



HAL
open science

Extending the PowerWatershed framework thanks to Γ -convergence *

Laurent Najman

► **To cite this version:**

Laurent Najman. Extending the PowerWatershed framework thanks to Γ -convergence *. 2017. hal-01428875v1

HAL Id: hal-01428875

<https://hal.science/hal-01428875v1>

Preprint submitted on 6 Jan 2017 (v1), last revised 29 Jun 2017 (v2)

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

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

Extending the PowerWatershed framework thanks to Γ -convergence*

Laurent Najman[†]

Abstract. In this paper, we provide a formal proof of the power-watershed framework relying on the Γ -convergence framework. The main ingredient for the proof is a concept of scale. The proof and the formalism introduced in this paper have the added benefit to clarify the algorithm, and to allow to extend the applicability of the power watershed algorithm to many other types of energy functions. Several examples of applications are provided, including Total Variation and Spectral Clustering.

Key words. example, L^AT_EX

AMS subject classifications. ?????, 68Q25, 68R10, 68U05

1. Problem statement: the power watershed framework. We are interested in the following problem. Given three strictly positive integers $p > 0$, $m > 0$, $n > 0$ and n real numbers $1 \geq \lambda_0 > \lambda_1 > \dots > \lambda_{n-1} > 0$, we set

$$(1) \quad Q^p(x) = \sum_{0 \leq k < n} \lambda_k^p Q_k(x)$$

where, for all $0 \leq k < n$, $Q_k : \mathbb{R}^m \rightarrow \mathbb{R}$ is a continuous function. We search $x^* \in \mathbb{R}^m$ such that

$$(2) \quad x^* \in \lim_{p \rightarrow \infty} \arg \min_{x \in \mathbb{R}^m} Q^p(x)$$

We are going to see that λ_k acts as a notion of *scale* for the problem at hands. Note that we are not interested in the limit of $Q^p(x)$ itself as $p \rightarrow +\infty$. Indeed, if $\lambda_0 < 1$, $\sum_k \lambda_k^p Q_k(x)$ tends to 0 uniformly on \mathbb{R}^m when $p \rightarrow +\infty$, and any $x \in \mathbb{R}^m$ is a minimizer of 0. Instead, we are interested in the limit x^* of the minimizers of Q^p as $p \rightarrow +\infty$. The whole question is thus the choice of an informative minimizing sequence. The study of these types of questions is the main objective of the Γ -theory [16, 7], which has been adapted to the case of space of graphs in [10, 19]. However, as the sequence of (continuous) functional is decreasing (*i.e.*, $Q^{p+1} \leq Q^p$) and converge pointwise to a continuous function (implying Γ -convergence), our formalism is simpler and does not require familiarity with the Γ -convergence framework. Furthermore, the theorems from the Γ -convergence theory are generally written with a coercivity¹ hypothesis (in our case, that would be on the Q_k) not applicable in our framework. We shall replace the coercivity hypothesis by a compactness argument. For the sake of completeness, we shall

*Submitted to the editors DATE.

Funding: This work has been partly funded by ANR-15-CE40-0006 CoMeDiC and ANR-14-CE27-0001 GRAPH-SIP research grants.

[†]Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, Équipe A3SI, ESIEE Paris, France (laurent.najman@esiee.fr, <http://www.laurentnajman.org/>).

¹A continuous function $g : \mathbb{R}^m \rightarrow \mathbb{R}$ is *coercive* iff for every $\alpha > 0$, the set $\{x \in \mathbb{R}^m \mid g(x) \leq \alpha\}$ is compact. Equivalently, this means that $\lim_{\|x\| \rightarrow +\infty} g(x) = +\infty$.

30 expose in this paper what is needed to understand the proof, without explicit reference to
 31 Γ -convergence.

32 As in many cases, the (first) limit of Q^p provides a functional with a lot of minimizers.
 33 However, a further “limit of higher order”, with a different scaling, bring more information
 34 (for a formalization of this idea in the Γ -convergence framework, see [2]).

35 The idea is thus the following. In order to observe what happens during the convergence
 36 process, one has to apply a change of scale. Roughly speaking, this amounts to dividing the
 37 generator Q^p by λ_0 for the first scale, and this provides some information; dividing it further
 38 by λ_1 for the second scale will provide more information, and so on. By doing so, we build a
 39 sequence of functionals the minimizers of which being the solution we are looking for.

40 Section 2 provides some motivations for solving (2), the main one being the power-
 41 watershed [13]. In section 3, the main theorem of this paper is demonstrated, with its associate
 42 generic algorithm. In section 4, a specific algorithm dedicated to a particular class of functions
 43 is given and proved. Section 5 clarifies the links of the proposed framework with the (union
 44 of) maximum spanning trees. Two examples of applications are then developed. Section 6
 45 shows how total-variation is related to watershed-based mosaic images. Section 7 exhibits an
 46 application to spectral clustering. Finally, in section 8, we propose some ideas for extension
 47 and future work.

48 **2. Motivation.** Although it is by no means necessary from a theoretical point of view,
 49 it is convenient for many practical purposes to think of \mathbb{R}^m as a graph. We shall adopt the
 50 following notations in this paper, which will allows us to clarify the links between Equation (2)
 51 and the classical minimum spanning tree problem.

52 **2.1. Notations.** A (simple) *graph* G consists of a pair $G = (V, E)$ with vertices v in a
 53 finite set V of cardinality $|V| = m$ and edges $e \in E \subseteq V \times V$ with cardinality $|E| \leq m^2$.
 54 An *edge*, e , spanning two vertices, v_i and v_j , is denoted by e_{ij} . In 2D image processing
 55 applications, each pixel is typically associated with a vertex of the graph and the vertices are
 56 connected locally via a 4 or 8-connected lattice. An edge-weighted graph assigns a real value
 57 to each edge, called a weight. In this work, the weights are assumed to be non-negative and
 58 bounded by 1. The weight of an edge e_{ij} is denoted by $w(e_{ij})$ or w_{ij} . We also denote by w_i the
 59 (unary) weights penalizing the observed configuration at node v_i . In the context of filtering,
 60 segmentation and clustering applications, the weights encode nodal affinity such that nodes
 61 connected by an edge with high weight are considered to be strongly connected and edges
 62 with a low weight represent nearly disconnected nodes.

