CrossMark

Optimal Micropatterns in 2D Transport Networks and Their Relation to Image Inpainting

Alessio Brancolini, Carolin Rossmanith
& Benedikt Wirth

Communicated by I. Fonseca

Abstract

We consider two different variational models of transport networks: the so-called branched transport problem and the urban planning problem. Based on a novel relation to Mumford–Shah image inpainting and techniques developed in that field, we show for a two-dimensional situation that both highly non-convex network optimization tasks can be transformed into a convex variational problem, which may be very useful from analytical and numerical perspectives. As applications of the convex formulation, we use it to perform numerical simulations (to our knowledge this is the first numerical treatment of urban planning), and we prove a lower bound for the network cost that matches a known upper bound (in terms of how the cost scales in the model parameters) which helps better understand optimal networks and their minimal costs.

1. Introduction

The optimization of transport networks, in particular the so-called urban planning [13] or the branched transport problem [6], is a nontrivial task. The corresponding cost functionals are highly non-convex, and their minimizers, the optimal networks, often exhibit a complicated, strongly ramified structure.

1.1. Convexification, Numerical Optimization, and Energy Scaling Laws

A better understanding of this can either be achieved by numerical or by analytical means.

Via numerical optimization one may, for instance, identify and explore the optimal network structures for different parameters or geometric settings. This then provides an intuition of the model behaviour. While for urban planning we are not aware of any numerical treatment, optimal networks in branched transport

have already been computed using a phase field approximation [37]. However, due to the complicated energy landscape it is highly nontrivial not to get stuck in local minima, and in general there is no guarantee to achieve the global optimum (which is known to exist).

A classical alternative, analytical way to achieve a better understanding of the network models is by proving how the optimal cost scales in the problem parameters. To this end one typically restricts things to a specific, simple situation and constructs a cost efficient network for that situation. If one can prove that, up to a constant factor, no network can attain a lower cost, then this provides information about the minimum achievable cost and the structure of near-optimal networks. In particular it can be learned how the minimum cost scales in the problem parameters.

Such energy scaling laws have already successfully been used to explain a number of complicated patterns seen in physical experiments such as martensite–austenite transformations [8,18,29–31], micromagnetics [22], intermediate states in type-I superconductors [19,23], membrane folding and blistering [9,10], or epitaxial growth [28]. They were also employed to find the optimal fine-scale structure of composite elastic materials [32,33] as well as of branched transport and urban planning networks [17].

In this article we perform both numerical optimization as well as energy scaling analysis. In particular, we compute optimal urban planning networks numerically for the first time, and we redo the energy scaling analysis of branched transport and urban planning networks. To these ends we follow a novel approach which is interesting on its own. For a two-dimensional situation we show that the problem can actually be convexified. This is achieved by exploiting a novel relation to image inpainting and techniques developed in that field.

We believe that the convex formulation is in many situations tight, that is, equivalent to the original network optimization, however, we only prove that it represents a lower bound. Exploring when it is also an upper bound will be more complicated and certainly requires quite technical constructions (similar to those involved in the regularity analysis of the Mumford–Shah functional; we refer to Remark 3.2.3 and the final discussion in Sect. 4). We will only briefly discuss the associated analytical problems and instead showcase numerical results that support the tightness of the convexification except for a few special cases.

Convex formulations have the great advantage that the local behaviour of the cost or energy provides global information of the full energy landscape, which can be used in many ways (for instance numerically to find the global energy minimizer or analytically to find lower energy bounds by convex duality). As applications we perform numerical network optimization, and we reprove the tight lower bound on the minimum achievable network cost, which is the essential ingredient of the energy scaling law.

1.2. Branched Transport and Urban Planning

Branched transport has been introduced in different formulations by Maddalena et al. [36] and by Xia [42] and then studied by various authors; a comprehensive treatment is found in the monograph [6], while regularity issues are

considered in [14,35,41], and [15]. It has been used to describe hierarchically branched networks such as the blood vessel system or the water supply network in plants. The efficiency of a network is judged by a functional, which measures the cost associated with the transport of mass from a given source distribution to a sink distribution. This functional has the feature that the transport cost per transport distance is not proportional to the transported mass, but increases sublinearly in the mass. In other words, the more mass is transported together, the lower is cost per single particle, which models an increase in transport efficiency. This feature leads to ramified networks, since it becomes more efficient to first collect all the mass via a hierarchically branched network and then transport it all together in bulk. The strength of this effect and thus the degree of ramification is governed by a model parameter $\varepsilon > 0$. For small ε, the effect is only weak, which leads to highly complicated structures with many pipes becoming optimal.

The urban planning model was introduced in [4] and is motivated by public transport networks. Here, the given source represents the distribution of homes in an urban area, and the sink represents the distribution of workplaces. Every commuter can travel on the network, paying a fixed cost of 1 per travel distance, or outside the network by own means, paying a slightly larger cost a. In contrast to branched transport, the resulting total transport cost is proportional to the transported mass. In addition, the urban planning model has maintenance costs which are proportional to the total network length with proportionality factor ε. Those maintenance costs again lead to the preference of hierarchically branched networks.

1.3. The Considered Setting

To better understand the model behaviour, an energy scaling law for the network costs has been derived in [17] (and will be reproved in this work, see Theorem 2.2.1) for the following simple problem geometry (Fig. 1 left). In two-dimensional Euclidean space one considers as source and sink distribution the following measures concentrated on two lines:

$$\mu_0 = \mathcal{L}^1 \llcorner [0, \ell] \times \{0\}, \qquad \mu_1 = \mathcal{L}^1 \llcorner [0, \ell] \times \{1\},$$

where $\ell > 0$, $\mathcal{L}^1$ denotes the one-dimensional Lebesgue measure and $\llcorner$ indicates its restriction onto a subset of $\mathbb{R}^2$. Note that the case of measures with a larger or smaller distance between their supports or with a different common total mass can easily be reduced to the above situation and that [17] actually also considers the analogous problem in more than two dimensions. One seeks the optimal network (with respect to the urban planning or branched transport costs) to transport the mass from μ_0 to μ_1. This situation may, for instance, be viewed as a strong simplification of the water transport within a single plant or a forest: μ_0 represents the groundwater reservoir and μ_1 the water consumption in the leaf canopy (Fig. 1 right).

The truly optimal transport network is too hard to find, but the previously described energy scaling analysis allows to examine how near-optimal network patterns look like. A qualitative sketch is given in Fig. 1 left: such a network is
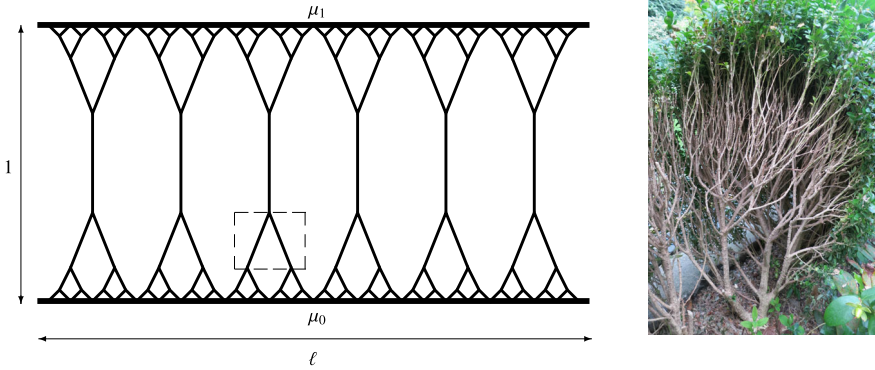
Fig. 1. Left: Sketch of the considered setting, two measures μ_0 and μ_1 supported on lines at distance 1, as well as an exemplary transport network in between composed of elementary cells (dashed line). Right: Photograph of a box tree exhibiting a branched network structure and a comparatively thin leaf canopy that may almost be thought of as a lower-dimensional manifold

composed of multiple hierarchical levels. Each level consists of an array of V-shaped elementary cells, with heights and aspect ratios depending on the level and on the problem parameters a and ε, the details are given in [17].

For numerical network optimization we will also consider different geometric settings such as sources and sinks distributed on a circle (see Fig. 8).

1.4. The Connection to the Mumford–Shah Problem

The Mumford–Shah problem is a variational formulation of an image segmentation or denoising task, originally introduced in [34]. The corresponding energy functional is designed to smooth out a given noisy image while preserving the edges or discontinuities visible inside the image. The work was very influential, and many techniques have been devised to deal with the problem numerically and analytically. Among these techniques there are so-called functional lifting methods that find a convex formulation of the highly complicated and non-convex Mumford–Shah segmentation at the expense of introducing an additional dimension to the problem [38].

The essential observation in the context of network optimization now is that a two-dimensional, divergence-free mass flux (the flux running through the network) becomes a gradient field after rotation by $\frac{\pi}{2}$. This gradient field can be interpreted as the gradient of an image, which suddenly allows us to reformulate network optimization as a Mumford–Shah-type image processing problem and thus to use the functional lifting techniques.

The outline of the article is as follows. We will introduce urban planning and branched transport in Sect. 2 as well as the energy scaling result, which is the main analytical application of our new convex formulation. The relation between network optimization and image processing is derived in Sect. 3, as is the convex reformulation. Section 3 further contains the applications, the lower bound of the

energy scaling as well as a numerical approach based on the reformulation. We close with a discussion in Sect. 4.

2. Transport Networks and Their Energy Scaling

We consider two different models for transport networks: so-called branched transport and urban planning. The formulation in which we state both models is originally due to XIA [42]; the fact that also urban planning can be formulated this way has been shown in [16].

In the following, $\mathcal{H}^n$ and $\mathcal{L}^n$ denote the n-dimensional Hausdorff and Lebesgue measure, fbm(Ω) is the set of non-negative finite Borel measures on $\Omega \subset \mathbb{R}^n$, and rca($\Omega; \mathbb{R}^n$) denotes the set of vector-valued regular countably additive measures (Radon measures). Both fbm(Ω) and rca($\Omega; \mathbb{R}^n$) are equipped with the notion of weak-$*$ convergence, denoted by $\overset{*}{\rightharpoonup}$. Finally, the Dirac delta distribution at a point $x \in \mathbb{R}^n$ will be denoted by δ_x.

2.1. Flux-Based Formulations of Urban Planning and Branched Transport

Definition 2.1.1. (*Discrete mass flux*) A *transport path* is a weighted directed graph G in $\mathbb{R}^n$ with vertices $V(G)$, straight edges $E(G)$, and weight function $w : E(G) \to [0, \infty)$.

Denote the initial and final point of an edge $e \in E(G)$ by e^+ and e^- so that the edge has direction $\hat{e} = \frac{e^- - e^+}{|e^- - e^+|}$. The edge can be identified with the vector measure $\mu_e = (\mathcal{H}^1 \llcorner e)\, \hat{e}$. The *discrete mass flux* corresponding to a transport path G is the vector measure

$$\mathcal{F}_G = \sum_{e \in E(G)} w(e)\mu_e.$$

Let $\mu_+ = \sum_{i=1}^{k} a_i \delta_{x_i}$, $\mu_- = \sum_{i=1}^{l} b_i \delta_{y_i}$ be discrete finite non-negative measures with $a_i, b_i > 0$, $x_i, y_i \in \mathbb{R}^n$ and equal mass $\|\mu_+\|_{\text{fbm}} = \|\mu_-\|_{\text{fbm}}$. G is called a *transport path between μ_+ and μ_-*, or equivalently $\mathcal{F}_G$ is called a *discrete mass flux between μ_+ and μ_-*, if div $\mathcal{F}_G = \mu_+ - \mu_-$ in the distributional sense.

The measure μ_+ and μ_- can be thought of as a mass source and sink, respectively, and the transport path G or equivalently the discrete mass flux $\mathcal{F}_G$ describe how the mass is transported from the source to the sink. Here, the edge weight $w(e)$ indicates how much mass is flowing along edge e, and the divergence condition div $\mathcal{F}_G = \mu_+ - \mu_-$ ensures that no mass gets lost on the way from μ_+ to μ_-.

Definition 2.1.2. (*Urban planning and branched transport cost function*) Let $a > 1$, $\varepsilon > 0$. The *urban planning cost* of a transport path G or discrete mass flux $\mathcal{F}_G$ is given by

$$E_{\text{F}}^{\varepsilon,a}(\mathcal{F}_G) = E_{\text{F}}^{\varepsilon,a}(G) = \sum_{e \in E(G)} \min(aw(e), w(e) + \varepsilon)\, l(e),$$

where $l(e)$ denotes the length of edge e. The *branched transport cost* for $\varepsilon \in (0, 1)$ is given by

$$M_{\mathrm{F}}^{\varepsilon}(\mathcal{F}_G) = M_{\mathrm{F}}^{\varepsilon}(G) = \sum_{e \in E(G)} w(e)^{1-\varepsilon} l(e).$$

The functionals $E_{\mathrm{F}}^{\varepsilon,a}$ and $M_{\mathrm{F}}^{\varepsilon}$ model the cost associated with the mass transport encoded by G. The cost contribution of each edge is proportional to the transport distance $l(e)$. The transport cost per particle and distance is $\min(a, 1 + \frac{\varepsilon}{w(e)})$ for urban planning and $w(e)^{-\varepsilon}$ for branched transport. Obviously, the cost per particle decreases the larger the total flux $w(e)$ through the edge, that is, the more particles travel together. These economies of scale lead to the fact that branched networks are more cost-efficient than non-branched ones. Indeed, it pays off to first collect mass together, then transport it inexpensively in bulk, and only later again divert the single particles to their final destinations. The parameters ε and a encode the strength of these economies of scale.

As already mentioned before, in urban planning the cost also has an alternative interpretation [17]: it can be viewed as the sum of the maintenance costs for the transport network (given by ε times the total network length) and the transport costs, which split up into a contribution from mass transport via the network (at cost 1 per particle and distance) as well as a contribution from mass transport outside the network (at a slightly larger cost a per particle and distance).

