C. R. Acad. Sci. Paris, Ser. I 352 (2014) 451–454

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Calculus of variations

A Modica–Mortola approximation for the Steiner Problem ☆

Une approximation à la Modica–Mortola pour le problème de Steiner

Antoine Lemenant [a], Filippo Santambrogio [b]

[a] *Université Paris-Diderot, Laboratoire Jacques-Louis-Lions, France*
[b] *Université Paris-Sud, Laboratoire de mathématiques d'Orsay, France*

ARTICLE INFO

Article history:
Received 13 November 2013
Accepted after revision 11 March 2014
Available online 2 April 2014

Presented by Yves Meyer

ABSTRACT

In this note we present a way to approximate the Steiner Problem by a family of elliptic energies of Modica–Mortola type, with an additional term relying on a weighted geodesic distance which takes care of the connectedness constraint.

RÉSUMÉ

Dans cette note, nous présentons une méthode d'approximation du problème de Steiner par une famille de fonctionnelles de type Modica–Mortola, avec un terme additionnel basé sur une distance géodésique à poids, pour prendre en compte la contrainte de connexité.

Version française abrégée

Le problème bien connu dit « de Steiner » consiste à trouver un compact connexe de longueur minimale qui contienne certains points du plan donnés au départ, en nombre fini. L'ensemble minimal est alors un arbre fini constitué de segments qui peuvent se joindre par nombre de 3 uniquement, formant des angles de 120° [6,12]. L'un des aspects qui a rendu ce problème si célèbre réside dans sa complexité de calcul, malgré une formulation simple en apparence, faisant partie de la liste des 21 problèmes NP-complets de Karp [7] (le temps polynomial étant évalué par rapport au nombre de points).

Dans cette note, nous proposons une méthode qui permet de trouver des solutions approchées du problème de Steiner. On se limite ici à présenter l'idée et les résultats mathématiques, en renvoyant à [4] pour les détails et les preuves. La stratégie repose sur l'emploi de fonctionnelles de type elliptique à la manière de Modica–Mortola [9], comme l'ont fait avant nous d'autres auteurs concernant des problèmes liés au périmètre ou à la longueur d'un fermé [2,10,13,11,1,8]. On remarque que, du fait de l'utilisation du terme de Modica–Mortola, qui approche une mesure $(n-1)$-dimensionnelle dans $\mathbb{R}^n$, souhaitant approcher une mesure de longueur, nous sommes forcés de nous restreindre à la dimension 2. Cela marque une différence par rapport à [11], où l'on pourrait espérer adapter la convergence au cas de la dimension supérieure ; les difficultés ne semblent pas ici seulement d'ordre technique.

☆ This work has been partially supported by the Agence Nationale de la Recherche, through the project ANR-12-BS01-0014-01 GEOMETRYA, and by The Gaspard Monge Program for Optimization and operations research (PGMO) via the project MACRO.
E-mail addresses: lemenant@ljll.univ-paris-diderot.fr (A. Lemenant), filippo.santambrogio@math.u-psud.fr (F. Santambrogio).

http://dx.doi.org/10.1016/j.crma.2014.03.008
1631-073X/

La nouveauté dans notre approche est l'ajout d'un terme permettant de gérer la contrainte de connexité sur l'ensemble à minimiser. Ce nouveau terme fait intervenir la fonction distance pondérée d_φ, définie en (2). Cette fonction peut être calculée numériquement sur une grille par une méthode, dite *fast-marching* [14], qui a été récemment améliorée dans [3], permettant le calcul à la fois de d_φ et de son gradient par rapport à φ. La fonctionnelle approximante que nous proposons est la suivante :

$$S_\varepsilon(\varphi) := \frac{1}{4\varepsilon}\int_\Omega (1-\varphi)^2\,\mathrm{d}x + \varepsilon\int_\Omega \|\nabla\varphi\|^2\,\mathrm{d}x + \frac{1}{c_\varepsilon}\sum_{i=1}^N d_\varphi(x_i, x_1),$$

où la constance c_ε satisfait $\lim_{\varepsilon\to 0} c_\varepsilon = 0$. Notre résultat principal stipule qu'étant donné une suite de minimiseurs (ou quasi-minimiseurs) φ_ε de S_ε et la suite de fonctions $d_{\varphi_\varepsilon}(\cdot, x_1)$ associée, ces fonctions convergent, à une sous-suite près, vers une fonction d, dont l'ensemble de niveau $\{d = 0\}$ est un minimiseur du problème de Steiner associés aux points $\{x_i\}$ (Théorème 2.1).