63 **2.2. Power-watershed with $q \geq 0$.** Let $q \geq 0$, we set

$$64 \quad (3) \quad W^p(x) = \sum_{e_{ij} \in E} w_{ij}^p |x_i - x_j|^q + \sum_{v_i \in V} w_i^p |x_i - f_i|^q$$

65 This problem was introduced for segmentation purposes in [13], with $q \geq 1$. In this case,
 66 Eq. (3) is a discrete formulation of the many possible variations on total variation denoising.
 67 In these formulations, w_{ij} are the *pairwise* weights, which can be interpreted as a weight
 68 on the gradient of the target configuration, such that the first term penalizes any unwanted
 69 high-frequency content in x and essentially forces x to vary smoothly within an object, while

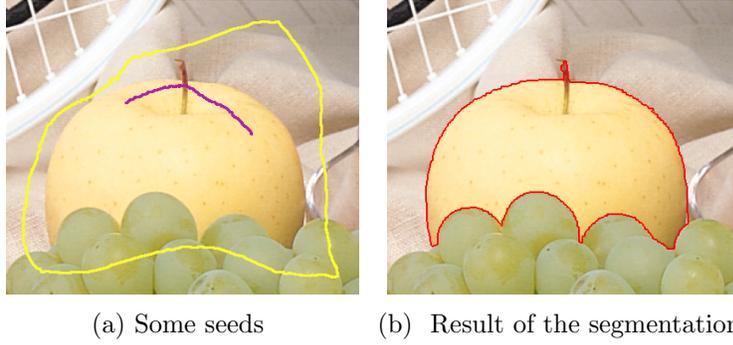


Figure 1: From [13], an example of segmentation in Figure 1(b) with the power-watershed watershed framework, by computing the Γ -limit of (3) when $q = 2$, using the seeds shown in Figure 1(a)

70 allowing large changes across the object boundaries. The second term enforces fidelity of x to
 71 a specified configuration f , w_i being the *unary* weights enforcing that fidelity. If $q > 1$, the
 72 function W^p in (3) is usually (depending on the w_{ij}) coercive, proper and strictly convex²,
 73 and a unique minimum $\arg \min_x W^p(x)$ exists for each $p > 0$. The existence and the unicity
 74 of the solution x^* to Eq. (2) thus depends on the convergence of these solutions. The earlier
 75 proof provided in [13] is, unfortunately, incomplete and difficult to understand. In [1], there
 76 is a proof corresponding to the case $q = 1$, where several solutions $\arg \min_x W^p(x)$ may exist.

77 We can rewrite Eq. (3) as follows:

$$78 \quad (4) \quad W^p(x) = \sum_{0 \leq k < n} \lambda_k^p \left\{ \sum_{e_{i,j} \in E_k} |x_i - x_j|^q + \sum_{v_i \in V_k} |x_i - f_i|^q \right\}$$

79 with $\lambda_0 > \lambda_1 > \dots > \lambda_{n-1}$, where $n \leq |V| + |E|$ is the number of different weights present in
 80 the graph G , be they pairwise or unary, E_k is the set of edges with weights equal to λ_k and
 81 V_k is the set of vertices with data-fidelity weights equal to λ_k .

82 For $0 \leq k < n$, we set

$$83 \quad (5) \quad W_k(x) = \sum_{e_{i,j} \in E_k} |x_i - x_j|^q + \sum_{v_i \in V_k} |x_i - f_i|^q$$

84 We have

$$85 \quad (6) \quad W^p(x) = \sum_{0 \leq k < n} \lambda_k^p W_k(x)$$

86 **Theorem 3.3** below ensures the convergence of the minimizers of W^p . In particular, when
 87 $q > 1$ and when the problem is strictly convex, we have unicity of the limit of the minimizers.

²See article by Combettes & Pesquet [11] for all the necessary hypotheses.

88 An example of application to seeded segmentation is shown in [Figure 1](#) (see [\[13\]](#) for more
89 details on this example.)

90 We can extend the power-watershed formulation, while keeping the same properties. Let
91 $q_1 > 0$, $q_2 > 0$, $a_{ij} \geq 0$ and $a_k \geq 0$. We set

$$92 \quad (7) \quad W^p(x) = \sum_{e_{ij} \in E} w_{ij}^p a_{ij} |x_i - x_j|^{q_1} + \sum_{v_i} w_i^p a_i |x_i - f_i|^{q_2}$$

93 With the notation of the present paper, we can rewrite this equation as

$$94 \quad (8) \quad W^p(x) = \sum_{0 \leq k < n} \lambda_k^p W_k(x)$$

95 with

$$96 \quad (9) \quad W_k = \sum_{e_{ij} \in E_k} a_{ij} |x_i - x_j|^{q_1} + \sum_{v_i \in V_k} a_i |x_i - f_i|^{q_2}$$

97 We remark that (loosely speaking) computing the limit of the minimizers amounts to solving
98 a weighted-graph variational problem on each one of the subgraph defined by (V_k, E_k) .

99 **2.3. Multi-scale regularization on weighted graphs.** Let $1 \geq \lambda_0 > \lambda_1 > \dots > \lambda_{n-1} > 0$.
100 For $0 \leq k < n$, let $q_k > 0$, and we set

$$101 \quad (10) \quad T_k(x) = \frac{1}{q_k} \sum_{v_i \in V_k} \left(\sum_{\{v_j | \{v_i, v_j\} \in E_k\}} a_{ij} |x_i - x_j|^2 \right)^{\frac{q_k}{2}} + \sum_{v_i \in V_k} \frac{a_i}{2} |x_i - f_i|^2$$

102 where (V_k, E_k) is a subgraph of G , $a_{ij} \geq 0$ and $a_k \geq 0$. Let us write

$$103 \quad (11) \quad T^p(x) = \sum_{0 \leq k \leq n} \lambda_k^p T_k(x)$$

104 We recognize in [Eq. \(10\)](#) a discrete Total-Variation-based regularization (more precisely, a dis-
105 crete weighted p -Dirichlet regularization) of the weighted subgraph (V_k, E_k) [\[21\]](#). [Theorem 3.3](#)
106 allows us to combine several graph regularizations into one unique formulation: minimizing
107 [Eq. \(11\)](#) can then be thought as a combination of several scales of Total-Variation regulariza-
108 tions.

109 **3. Scale-based approach for the Power Watershed framework.**

110 **3.1. A simple example.** For any $\varepsilon > 0$, let $Q_\varepsilon : \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined by

$$111 \quad (12) \quad Q_\varepsilon(x_0, x_1) = \varepsilon(x_0 - x_1)^2 + \varepsilon^2((x_0 - 1)^2 + x_1^2)$$

112 Note that $Q_\varepsilon \rightarrow 0$ when ε tends to 0. Also note that the functional Q_ε is quadratic positive
113 definite for any $\varepsilon > 0$. It is strongly convex with a single minimum (as we are going to verify
114 shortly).