Definition 2.1.3. (*Continuous mass flux and cost function*) Let $\mu_+, \mu_- \in \mathrm{fbm}(\mathbb{R}^n)$ of equal mass. A vector measure $\mathcal{F} \in \mathrm{rca}(\mathbb{R}^n; \mathbb{R}^n)$ is a *mass flux between μ_+ and μ_-*, if there exist sequences of discrete measures μ_+^k, μ_-^k with $\mu_+^k \overset{*}{\rightharpoonup} \mu_+$, $\mu_-^k \overset{*}{\rightharpoonup} \mu_-$ as $k \to \infty$, and a sequence of discrete mass fluxes $\mathcal{F}_{G_k}$ between μ_+^k and μ_-^k with $\mathcal{F}_{G_k} \overset{*}{\rightharpoonup} \mathcal{F}$. Note that $\mathrm{div}\, \mathcal{F} = \mu_+ - \mu_-$ follows by continuity with respect to weak-$*$ convergence.

A sequence $(\mu_+^k, \mu_-^k, \mathcal{F}_{G_k})$ satisfying the previous properties is called *approximating graph sequence*, and we write $(\mu_+^k, \mu_-^k, \mathcal{F}_{G_k}) \overset{*}{\rightharpoonup} (\mu_+, \mu_-, \mathcal{F})$.

If $\mathcal{F}$ is a mass flux between μ_+ and μ_-, the *urban planning* and *branched transport cost* are respectively defined as

$$E_{\mathrm{F}}^{\varepsilon,a}(\mathcal{F}) = \inf\left\{ \liminf_{k \to \infty} E_{\mathrm{F}}^{\varepsilon,a}(G_k) \ : \ (\mu_+^k, \mu_-^k, \mathcal{F}_{G_k}) \overset{*}{\rightharpoonup} (\mu_+, \mu_-, \mathcal{F}) \right\},$$

$$M_{\mathrm{F}}^{\varepsilon}(\mathcal{F}) = \inf\left\{ \liminf_{k \to \infty} M_{\mathrm{F}}^{\varepsilon}(G_k) \ : \ (\mu_+^k, \mu_-^k, \mathcal{F}_{G_k}) \overset{*}{\rightharpoonup} (\mu_+, \mu_-, \mathcal{F}) \right\}.$$

As before, $\mathcal{F} \in \mathrm{rca}(\mathbb{R}^n; \mathbb{R}^n)$ describes the mass transport from a source μ_+ to a sink μ_-. One is interested in the optimally cost-efficient mass transport, that is, for fixed μ_+ and μ_- one seeks the mass flux between μ_+ and μ_- with least cost $E_{\mathrm{F}}^{\varepsilon,a}$ or $M_{\mathrm{F}}^{\varepsilon}$.

Remark 2.1.4. (Upper bound on urban planning) For later reference, note that in urban planning the cost minimizing mass flux $\mathcal{F}$ between μ_+ and μ_- satisfies

$$E_{\mathrm{F}}^{\varepsilon,a}(\mathcal{F}) \leq a \mathrm{W}_1(\mu_+, \mu_-),$$

where
$$W_1(\mu_+, \mu_-) = \min_{\mu \in \Pi(\mu_+, \mu_-)} \int_{\mathbb{R}^n \times \mathbb{R}^n} |x - y| \, d\mu(x, y)$$

with $\Pi(\mu_+, \mu_-) = \{\mu \in \mathrm{fbm}(\mathbb{R}^n \times \mathbb{R}^n) \, : \, \mu(A \times \mathbb{R}^n) = \mu_+(A), \ \mu(\mathbb{R}^n \times B) = \mu_-(B)$ for A, B Borel $\subset \mathbb{R}^n\}$ denotes the Wasserstein distance.

Indeed, if μ_+, μ_- are discrete measures, let $\pi = \sum_{x \in \mathrm{spt}\mu_+, y \in \mathrm{spt}\mu_-} c_{x,y} \delta_{(x,y)} \in \Pi(\mu_+, \mu_-)$ be the optimal transport plan for the Wasserstein distance between μ_+, μ_-. Now introduce the transport path with vertices $V(G) = \mathrm{spt}\mu_+ \cup \mathrm{spt}\mu_-$, edges $E(G) = \{e_{x,y} \, : \, x \in \mathrm{spt}\mu_+, \ y \in \mathrm{spt}\mu_-\}$, and weights $w(e_{x,y}) = c_{x,y}$, where $e_{x,y}$ be the oriented line segment from x to y. Obviously,

$$E_{\mathrm{F}}^{\varepsilon,a}(\mathcal{F}) \leq E_{\mathrm{F}}^{\varepsilon,a}(G) = \sum_{e \in E(G)} \min\{aw(e), w(e) + \varepsilon\} l(e) \leq \sum_{e \in E(G)} aw(e)l(e)$$

$$= a \sum_{x \in \mathrm{spt}\mu_+, y \in \mathrm{spt}\mu_-} c_{x,y} |x - y| = aW_1(\mu_+, \mu_-).$$

For non-discrete measures μ_+, μ_-, let $(G_n, \mu_+^n, \mu_-^n) \overset{*}{\rightharpoonup} (\mathcal{F}, \mu_+, \mu_-)$ be an approximating graph sequence. We then finally have

$$E_{\mathrm{F}}^{\varepsilon,a}(\mathcal{F}) \leq \liminf_{n \to \infty} E_{\mathrm{F}}^{\varepsilon,a}(\mathcal{F}_{G_n}) \leq \liminf_{n \to \infty} aW_1(\mu_+^n, \mu_-^n) = aW_1(\mu_+, \mu_-).$$

2.2. Energy Scaling for a Simple, Two-Dimensional Situation

Given an initial and final measure μ_+ and μ_-, we look for the optimal mass fluxes between μ_+ and μ_-, that is, for minimizers of

$$\mathcal{E}^{\varepsilon,a,\mu_+,\mu_-}[\mathcal{F}] = \begin{cases} E_{\mathrm{F}}^{\varepsilon,a}(\mathcal{F}) & \text{if } \mathcal{F} \text{ is a mass flux between } \mu_+ \text{ and } \mu_-, \\ \infty & \text{else,} \end{cases}$$

$$\mathcal{M}^{\varepsilon,\mu_+,\mu_-}[\mathcal{F}] = \begin{cases} M_{\mathrm{F}}^{\varepsilon}(\mathcal{F}) & \text{if } \mathcal{F} \text{ is a mass flux between } \mu_+ \text{ and } \mu_-, \\ \infty & \text{else.} \end{cases}$$

To derive an energy scaling law we here consider the geometrically simple, two-dimensional ($n = 2$) situation with source and sink

$$\mu_+ := \mu_0 = \mathcal{L}^1 \llcorner [0, \ell] \times \{0\}, \qquad \mu_- := \mu_1 = \mathcal{L}^1 \llcorner [0, \ell] \times \{1\} \quad (2.2.1)$$

(cf. Fig. 1 left). It is not difficult to see that the minimum urban planning and branched transport cost for $\varepsilon = 0$ are given by

$$\mathcal{E}^{*,a,\mu_0,\mu_1} := \min_{\mathcal{F}} \mathcal{E}^{0,a,\mu_0,\mu_1}[\mathcal{F}] = W_1(\mu_0, \mu_1) = \ell,$$

$$\mathcal{M}^{*,\mu_0,\mu_1} := \min_{\mathcal{F}} \mathcal{M}^{0,\mu_0,\mu_1}[\mathcal{F}] = W_1(\mu_0, \mu_1) = \ell.$$

Indeed, both urban planning and branched transport cost are known to be bounded below by the Wasserstein distance (see [42, Lemma 4.2] for the case of branched transport; for the case of urban planning the equivalent formulation of [16] allows the same proof), and a minimizing flux achieving the above value is given by

$\mathcal{F} = \binom{0}{1}\mathcal{L}^2 \llcorner [0, \ell] \times [0, 1]$. Our aim is to prove via a novel relation to image inpainting that urban planning and branched transport satisfy the following energy scaling law:

Theorem 2.2.1. (Energy scaling) *There are constants $C_1, C_2, \varepsilon_0 > 0$ independent of ε, a, ℓ such that for $0 < \varepsilon < \min(\varepsilon_0, \ell^3)$ and $a > 1$ we have*

$$C_1 \ell \min\{a - 1, \varepsilon^{\frac{2}{3}}\} \leq \min_{\mathcal{F}} \mathcal{E}^{\varepsilon, a, \mu_0, \mu_1}[\mathcal{F}] - \mathcal{E}^{*, a, \mu_0, \mu_1} \leq C_2 \ell \min\{a - 1, \varepsilon^{\frac{2}{3}}\},$$

$$C_1 \ell \varepsilon |\log \varepsilon| \leq \min_{\mathcal{F}} \mathcal{M}^{\varepsilon, \mu_0, \mu_1}[\mathcal{F}] - \mathcal{M}^{*, \mu_0, \mu_1} \leq C_2 \ell \varepsilon |\log \varepsilon|.$$

The proof of the above result has already been performed in [17]. In particular, the upper bound was proven by providing a construction $\mathcal{F}$ with the desired energy scaling. The lower bound was obtained by a technique common in pattern analysis. One aim of this article is to establish the lower bound via a completely different route, by finding a convex relaxation of the problem and proving a lower bound via convex duality. The essential ingredient here is the identification of a relation between transport network optimization and image inpainting. The remainder of the article derives this relation and provides the lower bound proof as well as some numerical simulations as applications.

3. Optimal Networks in 2D via Convex Optimization

In this section we will transform the urban planning and the branched transport problem in two dimensions to variants of the Mumford–Shah segmentation problem from image processing. Since for the Mumford–Shah segmentation problem there are well-known convex reformulations via so-called functional lifting, the urban planning and branched transport problem in two dimensions each come with a convex formulation as well. While we show that the convex problems provide a lower bound to the original network optimization, we do not examine when they are equivalent. This would require to prove a limsup inequality which probably involves a difficult technical construction outside the scope of this article (cf. Remark 3.2.3 and Sect. 4). We will use the convex formulations to put forward a proof of the lower bound in Theorem 2.2.1 which is solely based on convex duality as well as to perform numerical network optimization that cannot get stuck in local minima.

3.1. Bijection Between Fluxes and Images

Denoting by $B_r(A)$ the open r-neighbourhood of a set $A \subset \mathbb{R}^2$, let $\Omega \subset \mathbb{R}^2$ and $V \subset \mathbb{R}^2$ be open bounded convex domains with $\Omega \subset\subset V$; for simplicity we shall assume

$$V \supset B_1(\Omega).$$

We will later only consider the case where the initial and final measure of the urban planning or branched transport problem are concentrated on the boundary $\partial\Omega$, that is,

$$\mathrm{spt}\,\mu_+ \subset \partial\Omega, \qquad \mathrm{spt}\,\mu_- \subset \partial\Omega. \tag{3.1.1}$$

For the specific choice (2.2.1) we may for instance take $\Omega = (0, \ell) \times (0, 1)$ and $V = B_1(\Omega)$. Now let $u \in \mathrm{BV}(V)$ represent a grey value image, where $\mathrm{BV}(V)$ denotes the space of scalar functions of bounded variation on V. The gradient of u can be decomposed into a part continuous with respect to the two-dimensional Lebesgue measure $\mathcal{L}^2$, a jump part concentrated on the discontinuity set S_u of u, and a Cantor part $D_c u$ [2, (3.89) and (3.90)],

$$Du = \nabla u \mathcal{L}^2 \llcorner V + [u] \nu \mathcal{H}^1 \llcorner S_u + D_c u.$$

Here ∇u is the Radon–Nikodym derivative of the Lebesgue-continuous part with respect to $\mathcal{L}^2$, ν is the unit normal to the (rectifiable) discontinuity set S_u, and $[u]$ is the jump in function values across S_u in direction ν.

Definition 3.1.1. (*Flux corresponding to an image*) Let $u \in \mathrm{BV}(V)$ be an image. The *flux* $\mathcal{F}_u \in \mathrm{rca}(V; \mathbb{R}^2)$ *associated to the image* u is defined by

$$\mathcal{F}_u = Du^\perp = \nabla u^\perp \mathcal{L}^2 \llcorner V + [u] \nu^\perp \mathcal{H}^1 \llcorner S_u + D_c u^\perp,$$

where superscript $\perp$ denotes a counterclockwise rotation by $\frac{\pi}{2}$.

Note that $\mathcal{F}_u$ is divergence-free in the distributional sense (in V). Indeed, let $\phi \in \mathcal{C}_c^\infty(V)$ be a smooth test function with compact support, then

$$\int_V \phi \, \mathrm{d}(\mathrm{div}\, \mathcal{F}_u) = -\int_V \nabla \phi \cdot \mathrm{d}\mathcal{F}_u = \int_V \nabla \phi^\perp \cdot \mathrm{d}Du = -\int_V \mathrm{div}(\nabla \phi^\perp) u \, \mathrm{d}\mathcal{L}^2 = 0$$

due to $\mathrm{div}(\nabla \phi^\perp) = \frac{\partial^2}{\partial x_2 \partial x_1} \phi - \frac{\partial^2}{\partial x_1 \partial x_2} \phi = 0$.

Remark 3.1.2. (Initial and final measure in the domain interior] Note that (3.1.1) imposes a strong restriction on the admissible initial and final measures that only allows specific applications. However, using the same ideas as in [12] one can in principle also treat the case of measures in the interior of Ω; a simple example of this approach is provided in Fig. 8. In essence, one can augment the sought network by a fixed artificial flux transporting all mass back from μ_- to μ_+. As a consequence, the combined flux can again be represented as the rotated gradient of an image, and one merely has to adapt the energy functional in order to ignore the contributions from the artificial flux. Throughout this article, however, we will restrict ourselves to the simpler setting of initial and final measures concentrated on $\partial \Omega$.