Les deux premiers termes de la fonctionnelle S_ε rappellent la fonctionnelle de Modica–Mortola. Le point essentiellement nouveau réside dans l'ajout du troisième terme, basé sur le fait suivant : si $\sum_{i=1}^N d_\varphi(x_i, x_1) = 0$, alors l'ensemble $\{d_\varphi = 0\}$ doit être connexe par arcs et contenir les $\{x_i\}$.

Dans l'article [4], écrit conjointement avec M. Bonnivard, nous utilisons cette technique pour approcher également certaines variantes du problème de Steiner, comme par exemple la fonctionnelle de distance moyenne. On peut y trouver des preuves détaillées, ainsi que des simulations numériques.

Les méthodes numériques envisagées, inspirées par le travail d'É. Oudet dans [10,11], se basent sur une méthode de gradient appliquée à chaque fonctionnelle S_ε (qui est convexe pour ε grand), en diminuant par étapes la valeur de ε et prenant comme initialisation à chaque étape le point de minimum approché trouvé à l'étape précédente. Cela ne garantit pas de converger vers un minimum global, mais permet en général de choisir un « bon » minimum local.

La preuve du résultat décrit plus haut est de type Γ-convergence. Plus précisément, différemment de ce qui a été fait dans [10,11] ainsi que dans les autres cas étudiés dans [4], il n'est pas possible de manière évidente d'exprimer notre résultat sous la forme d'un énoncé de Γ-convergence d'une suite de fonctionnelles vers une autre. Cependant, la démonstration en suit le même schéma. La Γ-limsup découle de techniques classiques que l'on peut trouver dans [2]. En revanche, la Γ-liminf est plus délicate. L'un des points difficiles à montrer est la rectifiabilité d'une limite Hausdorff d'ensembles de niveaux de fonctions d_{φ_ε} associées à des φ_ε d'énergies uniformément bornées. L'argument original de Modica–Mortola [9] est bien sûr essentiel, mais de nouvelles techniques nécessitent d'être introduites.

1. Introduction

Given a finite number of points $D := \{x_i\}_{i=1,\dots,N} \subset \Omega \subset \mathbb{R}^2$, the so-called Steiner Problem consists in solving

$$\min\bigl\{\mathcal{H}^1(K):\ K \subset \mathbb{R}^2 \text{ compact, connected, and containing } D\bigr\}. \tag{1}$$

Here, $\mathcal{H}^1(K)$ stands for the one-dimensional Hausdorff measure of K. It is known that minimizers for (1) do exist, need not to be unique, and are trees composed by a finite number of segments joining with only triple junctions at 120°, whereas computing a minimizer is very hard (some versions of the Steiner Problem belong to the original list of NP-complete problems by Karp [7]). We refer for instance to [6] for a history of the problem and to [12] for recent mathematical results on it.

In this note, we propose a way to approximate the problem, and we present a convergence result as some parameter ε goes to zero. Our strategy is to approximate the length by an elliptic energy of Modica–Mortola [9] type. This strategy was pursued before by many authors for similar problems involving the perimeter or the length of a closed set (see, e.g., [2,10,13,11,1,8]), but the novelty here is that we are able to add a term taking care of the connexity constraint. This term relies on the weighted geodesic distance d_φ, defined as follows. Given $\Omega \subset \mathbb{R}^2$, for any non-negative function $\varphi \in C^0(\overline{\Omega})$, we define the corresponding weighted geodesic distance through

$$d_\varphi(x, y) := \inf\left\{\int_\gamma \varphi(x)\,\mathrm{d}\mathcal{H}^1(x);\ \gamma \text{ curve in } \Omega \text{ connecting } x \text{ and } y\right\}. \tag{2}$$

Given a function φ and a point x_1, the distance $d_\varphi(\cdot, x_1)$ can be treated numerically by the so-called *fast-marching* method [14], since it is a solution of $\|\nabla u\| = \varphi$ with $u(x_1) = 0$ in the viscosity sense. A recent improvement of this algorithm (see [3]) is now able to compute at the same time d_φ and its gradient with respect to φ, which is useful every time one needs to optimize w.r.t. φ a functional involving d_φ. Our proposal to approximate the problem (1) is then to minimize

$$S_\varepsilon(\varphi) := \frac{1}{4\varepsilon}\int_\Omega (1-\varphi)^2\,\mathrm{d}x + \varepsilon\int_\Omega \|\nabla\varphi\|^2 + \frac{1}{c_\varepsilon}\sum_{i=1}^N d_\varphi(x_i, x_1),$$

among all functions $\varphi \in \mathcal{A} := H^1(\Omega) \cap C^0(\overline{\Omega}) \cap \{0 \leqslant \varphi \leqslant 1 \text{ and } \varphi = 1 \text{ on } \partial\Omega\}$. Here the constant c_ε is required to satisfy $\lim_{\varepsilon\to 0} c_\varepsilon = 0$.