115 **3.1.1. Direct approach.** By consideration of symmetry, we can reduce the problem to a
 116 single variable $\lambda \in \mathbb{R}$: we set $x_0 = \frac{1}{2} - \lambda$ and $x_1 = \frac{1}{2} + \lambda$. We then have

$$117 \quad (13) \quad Q_\varepsilon(\lambda) = 2\varepsilon^2\left(\lambda + \frac{1}{2}\right)^2 + 4\varepsilon\lambda^2 = \lambda^2(4\varepsilon + 2\varepsilon^2) + 2\lambda\varepsilon^2 + \frac{\varepsilon^2}{2}$$

118 A derivation with respect to λ leads to

$$119 \quad (14) \quad \frac{d}{d\lambda}Q_\varepsilon(\lambda) = 4\varepsilon(2 + \varepsilon)\lambda + 2\varepsilon^2$$

120 A second derivation yields:

$$121 \quad (15) \quad \frac{d^2}{d\lambda^2}Q_\varepsilon(\lambda) = 4\varepsilon(\varepsilon + 2),$$

122 which shows that the functional is strongly convex. As a minimum is reached for λ_ε such that
 123 $\frac{d}{d\lambda}Q_\varepsilon(\lambda_\varepsilon) = 0$, we get

$$124 \quad (16) \quad \lambda_\varepsilon = \frac{-\varepsilon}{2(2 + \varepsilon)}$$

125 which corresponds to the point $(x_0^\varepsilon, x_1^\varepsilon)$ such that

$$126 \quad (17) \quad x_0^\varepsilon = \frac{1}{2} - \frac{-\varepsilon}{2(2 + \varepsilon)} = \frac{1 + \varepsilon}{2 + \varepsilon}$$

$$127 \quad (18) \quad x_1^\varepsilon = \frac{1}{2} + \frac{-\varepsilon}{2(2 + \varepsilon)} = \frac{1}{2 + \varepsilon}$$

129 The distance of $(x_0^\varepsilon, x_1^\varepsilon)$ to $(\frac{1}{2}, \frac{1}{2})$ is equals to $\frac{\varepsilon}{\sqrt{2(2+\varepsilon)}}$, which proves the convergence of the
 130 sequence to $(\frac{1}{2}, \frac{1}{2})$ when ε tends to 0.

131 **3.1.2. Scale-based approach.** We first note that

$$132 \quad (19) \quad \frac{Q_\varepsilon(x)}{\varepsilon} = (x_0 - x_1)^2 + \varepsilon((x_0 - 1)^2 + x_1^2)$$

133 tends to $(x_0 - x_1)^2$ when ε tends to 0. Minimizing Q_ε , a first approximation at scale ε imposes
 134 $x_0 = x_1$. This corresponds to a restriction of the space on which Q_ε is defined. From a graph
 135 point of view, this corresponds to identifying v_0 to v_1 by (continuously) contracting the edge
 136 $e_{01} = \{v_1, v_0\}$. Minimizing the restriction of Q_ε to the space $\{v_0 = v_1\}$ leads to $x_0 = x_1 = \frac{1}{2}$.

137 The rest of the paper is dedicated to justifying that $x_0 = x_1 = \frac{1}{2}$ is indeed the limit of the
 138 minimizers of Q_ε , generalizing Q_ε to Q^p .

139 *Remark 3.1.* Using results from [19], Γ -theory allows to prove that

$$140 \quad (20) \quad (1, 0) = \lim_{\varepsilon \rightarrow \infty} \arg \min_x Q_\varepsilon(x).$$

3.2. Proof of the existence of the limit of the minimizers of Eq. (3).

Remark 3.2. The following theorem could also be proved with several applications of Claude Berge's maximum theorem [4] (well known in mathematical economics), which provides conditions for the continuity of an optimized function and the set of its maximizers as a parameter changes.

We shall prove the following.

Theorem 3.3. *Let $Q^p := \sum_{0 \leq k < n} \lambda_k^p Q_k$, where $(\lambda_k)_{0 \leq k < n} \in \mathbb{R}^n$ is such that $1 \geq \lambda_0 > \lambda_1 > \dots > \lambda_{n-1} > 0$, and $(Q_k)_{0 \leq k < n}$ are real-valued continuous functions defined on \mathbb{R}^m .*

Let M_0 be the set of minimizers of Q_0 , and for $0 < k < n$, M_k be recursively defined as follows:

$$(21) \quad M_0 = \arg \min_{x \in \mathbb{R}^m} Q_0(x)$$

$$(22) \quad \forall 1 \leq k < n, \quad M_k = \arg \min_{x \in M_{k-1}} Q_k(x)$$

Any convergent sequence $(x_p)_{p>0}$ of minimizers of Q^p converges to some point of M_{n-1} . In particular, if for all $p > 0$, $(x_p)_{p>0}$ is bounded (i.e. if there exists $C > 0$ such that for all $p > 0$, $\|x_p\|_\infty \leq C$), then, up to a subsequence, the sequence $(x_p)_p$ is convergent towards a point of M_{n-1} . Furthermore, we can then estimate the minimum of Q^p as follows:

$$(23) \quad \min_{x \in \mathbb{R}^m} Q^p(x) = \sum_{0 \leq k < n} \lambda_k^p m_k + o(\lambda_{n-1}^p)$$

where $m_k = \min_{x \in M_k} Q_k(x)$.

Note that there exist many applications where the minimizers of Q^p are bounded. This is in particular the case for elliptical problems, and for many problems where Total-Variation is used as a regularizer.

Thanks to [Theorem 3.3](#), we remark that if M_{n-1} is a singleton set, then any sequence of minimizers converges to the point in M_{n-1} . Thanks to [Theorem 3.3](#) and to the continuity of the application $p \rightarrow Q^p(\cdot)$, we can also remark that if for all $p > 0$, Q^p has only one minimizer, then there exists a unique $\bar{x} \in \mathbb{R}^m$ such that $\{\bar{x}\} \subseteq M_{n-1}$.

We are now ready to prove [Theorem 3.3](#).