As explained before, we consider the case of μ_+ and μ_- being any measures of equal mass satisfying (3.1.1). Without loss of generality we may assume $\partial \Omega$ to contain the origin and to lie in the right halfplane. The function $\gamma : [0, \mathcal{H}^1(\partial \Omega)) \to \partial \Omega$ shall parameterize $\partial \Omega$ counterclockwise by arclength, starting in $\gamma(0) = 0$, and its image of $[0, t)$ shall be denoted $\partial \Omega_t = \gamma([0, t))$. Finally, let us introduce the orthogonal projection

$$\pi_{\partial \Omega} : \mathbb{R}^2 \to \partial \Omega, \quad x \mapsto \arg\min \{|x - y| : y \in \partial \Omega\}$$

(in case of nonuniqueness we pick the lexicographically first point, which is well-defined since the set of closest points in $\partial \Omega$ is closed and bounded; the following

arguments do not depend on this specific choice, though). The next definition introduces a function whose gradient can be thought of as an extension of two given measures μ_+, μ_- on $\partial\Omega$ onto $\mathbb{R}^2$. Roughly speaking, the function is chosen such that the tangential derivative of its trace on $\partial\Omega$ is $\mu_+ - \mu_-$. The restriction of this function onto $V \backslash \overline{\Omega}$ will be used as a kind of boundary condition in order to avoid (equivalent but lengthier) Mumford–Shah energy formulations with additional contributions from the $\partial\Omega$-trace of $BV(\Omega)$-functions.

Definition 3.1.3. (*Admissible fluxes and images*) Given finite Borel measures μ_+, μ_- satisfying (3.1.1), we define the function

$$u(\mu_+, \mu_-) : \mathbb{R}^2 \to \mathbb{R}, \ x \mapsto (\mu_+ - \mu_-)(\partial\Omega_{\gamma^{-1}(\pi_{\partial\Omega}(x))}) \qquad (3.1.2)$$

(note that $u(\mu_+, \mu_-)$ is constant along rays orthogonal to $\partial\Omega$ up to its so-called medial axis).

The sets of *admissible fluxes* and of *admissible images* are given as

$$\mathcal{A}_{\mathcal{F}}(\mu_+, \mu_-) = \{\mathcal{F} \in \mathrm{rca}(V; \mathbb{R}^2) \ : \ \mathrm{spt}\mathcal{F} \subset \overline{\Omega}, \ \mathrm{div}\,\mathcal{F} = \mu_+ - \mu_-\},$$
$$\mathcal{A}_u(\mu_+, \mu_-) = \{u \in BV(V) \ : \ u = u(\mu_+, \mu_-) \text{ on } V \backslash \overline{\Omega}\}. \qquad (3.1.3)$$

The following lemma establishes a one-to-one relation between admissible fluxes and admissible images (its proof is deferred to the "Appendix"):

Lemma 3.1.4. (*Bijection between images and fluxes*) *The mapping* $u \mapsto \mathcal{F}_u \llcorner \overline{\Omega}$ *from* $\mathcal{A}_u(\mu_+, \mu_-)$ *to* $\mathcal{A}_{\mathcal{F}}(\mu_+, \mu_-)$ *is one-to-one. In particular, for any* $\mathcal{F} \in \mathcal{A}_{\mathcal{F}}(\mu_+, \mu_-)$ *there is a unique image* $u_{\mathcal{F}} \in \mathcal{A}_u(\mu_+, \mu_-)$ *with* $\mathcal{F} = \mathcal{F}_{u_{\mathcal{F}}} \llcorner \overline{\Omega}$.

Definition 3.1.5. (*Image corresponding to a flux*) Given a flux $\mathcal{F} \in \mathcal{A}_{\mathcal{F}}(\mu_+, \mu_-)$, the image $u_{\mathcal{F}}$ from the previous lemma is called the *image corresponding to* $\mathcal{F}$.

3.2. Image Inpainting Problem Induced by 2D Network Optimization

Let us now introduce functionals $\tilde{\mathcal{E}}^{\varepsilon,a,\mu_+,\mu_-}$, $\tilde{\mathcal{M}}^{\varepsilon,\mu_+,\mu_-}$ acting on images rather than mass fluxes.

Definition 3.2.1. (*Mumford–Shah-type functionals*) We set $\tilde{E}^{\varepsilon,a}$, $\tilde{M}^{\varepsilon} : BV(V) \to [0, \infty]$,

$$\tilde{E}^{\varepsilon,a}[u] = a \int_{\overline{\Omega} \backslash S_u} |Du| \, \mathrm{d}x + \int_{S_u \cap \overline{\Omega}} \min\{a[u], [u] + \varepsilon\} \, \mathrm{d}\mathcal{H}^1(x),$$

$$\tilde{M}^{\varepsilon}[u] = \int_{S_u \cap \overline{\Omega}} [u]^{1-\varepsilon} \, \mathrm{d}\mathcal{H}^1(x) + \iota_0((\nabla u \mathcal{L}^2 + D_c u) \llcorner \overline{\Omega}),$$

where the indicator function is given by

$$\iota_0(\mu) = \begin{cases} 0 & \text{if } \mu = 0 \\ \infty & \text{otherwise.} \end{cases}$$

Furthermore, for $\mu_+, \mu_- \in \mathrm{fbm}(\mathbb{R}^2)$ of equal mass and satisfying (3.1.1), we define the following functionals on $\mathrm{BV}(V)$:

$$\tilde{\mathcal{E}}^{\varepsilon,a,\mu_+,\mu_-}[u] = \begin{cases} \tilde{E}^{\varepsilon,a}[u] & \text{if } u \in \mathcal{A}_u(\mu_+, \mu_-), \\ \infty & \text{else}, \end{cases}$$

$$\tilde{\mathcal{M}}^{\varepsilon,\mu_+,\mu_-}[u] = \begin{cases} \tilde{M}^{\varepsilon}[u] & \text{if } u \in \mathcal{A}_u(\mu_+, \mu_-), \\ \infty & \text{else}. \end{cases}$$

Both functionals can be viewed as image inpainting functionals. Indeed, the image u is prescribed on $V \setminus \overline{\Omega}$, and inside $\overline{\Omega}$ it is to be reconstructed as the minimizer of $\tilde{\mathcal{E}}^{\varepsilon,a,\mu_+,\mu_-}$ or $\tilde{\mathcal{M}}^{\varepsilon,\mu_+,\mu_-}$. Furthermore, both functionals are of Mumford–Shah-type in that their integrand is convex in Du away from the jump set S_u, but subadditive on the jump set.

In this section we prove the following theorem, expressing the relation between the network costs and the image processing functionals:

Theorem 3.2.2. (Lower bound on transport problems via Mumford–Shah functionals) *For any flux $\mathcal{F} \in \mathcal{A}_{\mathcal{F}}(\mu_+, \mu_-)$ and the corresponding image $u_{\mathcal{F}} \in \mathcal{A}_u(\mu_+, \mu_-)$ (see Definition 3.1.5) we have*

$$\mathcal{E}^{\varepsilon,a,\mu_+,\mu_-}[\mathcal{F}] \geq \tilde{\mathcal{E}}^{\varepsilon,a,\mu_+,\mu_-}[u_{\mathcal{F}}], \qquad \mathcal{M}^{\varepsilon,\mu_+,\mu_-}[\mathcal{F}] \geq \tilde{\mathcal{M}}^{\varepsilon,\mu_+,\mu_-}[u_{\mathcal{F}}].$$

Remark 3.2.3. (Upper bound) We believe that the opposite inequality holds as well, but we do not attempt a proof here. It would imply that the urban planning and branched transport problem can in fact be formulated as the minimization of the above variants of the Mumford–Shah image segmentation functional. In view of Lemma A.0.2 below, a proof would require to show a certain limsup inequality, namely that the functionals $\tilde{E}^{\varepsilon,a,\mu_+,\mu_-}$ and $\tilde{M}^{\varepsilon,\mu_+,\mu_-}$ are actually the sequentially weakly-$*$ lower semi-continuous envelopes of their restrictions to images $u_{\mathcal{F}_G}$ corresponding to discrete graphs G. For the standard Mumford–Shah functional this is implied by [26, Theorem 0.6]. For other variants of the Mumford–Shah functional one would have to carefully study the regularity of the singular set of minimizers. The difficulties one would encounter are of similar type as those which characterize the regularity theory for the standard Mumford–Shah minimizers (which still is only partial despite several decades of investigation). In addition, existence theorems like [3, Theorem 4.1] or [2, Theorem 5.24] do not apply in our case.

The proof of Theorem 3.2.2 makes use of a few technical but standard density and continuity results, which are derived in Lemmas A.0.1 to A.0.3 in the "Appendix". In detail, Lemma A.0.1 shows that in Definition 2.1.3 of the cost function one may restrict to initial and final measures of the approximating graph sequences with support on $\partial\Omega$, Lemma A.0.2 shows the equivalence of flux- and image-based energy functionals for discrete fluxes, and Lemma A.0.3 proves the lower semi-continuity of the image-based energy functionals. We will furthermore make use of images corresponding to discrete fluxes, which can be described by as specific special functions of bounded variation as detailed in the following remark:

Remark 3.2.4. (Density of piecewise constant images) Let $SBV(V) \subset BV(V)$ denote the space of special functions of bounded variation, that is, those functions in $BV(V)$ whose derivative has no Cantor part. Furthermore, let $D \subset SBV(V)$ denote the set of piecewise constant images u whose discontinuity set is composed of finitely many straight lines.

D is dense in $BV(V)$ with respect to weak-$*$ convergence. Indeed, consider a quadrilateral grid $\mathcal{G}$ over V of sidelength $\frac{1}{n}$ and approximate $u \in BV(V)$ by the function $u_n \in D$ which on each square face of $\mathcal{G}$ equals the average of u on that face,

$$u_n(y) = \fint_{V_{ij}^n} u \, dx \quad \text{for all } y \in V_{ij}^n := \left(\left[\tfrac{i}{n}, \tfrac{i+1}{n} \right) \times \left[\tfrac{j}{n}, \tfrac{j+1}{n} \right) \right) \cap V.$$

We have $u_n \to u$ strongly in $L^1(V)$, which together with the boundedness of $\|Du_n\|_{\mathrm{rca}}$ implies $u_n \overset{*}{\rightharpoonup} u$ up to a subsequence.

Proof. (Proof of Theorem 3.2.2) By Lemmas 3.1.4, A.0.1, A.0.2, and A.0.3 we have

$$
\begin{aligned}
E_{\mathrm{F}}^{\varepsilon,a}(\mathcal{F}) &= \inf\{\liminf_{n\to\infty} E_{\mathrm{F}}^{\varepsilon,a}(G_n) \; : \; (\mathcal{F}_{G_n}, \mu_+^n, \mu_-^n) \overset{*}{\rightharpoonup} (\mathcal{F}, \mu_+, \mu_-), \\
&\qquad \mu_+^n, \mu_-^n \text{ satisfy } (3.1.1)\} \\
&= \inf\{\liminf_{n\to\infty} \tilde{E}^{\varepsilon,a}(u_{\mathcal{F}_{G_n}}) \; : \; (\mathcal{F}_{G_n}, \\
&\qquad \mu_+^n, \mu_-^n) \overset{*}{\rightharpoonup} (\mathcal{F}, \mu_+, \mu_-), \; u_{\mathcal{F}_{G_n}} \in \mathcal{A}_u(\mu_+^n, \mu_-^n)\} \\
&= \inf\{\liminf_{n\to\infty} \tilde{E}^{\varepsilon,a}(u_n) \; : \; (Du_n^\perp, \mu_+^n, \mu_-^n) \overset{*}{\rightharpoonup} (Du_{\mathcal{F}}^\perp, \mu_+, \mu_-), \\
&\qquad u_n \in D \cap \mathcal{A}_u(\mu_+^n, \mu_-^n)\} \\
&= \inf\{\liminf_{n\to\infty} \tilde{E}^{\varepsilon,a}(u_n) \; : \; u_n \overset{*}{\rightharpoonup} u_{\mathcal{F}} \text{ in } BV(V), \; u_n \in D \cap H\} \\
&\geq \tilde{E}^{\varepsilon,a}(u_{\mathcal{F}}),
\end{aligned}
$$

which is the desired result. In particular, the first equality is due to Lemma A.0.1, the second to Lemma A.0.2, the third to Lemma 3.1.4, and the fourth to the definition of $\mathcal{A}_u(\mu_+, \mu_-)$ in (3.1.3) and of H in (A.0.2). The final inequality is then due to the lower semi-continuity Lemma A.0.3.

The result for $\tilde{\mathcal{M}}^\varepsilon$ is obtained analogously. $\quad\square$

3.3. Functional Lifting of Network Optimization

In a series of articles [39, 40] Pock, Cremers, Bischof, and Chambolle showed that the minimization of certain variational problems admits a convex reformulation via so-called functional lifting. The underlying idea is based on [1] and [21], which consider functionals of the form

$$\mathcal{J}[u] = \int_\Omega g(x, u(x), \nabla u(x)) \, dx + \int_{S_u} \psi(x, u^+, u^-, \nu) \, d\mathcal{H}^1(x),$$

where u^+ and u^- are the function values of u on either side of the discontinuity set S_u, and $g : \Omega \times \mathbb{R} \times \mathbb{R}^2 \to [0, \infty]$ and $\psi : \Omega \times \mathbb{R} \times \mathbb{R} \times S^1 \to [0, \infty]$. Introducing the characteristic function of the subgraph of the image u,

$$1_u : \Omega \times \mathbb{R} \to \{0, 1\}, \ (x, s) \mapsto \begin{cases} 1 & \text{if } u(x) > s, \\ 0 & \text{else,} \end{cases}$$

the authors show

$$\mathcal{J}[u] \geq \sup_{\phi \in \mathcal{K}} \int_{\Omega \times \mathbb{R}} \phi \cdot dD 1_u$$

$$\text{with } \mathcal{K} = \Big\{ \phi = (\phi_x, \phi_s) \in \mathcal{C}_0^\infty(\Omega \times \mathbb{R}; \mathbb{R}^2 \times \mathbb{R}) \ :$$

$$\phi_s(x, s) \geq g^*(x, s, \phi_x(x, s)) \quad \forall (x, s) \in \Omega \times \mathbb{R},$$

$$\left| \int_{s_1}^{s_2} \phi_x(x, s)\, ds \right| \leq \psi(x, s_1, s_2, \nu) \quad \forall x \in \Omega, s_1 < s_2, \nu \in S^1 \Big\}.$$

Here, g^* denotes the Legendre–Fenchel dual of g with respect to its last argument. Consequently, instead of minimizing the potentially non-convex functional $\mathcal{J}[u]$ for u one can find a solution to the saddle point problem $\min_{1_u} \sup_{\phi \in \mathcal{K}} \int_{\Omega \times \mathbb{R}} \phi \cdot dD 1_u$.