The first two terms are a simple variant of the standard Modica–Mortola functional, already used in [2]: as $\varepsilon \to 0$, they force φ to tend to 1 a.e. and, if φ_ε stays small (close to 0) on a thin region (with measure tending to 0), they force to pay the transition between the value 1 and the value 0 by means of the length of the transition set. Notice that, in arbitrary dimension, these very terms would converge rather to an $(n-1)$-dimensional measure, which is the reason to stick to $\mathbb{R}^2$ if we want to approximate a length term.

In order to see the connection between the two first terms and the measure $\mathcal{H}^1$, a crucial point is a construction coming from [2, Theorem 3.1], which states that for any well-behaved subset $K \subset \mathbb{R}^2$ (in particular for any compact connected K with $\mathcal{H}^1(K) < +\infty$), there exists a sequence of functions ψ_ε such that $0 \leqslant \psi_\varepsilon \leqslant 1$, $\psi_\varepsilon = 1$ on $\partial\Omega$, $\psi_\varepsilon = 0$ on a neighborhood of K, $\|\nabla\psi_\varepsilon\| \leqslant C\varepsilon^{-1}$, and

$$\limsup_{\varepsilon} S_\varepsilon(\psi_\varepsilon) \leqslant \mathcal{H}^1(K) \tag{3}$$

(to be precise, our function ψ_ε is the same as [2] multiplied times $(1-\varepsilon)^{-1}$).

As far as the last term is concerned, it disappears at the limit, but tends to enforce connectedness. The key point is that whenever $\sum_{i=1}^N d_\varphi(x_i, x_1) = 0$, the set $\{d_\varphi = 0\}$ must be path-connected, must contain all the points $\{x_i\}$, and the path connecting them inside this set are such that $\varphi = 0$ a.e. on these curves, w.r.t. $\mathcal{H}^1$. Our result is of Γ-convergence type, even though the adequate framework to state it rigorously as a family of Γ-converging functionals is not clear.

In the paper [4], the authors together with M. Bonnivard used this idea to approximate some variant of the Steiner Problem, as the Average distance and p-Compliance problem. In the present paper we only give a detailed description of the idea and of the results, and we refer to [4] for full details of the proofs and some numerical experiments. The main idea for numerics is based on the work by É. Oudet in [10,11]: for every ε one can run a gradient descent for S_ε (which is convex for large ε), and a candidate minimizer for the limit problem is obtained by reducing at each step the value ε and initializing the gradient with the critical point obtained at the previous step. There is no guarantee that this converges to a global minimum, but at least a "well-chosen" local minimum is selected.

Existence of minimizers for S_ε. The existence of minimizers for the functional S_ε is a delicate matter. This depends on the fact that H^1 does not inject into C^0 and on the behavior of the map $\varphi \mapsto d_\varphi$. First, notice that we only restricted our attention to $\varphi \in C^0(\overline{\Omega})$ for the sake of simplicity. Indeed, it is possible to define d_φ as a continuous function as soon as $\varphi \in L^p$ for an exponent p larger than the dimension (here, $p > 2$, see [5]). The difficult question is which kind of convergence on φ provides pointwise convergence for d_φ. An easy result is the following: if $\varphi_n \to \varphi$ uniformly and a uniform lower bound $\varphi_n \geqslant c > 0$ holds, then $d_{\varphi_n}(x, x_1) \to d_\varphi(x, x_1)$. Counterexamples are known if the lower bound is omitted. On the contrary, replacing the uniform convergence with a weak H^1 convergence (which would be natural in the minimization of S_ε) is a delicate matter (by the way, the continuity seems to be true and it is not known whether the lower bound is necessary or not; this is the object of an ongoing work with T. Bousch).

For the sake of our paper, one could enforce existence of minimizers for fixed $\varepsilon > 0$ by adding an extra term of the form $\varepsilon^{p+1} \int \|\nabla\varphi\|^p$ with $p > 2$ (which enforces continuity and uniform convergence), and imposing a constraint $\phi \geqslant c_\varepsilon^2$ (the choice of c_ε^2 is made in order to preserve the vanishing property of the last term of the functional). Anyway, as we will see, a sequence of quasi-minimizers is sufficient instead of an exact minimizing sequence and, from the point of view of the approximation result and of the numerical applications, this is not crucial.