3.2.1. Scale λ_0 . We write

$$(24) \quad \frac{Q^p(x)}{\lambda_0^p} = Q_0(x) + \sum_{k>0} \left(\frac{\lambda_k}{\lambda_0} \right)^p Q_k(x)$$

As $\lambda_0 > \lambda_k$ for any $k > 0$, the second term of Eq. (24) is negligible with respect to $Q_0(x)$ as soon as p is large enough. Let us write $M_0 := \arg \min_x Q_0(x) = \{y \mid y \in \arg \min_x Q_0(x)\}$. Let us set $m_0 = \min_x Q_0(x)$. By definition of M_0 , for any $x \in M_0$, we have $Q_0(x) = m_0$. If the minimizers of Q^p are bounded, i.e., if they live in a compact, then we can also bound the $Q_k(x)$ for any x in the same compact and for all k . We thus have

$$(25) \quad \min_x Q^p(x) = \lambda_0^p m_0 + o(\lambda_0^p)$$

176 **3.2.2. Scale λ_1 .** The process can be repeated with the next scale λ_1 . We set

$$177 \quad (26) \quad R_0^p(x) = Q^p(x) - \lambda_0^p Q_0(x)$$

178 and we note that $\frac{R_0^p(x)}{\lambda_1^p} = Q_1(x) + \sum_{k>1} (\frac{\lambda_k}{\lambda_1})^p Q_k(x)$. We set $M_1 := \arg \min_{x \in M_0} Q_1(x)$, and
179 we have

$$180 \quad (27) \quad \min_x Q^p(x) = \lambda_0^p m_0 + \lambda_1^p m_1 + o(\lambda_1^p)$$

181 where $m_1 = \min_{x \in M_0} Q_1(x)$.

182 **3.2.3. Scale λ_k .** More generally, the step corresponding to scale λ_k is

$$183 \quad (28) \quad R_k^p(x) = R_{k-1}^p(x) - \lambda_k^p Q_k(x)$$

184 and the whole process is repeated for all k .

185 Repeating the process for all scales λ_k , we obtain

$$186 \quad (29) \quad \min_x Q^p(x) = \sum_{0 \leq k < n} \lambda_k^p m_k + o(\lambda_{n-1}^p)$$

187 where $m_k = \min_{x \in M_k} Q_k(x)$.

188 Now, let us take a sequence $(x^p)_p$ of minimizers of Q^p converging to some $x^* \in \mathbb{R}^m$.
189 Thanks to the continuity of the application $(x, p) \mapsto Q^p(x)$ and to the continuity of Q_0 , we
190 obtain $x^* \in M_0$. By reiterating the argument, we get that $x^* \in M_1$, and so on, until we get
191 $x^* \in M_{n-1}$.

192 Hence the [Theorem 3.3](#).

193 **3.3. A generic algorithm.** The [algorithm 1](#) is a direct application of [Theorem 3.3](#). How-
194 ever, in its generic form, it is not easy to implement. In the next section, we are going to
195 particularize the function Q_k , so that we can provide a more specific implementation and so
196 that we can link this implementation with minimum spanning tree algorithms.

Data: A set of n continuous functions $(Q_k)_{0 \leq k < n}$ from \mathbb{R}^m to \mathbb{R} , together with their
scale $\lambda_0 > \lambda_1 > \dots > \lambda_{n-1}$.

Result: x solution to Eq. (2)

forall scales λ_k *by decreasing value do*

| Compute $M_k = \arg \min_{x \in M_{k-1}} Q_k$.

end

return some $x \in M_{n-1}$

Algorithm 1: Generic hierarchical optimization algorithm, optimizing Eq. (2)

197 **4. Algorithm for the Power Watershed.**

198 **4.1. Notations and preliminary concepts.** We are going to detail how the limit of the
 199 minimizers in Eq. (2) is obtained for a specific class of functions Q^p that extends the power-
 200 watershed formulation given by Eq. (9).

201 We suppose that the graph G is weighted by w , and that we are given a family $(\varphi_{ij})_{e_{ij} \in E}$
 202 of positive functions from \mathbb{R} to \mathbb{R}^+ such that $\varphi_{ij}(z) = 0$ if and only if $z = 0$. We set

$$203 \quad (30) \quad Q^p(x) = \sum_{e_{ij} \in E} w_{ij}^p \varphi_{ij}(x_i - x_j)$$

204 Let $V^f \subseteq V$, $V^f \neq \emptyset$. The set V^f is the set of boundary conditions for which the value $x_i = f_i$
 205 of a point $v_i \in V^f$ is known (fixed). If there exists (at least) $w_{ij} > 0$ for which $v_i \in V^f$, then
 206 one can show that the minimum of Q^p is bounded.

207 Let n be the number of different weights w_{ij} , and let $\lambda_0 > \lambda_1 > \dots > \lambda_{n-1}$ be those
 208 different weights. We set $E_k = \{e_{ij} \mid w_{ij} = \lambda_k\}$. We write

$$209 \quad (31) \quad Q_k(x) = \sum_{e_{ij} \in E_k} \varphi_{ij}(x_i - x_j).$$

210 Then Q^p can be written as

$$211 \quad (32) \quad Q^p(x) = \sum \lambda_k^p Q_k(x)$$

212 Recall that an *edge-induced subgraph* of a graph G is a subset of the edges of G together
 213 with any vertices that are their endpoints. A *path* in a graph G is a sequence of edges of the
 214 graph which connect a sequence of vertices. A *connected component* is a subgraph of G in
 215 which any two vertices are connected to each other by paths, and which is connected to no
 216 additional vertices in G . We denote by $CC(G)$ the set of connected components of a graph G .

217 In the rest of the paper, we suppose that the graph G is connected. For any $t \geq 0$, the
 218 *level set* $[w]_t$ is the graph induced by $\{e_{ij} \in E \mid w_{ij} \geq t\}$. We remark that any two connected
 219 components of $([w]_t)_{t>0}$ are either disjoint or nested. Thus, these connected components can
 220 be organized in a tree structure³. $[w]_t$ is a subgraph of G . If a connected component C of
 221 $CC([w]_t)$ is such that $V(C) \cap V^f \neq \emptyset$, we say that C is *seeded*. As $V^f \neq \emptyset$, at least one of the
 222 connected component of one the level sets is seeded.

223 **Lemma 4.1.** *Let $t > 0$ and $C \in CC([w]_t)$. We have $\sum_{e_{ij} \in E(C)} \varphi_{ij}(x_i - x_j) = 0$ if and only*
 224 *if $x_i = x_j$ for any $v_i, v_j \in V(C)$.*

225 *In particular, if C is not seeded, then $\arg \min_x \sum_{e_{ij} \in E(C)} \varphi_{ij}(x_i - x_j) = 0$, and any x that*
 226 *achieves the minimum is constant on C .*

227 **Proof.** As for any $z \in \mathbb{R}$, $\varphi_{ij}(z) \geq 0$ a the sum can only vanish if all terms $\varphi_{ij}(x_i - x_j)$
 228 vanish. Thus, if two vertices v_i and v_j are connected by an edge of C , then x_i needs to be equal
 229 to x_j . As $\varphi_{ij}(z) = 0$ if and only if $z = 0$, we see that x needs to be constant for all vertices
 230 which can be connected by a path in the graph. By definition of a connected component, all
 231 the vertices of C are connected by a path, hence x needs to be constant on C .