Unfortunately, the set $\{ v \in \mathrm{BV}(\Omega \times \mathbb{R}; \{0, 1\}) \ : \ v = 1_u \text{ for some } u \in \mathrm{BV}(\Omega) \}$ of characteristic functions, over which one minimizes, is non-convex. However, this set can be relaxed to the larger set

$$\mathcal{C} = \Big\{ v \in \mathrm{BV}(\Omega \times \mathbb{R}; [0, 1]) \ : \ \lim_{s \to -\infty} v(x, s) = 1, \ \lim_{s \to \infty} v(x, s) = 0 \Big\}.$$

With this relaxation,

$$\inf_{u \in \mathrm{BV}(\Omega)} \mathcal{J}[u] \geq \inf_{v \in \mathcal{C}} \sup_{\phi \in \mathcal{K}} \int_{\Omega \times \mathbb{R}} \phi \cdot dD v, \tag{3.3.1}$$

where the right-hand side is a convex problem.

For our setting, given $\mu_+, \mu_- \in \mathrm{fbm}(\mathbb{R}^2)$ of equal mass and with (3.1.1), let $1_{u(\mu_+, \mu_-)}$ be the characteristic function of the subgraph of the function defined in (3.1.2), and let us introduce the sets

$$\tilde{\mathcal{C}} = \Big\{ v \in \mathrm{BV}(V \times \mathbb{R}; [0, 1]) \ :$$

$$\lim_{s \to -\infty} v(x, s) = 1, \ \lim_{s \to \infty} v(x, s) = 0, \ v = 1_{u(\mu_+, \mu_-)} \text{ on } (V \setminus \overline{\Omega}) \times \mathbb{R} \Big\},$$

$$\mathcal{K}_1 = \Big\{ \phi = (\phi_x, \phi_s) \in \mathcal{C}_0^\infty(V \times \mathbb{R}; \mathbb{R}^2 \times \mathbb{R}) \ :$$

$$\phi_s \geq 0, \ |\phi_x| \leq a, \ \left| \int_{s_1}^{s_2} \phi_x(x, s)\, ds \right|$$

$$\leq \min\{|s_2 - s_1| + \varepsilon, a|s_2 - s_1|\} \, \forall x \in V, \ s_1, s_2 \in \mathbb{R} \Big\},$$

$$\mathcal{K}_2 = \Big\{ \phi = (\phi_x, \phi_s) \in \mathcal{C}_0^\infty(V \times \mathbb{R}; \mathbb{R}^2 \times \mathbb{R}) \ :$$

$$\phi_s \geq 0, \ \left| \int_{s_1}^{s_2} \phi_x(x, s)\, ds \right| \leq |s_2 - s_1|^{1-\varepsilon} \, \forall x \in V, \ s_1, s_2 \in \mathbb{R} \Big\}.$$

Note that $\mathcal{K}_1$ is $\mathcal{K}$ for the choice corresponding to urban planning,

$$g(x, u, p) = ap, \qquad \psi(x, u^+, u^-, v) = \min\{a|u^+ - u^-|, |u^+ - u^-| + \varepsilon\},$$

and that $\mathcal{K}_2$ is $\mathcal{K}$ for the choice corresponding to branched transport,

$$g(x, u, p) = \iota_0(p), \qquad \psi(x, u^+, u^-, v) = |u^+ - u^-|^{1-\varepsilon}.$$

Applying the previous discussion, we have

$$\min_{\mathcal{F} \in \mathcal{A}_\mathcal{F}(\mu_+, \mu_-)} \mathcal{E}^{\varepsilon, a, \mu_+, \mu_-}[\mathcal{F}] \geq \inf_{u \in \mathcal{A}_u(\mu_+, \mu_-)} \tilde{\mathcal{E}}^{\varepsilon, a, \mu_+, \mu_-}[u]$$

$$\geq \inf_{v \in \tilde{\mathcal{C}}} \sup_{\phi \in \mathcal{K}_1} \int_{\overline{\Omega} \times \mathbb{R}} \phi \cdot dDv,$$

$$\min_{\mathcal{F} \in \mathcal{A}_\mathcal{F}(\mu_+, \mu_-)} \mathcal{M}^{\varepsilon, \mu_+, \mu_-}[\mathcal{F}] \geq \inf_{u \in \mathcal{A}_u(\mu_+, \mu_-)} \tilde{\mathcal{M}}^{\varepsilon, a, \mu_+, \mu_-}[u]$$

$$\geq \inf_{v \in \tilde{\mathcal{C}}} \sup_{\phi \in \mathcal{K}_2} \int_{\overline{\Omega} \times \mathbb{R}} \phi \cdot dDv.$$

The right-hand sides are convex optimization problems in v. Furthermore, in some cases the optimal v can be shown to have the form 1_u for the solution u of the generalised Mumford–Shah problem ([21, Theorem 8.1] for the one-dimensional problem). The optimal mass flux can then be readily read off as $Du^\perp$.

3.4. Lower Bound on Network Costs by Convex Duality

As an application of the convex reformulation from the previous section we finally obtain a proof of the lower bound in Theorem 2.2.1 just based on convex duality. In particular, we show the following:

Theorem 3.4.1. *There exist $c, \varepsilon_0 > 0$ such that for $n = 2$ and $\varepsilon < \varepsilon_0$ we have*

$$\min_{\mathcal{F}} \mathcal{E}^{\varepsilon, a, \mu_0, \mu_1}[\mathcal{F}] - \mathcal{E}^{*, a, \mu_0, \mu_1} \geq c\ell \min\{\varepsilon^{\frac{2}{3}}, a - 1\},$$

$$\min_{\mathcal{F}} \mathcal{M}^{\varepsilon, \mu_0, \mu_1}[\mathcal{F}] - \mathcal{M}^{*, \mu_0, \mu_1} \geq c\ell\varepsilon|\log \varepsilon|.$$

Proof. Here we choose $\Omega = (0, \ell) \times (0, 1)$ and $V = B_1(\Omega)$. Note that for μ_0, μ_1 from (2.2.1) we have $u(\mu_0, \mu_1)(x) = x_1$ for $x \in \partial\Omega$. Swapping the infimum and supremum in the convex formulation we obtain

$$\min_{\mathcal{F} \in \mathcal{A}_\mathcal{F}(\mu_0, \mu_1)} \mathcal{E}^{\varepsilon, a, \mu_0, \mu_1}[\mathcal{F}] \geq \sup_{\phi \in \mathcal{K}_1} \inf_{v \in \tilde{\mathcal{C}}} \int_{\overline{\Omega} \times \mathbb{R}} \phi \cdot dDv$$

$$= \sup_{\phi \in \mathcal{K}_1} \inf_{v \in \tilde{\mathcal{C}}} \int_P \phi \cdot v \, d\mathcal{H}^2(x, s) - \int_{\Omega \times \mathbb{R}} v \operatorname{div} \phi \, dx \, ds$$

$$= \sup_{\phi \in \mathcal{K}_1} \int_P \phi \cdot v \, d\mathcal{H}^2(x, s) - \int_{\Omega \times \mathbb{R}} \max(0, \operatorname{div} \phi) \, dx \, ds,$$
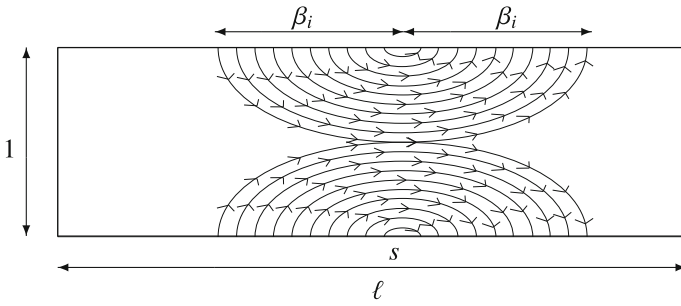
Fig. 2. Sketch of the flow ϕ^i in cross-section $\Omega \times \{s\}$

where $P = \{(x, s) \in \partial\Omega \times \mathbb{R} : 1_{u(\mu_0, \mu_1)}(x, s) = 1\} = \{(x, s) \in \partial\Omega \times \mathbb{R} : x_1 \geq s\}$, ν is the unit outward normal to P, as well as

$$\min_{\mathcal{F} \in \mathcal{A}_{\mathcal{F}}(\mu_0, \mu_1)} \mathcal{M}^{\varepsilon, \mu_0, \mu_1}[\mathcal{F}] \geq \sup_{\phi \in \mathcal{K}_2} \int_P \phi \cdot \nu \, d\mathcal{H}^2(x, s) - \int_{\Omega \times \mathbb{R}} \max(0, \operatorname{div} \phi) \, dx \, ds.$$

All that remains to do is to construct a proper test function ϕ to be used in the above convex duality estimate. Define the clockwise circular flow

$$\theta(x_1, x_2) = \tfrac{1}{r} \left(\begin{smallmatrix} x_2 \\ -x_1 \end{smallmatrix} \right) \quad \text{if } r = \left| \left(\begin{smallmatrix} x_2 \\ -x_1 \end{smallmatrix} \right) \right| \leq \tfrac{1}{2} \text{ and } \theta(x_1, x_2) = 0 \text{ else.}$$

Obviously, $\operatorname{div} \theta = 0$ in the sense of distributions. Now we stretch this circular flow horizontally by a factor $2\beta_i > 2$ (which will be specified later), $i = 1, 2$, and divide by two,

$$\varphi_i(x) = \tfrac{1}{2} B_i \theta(B_i^{-1} x) \quad \text{for } B_i = \left(\begin{smallmatrix} 2\beta_i & 0 \\ 0 & 1 \end{smallmatrix} \right).$$

It is readily checked that also $\operatorname{div} \varphi_i = 0$. Now define $\phi^i = (\phi_x^i, 0)$ as (cf. Fig. 2)

$$\phi_x^i(x, s) = \begin{cases} 0 & \text{if } s \notin [0, \ell], \\ \varphi_i(x_1 - s, x_2) & \text{if } x_2 \leq \tfrac{1}{2}, \ s \in [0, \ell], \\ -\varphi_i(x_1 - s, 1 - x_2) & \text{else.} \end{cases}$$

Denoting the unit outward normal to $\partial\Omega$ by n and using $\operatorname{div} \phi_x^i(\cdot, s) = 0$ as well as $\phi_x^i((s, x_2), s) = \varphi_i(0, x_2) = \left(\begin{smallmatrix} \beta_i \\ 0 \end{smallmatrix} \right)$ for $x_2 \in [0, 1]$ we can calculate

$$\int_P \phi^i \cdot \nu \, d\mathcal{H}^2(x, s) = \int_0^\ell \int_{\{x \in \partial\Omega : x_1 \geq s\}} \phi_x^i \cdot n \, d\mathcal{H}^1(x) \, ds$$

$$= \int_0^\ell \left(\int_{[s, \ell] \times [0, 1]} \operatorname{div} \phi_x^i(\cdot, s) \, dx - \int_{\{s\} \times [0, 1]} \phi_x^i(x, s) \cdot \left(\begin{smallmatrix} -1 \\ 0 \end{smallmatrix} \right) d\mathcal{H}^1(x) \right) ds = \beta_i \ell,$$

where in the second step we used the divergence theorem. Now the test functions ϕ^1, ϕ^2 can be approximated arbitrarily well by divergence-free functions in $\mathcal{C}_0^\infty(V \times \mathbb{R}; \mathbb{R}^2 \times \mathbb{R})$ so that the above convex duality estimates imply

$$\min_{\mathcal{F} \in \mathcal{A}_{\mathcal{F}}(\mu_0, \mu_1)} \mathcal{E}^{\varepsilon, a, \mu_0, \mu_1}[\mathcal{F}] \geq \int_P \phi^1 \cdot v \, d\mathcal{H}^2(x, s) = \beta_1 \ell = \mathcal{E}^{*, a, \mu_0, \mu_1} + \ell(\beta_1 - 1),$$

$$\min_{\mathcal{F} \in \mathcal{A}_{\mathcal{F}}(\mu_0, \mu_1)} \mathcal{M}^{\varepsilon, \mu_0, \mu_1}[\mathcal{F}] \geq \int_P \phi^2 \cdot v \, d\mathcal{H}^2(x, s) = \beta_2 \ell = \mathcal{M}^{*, \mu_0, \mu_1} + \ell(\beta_2 - 1),$$

where β_1, β_2 will be chosen such that

$$\left| \int_{s_1}^{s_2} \phi_x^1(x, s) \, ds \right| \leq \min\{|s_2 - s_1| + \varepsilon, a|s_2 - s_1|\} \quad \forall x \in V, s_1, s_2 \in \mathbb{R},$$

$$\left| \int_{s_1}^{s_2} \phi_x^2(x, s) \, ds \right| \leq |s_2 - s_1|^{1-\varepsilon} \quad \forall x \in V, s_1, s_2 \in \mathbb{R}$$

as well as $\beta_1 \leq a$ in order to satisfy $|\phi_x^1| \leq a$. We are going to show that $\beta_1 = \min(1 + c\varepsilon^{\frac{2}{3}}, a)$ and $\beta_2 = 1 + c\varepsilon|\log \varepsilon|$ are admissible choices for some $c > 0$, which concludes the proof.