2. The main result

Let us first define a sequence of quasi-minimizers for S_ε as a sequence $(\varphi_\varepsilon)_\varepsilon \subset \mathcal{A}$ such that $\lim_{\varepsilon\to 0}(S_\varepsilon(\varphi_\varepsilon) - \inf_{\varphi\in\mathcal{A}} S_\varepsilon(\varphi)) = 0$.

Theorem 2.1. *Let Ω be a bounded open convex set containing the $\{x_i\}$. Let φ_ε be a sequence of quasi-minimizers for S_ε. Consider a sequence of functions $d_{\varphi_\varepsilon}(\cdot, x_1)$, which converge uniformly to a certain function d. Then the set $K := \{d = 0\}$ is compact, connected and is a solution to the Steiner Problem (1).*

Remark. Notice that assuming that the functions $d_{\varphi_\varepsilon}(\cdot, x_1)$ converge uniformly to some d is not restrictive since they are equi-Lipschitz (with constant 1).

Proof of Theorem 2.1. It is easy to see that the set $K = \{d = 0\}$, is a compact and connected set as a Hausdorff limit of sub level sets of $d_{\varphi_\varepsilon}(\cdot, x_1)$, which are all compact connected sets.

Let now K' be any competitor in the Steiner Problem, that we can assume contained in Ω. Let ψ_ε be the family of functions from [2, Theorem 3.1] satisfying (3). Notice that the last term $\frac{1}{c_\varepsilon}\sum_{i=1}^N d_{\psi_\varepsilon}(x_i, x_1)$ vanishes for every $\varepsilon > 0$, since $\psi_\varepsilon = 0$ on a neighborhood of K', i.e. on a connected open set containing all the points x_i.

On the other hand, it is clear from the quasi-minimizing property of φ_ε that

$$\liminf_{\varepsilon} S_\varepsilon(\varphi_\varepsilon) \leqslant \limsup_{\varepsilon} S_\varepsilon(\psi_\varepsilon). \tag{4}$$

First of all, this provides a uniform bound $S_\varepsilon(\varphi_\varepsilon) \leqslant C$, which implies that we have $\lim_{\varepsilon\to 0}\sum_{i=1}^N d_{\varphi_\varepsilon}(x_i, x_1) = 0$. This gives $d(x_i) = 0$ and thus $x_i \in K$ for every i. Hence, the proof will be finished provided that we show the following claim

$$\mathcal{H}^1(K) \leqslant \liminf_{\varepsilon} S_\varepsilon(\varphi_\varepsilon). \tag{5}$$

The full details of this fact can be found in [4, Lemma 3.1]. We shall describe here only the ideas of the proof, which is achieved within two main steps. The first one consists in finding a bound $\mathcal{H}^1(K) \leqslant C$ when $\liminf S_\varepsilon(\varphi_\varepsilon) < +\infty$ (which is the case here).

The main tool is the definition of the following geometric quantity: for each set $\Gamma \subset \mathbb{R}^2$, each unit vector $\nu \in \mathbb{S}^1$ and each $\lambda > 0$, we set

$$\Gamma_{\lambda,\nu} := \left\{x \in \mathbb{R}^2\text{: there exists } t \in [-\lambda, \lambda] \text{ with } x - t\nu \in \Gamma\right\}$$

and we define

$$I_\lambda(\Gamma) := \frac{1}{2\pi\lambda}\int_{\mathbb{S}^1} \mathscr{L}^2\big((\Gamma)_{\lambda,\nu}\big)\,\mathrm{d}\nu.$$

The following geometrical estimate [4, Lemma 2.6] is one of our key ingredients and is of independent interest: whenever Γ_ε are compact connected sets converging to Γ as $\varepsilon \to 0$ in the Hausdorff distance, then

$$\exists\lambda, \varepsilon_0 > 0; \quad I_\lambda(\Gamma_\varepsilon) \geqslant C\mathcal{H}^1(\Gamma_0), \quad \forall\varepsilon \leqslant \varepsilon_0, \tag{6}$$

where the constant C is universal.