232 In particular, if C is not seeded, then any x that achieves the minimum is constant on
 233 C . ■

³This tree structure is called the *Max-tree* in the litterature. There exist fast algorithms for building this
 Max-tree [8].

234 **4.2. Algorithm for the power watershed..** To provide an algorithm computing Eq. (2)
 235 where Q^p is given by Eq. (30), we need to define the contract operation. We say that we
 236 *contract* two connected vertices v_1 and v_2 when we remove the edge $\{v_1, v_2\}$ linking v_1 to v_2
 237 while simultaneously merging the two vertices into a unique vertex. We say that we *contract* a
 238 connected subgraph C when we repeat the contraction until there is no edge in C , *i.e.* when C
 239 is reduced to a unique vertex. We remark that the order of the sequence of edge removal has
 240 no consequence on the result. We also remark that when we contract a connected subgraph
 241 C of a simple graph G , the resulting graph G' might no longer be a simple graph, but G' can
 242 be a *multi-graph*, *i.e.*, a graph in which an edge is repeated several times⁴.

243 We are now ready to state [algorithm 2](#). This algorithm is essentially the same as the one
 244 proposed in [13], although the presentation is simplified by the formalism we introduce in this
 245 paper. As stated previously, the various λ_k act as a notion of scale for the problem. Roughly
 246 speaking, at a given scale λ_k , this algorithm considers the level set of the graph corresponding
 247 to λ_k , and find the minimum of Q_k on this part of the graph. In the subsequent scales, the
 248 potentials fixed at previous scales are used as initial conditions.

Data: A weighted graph $G' = (V', E', w)$ and the functions Q_k
Result: x solution to Eq. (2)
forall scales λ_k by decreasing value **do**
 forall connected components C of $[w]_{\lambda_k}$ **do**
 if C is seeded **then**
 Fix the unknown potential x_i of the vertices $v_i \in V(C)$ by minimizing Q_k
 on C .
 end
 else Contract C . ;
 end
end

Algorithm 2: Power watershed algorithm, optimizing Eq. (2)

249 We have the following theorem.

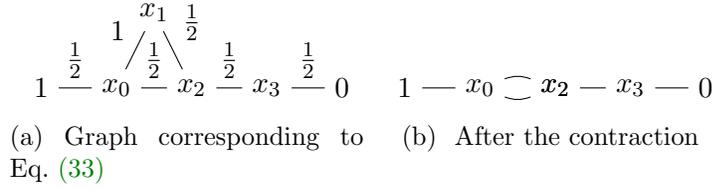
250 **Theorem 4.2.** *Let Q^p given by Eq. (30), and such that, for any $p \geq 0$, $\arg \min_x Q^p(x)$ is*
 251 *bounded. Then [algorithm 2](#) computes the solution to Eq. (2).*

252 Remark that Eq. (30) includes $\varphi_{ij}(z) = a_{ij}z^{q_{ij}}$, with $a_{ij} > 0$ and $q_{ij} > 0$, which itself
 253 generalizes the power-watershed equation (3). When the solutions to $\arg \min_x Q^p(x)$ can be
 254 bounded (which is the case for the power-watershed), then, by application of [Theorem 3.3](#),
 255 there exists at least one minimizer $\bar{x} \in M_{n-1}$ such that $\bar{x} \in \lim_p \arg \min_x Q^p(x)$.

256 The rest of the section is dedicated to proving [Th. 4.2](#). We will proceed by succesively
 257 considering the various scales λ_k .

258 **4.2.1. Scale λ_0 .** As the restriction of Q_0 to any connected component of $[w]_{\lambda_0}$ is positive,
 259 we can independently optimise Q_0 on each element $C \in CC([w]_{\lambda_0})$.

⁴As in the contraction of a connected set C we remove any edge connecting two (not necessarily different) vertices of the set, there is no loop in the resulting contracted set.

Figure 2: Applying [algorithm 2](#) on Eq. (33)

260 If C is seeded, the data attachment term provides a boundary condition for finding the
 261 minimum, and a potential x_i for any $v_i \in V(C)$ is fixed at this scale.

262 Otherwise, if C is not seeded (*i.e.*, there is no data attachment for C and only the regu-
 263 larisation term is present), we obtain $x_i = x_j$ for any $v_i, v_j \in V(C)$ thanks to [Lemma 4.1](#). In
 264 other word, we contract C .

265 For the sake of simplicity, in the sequel, we still denote by G the resulting contracted
 266 multi-graph.

267 **4.2.2. Scale λ_1 .** At scale λ_1 , we want to find a minimizer of Q_1 restricted to the minimiz-
 268 ers of Q_0 . In other words, we want to obtain the potential x_i of any v_i in $[w]_{\lambda_1}$, constrained
 269 by the information on any $v_j \in V([w]_{\lambda_0})$ obtained at the previous scale. The x_i of any
 270 $v_i \in V([w]_{\lambda_0})$ plays the role of a data attachment for the vertices of $[w]_{\lambda_1} \setminus [w]_{\lambda_0}$.

271 As in the previous step, the restriction of Q_1 to any connected component of $[w]_{\lambda_1}$ being
 272 positive, we can independently optimise Q_1 on each element $C \in CC([w]_{\lambda_1})$. If C is seeded,
 273 then the data attachment allows us to fix the potential x_i of any $v_i \in [w]_{\lambda_1}$ not already fixed
 274 at the previous scale. Otherwise, if C is not seeded, (*i.e.*, there is no data attachment for C),
 275 [Lemma 4.1](#) ensures that $x_i = x_j$ for any $v_i, v_j \in V(C)$. We thus contract C onto a unique
 276 vertex. As before, we continue to denote by G the resulting contracted multi-graph.

277 **4.2.3. Other scales λ_k , $1 < k < n$.** The process is straightforwardly repeated for all the
 278 other scales λ_k , $1 < k < n$. At each step of the scaling process, the potential x_i of some
 279 vertices v_i becomes known. At the end, as the graph G is connected and as at least one of
 280 component of the level set is seeded, a complete potential $\bar{x} \in R^m$ has been obtained.