Fix a position $x = (x_1, x_2)$, where due to symmetry it suffices to consider $x_2 \in [0, \frac{1}{2}]$. Let us abbreviate $\hat{s} = s - x_1$ and $r_i(\hat{s}) = |B_i^{-1}\binom{\hat{s}}{x_2}| = \sqrt{x_2^2 + (\frac{\hat{s}}{2\beta_i})^2}$. We have

$$|\phi^i(x, s)| = |\varphi_i(-\hat{s}, x_2)| = \frac{1}{r_i(\hat{s})} \sqrt{\beta_i^2 x_2^2 + (\frac{\hat{s}}{4\beta_i})^2} = \sqrt{\frac{\beta_i^2 + (\hat{s}/(4x_2 \beta_i))^2}{1 + (\hat{s}/(2x_2 \beta_i))^2}},$$

where ϕ^i is nonzero. Note that for $\bar{s}_i = 4x_2 \beta_i \sqrt{(\beta_i^2 - 1)/3}$,

$$|\phi^i(x, s)| \leq \begin{cases} \beta_i & \text{if } \hat{s} \in [-\bar{s}_i, \bar{s}_i], \\ 1 & \text{else,} \end{cases}$$

thus, in particular, $|\phi^i(x, s)| \leq \beta_i$.

We first consider the case $i = 1$, in which we have (assuming $s_2 \geq s_1$ without loss of generality)

$$\left| \int_{s_1}^{s_2} \phi^1(x, s) \, ds \right| - |s_2 - s_1| \leq \int_{s_1}^{s_2} |\phi^1(x, s)| - 1 \, ds$$

$$\leq 2\bar{s}_1(\beta_1 - 1) \leq 4\beta_1 \sqrt{\frac{\beta_1^2 - 1}{3}}(\beta_1 - 1) \leq 4\sqrt{\frac{\beta_1^2 - 1}{3}}(\beta_1^2 - 1),$$

which is indeed smaller than ε, as required, if we choose $\beta_1 = \min(1 + \frac{1}{4}(\frac{3}{2})^{\frac{2}{3}} \varepsilon^{\frac{2}{3}}, a)$. Likewise, this choice of β_1 also satisfies

$$\left| \int_{s_1}^{s_2} \phi^1(x, s) \, ds \right| - a|s_2 - s_1| \leq 2\bar{s}_1(\beta_1 - a) \leq 0,$$

as required.

For $i = 2$, assume first $\hat{s}_1, \hat{s}_2 \in [-\bar{s}_i, \bar{s}_i]$. In that case,

$$\left| \int_{s_1}^{s_2} \phi^2(x, s) \, ds \right| / |s_2 - s_1| \leq \int_{s_1}^{s_2} |\phi^2(x, s)| \, ds / |s_2 - s_1| \leq \beta_2.$$

Choosing $\beta_2 = 1 + \frac{\varepsilon|\log \varepsilon|}{2}$, this is smaller than $(2\bar{s}_2)^{-\varepsilon} \leq |s_2 - s_1|^{-\varepsilon}$, as required. Indeed, for $\varepsilon \leq 1$, $\beta_2 = 1 + \frac{\varepsilon|\log \varepsilon|}{2} \leq \sqrt{\varepsilon}^{-\varepsilon} \leq \left(4\sqrt{\frac{\varepsilon|\log \varepsilon|}{3}}\right)^{-\varepsilon} \leq (2\bar{s}_2)^{-\varepsilon}$. It remains to show $\left|\int_{s_1}^{s_2} \phi(x, s) \, ds\right| \leq |s_2 - s_1|^{1-\varepsilon}$ for $\hat{s}_1$ or $\hat{s}_2$ outside $[-\bar{s}_2, \bar{s}_2]$. In the next paragraph we will show

$$|\phi^2(x, s)| \leq (1 - \varepsilon)|2\hat{s}|^{-\varepsilon} \quad \text{for } \hat{s} \in [-\beta_2, \beta_2] \backslash [-\bar{s}_2, \bar{s}_2], \qquad (3.4.1)$$

which then yields for a given $\Delta s = |s_2 - s_1|$ (exploiting that $|\phi^2|$ decreases to both sides of $\hat{s} = 0$)

$$\left|\int_{s_1}^{s_2} \phi^2(x, s) \, ds\right| \leq \int_{s_1}^{s_2} |\phi^2(x, s)| \, ds \leq \int_{x_1 - \frac{\Delta s}{2}}^{x_1 + \frac{\Delta s}{2}} |\phi^2(x, s)| \, ds$$

$$\leq \begin{cases} (\Delta s)^{1-\varepsilon} & \text{if } \frac{\Delta s}{2} \leq \bar{s}_2, \\ (2\bar{s}_2)^{1-\varepsilon} + 2\int_{\bar{s}_2}^{\frac{\Delta s}{2}} (1 - \varepsilon)|2\hat{s}|^{-\varepsilon} \, d\hat{s} = (\Delta s)^{1-\varepsilon} & \text{else,} \end{cases}$$

as required.

To prove inequality (3.4.1), let us abbreviate $t = \frac{(1-\varepsilon)^{1/\varepsilon}}{2} \approx \frac{1}{2e}$. Obviously,

$$(1 - \varepsilon)|2\hat{s}|^{-\varepsilon} = |\tfrac{t}{\hat{s}}|^{-\varepsilon} \geq 1 \geq |\phi^2(x, s)| \quad \text{for } \hat{s} \in [-t, t] \backslash [-\bar{s}_2, \bar{s}_2],$$

while for $\hat{s} \in [-\beta_2, \beta_2] \backslash [t, t]$ we have

$$|\phi^2(x, s)| \leq \sqrt{\frac{\beta_2^4 + t^2/4}{\beta_2^2 + t^2}} \leq (1 - \varepsilon)|2\beta_2|^{-\varepsilon} \leq (1 - \varepsilon)|2\hat{s}|^{-\varepsilon},$$

where the middle inequality holds for ε small enough due to $\sqrt{\frac{\beta_2^4 + t^2/4}{\beta_2^2 + t^2}} \rightarrow$ $\sqrt{\frac{(4e)^2 + 1}{(4e)^2 + 4}} < 1$ and $(1 - \varepsilon)|2\beta_2|^{-\varepsilon} \rightarrow 1$ as $\varepsilon \rightarrow 0$. $\square$

3.5. Numerical Optimization of Urban Planning and Branched Transport Networks

The proposed convex reformulation of the branched transport and urban planning model as variants of the Mumford–Shah segmentation problem allows a numerical network optimization that cannot get stuck in local minima due to the convexity. We shall use a finite difference discretization of the right-hand side in (3.3.1) (similarly to [38]) and apply a simple primal-dual algorithm to numerically find the saddle point.

In the following, for simplicity we assume without loss of generality a rectangular domain $V \subset \mathbb{R}^2$ with bottom left corner at the origin. We discretize the domain $V \times \mathbb{R}$ of the lifted function v by a finite three-dimensional $n \times m \times p$ grid

$$\mathcal{G} = \{(ih_1, jh_2, lh_s) : i = 0, \ldots, n, \ j = 0, \ldots, m, \ l = 0, \ldots, p\},$$

where $h_1, h_2, h_s > 0$ denote the grid size in each direction. Hence, the discrete counterparts of $v \in \mathrm{BV}(V \times \mathbb{R}; [0, 1])$ and $\phi \in C_0^\infty(V \times \mathbb{R}; \mathbb{R}^2 \times \mathbb{R})$

are given by $v^h : \mathcal{G} \to [0,1]$ and $\phi^h : \mathcal{G} \to \mathbb{R}^2 \times \mathbb{R}$. For the sake of simplicity, for every $(ih_1, jh_2, lh_s) \in \mathcal{G}$, we write $v^h_{ijl} = v^h(ih_1, jh_2, lh_s)$ and $\phi^h_{ijl} = \phi^h(ih_1, jh_2, lh_s)$. Defining the finite difference gradient operator $D = (D_1, D_2, D_s)^T$ by

$$(D_1 v^h)_{ijl} = \frac{v^h_{i+1,j,l} - v^h_{i,j,l}}{h_1}, \quad (D_2 v^h)_{ijl} = \frac{v^h_{i,j+1,l} - v^h_{i,j,l}}{h_2}, \quad (D_s v^h)_{ijl} = \frac{v^h_{i,j,l+1} - v^h_{i,j,l}}{h_s},$$

the discretized form of saddle point problem (3.3.1) reads

$$\min_{v^h \in \mathcal{C}^h} \max_{\phi^h \in \mathcal{K}^h} \sum_{i,j,l} \phi^h_{ijl} (Dv^h)_{ijl}, \tag{3.5.1}$$

in which the discrete versions of the convex sets $\tilde{\mathcal{C}}$ and $\mathcal{K} = \mathcal{K}_1$ (for urban planning) or $\mathcal{K}_2$ (for branched transport) are given by

$$\mathcal{C}^h = \Big\{ v^h : \mathcal{G} \to [0,1] : v^h_{ij0} = 1, \ v^h_{ijp} = 0 \ \forall i, j, \ v^h$$

$$= 1^h_{u(\mu_+, \mu_-)} \text{ on } \mathcal{G} \backslash (\overline{\Omega} \times \mathbb{R}) \Big\},$$

$$\mathcal{K}^h_1 = \Big\{ \phi^h = (\phi^h_x, \phi^h_s) : \mathcal{G} \to \mathbb{R}^2 \times \mathbb{R} :$$

$$\phi^h_s \geq 0, \ |\phi^h_x| \leq a, \ |h_s \sum_{l=s_1}^{s_2} (\phi^h_x)_{ijl}|$$

$$\leq \min\{|s_2 - s_1| + \varepsilon, a|s_2 - s_1|\} \ \forall i, j, s_1 < s_2 \Big\},$$

$$\mathcal{K}^h_2 = \Big\{ \phi^h = (\phi^h_x, \phi^h_s) : \mathcal{G} \to \mathbb{R}^2 \times \mathbb{R} : \phi^h_s \geq 0, \ |h_s \sum_{l=s_1}^{s_2} (\phi^h_x)_{ijl}|$$

$$\leq |s_2 - s_1|^{1-\varepsilon} \ \forall i, j, s_1 < s_2 \Big\}.$$

Above, $1^h_{u(\mu_+, \mu_-)}$ denotes the restriction of $1_{u(\mu_+, \mu_-)}$ onto the grid $\mathcal{G}$.

The discretized saddle point problem is solved using the primal-dual algorithm from [24], which alternatingly performs a primal gradient descent and a dual gradient ascent step with step length τ (respectively σ), accompanied by an over-relaxation with parameter θ. Denoting the kth approximation to v^h and ϕ^h by v^k and ϕ^k, respectively, the iterative algorithm reads

$$\begin{cases} v^{k+1} = P_{\mathcal{C}^h}(v^k - \tau D^* \phi^k), \\ \overline{v}^{k+1} = v^{k+1} + \theta(v^{k+1} - v^k), \\ \phi^{k+1} = P_{\mathcal{K}^h}(\phi^k + \sigma D\overline{v}^{k+1}), \end{cases}$$

starting from an arbitrary initialization v^0, ϕ^0. The orthogonal projection $P_{\mathcal{K}^h}$ onto $\mathcal{K}^h$ is performed by Dykstra's method [7], while the projection $P_{\mathcal{C}^h}$ onto the set $\mathcal{C}^h$ is straightforward. When v^k has sufficiently converged, it is typically close to binary. The rounding of v^k to $\{0,1\}$ is then interpreted as the sought function 1_u, and the optimal flux is identified as $Du^\perp$. While the primal-dual algorithm is very
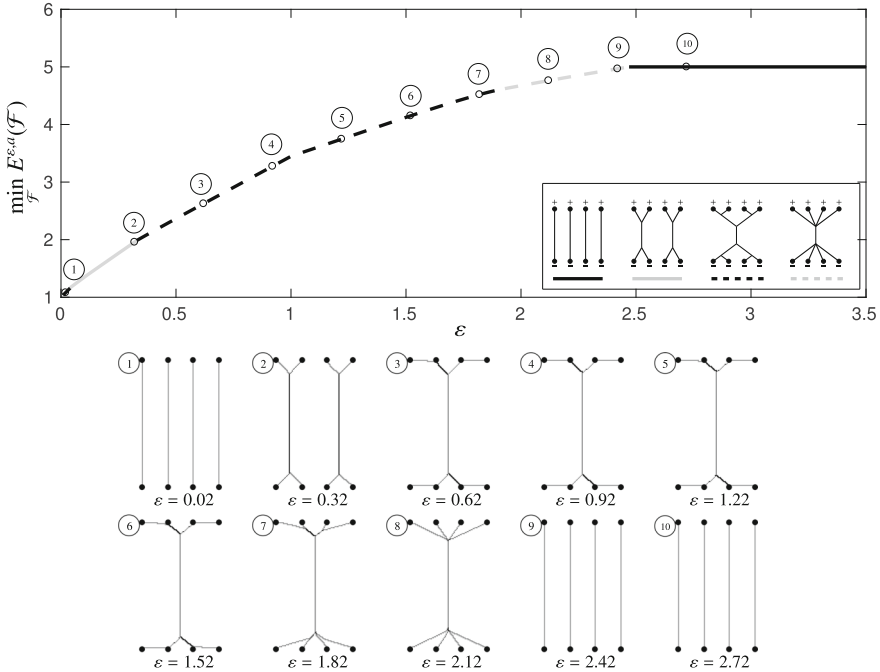
Fig. 3. Parameter study for urban planning. Top: plot of the manually computed minimal energy for different values of ε and fixed $a = 5$. The line type indicates the optimal network topology. Bottom: numerically computed optimal fluxes for evenly spaced values of ε in the same range. The numerically obtained network topologies match the predicted ones except for example ⑨

simple and straightforward to implement, there may be more efficient optimization approaches that better exploit the structure of the optimization problem. In particular, (3.5.1) has the form of a second order cone program in ϕ^h (which can be easily seen by first performing the minimization in v^h) with conic constraints $\phi^h \in \mathcal{K}^h$, for which there are efficient interior point algorithms. Note, though, that the primal-dual algorithm is more straightforward to parallelize, for instance on graphics cards.

