set of liner constraints $A\mathbf{v} - \mathbf{b} = 0$ we can now associate a dual variable $\boldsymbol{\lambda} \in \mathbb{R}^{n_\lambda}$ so that they can be incorporated into the objective as

$$\min_{\mathbf{v}} \max_{\boldsymbol{\varphi}, \boldsymbol{\lambda}} \langle \boldsymbol{\varphi}, B\mathbf{v} \rangle - \chi_{K^\alpha}(\boldsymbol{\varphi}) + \langle \boldsymbol{\lambda}, A\mathbf{v} - b \rangle.$$

The problem, written this way, turns into an instance of a general inf-sup problem of the form

$$\min_{x \in \mathbb{R}^n} \max_{y \in \mathbb{R}^m} \langle y, Kx \rangle + G(x) - F^*(y) \tag{4.5}$$

with K an $m \times n$ matrix and $G \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, $F^* \colon \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ convex lsc functions. Among the possible numerical schemes which have been developed in the literature for the resolution of (4.5) we choose here the preconditioned primal-dual scheme presented in [27]. The scheme can be summarized as follows: let $\gamma \in [0, 2]$, $T = \operatorname{diag}(\tau_1, \dots, \tau_n)$ and $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_m)$, with

$$\tau_j = \frac{1}{\sum_{i=1}^m |K_{ij}|^{2-\gamma}} \quad \text{and} \quad \sigma_i = \frac{1}{\sum_{j=1}^n |K_{ij}|^\gamma},$$

fix $x^0 \in \mathbb{R}^n$, $y^0 \in \mathbb{R}^m$, and iterate for any $k > 0$

$$\begin{cases} x^{k+1} = (I + T\partial G)^{-1}(x^k - TK^T y^k) \\ y^{k+1} = (I + \Sigma \partial F^*)^{-1}(y^k + \Sigma K(2x^{k+1} - x^k)) \end{cases} \tag{4.6}$$

In this context the proximal mappings are defined as

$$(I + T\partial G)^{-1}(\hat{x}) = \arg\min_x \left[G(x) + \frac{1}{2}\langle T^{-1}(x - \hat{x}),\, x - \hat{x}\rangle \right]$$

and represent the extension of the classical definition with constant step size to this situation with "variable dependent" step sizes.

In our specific use case the scheme takes the following form: define $T = \mathrm{diag}(\tau_1, \dots, \tau_{n_v})$, $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_{n_\varphi})$ and $\tilde{\Sigma} = \mathrm{diag}(\tilde{\sigma}_1, \dots, \tilde{\sigma}_{n_\lambda})$, with

$$\tau_j = \frac{1}{\sum_{i=1}^{n_\varphi} |B_{ij}|^{2-\gamma} + \sum_{i=1}^{n_\lambda} |A_{ij}|^{2-\gamma}}, \quad \sigma_i = \frac{1}{\sum_{j=1}^{n_v} |B_{ij}|^{\gamma}}, \quad \tilde{\sigma}_i = \frac{1}{\sum_{j=1}^{n_v} |A_{ij}|^{\gamma}},$$

given $\mathbf{v}^0, \boldsymbol{\varphi}^0, \boldsymbol{\lambda}^0$ iterate for $k > 0$

$$\begin{cases} \mathbf{v}^{k+1} = \mathbf{v}^k - T(B^T \boldsymbol{\varphi}^k + A^T \boldsymbol{\lambda}^k) \\ \boldsymbol{\varphi}^{k+1} = \mathrm{proj}(\boldsymbol{\varphi}^k + \Sigma B(2\mathbf{v}^{k+1} - \mathbf{v}^k) \mid K^\alpha) \\ \boldsymbol{\lambda}^{k+1} = \boldsymbol{\lambda}^k + \tilde{\Sigma}(A(2\mathbf{v}^{k+1} - \mathbf{v}^k) - \mathbf{b}) \end{cases} \tag{4.7}$$

The computational bottleneck for this simple iterative procedure resides in the projection of a given vector $\bar{\boldsymbol{\varphi}} \in \mathbb{R}^{n_\varphi}$ onto the convex set K^α. By definition this operation reduces to the cell-wise projection on K^α of the matrices $\varphi^{k,\ell}$, and so we fix a $d \times (N-1)$ matrix $q = (q_1, \dots, q_{N-1})$ and split the discussion into two sub-steps.

Projection on individual sets: for each fixed subset $J \subset \{1, \dots, N-1\}$ we define the convex set

$$K_J^\alpha = \left\{ p \in \mathbb{R}^{d \times (N-1)}, \quad \left\| \sum_{j \in J} p_j \right\|_2 \leq |J|^\alpha \right\}.$$

The projection of q over K_J^α can be computed explicitly: define $v_J = \sum_{j \in J} q_j$, then the projection $p = \mathrm{proj}(q \mid K_J^\alpha) = (p_1, \dots, p_{N-1})$ has columns defined as $p_j = q_j$ if $j \notin J$ and

$$p_j = q_j - 1/|J|\, (\|v\|_2 - |J|^\alpha)^+ \frac{v}{\|v\|_2} \quad \text{if } j \in J.$$