Now, fix $\delta_0, \tau_0 > 0$, and let $\{z_1, z_2, \ldots, z_N\} \subseteq K$ be a τ_0-network in K, i.e. $K \subseteq \bigcup_{1\leqslant i\leqslant N} B(z_i, \tau_0)$. Due to the convergence $d_{\varphi_\varepsilon}(z_i, x_1) \to d(z_i) = 0$, for small ε we can build a set $\Gamma_\varepsilon = \bigcup_{1\leqslant i\leqslant N}\Gamma_i^\varepsilon$ where each Γ_i^ε is a C^1 curve connecting z_i to x_1 and satisfying $\int_{\Gamma_i^\varepsilon}\varphi_\varepsilon(s)\,\mathrm{d}\mathcal{H}^1(s) < \delta_0$.

We use the usual estimate $\frac{1}{4\varepsilon}(1-\varphi_\varepsilon)^2 + \varepsilon\|\nabla\varphi_\varepsilon\|^2 \geqslant \|\nabla(P(\varphi_\varepsilon))\|$, where $P(t) = t - t^2/2$ is a primitive of $(1-t)$, and compute the total variation of $P(\varphi_\varepsilon)$ in the direction ν on a set $(\Gamma_\varepsilon)_{\lambda,\nu}$. Using that $P(\varphi_\varepsilon)$ is almost 0 on Γ_ε (by definition of Γ_ε and using $P(t) \leqslant t$) and that, on the contrary, $P(\varphi_\varepsilon) \to P(1) = 1/2$ a.e., we get an estimate on $I_\lambda(\Gamma_\varepsilon)$. Thanks to (6), this turns into an estimate on the $\mathcal{H}^1$ measure of the Hausdorff limit of Γ_ε. By taking then the limit $\delta_0 \to 0$, and finally $\tau_0 \to 0$, one gets an estimate on $\mathcal{H}^1(K)$ and concludes the first step.

The second step is a refinement of the first one: once we have established the rectifiability of K, we can use the existence of tangent line $\mathcal{H}^1$-a.e. on K. Using a similar argument as the one above but adapted locally around each point of K (i.e. choosing the direction ν orthogonal to the tangent to K instead of taking an average over all directions), we are able to prove the better estimate (5) and this finishes the proof. □

Acknowledgements

This work has been partially supported by the Agence Nationale de la Recherche, through the project ANR-12-BS01-0014-01 GEOMETRYA, and by The Gaspard Monge Program for Optimization and operations research (PGMO) via the project MACRO.

References

[1] L. Ambrosio, A. Lemenant, G. Royer-Carfagni, A variational model for plastic slip and its regularization via gamma-convergence, J. Elasticity 110 (2) (2013) 201–235.
[2] L. Ambrosio, V.M. Tortorelli, On the approximation of free discontinuity problems, Boll. Un. Mat. Ital. B (7) 6 (1) (1992) 105–123.
[3] F. Benmansour, G. Carlier, G. Peyré, F. Santambrogio, Derivatives with respect to metrics and applications: subgradient marching algorithm, Numer. Math. 116 (3) (2010) 357–381.
[4] M. Bonnivard, A. Lemenant, F. Santambrogio, Approximation of length minimization problems among compact connected sets, preprint available on CVGMT.
[5] G. Carlier, C. Jimenez, F. Santambrogio, Optimal transportation with traffic congestion and wardrop equilibria, SIAM J. Control Optim. 47 (2008) 1330–1350.
[6] E.N. Gilbert, H.O. Pollak, Steiner minimal trees, SIAM J. Appl. Math. 16 (1968) 1–29.
[7] R. Karp, Reducibility among combinatorial problems, in: Complexity of Computer Computations, Plenum Press, 1972, pp. 85–103.
[8] G. Dal Maso, F. Iurlano, Fracture models as Γ-limits of damage models, Comm. Pure Appl. Anal. 12 (4) (2013) 1657–1686.
[9] L. Modica, S. Mortola, Il limite nella Γ-convergenza di una famiglia di funzionali ellittici, Boll. Un. Mat. Ital. A (5) 14 (3) (1977) 526–529.
[10] É. Oudet, Approximation of partitions of least perimeter by Γ-convergence: around Kelvin's conjecture, Exp. Math. 20 (3) (2011) 260–270.
[11] É. Oudet, F. Santambrogio, A Modica–Mortola approximation for branched transport and applications, Arch. Ration. Mech. Anal. 201 (1) (2011) 115–142.
[12] E. Paolini, E. Stepanov, Existence and regularity results for the Steiner problem, Calc. Var. Partial Differential Equations 46 (3) (2013) 837–860.
[13] F. Santambrogio, A Modica–Mortola approximation for branched transport, C. R. Acad. Sci. Paris, Ser. I 348 (15–16) (2010) 941–945.
[14] J.A. Sethian, Level Set Methods and Fast Marching Methods, Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 1999.