Figures 3 and 4 show numerically optimized urban planning and branched transportation networks or rather fluxes between four evenly-spaced point sources at the top of a rectangular domain and four evenly spaced point sinks at the bottom of the domain. For such a simple geometric setting one can still calculate the true optimal solution manually by enumerating all possible network topologies and optimizing their vertex positions by hand. In both figures, the top graph shows the manually computed optimal energy and corresponding network topology for different parameter values. Below, numerical solutions are shown that uniformly sample the same parameter range. Except for example ⑨ in Fig. 3 and example ③ in Fig. 4, the numerical solutions coincide with the predicted, truly optimal network topology. The discrepancy between the numerical and the true solution in examples ⑨ and ③ (which both lie close to a bifurcation point) may be due to our discretization grid,
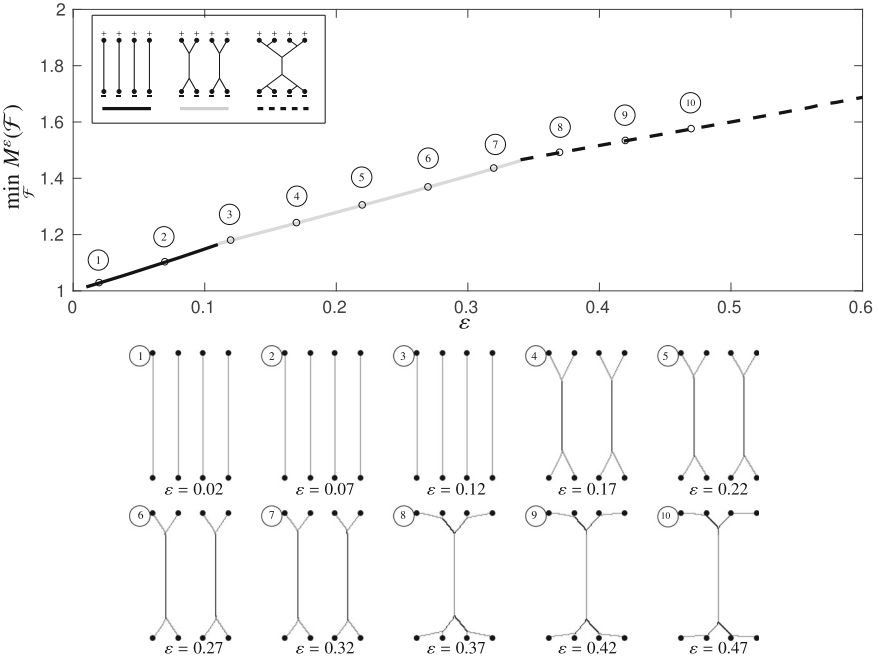
Fig. 4. Parameter study for branched transport. Top: plot of the manually computed minimal energy for different values of ε. The line type indicates the optimal network topology. Bottom: numerically computed optimal fluxes for evenly spaced values of ε in the same range. The numerically obtained network topologies match the predicted ones except for example ③
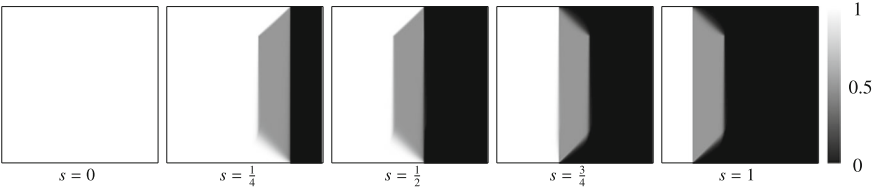

Fig. 5. Example of a numerical optimization for urban planning, resulting in a non-binary solution v (the images show different cross-sections). This indicates a lack of tightness of the convex reformulation for the chosen parameters ($a = 2.13$ and $\varepsilon = 0.5$). The parameters lie close to a bifurcation at which the truly optimal network topology changes from four vertical pipes to a single tree

which is vertically aligned and thus slightly favours vertical structures, since at fixed grid resolution diagonal structures cannot be represented as exactly. Note that unlike for other numerical approaches, the shown result cannot just be a suboptimal local optimum, since the numerical method always leads to the global optimum due to convexity.

If the proposed convex reformulation of urban planning and branched transport were always tight, the optimal solution v to (3.3.1) would be binary and only take values 0 or 1. However, in some situations this is not the case. For the case of urban planning, Fig. 5 shows a numerically optimized function v which distinctly takes on
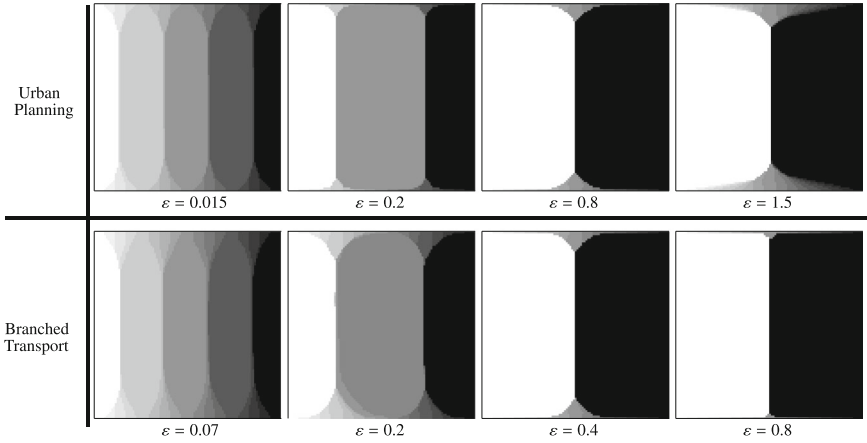
Fig. 6. Numerical optimization results for transport from 16 more or less evenly spaced point sources of same mass at the top to 16 evenly spaced point sinks (of same mass as well) at the bottom ($a = 5$ in case of urban planning). Instead of the optimal flux we show the corresponding optimal image u
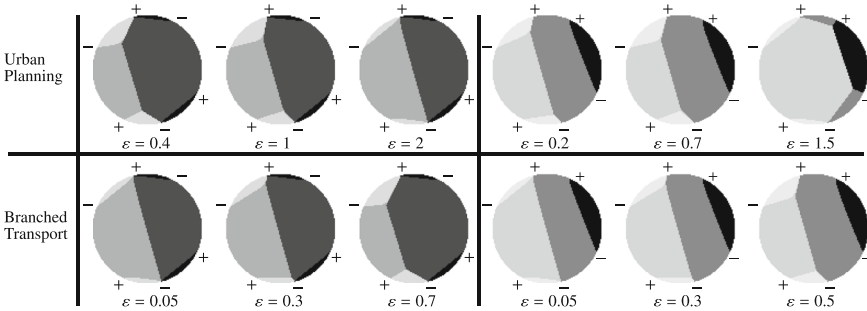

Fig. 7. Numerical optimization results for urban planning and branched transport with different parameters ($a = 5$ in case of urban planning). In the left column, the prescribed masses are $+\frac{1}{2}, -\frac{1}{8}, +\frac{1}{8}, -\frac{1}{2}, +\frac{1}{8}, -\frac{1}{8}$ (counterclockwise from top), in the right column $+\frac{1}{2}, -\frac{1}{8}, +\frac{1}{8}, -\frac{1}{2}, -\frac{1}{2}, +\frac{1}{2}$

three values, 0, $\frac{1}{2}$, and 1. In our numerical experience, this may sometimes happen close to a bifurcation point, where the optimal network topology changes. Indeed, the parameters of Fig. 5 lie almost exactly at a bifurcation where the truly optimal solution changes from four vertical pipes to a single tree.

Figure 6 shows numerical results obtained on a rectangular domain with 16 evenly spaced point sources of same mass at the top and as many point sinks (of same mass as well) at the bottom (approximating a continuous line measure discretely). Instead of the optimal flux we here show the corresponding optimal image u, whose jump set represents the network. Depending on the chosen parameters, mass is preferentially collected in one or more tree-like networks before being separated again. With decreasing parameter ε (which in a way encodes the efficiency of transporting mass in bulk) the number of trees increases.
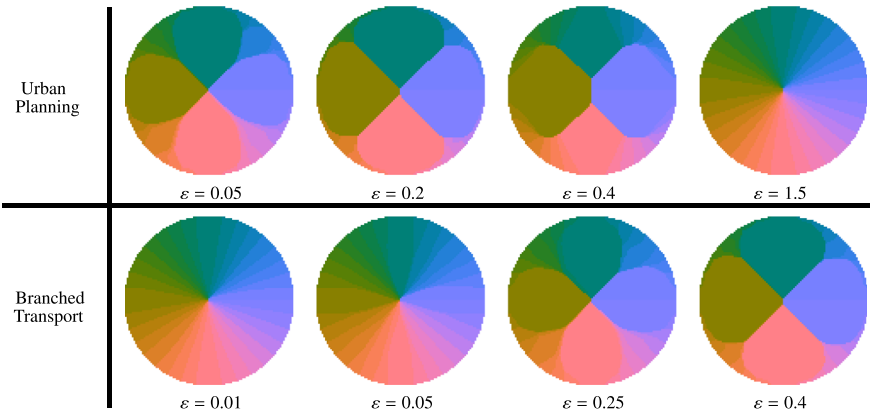
Fig. 8. Numerical optimization results for urban planning and branched transport for a single point source at the centre of the circular domain to 32 point sinks on the boundary ($a = 5$ in case of urban planning). The discontinuity set of the image corresponds to the optimal network. For this geometry, the image u takes values in S^1, which is here indicated by the periodic colour scale

Figure 7 shows computational examples in which the underlying domain is non-rectangular and in which sources and sinks of different weights alternate on the domain boundary. A more complex case is displayed in Fig. 8, where we simulated the transport from a point source at the centre of the circular domain to 32 point sinks of equal mass on the boundary (approximating a uniform measure). Such a geometry can be achieved by a trick using a periodic covering of the disk. In detail, we chose the domain $\Omega = B_1(0)\backslash\{0\}$ and prescribed the sought image u outside Ω taking values in S^1 rather than $\mathbb{R}$ (in between each two point sinks the image is prescribed to equal the midpoint between both sink positions on S^1). Now $\mathbb{R}$ is interpreted as a covering of S^1 via the mapping $s \mapsto (\cos s, \sin s)$ so that the calculations can be performed as before (similarly to the approach in [25]).

4. Discussion

The key idea to arrive at the convexification of network optimization in this work was the identification of network optimization with particular image inpainting problems, for which convexifications are known.

Let us stress that, even though we considered a simple rectangular or circular problem domain Ω, the same approach can easily be applied for more general convex domains.

However, the formulation as image inpainting imposes two major restrictions on the network optimization problem. First of all, such a formulation is only possible in two spatial dimensions because otherwise there cannot be any interpretation of rotated mass fluxes as image gradients. Secondly, since image gradients are always curl-free, the corresponding mass fluxes must be divergence-free in the domain interior. As a consequence, the source and sink distribution μ_+ and μ_- must lie on

the boundary of the domain. Deviations from this situation may be possible (see Fig. 8), but introduce additional complications.

The convexification derived in this work has only been shown to represent a lower bound to the original network optimization. It still remains an open question when it is actually equivalent. This question naturally decomposes into two separate issues: First one has to investigate whether the network optimization problems and their Mumford–Shah versions are equivalent. As explained in Sect. 3.2, this leads to analysing the regularity of the singular set of minimizers of Mumford–Shah-type functionals, which is notoriously difficult and for which results are so far only partial. For instance, the Mumford–Shah conjecture for the standard Mumford–Shah functional has been proved in two dimensions assuming that the singular set is connected by Bonnet in [11]. The second issue concerns the tightness of the convexification via functional lifting. For some lifted functionals there exist thresholding results showing that the solution of the original problem can be recovered from the convexified problem via thresholding (see for instance [20]), however, our numerical experiments indicate that this may sometimes not be the case. Nevertheless, our proposed convexification turned out to be tight enough to derive the sharp lower bound in the energy scaling law for complicated networks, and also in most numerical simulations it yielded the truly optimal network geometry.

A. Auxiliary Results

In this section we first provide the deferred proof of Lemma 3.1.4, after which we collect some density and continuity results needed for the the central statement of this article, Theorem 3.2.2.