Projection on the intersection: observe that $K^\alpha = \cap_J K_J^\alpha$, i.e. K^α is the intersection of a family of convex sets. In order to get an approximation of $\mathrm{proj}(q \mid K^\alpha)$ we can apply the Dykstra's projection algorithm (see [18]). The scheme in our setting is the following: let $J_1, \dots, J_{2^{N-1}}$ be all the subsets of $\{1, \dots, N-1\}$, let $\{y_j^0\}_{j=1}^{2^{N-1}}$ be 2^{N-1} null matrices of size $d \times (N-1)$, $p^0 = q$, then for any $k \geq 1$ iterate

$$\begin{cases} p_0^k = p^{k-1} \\ \text{for } j = 1, \dots, 2^{N-1} \\ \qquad p_j^k = \mathrm{proj}(p_{j-1}^k + y_j^{k-1} \mid K_{J_j}^\alpha) \\ \qquad y_j^k = y_j^{k-1} + p_{j-1}^k - p_j^k \\ \text{end for} \\ p^k = p_{2^{N-1}}^k \end{cases}$$

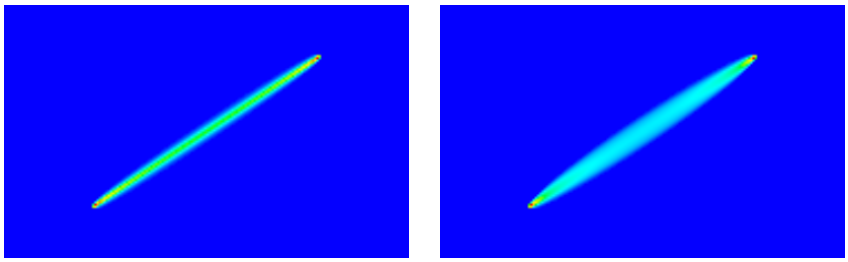


Figure 3: Energy concentration for the minimizer of $\mathcal{R}^0$ for $P_1 = (1/4, 1/3)$, $P_2 = (3/4, 2/3)$. Left: solution obtained via a conic solver, final energy ≈ 0.606307. Right: solution obtained after 200000 iterations of the primal-dual scheme, $\gamma = 0.6$, final energy ≈ 0.606765.

We then have $p^k \to \text{proj}(q \mid K^\alpha)$ as $k \to +\infty$.

Remark 4.1. Each step of the previous iterative projection procedure requires 2^{N-1} sub-projections and thus the scheme is intrinsically time-consuming. Up to our knowledge there seems to be no immediate simplifications to avoid some of the 2^{N-1} inner projections: for example the restriction of the inner loop over sets $K^\alpha_{J_j}$ such that $q \notin K^\alpha_{J_j}$ is not going to work in general. At the same time we observe that established convergence rates for (4.6) do not apply in this case because our projection, which represents one of the two proximal mappings, is only approximated and not exact, making us falling back in a context like [30].

5 Numerical details

The two resolution paths presented above allow us to overcome some shortcomings of the simplified framework of Section 3 but introduces at the same time an higher computational cost, mainly depending on the combinatorial nature of the set K^α, which reflects in the high number of variables involved in (4.3) and in the complicated projection in (4.7).

Generally speaking the primal-dual scheme is the cheapest of the two in terms of computational resources: it can be implemented so that every operation is done in-place, reducing to almost zero any further memory requirement apart from initialization, while the interior point approach used by a conic solver is extremely demanding in terms of memory due to the 2^{N-1} additional variables needed to define (4.3). However, since we are looking for 1d structures, our solver also needs to be able to provide very localized optima, and with this regards the primal-dual approach is not very satisfactory. As we can see in figure 3, where we use the two schemes for the same regular 201×201 grid over $[0, 1]^2$, the solution provided by the primal-dual scheme is more diffused than the one obtained using the conic approach. For this reason we would like to use for our experiments the conic formulation (4.3) and to do so, in order to be able to treat medium scale problems, we need to find a way to reduce a-priori the huge number of additional variables that are introduced: this can be done both via a classical grid refinement and via a variables "selection".

5.1 Grid refinement

The numerical solution is expected to concentrate on a 1-dimensional structure, and so the grid needs to be fine only on a relatively small region of the domain. This suggests the implementation of a refinement strategy able to localize in an automatic way the region of interest. For doing so we use non-conformal quadtree type meshes (see e.g. [29, 6]), which are a particular class of grids where the domain is partitioned using M square cells as $\Omega = \cup_m S_m$ and each square cell S_m can be obtained by recursive subdivision of the box $[0,1]^2$ (see figure 4 for examples of such grids). As in the case of uniform regular meshes we employ for the discretization a staggered approach: we set the degrees of freedom of vector fields on faces of each element, with the additional requirement that whenever a face is also a subsegment of another face then the two associated degrees of freedom are equal (this is to maintain continuity of the normal components of the discrete fields across edges and guarantees that fluxes are globally well behaved). The matrix valued function φ is again defined to be constant on each element of the grid so that the nature of the discrete problem we need to solve remains the same.

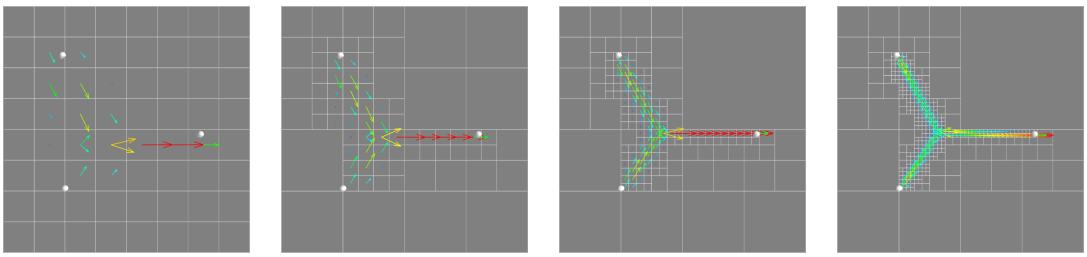


Figure 4: Refinement example for 3 points. At each iterate we plot the grid and the two fields V_1, V_2, which are then used to build the next grid.