281 This concludes the proof.

282 **4.3. An illustration of the application of [Theorem 4.2](#).** Let us consider the following
 283 equation.

$$284 \quad (33) \quad Q^p(x) = (x_0 - x_1)^2 + \left(\frac{1}{2}\right)^p((x_1 - x_2)^2 + (x_0 - x_2)^2 + (x_2 - x_3)^2 + (x_0 - 1)^2 + x_3^2)$$

285 The graph corresponding to Eq. (33) is depicted in [Figure 2\(a\)](#). After the contraction,
 286 we obtain the unweighted multi-graph depicted in [Figure 2\(b\)](#), which corresponds to the
 287 polynomial

$$288 \quad (34) \quad 2(x_0 - x_2)^2 + (x_2 - x_3)^2 + (x_0 - 1)^2 + x_3^2.$$

289 Hence, the solution of the minimzation leads to $x_0 = x_1 = \frac{5}{7}$, $x_3 = \frac{2}{7}$ and $x_2 = \frac{4}{7}$.

290 **5. Power-watershed and the union of all maximum spanning trees.** Algorithm 2 has
 291 the structure of Kruskal's algorithm for the maximum spanning tree [24], and thus this invites
 292 us to look precisely at the relation between Eq. (2) and the maximum spanning tree problem.

293 A *tree* T is a connected graph such $|E(T)| = |V(T)| - 1$. A *spanning tree* T of a graph G
 294 is a tree such that $V(T) = V(G)$. The weight of a weighted graph (G, w) is the number

$$295 \quad (35) \quad W(G) := \sum_{e_{ij} \in E(G)} w_{ij}$$

296 A maximum spanning tree of a graph G is a spanning tree T such that the weight of T is
 297 greater or equal to the weight of any other spanning tree of G .

298 The problem of finding a maximum spanning tree is the oldest problem in combinatorial
 299 optimization [29].

300 In general, there exist several maximum spanning trees of a given weighted graph. We
 301 denote by $MST(G)$ the union of all the maximum spanning trees of the graph G . We remark
 302 that $MST(G)$ is a subgraph of G such that $|E(MST(G))| \leq |E(G)|$.

303 The following lemma is easily deduced from the proof of Kruskal's algorithm for maximum
 304 spanning tree, and thus its proof is left for the reader. To ease the writing of the lemma, we
 305 note by convention $[w]_{\lambda_{-1}} = \emptyset$.

306 **Lemma 5.1.** *Let $e_{ij} = \{x_i, x_j\}$ be an edge of $[w]_{\lambda_k}$ such that $w_{ij} = \lambda_k$. Then one of the*
 307 *two situations holds:*

308• *either e_{ij} links two different components of $[w]_{\lambda_{k-1}}$. In this case, e_{ij} belongs to a maximum*
 309 *spanning tree of G , i.e., $e_{ij} \in E(MST(G))$.*

310• *or x_i and x_j belong to the same component of $[w]_{\lambda_{k-1}}$. In that case $e_{ij} \notin E(MST(G))$.*

311 **Theorem 5.2.** *Solving Eq. (2) on G is equivalent to solving it on $MST(G)$.*

312 *Proof.* It is a straightforward consequence of Theorem 4.2 and of Lemma 5.1. Indeed, let
 313 $e_{ij} \notin MST(G)$, with $w_{ij} = \lambda_k$. According to Lemma 5.1, e_{ij} links two vertices of a component
 314 C of $[w]_{\lambda_{k-1}}$. Hence, by Theorem 4.2 (more precisely, by application of algorithm 2), either
 315 all the potential of the vertices of C have been fixed at a previous scale, or C has been
 316 contracted. ■

317 We denote by $R(G)$ the graph obtained by contracting any non-seeded component of $[w]_t$
 318 for any $t > 0$.

319 **Theorem 5.3.** *Solving Eq. (2) on G is equivalent to solving it on $MST(R(G))$.*

320 *Proof.* This theorem is also a straightforward consequence of Theorem 4.2, Theorem 5.2,
 321 and Lemma 5.1. ■

322 **6. An application to Total Variation: the watershed-based mosaic image.** Let $p \geq 0$,
 323 and $q \in \{0, 1, 2\}$. We set

$$324 \quad (36) \quad TV_q^p(x) = \sum_{e_{ij} \in E} w_{ij}^p |x_i - x_j| + \sum_{v_i \in V} w_i^p |x_i - f_i|^q$$

325 In this section, we study the dependence of the solution x^* of Eq. (2) with respect to data
 326 f when $Q^p = TV_q^p$ is given by Eq. (36). This corresponds to a discrete version of the Total

327 Variation problem [30, 18]. More precisely, depending on the value of q , it is a discrete Total
 328 Variation model with a data attachment term in L_0 (with the convention that $0^0 = 0$), L_1
 329 (TV-L1) or L_2 (the ROF model) norm.

330 Let T_q be the operator that maps f to a solution x^* of Eq. (2).

331 **6.1. Inheriting some properties of TV-L1.** The following properties are inherited from
 332 the properties of the classical TV-L1 model, see for example [20]. Formal proofs are omitted
 333 for brevity.

334 **Property 6.1.** *The operator T_1 is idempotent, i.e.. $T_1(T_1(f)) = T_1(f)$.*

335 **Property 6.2.** *The operator T_1 commutes with the addition of constants (i.e., $T_1(f + C) =$
 336 $T_1(f) + C$) and is self-dual (i.e., $T_1(1 - f) = 1 - T_1(f)$).*

337 **Property 6.3.** *The operator T_1 is a contrast-invariant operator, i.e., $T_1(g \circ f) = g(T_1(f))$,
 338 where g is an bounded increasing C^1 diffeomorphism.*

339 **Property 6.4.** *The operator T_1 satisfies the maximum principle ($m \leq f \leq M \implies m \leq$
 340 $T_1(f) \leq M$).*

341 **6.2. Total Variation and the flat-zone hierarchy.** In this section, we fix $f \in [0, 1]^m$, and
 342 we set $w_{ij} = \exp(-|f_i - f_j|)$. For the sake of simplicity, we set $w_i = \exp(-\mu)$ for all i , with
 343 $\mu \in \mathbb{R}$ (so that μ varies in the same range as the gradient). For any $0 \leq \mu \leq 1$, we note x_q^μ a
 344 solution to Eq. (2), i.e., $x_q^\mu = T_q(f)$.