Proof. (Proof of Lemma 3.1.4) We first verify that indeed $\mathcal{F}_u \llcorner \overline{\Omega} \in \mathcal{A}_{\mathcal{F}}(\mu_+, \mu_-)$. Denoting by n the unit outward normal to $\partial(V \backslash \overline{\Omega})$ and by T the trace operator on $\mathrm{BV}(V \backslash \overline{\Omega})$ [27, S 5.3], for $\phi \in \mathcal{C}_c^\infty(V)$ we have

$$
\int_V \phi \, \mathrm{d} \operatorname{div}(\mathcal{F}_u \llcorner \overline{\Omega}) = -\int_{\overline{\Omega}} \nabla\phi \cdot \mathrm{d}\mathcal{F}_u = \int_{\overline{\Omega}} \nabla\phi^\perp \cdot \mathrm{d}Du
$$

$$
= \int_V \nabla\phi^\perp \cdot \mathrm{d}Du - \int_{V \backslash \overline{\Omega}} \nabla\phi^\perp \cdot \mathrm{d}Du
$$

$$
= -\int_V u \operatorname{div}(\nabla\phi^\perp) \, \mathrm{d}\mathcal{L}^2 + \int_{V \backslash \overline{\Omega}} u \operatorname{div}(\nabla\phi^\perp) \, \mathrm{d}\mathcal{L}^2
$$

$$
- \int_{\partial\Omega} Tu\nabla\phi^\perp \cdot n \, \mathrm{d}\mathcal{H}^1 = -\int_{\partial\Omega} Tu\nabla\phi^\perp \cdot n \, \mathrm{d}\mathcal{H}^1
$$

$$= -\int_{[0,\mathcal{H}^1(\partial\Omega))} Tu(\gamma(t))\tfrac{d}{dt}\phi(\gamma(t))\,dt$$

$$= \int_{[0,\mathcal{H}^1(\partial\Omega))} \phi(\gamma(t))\,d\tfrac{d}{dt}Tu(\gamma(t))$$

$$= \int_{\partial\Omega} \phi\,d(\mu_+ - \mu_-) = \int_V \phi\,d(\mu_+ - \mu_-),$$

where we have used $Tu(\gamma(t)) = u(\mu_+, \mu_-) \circ \gamma(t)$ with $\tfrac{d}{dt}u(\mu_+, \mu_-) \circ \gamma = \mu_+ - \mu_-$.

Next, let $\mathcal{F} \in \mathcal{A}_{\mathcal{F}}(\mu_+, \mu_-)$ be given. We extend this flux to a flux $\tilde{\mathcal{F}}$ on $\mathbb{R}^2$ as follows:

$$\tilde{\mathcal{F}} = \mathcal{F} + (Du(\mu_+, \mu_-)^\perp) \llcorner (V \backslash \overline{\Omega}).$$

It is straightforward to check $\operatorname{div}\tilde{\mathcal{F}} = 0$ in the distributional sense. Indeed, for $\phi \in \mathcal{C}_c^\infty(\mathbb{R}^2)$ and n the inner normal to $\partial\Omega$, we have

$$\int_{\mathbb{R}^2} \nabla\phi \cdot d\tilde{\mathcal{F}}(x) = \int_{\overline{\Omega}} \nabla\phi \cdot d\mathcal{F}(x) + \int_{\mathbb{R}^2 \backslash \overline{\Omega}} \nabla\phi \cdot dDu(\mu_+, \mu_-)^\perp$$

$$= -\int_{\overline{\Omega}} \phi\,d\operatorname{div}\mathcal{F} - \int_{\mathbb{R}^2 \backslash \overline{\Omega}} \nabla\phi^\perp \cdot dDu(\mu_+, \mu_-)$$

$$= \int_{\overline{\Omega}} \phi\,d(\mu_- - \mu_+) - \int_{\partial\Omega} Tu\nabla\phi^\perp \cdot n\,d\mathcal{H}^1$$

$$+ \int_{\mathbb{R}^2 \backslash \overline{\Omega}} u(\mu_+, \mu_-)\operatorname{div}(\nabla\phi^\perp)\,d\mathcal{L}^2$$

$$= \int_{\overline{\Omega}} \phi\,d(\mu_- - \mu_+) + \int_{\partial\Omega} \phi\,d(\mu_+ - \mu_-) = 0.$$

Next, we mollify $\tilde{\mathcal{F}}$ using a smooth Dirac sequence $\eta_\delta \in \mathcal{C}_c^\infty(\mathbb{R}^2)$ with support in the δ-ball around the origin, $\operatorname{spt}\eta_\delta \subset B_\delta(0)$,

$$\mathcal{F}_\delta = \tilde{\mathcal{F}} * \eta_\delta.$$

For any $\delta > 0$, $\mathcal{F}_\delta$ is smooth and divergence-free (with a slight misuse of notation we interpret $\mathcal{F}_\delta$ as the density with respect to $\mathcal{L}^2$). Now define

$$\mathcal{G}_\delta = -\mathcal{F}_\delta^\perp, \quad u_\delta(x) = (u(\mu_+, \mu_-) * \eta_\delta)(-1, 0) + \int_\gamma \mathcal{G}_\delta \cdot d\gamma,$$

where γ is any Lipschitz-continuous path connecting $(-1, 0)^T$ with x. Since $\mathcal{G}_\delta$ is curl-free (recall that the curl of a two-dimensional vector field is defined via the curl of the embedded vector field in $\mathbb{R}^3$) and thus conservative, u_δ is independent of the particular path γ chosen. We have

$$\nabla u_\delta = \mathcal{G}_\delta = -\mathcal{F}_\delta^\perp \overset{*}{\rightharpoonup} -\tilde{\mathcal{F}}^\perp$$

in $\mathrm{rca}(V)$ as $\delta \to 0$. Furthermore, due to $\nabla u_\delta = -\mathcal{F}_\delta^\perp = (Du(\mu_+, \mu_-)) * \eta_\delta = \nabla(u(\mu_+, \mu_-) * \eta_\delta)$ in $V \backslash B_\delta(\Omega)$ and $u_\delta(-1, 0) = (u(\mu_+, \mu_-) * \eta_\delta)(-1, 0)$ we have $u_\delta = u(\mu_+, \mu_-) * \eta_\delta$ on $V \backslash B_\delta(\Omega)$ so that

$$\sup_{x \in V \backslash B_\delta(\Omega)} |u_\delta(x)| \leq \sup_{x \in \mathbb{R}^2} |u(\mu_+, \mu_-)(x)| \leq \mu_+(\partial\Omega) + \mu_-(\partial\Omega).$$

Thus, $\|u_\delta\|_{L^1(V)}$ is uniformly bounded by virtue of Poincaré's inequality and the uniform boundedness of $\|\nabla u_\delta\|_{L^1(V)}$. Combining the above properties, we see that as $\delta \to 0$ a subsequence of u_δ converges weakly-$*$ in $\mathrm{BV}(V)$ against some function $u_\mathcal{F} \in \mathrm{BV}(V)$ which satisfies all conditions from the definition of $\mathcal{A}_u(\mu_+, \mu_-)$ as well as $Du_\mathcal{F}^\perp \llcorner \overline{\Omega} = \mathcal{F}$ (recall that u_n is said to converge weakly-$*$ in $\mathrm{BV}(V)$ against u, $u_n \overset{*}{\rightharpoonup} u$, if $u_n \to u$ in $L^1(V)$ and $Du_n \overset{*}{\rightharpoonup} Du$ in $\mathrm{rca}(V)$ [2, Definition 3.11]). The uniqueness of $u_\mathcal{F}$ follows in the standard way: assume, $u_i \in \mathcal{A}_u(\mu_+, \mu_-)$, $i = 1, 2$, satisfy $\mathcal{F} = Du_i^\perp \llcorner \overline{\Omega}$, then $u = u_2 - u_1$ satisfies $Du = 0$ as well as $u = 0$ on $V \backslash \Omega$ so that $u \equiv 0$. $\square$

Next we provide some standard results required for the proof of Theorem 3.2.2, which establishes a relation between flux-based and Mumford–Shah-type functionals. We first show that in Definition 2.1.3 of the cost function one may restrict to initial and final measures of the approximating graph sequences with support on $\partial\Omega$. This will allow us to represent each discrete flux of an approximating graph sequence by a corresponding image.

Lemma A.0.1. (Reduction of measure support) *Let* $\mu_+, \mu_- \in \mathrm{fbm}(\mathbb{R}^2)$ *of equal mass, satisfying* (3.1.1)*, and let* $\mathcal{J}$ *either denote the branched transport functional* M_F^ε *or the urban planning functional* $E_\mathrm{F}^{\varepsilon, a}$. *We have*

$$\mathcal{J}(\mathcal{F}) = \inf \left\{ \liminf_{n \to \infty} \mathcal{J}(G_n) : (\mathcal{F}_{G_n}, \mu_+^n, \mu_-^n) \overset{*}{\rightharpoonup} (\mathcal{F}, \mu_+, \mu_-), \right.$$
$$\left. \mu_+^n, \mu_-^n \text{ satisfy } (3.1.1) \right\}.$$

Proof. Let $(\mathcal{F}_{G_n}, \mu_+^n, \mu_-^n) \overset{*}{\rightharpoonup} (\mathcal{F}, \mu_+, \mu_-)$ be an approximating graph sequence. Then, we have $\mathcal{J}(\mathcal{F}) = \lim_{n \to \infty} \mathcal{J}(G_n)$. We will construct another approximating graph sequence $(\mathcal{F}_{\tilde{G}_n}, \tilde{\mu}_+^n, \tilde{\mu}_-^n)$ with $\mathrm{spt}\tilde{\mu}_+^n$, $\mathrm{spt}\tilde{\mu}_-^n \subseteq \partial\Omega$ and the same limit energy $\mathcal{J}(\mathcal{F}) = \lim_{n \to \infty} \mathcal{J}(\tilde{G}_n)$. Indeed, for $\mu_+^n = \sum_{i=1}^{N_n} a_i^n \delta_{x_i^n}$ and $\mu_-^n = \sum_{i=1}^{M_n} b_i^n \delta_{y_i^n}$ we set

$$\tilde{\mu}_+^n = \sum_{i=1}^{N_n} a_i^n \delta_{\pi_{\partial\Omega}(x_i^n)}, \qquad \tilde{\mu}_-^n = \sum_{i=1}^{M_n} b_i^n \delta_{\pi_{\partial\Omega}(y_i^n)},$$

where $\pi_{\partial\Omega} : \mathbb{R}^2 \to \partial\Omega$ is the closest point projection (as before, in case of non-uniqueness an arbitrary closest point is chosen). Since $\tilde{\mu}_+^n$ is the measure μ_+^n projected onto $\partial\Omega$, which is Lipschitz regular and contains the support of μ_+, it is clear that $\mu_+^n \overset{*}{\rightharpoonup} \mu_+$ implies $\tilde{\mu}_+^n \overset{*}{\rightharpoonup} \mu_+$ (analogously for $\tilde{\mu}_-^n$). Let further G_+^n be a $\mathcal{J}$-minimizing transport path between $\tilde{\mu}_+^n$ and μ_+^n and G_-^n a $\mathcal{J}$-minimizing transport path between μ_-^n and $\tilde{\mu}_-^n$. Such transport paths exist, since

optimal mass fluxes between discrete measures are known to be discrete graphs (in the urban planning case this follows from the same argument as in [16, proof of Prop. 4.3.1]; in the branched transport case this is contained in [5, Theorem 4.7] or [43, Theorem 4.10]). Moreover, note that by Remark 2.1.4 the minimum urban planning cost is bounded above by a times the Wasserstein-1 distance W_1, which metrises weak-$*$ convergence, and that the minimum branched transport cost is continuous with respect to weak-$*$ convergence [42, Lemma 4.1]. Thus we have

$$\mathcal{J}(G_\pm^n) \to 0 \text{ as } n \to \infty \text{ due to } W_1(\mu_\pm^n, \tilde{\mu}_\pm^n) \to 0.$$

We now set $\tilde{G}_n = (V(G_n) \cup V(G_+^n) \cup V(G_-^n), E(G_n) \cup E(G_+^n) \cup E(G_-^n))$. It is clear that $\tilde{G}_n$ is a transport path between $\tilde{\mu}_+^n$ and $\tilde{\mu}_-^n$. Furthermore, $\mathcal{J}(G_\pm^n) \to_{n\to\infty} 0$ implies $\mathcal{F}_{G_\pm^n} \overset{*}{\rightharpoonup} 0$ and thus

$$\mathcal{F}_{\tilde{G}_n} = \mathcal{F}_{G_n} + \mathcal{F}_{G_+^n} + \mathcal{F}_{G_-^n} \overset{*}{\rightharpoonup} \mathcal{F} \text{ as } n \to \infty.$$

Finally, as desired, we have

$$\mathcal{J}(\tilde{G}_n) \leq \mathcal{J}(G_n) + \mathcal{J}(G_+^n) + \mathcal{J}(G_-^n) \to \mathcal{J}(\mathcal{F}) \quad \text{and}$$
$$\mathcal{J}(\tilde{G}_n) \geq \mathcal{J}(G_n) \to \mathcal{J}(\mathcal{F})$$

as $n \to \infty$. $\square$

The next lemma establishes the desired equivalence between flux-based and Mumford–Shah-type energies on the level of discrete fluxes.

Lemma A.0.2. (Equivalence for discrete fluxes) *Let $\mu_+, \mu_- \in \text{fbm}(\mathbb{R}^2)$ of equal mass, satisfying* (3.1.1)*, and let G be a transport path between μ_+ and μ_-. Then*

$$E_F^{\varepsilon,a}[G] = \tilde{E}^{\varepsilon,a}[u_{\mathcal{F}_G}], \qquad M_F^\varepsilon[G] = \tilde{M}^\varepsilon[u_{\mathcal{F}_G}]$$

for $\mathcal{F}_G$ from Definition 2.1.1 and $u_{\mathcal{F}_G}$ from Definition 3.1.5.