A refinement procedure can then be described as follows: fix a coarse quadtree grid $\mathcal{T}$, for example a regular 8×8 one, and then

- solve the problem on the given grid $\mathcal{T}$;
- identify elements of the grid where the solution concentrates the most and label them as "used", identify elements of the grid where the solution is almost zero and label them as "unused";
- refine the grid subdividing each "used" element into 4 equal sub-elements and try to merge "unused" elements into bigger ones (the merging will occur if four elements labelled as "unused" have the same father in the quadtree structure);
- repeat.

As we can see in figure 4 this procedure allows us to localize computations in a neighbourhood of the optimal structure we are looking for. This way we can attain a good level of fineness around the solution without being forced to employ a full grid which would require the introduction of a lot of useless degrees of freedom.

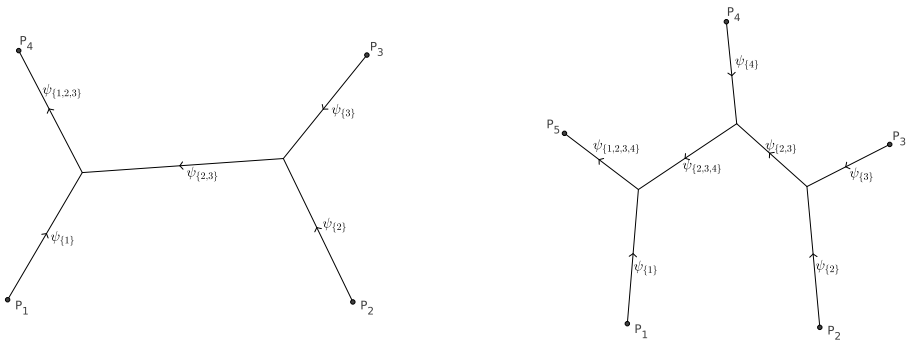


Figure 5: Expected behaviour of the variables $\{\psi_J\}_J$. We can see how each V_i can be reconstructed as the union of the ψ_J such that $i \in J$ and that only a subset of the ψ_J is used.

5.2 Variables selection

Generally speaking, in an optimum for (4.3) most of the variables ψ_J will turn out to be identically 0 while the ones that are not 0 everywhere will be concentrated on small regions of the domain. Indeed each ψ_J can be seen as a possible building block of the final solution because, due to formula (4.4), the vector field $\psi_{\{j_1,\dots,j_k\}}$, $\{j_1,\dots,j_k\} \subset \{1,\dots,N-1\}$, represents the portion of the graph where the fields $V_{j_1},\dots,V_{j_k}$ coincide (see for example figure 5 for a visual depiction in two cases). This means that we expect only a few ψ_J to be non zero on each element of the grid. With this in mind we can add the following selection procedure to the previous refinement scheme: given an approximate solution on a grid $\mathcal{T}$, we identify for each square element S_m which are the non zero variables $\psi_{J_1^m},\dots,\psi_{J_{km}^m}$ on that element and then, at the next step, we introduce only these variables in that particular region (in case the element S_m is one of those labelled as "used" this means that in the next optimization we will use only $\psi_{J_1^m},\dots,\psi_{J_{km}^m}$ within its 4 children).

The main advantage of this procedure is clear: once we are able to identify the regions where each variable ψ_J concentrates (if any) we can dramatically reduce the number of unknowns we need to introduce, passing from 2^{N-1} vector fields to be defined on each element to only a few of them. Thanks to these two refinement procedures we are now in a position to efficiently tackle the optimization of $\mathcal{R}^\alpha$ using accurate conic solvers.

6 Results in flat cases

We present in this section different results obtained using the outlined scheme integrated with the two refinement procedures described above.

In figure 6 we compute minimizers of the relaxed energy $\mathcal{R}^0$ for regular configurations of terminal points placed on the vertices of a triangle, a square, a pentagon and an hexagon. In all cases we start with a regular 32×32 mesh and then apply the previous refinement procedures 5 times, ending up with a grid size of $1/1024$ around the optimal

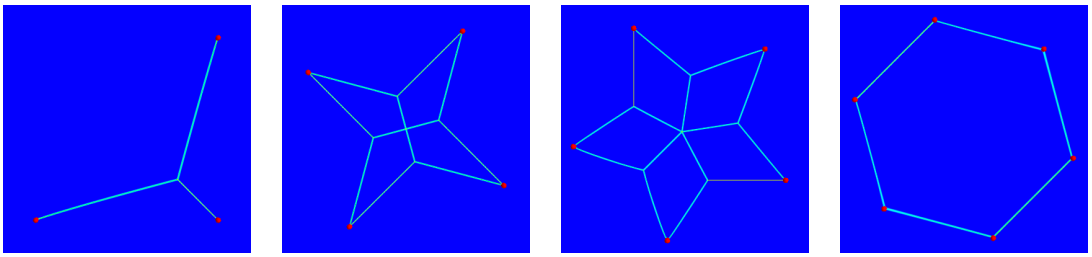


Figure 6: Optima of $\mathcal{R}^0$ in $\mathbb{R}^2$ for $3, 4, 5, 6$ terminal points on the vertices of regular polygons.

structure. In the first example we are able to retrieve the unique minimizer while in the second example we obtain a convex combination of the two possible minimizers for (STP). In the latter case this behaviour is expected because for this particular configuration of points the relaxation is sharp due to the calibration argument presented in [23]. In the third experiment we recover the star-shaped counterexample of figure 1 which seems to be the actual minimizer of the relaxed problem and in the last picture we get a convex combination of the six possible minimizers. We remark that the hexagon case is not a calibrated example in the work of Marchese–Massaccesi but our numerical result suggests the existence of a calibration because the relaxation seems to be sharp.