345 We remark that for $\mu = 0$ (strong data attachment), $x_q^\mu = f$ and that for $\mu \geq 1$ (no
 346 data attachment), x_q^μ is constant. For $0 \leq \mu \leq 1$, let us denote by $CC_q^\mu(v_i)$ the connected
 347 component that contains v_i such that x_q^μ is constant on the component. We call $CC_q^\mu(v_i)$ a
 348 flat zone of x_q^μ .

349 It is easy to see, thanks to a simple computation that, for $e_{ij} \in E$, as soon as $\mu > |f_i - f_j|$,
 350 then $x_i = x_j$. Conversely, if $\mu < |f_i - f_j|$, then $x_i = f_i$. Thus the class CC_q^μ of a point v_i is
 351 given by

$$\begin{aligned}
 352 \quad CC_q^\mu(v_i) &= \{v_i\} \cup \left\{ v_j \mid \text{there exists a path } \langle v_i = v^1, \dots, v^n = v_j \rangle, \right. \\
 353 \quad &\quad \left. \text{such that } |f_k - f_{k+1}| < \mu \right. \\
 354 \quad (37) \quad &\quad \left. \forall 1 \leq k < n \right\} \\
 355
 \end{aligned}$$

356 It is the quasi-flat zones hierarchy, introduced by Nagao *et al.* in 1979 [26], recently revived
 357 in [32] and shown to be equivalent to an edge-based topological watershed of the gradient in
 358 [27].

359 Let H_q be the the family of all flat zones of the solutions x_q^μ for all μ . The family H_q is
 360 hierarchical in the sense that:

- 361 • For any $q \in \{0, 1, 2\}$, x_q^1 is totally flat. Indeed, we have x_0^1 is the mode of the vector
 362 f , x_1^1 is the median of f and $x_2^1 = \frac{1}{N} \sum_{v_i} f_i$ is the mean of f . Hence the whole space
 363 is an element of H ,
- 364 • x_q^0 is exactly f , hence all the points are elements of H ,
- 365 • two elements of H are either disjoint or nested.

366 Thus, we have the following remarkable result. A classical morphological idea [5, 6] consists
 367 in computing a watershed of the gradient and building a *mosaic image* by setting the value of
 368 the mosaic image to be the mode for $q = 0$ (resp. the median for $q = 1$, the mean for $q = 2$) of
 369 the original image on each one of the regions of the watershed segmentation. Such a mosaic
 370 image is the optimal solution of Eq. (2).

371 As an example of application, let us consider the graph depicted in Figure 3(a), that
 372 corresponds to the following equation

373 (38)
$$TVL_1^p(x) = e^{-9p}|x_0 - x_1| + e^{-4p}|x_0 - x_2| + e^{-3p}|x_1 - x_3| + e^{-2p}|x_2 - x_3|$$

 374
$$+ e^{-\mu} [|x_0 - 1| + |x_1 - 10| + |x_2 - 5| + |x_3 - 7|].$$

376 The hierarchy H_1 corresponding to all flat zones of the solution x_1^μ to Eq. (2) for all μ is
 377 depicted in Figure 3(b). Some x_1^μ are given in Figures 3(c) to 3(f) depending on different
 378 values for μ .

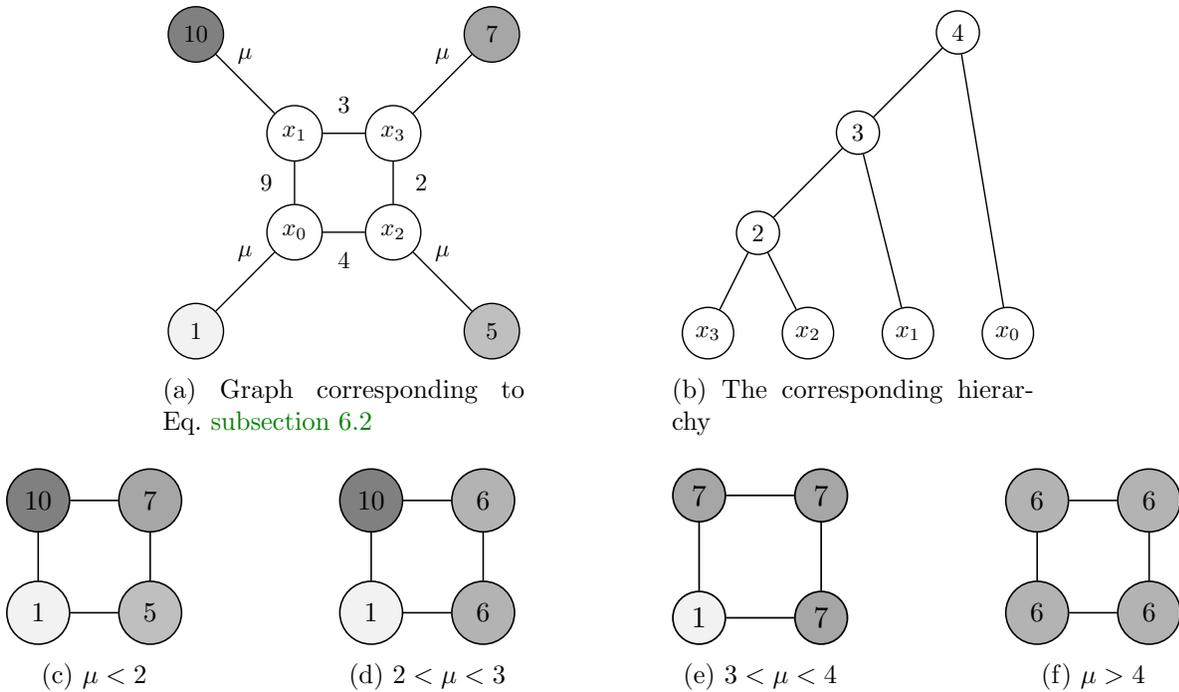


Figure 3: An example of application of the Γ -limit of the TV-L1 scheme corresponding to Eq. subsection 6.2.

379 For fast algorithms computing the complete hierarchy of mosaic images, we refer to the
 380 ultrametric/saliency framework [27, 15, 28, 14]. Such algorithms allows one to compute the
 381 whole family of x_q^μ for $\mu \in [0, 1]^m$ in linear or quasi-linear time.