Proof. For simplicity let us abbreviate $u \equiv u_{\mathcal{F}_G}$. By Lemma 3.1.4, $u \in \mathcal{A}_u(\mu_+, \mu_-)$ as well as

$$Du \llcorner \overline{\Omega} = - \sum_{e \in E(G)} w(e)(\mathcal{H}^1 \llcorner e)\hat{e}^\perp$$

for $\hat{e}$ the unit vector parallel to the edge e. In particular, u is piecewise constant with a discontinuity set $S_u \cap \overline{\Omega} = \bigcup_{e \in E(G)} e$ on which it has jump size $[u] = w(e)$. Inserting this into the urban planning and branched transport energy, we obtain

$$\tilde{E}^{\varepsilon,a}[u] = \int_{S_u \cap \overline{\Omega}} \min\{a[u], [u] + \varepsilon\} \, d\mathcal{H}^1(x)$$
$$= \sum_{e \in E(G)} l(e) \min\{aw(e), w(e) + \varepsilon\} = E_F^{\varepsilon,a}[G],$$
$$\tilde{M}^\varepsilon[u] = \int_{S_u \cap \overline{\Omega}} [u]^{1-\varepsilon} \, d\mathcal{H}^1(x) = \sum_{e \in E(G)} l(e)w(e)^{1-\varepsilon} = M_F^\varepsilon[G],$$

which is the desired result. $\square$

The previous energy relation is extended from discrete fluxes to general ones by continuity, for which reason we need to show a continuity property of the Mumford–Shah-type functionals. Since all functions $u \in \mathcal{A}_u(\mu_+, \mu_-)$ satisfy $u(x) = u(\pi_{\partial\Omega}(x)) \forall x \in V\backslash\overline{\Omega}$, it suffices to examine continuity on the set of functions

$$H = \{u \in \mathrm{BV}(V) : u(x) = u(\pi_{\partial\Omega}(x)) \forall x \in V\backslash\overline{\Omega}\}. \tag{A.0.2}$$

Note that H is closed in $\mathrm{BV}(V)$ with respect to weak-$*$ convergence (even with respect to $L^1(V)$-convergence). Indeed, let $u_n \in H$ with $u_n \overset{*}{\rightharpoonup} u$ in $\mathrm{BV}(V)$, then a subsequence converges pointwise almost everywhere against u, and since the u_n are constant along rays orthogonal to $\partial\Omega$, the pointwise limit must be so as well.

Lemma A.0.3. (Lower semi-continuity) $\tilde{E}^{\varepsilon,a}$ and $\tilde{M}^{\varepsilon}$ are sequentially weakly-$*$ lower semi-continuous on H.

Proof. For $u \in H$, $n \in \mathbb{N}$, and any measurable set $B \subset V$ define

$$E_B[u] = a|Du|(B\backslash S_u) + \int_{S_u \cap B} \min\{a[u], [u] + \varepsilon\} \, d\mathcal{H}^1,$$

$$M_B^n[u] = \int_{S_u \cap B} \min\{n[u], [u]^{1-\varepsilon}\} \, d\mathcal{H}^1 + n|Du|(B\backslash S_u).$$

We have $\tilde{E}^{\varepsilon,a}[u] = E_{\overline{\Omega}}[u]$ as well as $\tilde{M}^{\varepsilon}[u] = \sup_n M_{\overline{\Omega}}^n[u]$. Note that by the monotone convergence theorem one has

$$\sup_n \int_{S_u \cap B} \min\{n[u], [u]^{1-\varepsilon}\} \, d\mathcal{H}^1 = \lim_{n\to\infty} \int_{S_u \cap B} \min\{n[u], [u]^{1-\varepsilon}\} \, d\mathcal{H}^1$$

$$= \int_{S_u \cap B} [u]^{1-\varepsilon} \, d\mathcal{H}^1.$$

Thus it suffices to show the sequential weak-$*$ lower semi-continuity of $E_{\overline{\Omega}}$ and $M_{\overline{\Omega}}^n$ for all n.

Now let $B_\delta = B_\delta(\Omega)$ be the open δ-neighbourhood of Ω. By [2, Theorem 5.4], E_{B_δ} and $M_{B_\delta}^n$ are sequentially weakly-$*$ lower semi-continuous on $\mathrm{BV}(B_\delta)$. Consider a sequence $u_k \overset{*}{\rightharpoonup} u$ in H and let $C \in \mathbb{R}$ be an upper bound on its total variation, $C > \|Du_k\|_{\mathrm{rca}}, \|Du\|_{\mathrm{rca}}$. We have

$$E_{\overline{\Omega}}(u) \leq E_{B_\delta}(u) \leq \liminf_{k\to\infty} E_{B_\delta}(u_k) \leq \liminf_{k\to\infty} E_{\overline{\Omega}}(u_k)$$

$$+a|Du|(B_\delta\backslash\overline{\Omega}) \leq \liminf_{k\to\infty} E_{\overline{\Omega}}(u_k) + a\delta C.$$

The last inequality uses the fact that outside Ω, u is constant along rays orthogonal to $\partial\Omega$ so that the total variation of u on each set $S_s = \{x \in \mathbb{R}^2\backslash\overline{\Omega} : \mathrm{dist}(x, \partial\Omega) = s\}$ is the same for all $s > 0$, where dist denotes the Euclidean distance. Indeed, writing ∂_t for the tangential derivative, we have

$$|Du|(B_\delta\backslash\overline{\Omega}) = \int_{(0,\delta)} |Du|(S_s) \, ds = \int_{(0,\delta)} |\partial_t u|(S_s) \, ds = \delta|\partial_t u|(\partial\Omega)$$

$$\leq \delta|Du|(\partial\Omega) \leq \delta C.$$

Since δ is arbitrary, the above implies the sequential weak-$*$ lower semi-continuity of $E_{\overline{\Omega}}$. Analogously,

$$M_{\overline{\Omega}}^n(u) \le M_{B_\delta}^n(u) \le \liminf_{k \to \infty} M_{B_\delta}(u_k) \le \liminf_{k \to \infty} M_{\overline{\Omega}}(u_k)$$
$$+ n|Du|(B_\delta \setminus \overline{\Omega}) \le \liminf_{k \to \infty} M_{\overline{\Omega}}(u_k) + n\delta C,$$

again implying the sequential weak-$*$ lower semi-continuity of $M_{\overline{\Omega}}^n$ by the arbitrariness of δ. $\quad\square$

References

1. ALBERTI, G., BOUCHITTÉ, G., DAL MASO, G.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. *Calc. Var. Partial Differ. Equ.* **16**(3), 299–333, 2003

2. AMBROSIO, L., FUSCO, N., PALLARA, D.: *Functions of Bounded Variation and Free Discontinuity Problems*. Oxford Mathematical Monographs. Oxford University Press, New York, 2000

3. AMBROSIO, L.: Existence theory for a new class of variational problems. *Arch. Rational Mech. Anal.* **111**(4), 291–322, 1990

4. BRANCOLINI, A., BUTTAZZO, G.: Optimal networks for mass transportation problems. *ESAIM Control Optim. Calc. Var.* 11(1), 88–101 (electronic), 2005

5. BERNOT, M., CASELLES, V., MOREL, J.-M.: The structure of branched transportation networks. *Calc. Var. Partial Differ. Equ.*, **32**(3), 279–317, 2008

6. BERNOT, M., CASELLES, V., MOREL, J.-M.: *Optimal Transportation Networks, volume 1955 of Lecture Notes in Mathematics*. Springer, Berlin, 2009; Models and theory

7. BOYLE, J.P., DYKSTRA, R.L.: A method for finding projections onto the intersection of convex sets in hilbert spaces. *Adv. Order Restricted Stat. Inference* **37**, 28–47, 1986

8. BELLA, P., GOLDMAN, M.:L Nucleation barriers at corners for a cubic-to-tetragonal phase transformation. *Proc. R. Soc. Edinb. Sect. A Math.* **145**(4), 715–724, 2015

9. BELLA, P., KOHN, R.V.: Wrinkles as the result of compressive stresses in an annular thin film. *Comm. Pure Appl. Math.* **67**(5), 693–747, 2014

10. BEDROSSIAN, J., KOHN, R.V.: Blister patterns and energy minimization in compressed thin films on compliant substrates. *Comm. Pure Appl. Math.* **68**(3), 472–510, 2015

11. BONNET, A.: On the regularity of edges in image segmentation. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **13**(4), 485–528, 1996

12. BONAFINI, M., ORLANDI, G., OUDET, E.: Variational Approximation of Functionals Defined on 1-Dimensional Connected Sets: The Planar Case. https://arxiv.org/abs/1610.03839, 2017

13. BUTTAZZO, G., PRATELLI, A., SOLIMINI, S., STEPANOV, E.: *Optimal Urban Networks via Mass Transportation, volume 1961 of Lecture Notes in Mathematics*. Springer, Berlin, 2009

14. BRANCOLINI, A., SOLIMINI, S.: On the Hölder regularity of the landscape function. *Interfaces Free Bound.* **13**(2), 191–222, 2011

15. BRANCOLINI, A., SOLIMINI, S.: Fractal regularity results on optimal irrigation patterns. *J. Math. Pures Appl. (9)* **102**(5), 854–890, 2014

16. BRANCOLINI, A., WIRTH, B.: Equivalent formulations for the branched transport and urban planning problems. *J. Math. Pures Appl.* **106**(4), 695–724, 2016

17. BRANCOLINI, A., WIRTH, B.: Optimal energy scaling for micropatterns in transport networks. *SIAM J. Math. Anal.* **49**(1), 311–359, 2017

18. CHAN, A., CONTI, S.: Energy scaling and branched microstructures in a model for shape-memory alloys with SO(2) invariance. *Math. Models Methods Appl. Sci.* **25**(6), 1091–1124, 2015

19. CHOKSI, R., CONTI, S., KOHN, R.V., OTTO, F.: Ground state energy scaling laws during the onset and destruction of the intermediate state in a type I superconductor. *Comm. Pure Appl. Math.* **61**(5), 595–626, 2008

20. CHAN, T.F., ESEDOḠLU, S., NIKOLOVA, M.: Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* **66**(5), 1632–1648, 2006

21. CHAMBOLLE, A.: Convex representation for lower semicontinuous envelopes of functionals in L^1. *J. Convex Anal.* **8**(1), 149–170, 2001

22. CHOKSI, R., KOHN, R.V., OTTO, F.: Domain branching in uniaxial ferromagnets: a scaling law for the minimum energy. *Comm. Math. Phys.* **201**(1), 61–79, 1999

23. CHOKSI, R., KOHN, R.V., OTTO, F.: Energy minimization and flux domain structure in the intermediate state of a type-I superconductor. *J. Nonlinear Sci.* **14**(2), 119–171, 2004

24. CHAMBOLLE, A., POCK, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.*, **40**(1), 120–145, 2010

25. CREMERS, D., STREKALOVSKIY, E.: Total cyclic variation and generalizations. *J. Math. Imaging Vis.* **47**(3), 258–277, 2013

26. DAL MASO, G., MOREL, J.-M., SOLIMINI, S.: A variational method in image segmentation: existence and approximation results. *Acta Math.* **168**(1–2), 89–151, 1992

27. EVANS, L.C., GARIEPY, R.F.: *Measure Theory and Fine Properties of Functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, 1992

28. GOLDMAN, M., ZWICKNAGL, B.: Scaling law and reduced models for epitaxially strained crystalline films. *SIAM J. Math. Anal.* **46**(1), 1–24, 2014

29. KNÜPFER, H., KOHN, R.V., OTTO, F.: Nucleation barriers for the cubic-to-tetragonal phase transformation. *Comm. Pure Appl. Math.* **66**(6), 867–904, 2013

30. KOHN, R.V., MÜLLER, S.: Branching of twins near an austenite-twinned-martensite interface. *Philos. Mag. A*, 1992

31. KOHN, R.V., MÜLLER, S.: Surface energy and microstructure in coherent phase transitions. *Comm. Pure Appl. Math.* **47**(4), 405–435, 1994

32. KOHN, R.V., WIRTH, B.: Optimal fine-scale structures in compliance minimization for a uniaxial load. *Proc. R. Soc. Lond. A: Math. Phys. Eng. Sci.* **470**(2170), 2014

33. KOHN, R.V., WIRTH, B.: Optimal fine-scale structures in compliance minimization for a shear load. *Comm. Pure Appl. Math.* 2015

34. MUMFORD, D., SHAH, J.: Optimal approximation by piecewise smooth functions and associated variational problems. *Commun. Pure Appl. Math.* **42**(5), 577–685, 1989

35. MOREL, J.-M, SANTAMBROGIO, F.: The regularity of optimal irrigation patterns. *Arch. Ration. Mech. Anal.* 195(2), 499–531, 2010

36. MADDALENA, F., SOLIMINI, S., MOREL, J.-M.: A variational model of irrigation patterns. *Interfaces Free Bound.* **5**(4), 391–415, 2003

37. OUDET, E., SANTAMBROGIO, F.: A Modica-Mortola approximation for branched transport and applications. *Arch. Ration. Mech. Anal.* **201**(1), 115–142, 2011

38. POCK, T., CHAMBOLLE, A., BISCHOF, H., CREMERS, D.: A convex relaxation approach for computing minimal partitions. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009

39. POCK, T., CREMERS, D., BISCHOF, H., CHAMBOLLE, A.: An algorithm for minimizing the mumford-shah functional. In: *2009 IEEE 12th International Conference on Computer Vision*, pp. 1133–1140, 2009

40. POCK, T., CREMERS, D., BISCHOF, H., CHAMBOLLE, A.: Global solutions of variational models with convex regularization. *SIAM J. Imaging Sci.* **3**(4), 1122–1145, 2010

41. SANTAMBROGIO, F.: Optimal channel networks, landscape function and branched transport. *Interfaces Free Bound.* **9**(1), 149–169, 2007

42. XIA, Q.: Optimal paths related to transport problems. *Commun. Contemp. Math.* **5**(2), 251–279, 2003

43. XIA, Q.: Interior regularity of optimal transport paths. *Calc. Var. Partial Differ. Equ.* **20**(3), 283–299, 2004

Alessio Brancolini, Carolin Rossmanith, and Benedikt Wirth
Institute for Numerical and Applied Mathematics,
University of Münster,
Einsteinstraße 62,48149 Münster,
Germany.
Alessio Brancolini
e-mail: alessio.brancolini@uni-muenster.de

Carolin Rossmanith
e-mail: carolin.rossmanith@uni-muenster.de

Benedikt Wirth
e-mail: benedikt.wirth@uni-muenster.de