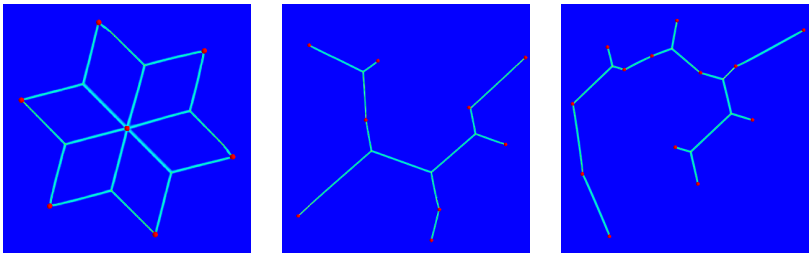


Figure 7: Optima of $\mathcal{R}^0$ in $\mathbb{R}^2$ for $7, 9$ and 13 terminal points.

In figure 7 we first compute a minimizer for a 7 points configuration (6 vertices of the hexagon plus the center) and observe how we are able to obtain a convex combination of the two Steiner trees (again this was expected due to a calibration argument). We observe that in this example the points do not lie on the boundary of a convex set, meaning that the problem cannot be simplified into an optimal partition problem as it is done for example in [14]. We then move to some non symmetric distributions of terminal points: in the second picture we see the result for 9 randomly selected points while in the third one we increase the number of terminals up to 13. In this last case an ad-hoc approach is necessary. Due to the high number of variables introduced in (4.1) a direct minimization using a conic solver is unfeasible even for very coarse grids (the amount of memory required to just set up the interior point solver is too much). To circumvent this problem we first compute a rough solution either optimizing $\mathcal{R}^0$ on a coarse grid using the primal-dual minimization scheme or applying the augmented graph idea presented in

section 3 (see picture 2), and then we use this approximation for deducing which are the variables ψ_J active at a given point: for every cell (k, ℓ) of a uniform grid we introduce $\psi_{\{j_1,\dots,j_k\}}$ on that cell only if in the approximate solution every field $V_{j_1}, \dots, V_{j_k}$ is not identically zero in a suitable neighbourhood of the cell. This way we rule out a huge amount of the ψ_J obtaining a problem which is now tractable through interior point schemes.

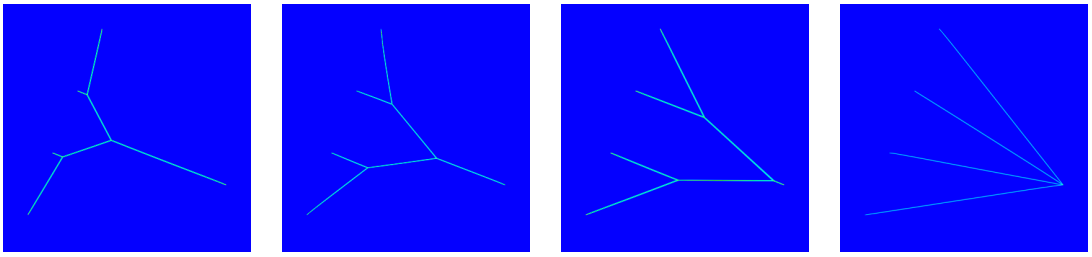


Figure 8: Irrigation networks minimizing $\mathcal{R}^\alpha$ and moving 4 masses to a unique sink, $\alpha = 0.6, 0.8, 0.95, 1$.

In figure 8 we test the relaxation $\mathcal{R}^\alpha$ for a simple irrigation problem where we approximate the shape of the optimal network moving 4 unit masses located at $S_1 = (0.4, 0.9)$, $S_2 = (0.3, 0.65)$, $S_3 = (0.2, 0.4)$, $S_4 = (0.1, 0.15)$, to the unique sink $T = (0.9, 0.27)$. We can see how for small α the optimal shape is close to the optimal Steiner tree while for higher values of α the network approaches more and more the configuration for an optimal Monge–Kantorovitch transport attaining it for $\alpha = 1$ as expected.

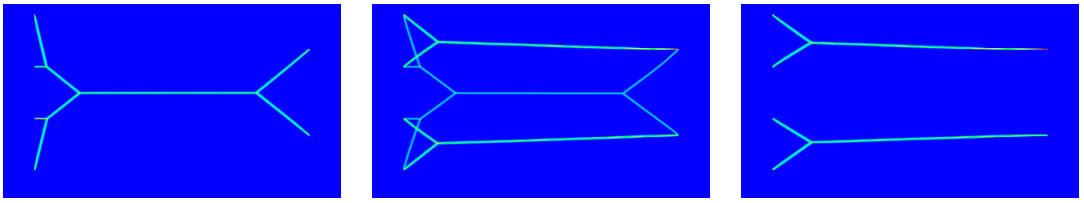


Figure 9: Optima of $\mathcal{R}^\alpha$ for moving 4 masses from left to right, $\alpha = 0.65, 0.7, 0.75$. The pairings realizing the first infimum and the third one are different.

We turn next in figure 9 to an example where 4 unit masses located at 4 sources on the left ($S_1 = (0.1, 0.55)$, $S_2 = (0.1, 0.4)$, $S_3 = (0.1, 0.25)$, $S_4 = (0.1, 0.1)$) has to be moved to 2 sinks of magnitude 2 on the right ($T_1 = (0.9, 0.2)$, $T_2 = (0.9, 0.45)$). Since for each mass we have two possible destinations we need to loop over all feasible combinations of source/sink couples, solve the corresponding "who goes where" problem and then choose the one giving the optimizer with less energy. In the examples the optimal couplings are $\{(S_1, T_1), (S_2, T_1), (S_3, T_2), (S_4, T_2)\}$ for $\alpha = 0.65$ and $\{(S_1, T_2), (S_2, T_2), (S_3, T_1), (S_4, T_1)\}$ for $\alpha = 0.75$. In the case $\alpha = 0.7$ we are at the switching point between a connected and a disconnected optimal structure and our relaxed optimum concentrates on both.

The numerical scheme we have described for the two dimensional case can be extended directly to the three dimensional context for addressing the optimization of $\mathcal{R}^\alpha$ in $\mathbb{R}^3$.

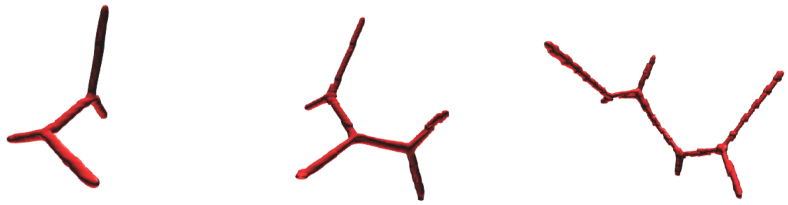


Figure 10: Optima of $\mathcal{R}^0$ for $4, 5$ and 7 points in $\mathbb{R}^3$.

Non-conformal quadtree type grids are replaced by non-conformal octree type grids (see [29]) and a staggered approach is employed placing the degrees of freedom on faces of each cubic element composing the grid. The underlying structure of the discrete optimization we end up with remains the same and the two refinement procedures can be extended as they are, without any major change. We see in figure 10 the results for $4, 5$ and 7 points configurations. All the examples are purely 3-dimensional and in the first two cases we have a maximum number of Steiner points (respectively 2 and 3), while in the last case the optimal structure consists of two "disjoint" optimal sub-trees connected through a central terminal point.

7 Extension to surfaces

As already observed in Section 2 the proposed relaxation can also be used to address (STP) and α-irrigation problems on surfaces. Up to our knowledge, even in the Steiner tree case, this is the first numerical approximation of these problems covering the manifold framework. Theoretically speaking what we need to do is to solve problem (2.4) on a manifold S embedded in $\mathbb{R}^3$, where now a candidate minimizer Λ is a matrix valued measure defined on the manifold and divergence constraints translate accordingly. From a numerical point of view our unknowns are again vector fields $(V_1, \dots, V_{N-1})$ living on the surface and the domain will be approximated by means of a triangulated surface $\mathcal{T}_h$. We first discuss the direct extension of the staggered grid idea to $\mathcal{T}_h$ and then present a more accurate discretization, eventually used in our experiments.

7.1 Raviart–Thomas approach

The staggered approach presented for quadrilateral grids can be extended to triangular meshes considering a discretization based on the so-called Raviart–Thomas basis functions, which are vector valued functions whose degrees of freedom are related to the flux of the given basis function across edges (see [13]).

Let $\mathcal{T}_h$ be a regular triangulation of S, with n vertices and m edges, and consider the lowest order Raviart–Thomas basis functions over $\mathcal{T}_h$: for each edge e in the triangulation we call K_- the "left" triangle adjacent to e and K_+ the "right" triangle adjacent to e

(according to a given fixed orientation) and define the vector function

$$\Phi_e(x) = \begin{cases} \dfrac{\ell_e}{2A_e^+}(x - p_+) & \text{if } x \in K_+ \\ -\dfrac{\ell_e}{2A_e^-}(x - p_-) & \text{if } x \in K_- \\ (0,0,0) & \text{otherwise} \end{cases}$$

where ℓ_e is the length of the edge, $A_e^\pm = |K_\pm|$ are the areas of the triangles and p_+, p_- are the opposite corners (with the obvious modification for boundary edges). We then approximate each V_i, $i = 1, \ldots, N-1$, as

$$V_i(x) = \sum_{e=1}^{m} V_i^e \, \Phi_e(x)$$

and as before matrix valued variables $\varphi = (\varphi_1, \ldots, \varphi_{N-1})$ are considered to be piecewise constant over each element of the triangulation, i.e. $\varphi_i|_K = \varphi_i^K = (\varphi_{i,1}^K, \varphi_{i,2}^K, \varphi_{i,3}^K) \in \mathbb{R}^3$ for all $K \in \mathcal{T}_h$, $i = 1, \ldots, N-1$. The unknowns are then the family of parameters (V_i^e) and $(\varphi_{i,d}^K)$. Looking at $\mathcal{R}^\alpha$ the integral we need to compute becomes

$$\sum_{K \in \mathcal{T}_h} \sum_{i=1}^{N-1} \int_K \left(\sum_{e=1}^{m} V_i^e \, \Phi_e(x) \right) \cdot \varphi_i^K \, dx, \tag{7.1}$$

and can be made explicit as follows: let e_j^K be the edge of triangle K opposite to point P_j^K (j-th point of triangle K) and $s^{K,e_j^K} = \pm 1$ the position of that triangle with respect to the edge e_j^K, then (7.1) yields

$$\frac{1}{6} \sum_{K \in \mathcal{T}_h} \sum_{i=1}^{N-1} \Big[s^{K,e_1^K} \ell_{e_1^K} V_i^{e_1^K} (P_2^K + P_3^K - 2P_1^K) \varphi_{i,1}^K + s^{K,e_2^K} \ell_{e_2^K} V_i^{e_2^K} (P_1^K + P_3^K - 2P_2^K) \varphi_{i,2}^K + s^{K,e_3^K} \ell_{e_3^K} V_i^{e_3^K} (P_1^K + P_2^K - 2P_3^K) \varphi_{i,3}^K \Big].$$

The structure of the discrete energy is the same as the one obtained in the Euclidean setting (the conditions on φ translates again in the element-wise constraint $\varphi^K \in K^\alpha$ for all $K \in \mathcal{T}_h$). Furthermore within this Raviart–Thomas framework we have two advantages: fields V_i are by construction surface vector fields (i.e. they live in the tangent space to the surface) and divergence constraints translate into simple flux conditions of the form

$$s^{K,e_1^K} \ell_{e_1^K} V_i^{e_1^K} + s^{K,e_2^K} \ell_{e_2^K} V_i^{e_2^K} + s^{K,e_3^K} \ell_{e_3^K} V_i^{e_3^K} = 0 \text{ or } \pm 1$$

depending on K containing P_i, P_N or none of them, and $V_i^e = 0$ whenever e is a boundary edge. The price to pay for such simplicity resides in the fact that this Raviart–Thomas approximation is a low-order scheme. The objects we would like to approximate are singular vector fields concentrated on 1-dimensional structures but with this approach we

generally obtain solutions that are quite diffused and can only give us an approximate idea of the underlying optimal set. At the same time this diffusivity prevents a good refinement because the refined region turns out to be too large. For this reason a better approximation space is needed, even if we will end up with a more complex discrete problem.

7.2 $\mathbb{P}_2$-based approach

Let $\mathcal{T}_h$ be a regular triangulation of S. We consider the standard discrete space

$$X_h^2 = \{v_h \in C^0(\mathcal{T}_h) \,:\, v_h|_K \in \mathbb{P}_2(K), \text{ for all } K \in \mathcal{T}_h\}$$

and take vector fields $V_i \in (X_h^2)^3$ for all $i = 1, \dots, N-1$. As in the staggered case matrix valued variables $\varphi = (\varphi_1, \dots, \varphi_{N-1})$ are defined to be piecewise constant over each element of the triangulation, i.e. $\varphi_i|_K = \varphi_i^K = (\varphi_{i,1}^K, \varphi_{i,2}^K, \varphi_{i,3}^K)$ for all $K \in \mathcal{T}_h$, $i = 1, \dots, N-1$. The energy $\mathcal{R}^\alpha$ is then

$$\sup \left\{ \sum_{K \in \mathcal{T}_h} \sum_{i=1}^{N-1} \int_K V_i \cdot \varphi_i^K \, dx, \quad \varphi^K \in K^\alpha \text{ for all } K \in \mathcal{T}_h \right\}$$

and the integral over each triangle K can be computed explicitly in terms of the degrees of freedom associated to V_i and φ_i, $i = 1, \dots, N-1$ (the integrand reduces to a polynomial of degree 2). We are left with the specification of how we impose divergence and tangency constraints on each V_i, $i = 1, \dots, N-1$.

Divergence constraints: for each vector field V_i we have to impose $\operatorname{div} V_i = \delta_{P_i} - \delta_{P_N}$, where this time the divergence has to be interpreted as the tangential divergence operator on surfaces (see for instance [28]). We observe that $\operatorname{div} V_i$ is piecewise linear over each element of the triangulation and thus, for $K \in \mathcal{T}_h$ not containing P_i or P_N we impose $(\operatorname{div} V_i)|_K = 0$ requiring it to be 0 at the three vertices of K. On the other hand, if $K \in \mathcal{T}_h$ contains P_i (or P_N) we require the flux of V_i over ∂K to be $+1$ (or -1). Eventually, for each boundary edge e_b of the triangulation we request the flux of V_i through e_b to be 0.

Tangency constraints: while for the Raviart–Thomas approach the approximate fields are surface vector fields by construction, for this $\mathbb{P}_2$ approach we need to impose this constraint as an additional condition. For doing so we require tangency of V_i at each node of the triangulation and at the mid-point of each edge. Normals at these points are approximated as a weighted average of normals of surrounding elements.

The above constraints, as it happens in the staggered case, translate into linear constraints over the degrees of freedom of $V_1, \dots, V_{N-1}$, and the discrete problem we end up with can be solved using the same strategies presented in Section 4. Eventually we observe that we can extend the refinement procedures of Section 5 also on triangulated surfaces taking advantage of the re-meshing functionalities of the Mmg Platform [1]: at each step we identify the region where the solution concentrates the most and then remesh the surface requiring the new mesh to be finer in that region and coarser elsewhere.

7.3 Results

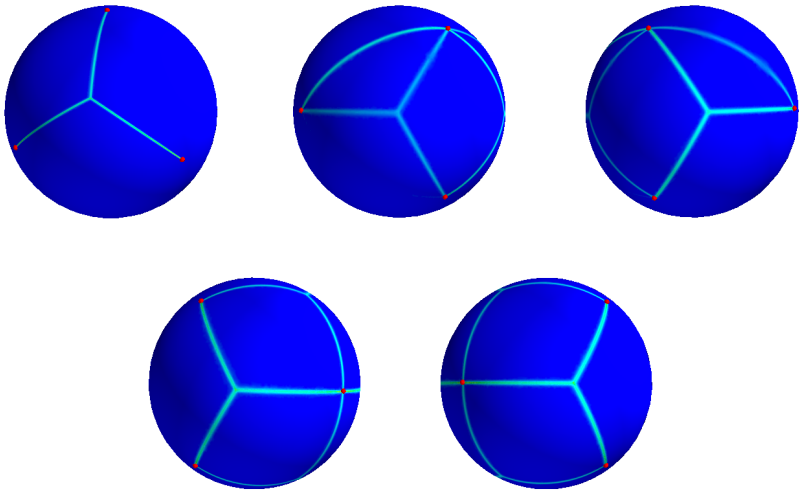


Figure 11: Optima of $\mathcal{R}^0$ for $3, 4, 5$ points on a sphere, single view for 3 terminals and different view angles for $4, 5$ terminals.

In figure 11 we see the results obtained through the $\mathbb{P}_2$-based approach for 3 instances of (STP) on the sphere. In the first case (upper-left) we approximate the Steiner tree associated to the terminal points $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, and observe how we get a classical triple junction. In the second example (upper-middle and upper-right) we add a fourth point, $(0, -1, 0)$, and obtain a convex combination of minimizers: in this case a possible minimizer can be constructed using the structure of the first picture completed with an geodesic arc connecting $(0, 0, 1)$ to $(0, -1, 0)$. We also observe that due to the refinement steps energy concentrates only on two of the possible four minimizers, the two around which the mesh gets refined. In the third example (second row) we add a fifth point, $(-1, 0, 0)$, and obtain a convex combination of the two minimizers.

As we change the topological nature of the surface results become more interesting. We approximate in figure 12 minimizers of $\mathcal{R}^0$ for some points configurations on the torus. In the first example (upper-left) we fix two terminal points opposite to each other on the largest equator and observe an energy concentration on four different paths (each one a geodesic connecting the two points). For certain 3 points configurations we obtain a unique structure with a triple junction (upper-right), while for 3 points in a symmetric disposition on the largest equator we observe as solution a convex combination of the 6 possible minimizers (bottom-left). In the last example (bottom-right) we increase the number of holes of our torus and obtain for a symmetrical 3 points configuration a minimizer which cannot be seen as a convex combination of Steiner trees (i.e. another non sharpness example).

Finally, in figure 13, we test our relaxation on some surfaces with boundary. In the first example we connect three given points on the graph of a function while in the last

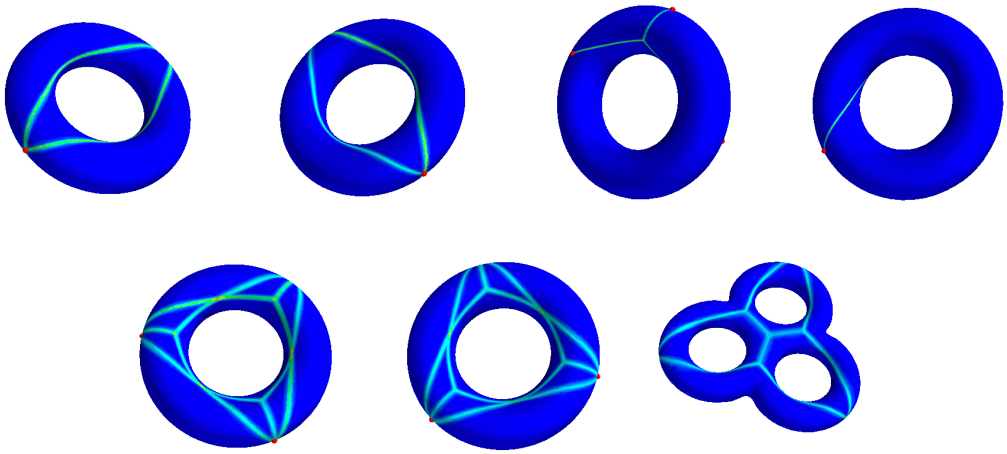


Figure 12: Optima of $\mathcal{R}^0$ for $2, 3$ points on different tori (front/back views).

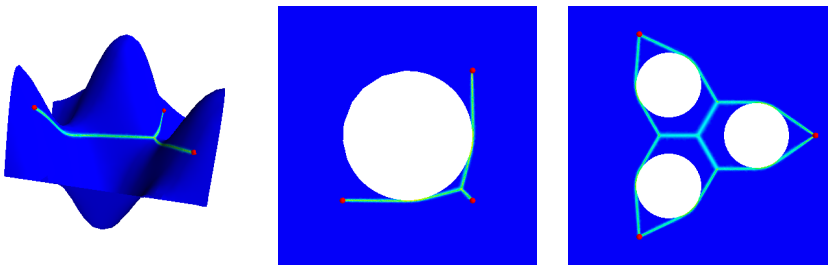


Figure 13: Optima of $\mathcal{R}^0$ for 3 points on the graph of a function and on some punctured domains in $\mathbb{R}^2$.

two we use flat surfaces with holes, which can be seen as the flat version of the previous tori. In this case solutions can adhere to the interior boundary of the domain as long as this is energetically favourable. Observe that, similarly to counter example of figure 1, we obtain a profile which is not a convex combination of optimal trees. As in figure 1, we suspect this solution to illustrate the fact that our convexification may be not sharp in specific situations.

Acknowledgements

The second author gratefully acknowledges the support of the ANR through the project GEOMETRYA, the project COMEDIC and the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01).

References

[1] *Online at https://www.mmgtools.org/.*

[2] Luigi Ambrosio and Andrea Braides. Functionals defined on partitions in sets of finite perimeter. I. Integral representation and Γ-convergence. *J. Math. Pures Appl. (9)*, 69(3):285–305, 1990.

[3] Luigi Ambrosio and Andrea Braides. Functionals defined on partitions in sets of finite perimeter. II. Semicontinuity, relaxation and homogenization. *J. Math. Pures Appl. (9)*, 69(3):307–333, 1990.

[4] Sanjeev Arora. Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *J. ACM*, 45(5):753–782, 1998.

[5] Sanjeev Arora. Approximation schemes for NP-hard geometric optimization problems: a survey. *Math. Program.*, 97(1-2, Ser. B):43–69, 2003. ISMP, 2003 (Copenhagen).

[6] J. M. Bass and J. T. Oden. Adaptive finite element methods for a class of evolution problems in viscoplasticity. *Internat. J. Engrg. Sci.*, 25(6):623–653, 1987.

[7] Ahron Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. Siam, 2001.

[8] Marc Bernot, Vicent Caselles, and Jean-Michel Morel. *Optimal transportation networks: models and theory*, volume 1955. Springer Science & Business Media, 2009.

[9] Mauro Bonafini, Giandomenico Orlandi, and Édouard Oudet. Variational approximation of functionals defined on 1-dimensional connected sets: the planar case. *SIAM J. Math. Anal.*, accepted.

[10] Matthieu Bonnivard, Elie Bretin, and Antoine Lemenant. Numerical approximation of the steiner problem in dimension 2 and 3. 2018.

[11] Matthieu Bonnivard, Antoine Lemenant, and Filippo Santambrogio. Approximation of length minimization problems among compact connected sets. *SIAM J. Math. Anal.*, 47(2):1489–1529, 2015.

[12] Guy Bouchitté and Michel Valadier. Integral representation of convex functionals on a space of measures. *Journal of functional analysis*, 80(2):398–420, 1988.

[13] Franco Brezzi and Michel Fortin. *Mixed and hybrid finite element methods*, volume 15. Springer Science & Business Media, 2012.

[14] Antonin Chambolle, Daniel Cremers, and Thomas Pock. A convex approach to minimal partitions. *SIAM J. Imaging Sci.*, 5(4):1113–1158, 2012.

[15] Antonin Chambolle, Luca Alberto Davide Ferrari, and Benoit Merlet. A phase-field approximation of the steiner problem in dimension two. *Advances in Calculus of Variations*, 2017.

[16] Antonin Chambolle, Luca Alberto Davide Ferrari, and Benoit Merlet. Variational approximation of size-mass energies for k-dimensional currents. *arXiv preprint arXiv:1710.08808*, 2017.

[17] Iain Dunning, Joey Huchette, and Miles Lubin. JuMP: A Modeling Language for Mathematical Optimization. *SIAM Review*, 59(2):295–320, 2017.

[18] Richard L Dykstra. An algorithm for restricted least squares regression. *Journal of the American Statistical Association*, 78(384):837–842, 1983.

[19] Claudia DAmbrosio, Marcia Fampa, Jon Lee, and Stefan Vigerske. On a nonconvex minlp formulation of the euclidean steiner tree problem in n-space. In *International Symposium on Experimental Algorithms*, pages 122–133. Springer, 2015.

[20] Edgar N Gilbert. Minimum cost communication networks. *Bell Labs Technical Journal*, 46(9):2209–2227, 1967.

[21] Richard M Karp. Reducibility among combinatorial problems. In *Complexity of computer computations*, pages 85–103. Springer, 1972.

[22] Andrea Marchese and Annalisa Massaccesi. An optimal irrigation network with infinitely many branching points. *ESAIM Control Optim. Calc. Var.*, 22(2):543–561, 2016.

[23] Andrea Marchese and Annalisa Massaccesi. The Steiner tree problem revisited through rectifiable G-currents. *Adv. Calc. Var.*, 9(1):19–39, 2016.

[24] Annalisa Massaccesi, Édouard Oudet, and Bozhidar Velichkov. Numerical calibration of Steiner trees. *Applied Mathematics & Optimization*, pages 1–18, 2017.

[25] APS Mosek. The MOSEK optimization software. *Online at http://www. mosek. com*, 54, 2010.

[26] Edouard Oudet and Filippo Santambrogio. A Modica-Mortola approximation for branched transport and applications. *Arch. Ration. Mech. Anal.*, 201(1):115–142, 2011.

[27] Thomas Pock and Antonin Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1762–1769. IEEE, 2011.

[28] Marie E. Rognes, David A. Ham, Colin J. Cotter, and Andrew T.T. McRae. Automating the solution of PDEs on the sphere and other manifolds in FEniCS 1.2. *Geoscientific Model Development*, 6(6):2099–2119, 2013.

[29] Hanan Samet. An overview of quadtrees, octrees, and related hierarchical data structures. In *Theoretical Foundations of Computer Graphics and CAD*, pages 51–68. Springer, 1988.

[30] Mark Schmidt, Nicolas L Roux, and Francis R Bach. Convergence rates of inexact proximal-gradient methods for convex optimization. In *Advances in neural information processing systems*, pages 1458–1466, 2011.

[31] DM Warme, Pawel Winter, and Martin Zachariasen. GeoSteiner 3.1. *Department of Computer Science, University of Copenhagen (DIKU)*, 2001.

[32] Qinglan Xia. Optimal paths related to transport problems. *Commun. Contemp. Math.*, 5(2):251–279, 2003.