382 **7. An application to spectral clustering: the power-ratio cut.** Spectral clustering [35]
 383 is a successful approach to clustering and many different variations do exist. One of those

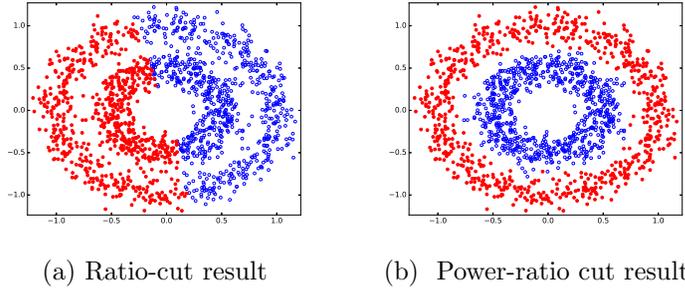


Figure 4: From [9], an example of comparison between ratio-cut and power-ratio cut spectral clusterings on two nested noisy circular sets of 2D points.

384 variations is called *ratio-cut* [36], and can be described thanks to the following optimization
 385 problem. For finding k clusters, compute the solution to

$$\begin{aligned}
 & \text{minimize } Tr(H^t L H) \\
 & H \in \mathbb{R}^{m \times k} \\
 & \text{subject to } H^t H = \mathbb{I}
 \end{aligned}
 \tag{39}$$

387 where L is the graph-laplacian⁵, Tr is the trace operator and \mathbb{I} is the identity matrix.

388 Without restriction in generality, let us suppose that the graph has distinct weights $w_1 <$
 389 $w_2 < \dots < w_j$ with $j \leq m$. Let us define L_k as the graph-laplacian of the subgraph induced
 390 by the edges whose weight is exactly w_k . As described in [9], it can be easily seen that (39)
 391 is equivalent to

$$\begin{aligned}
 & \text{minimize } \sum_{k=1}^j w_k Tr(H^t L_k H) \\
 & H \in \mathbb{R}^{m \times k} \\
 & \text{subject to } H^t H = \mathbb{I}
 \end{aligned}
 \tag{41}$$

393 By raising the w_k at the power p , we can see that (41) fits into the power-watershed
 394 framework. Thus, we can define the *power-ratio cut* as the solution to the γ -limit of (41)
 395 when raising w_k at the power p and letting p tends to infinity. Algorithm 2 is not applicable
 396 to this situation, but we can apply some variation of algorithm 1. A complete theory, together
 397 with an efficient algorithm for solving this problem, as well as illustrations (such as the one
 398 of Figure 4) and experiments, can be found in [9].

⁵The elements of L are given by

$$L_{i,j} := \begin{cases} \deg(v_i) & \text{if } i = j \\ -w_{ij} & \text{if } i \neq j \text{ and } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise} \end{cases}
 \tag{40}$$

where $\deg(v_i) = \sum_j w_{ij}$ is the *degree* of the vertex

399 **8. Conclusion: some ideas for extensions and applications.** The framework developed
 400 in this paper casts a new light on the links and differences between the mathematical morphol-
 401 ogy theory for segmentation and filtering, and the more classical one based on optimization.
 402 We would like to highlight here some possible research directions for the future.

403 Some of those ideas are natural extension of the current paper.

- 404 • Exploring the extension of the classical power-watershed, as proposed in Eq. (7), is a
 405 must.
- 406 • We showed in section 6 how to express the morphological mosaic images as a Γ -limit
 407 of some Total-Variation operators. The next steps would be to look at the theory
 408 of *scale-set image analysis* [22, 23] and find if it would be possible to obtain optimal
 409 hierarchies composed of optimal partitions. This possibly would allows us to think the
 410 watershed operator as a projector, and would probably allows to strenghten the ideas
 411 developed in [12] regarding anisotropic diffusion.
- 412 • Regarding spectral classification, we hint in section 7 at what can be done with the
 413 ratio-cut criterion [9]. Other criterions, such as normalized cuts [31], should of course
 414 be studied, but this could be more difficult.
- 415 • Application of the proposed framework for morphological filtering, in order to clarify
 416 links and differences between for example MST-based filtering [33, 3] and distance-
 417 based filtering (such as the amoeba framework [25]) is also of interest. A first step in
 418 this direction can be found in [17].

419 We list below some other open problems.

420 **8.1. Is x^* useful for estimating x^p ?** Such an estimation would be most useful. For
 421 example, if we know that x^p is close to x^* , then x^* can be used to initialize a convergence
 422 process to x^p .

423 Under certain conditions, we can obtain such an estimation. For example, with strong
 424 convexity: for any strongly convex function γ , there exists $\nu > 0$ such that

$$425 \quad (42) \quad \|x - x_0\|_2^2 \leq \frac{2}{\nu}(\gamma(x) - \gamma(x_0)).$$

426 where x_0 is the unique minimum of γ . Hence we can approximate x^p with x^* . However, ν
 427 would also tend to 0 when $p \rightarrow \infty$, so this has to be dig in.

428 **8.2. Link between the continuous world and the discrete one.** When $\varphi_{ij}(x) = \varphi_k(x) =$
 429 $|x|$, Eq. (3) is equivalent to a min-cut/max-flow problem. This problem has a continuous
 430 formulation, that was proposed by G. Strang [34]. But the maximum spanning tree prob-
 431 lem, being a tree problem, is purely discrete, and we do not see any obvious formulation in
 432 continuous terms.

433 Can the link between max-flow/min-cut and maximum spanning tree described in the
 434 present paper be used for proposing a continuous notion of maximum spanning tree?

435 **8.3. Extension to quantum mechanics.** Everything here should extend to Schrödinger
 436 operators. This might open the door for novel applications in physics.

437 **Acknowledgments.** The author would like to thank A. Challa, S. Danda, F. Malgouyre,
 438 J.-C. Pesquet and H. Talbot for their insightful comments of previous versions of this paper.

439 Dedicated to the memory of Monique Benayoun (14 novembre 1944 - 24 janvier 2014), in
 440 whose apartment I wrote the very first draft of this paper.

441

REFERENCES

- 442 [1] C. ALLÈNE, J.-Y. AUDIBERT, M. COUPRIE, AND R. KERIVEN, *Some links between extremum spanning*
 443 *forests, watersheds and min-cuts*, Image and Vision Computing, 28 (2010), pp. 1460–1471, <https://doi.org/10.1016/j.imavis.2009.06.017>, <http://hal-upec-upem.archives-ouvertes.fr/hal-00622507>.
 444
 445 [2] G. ANZELLOTTI AND S. BALDO, *Asymptotic development by γ -convergence*, Applied mathematics and
 446 optimization, 27 (1993), pp. 105–123.
 447 [3] L. BAO, Y. SONG, Q. YANG, H. YUAN, AND G. WANG, *Tree filtering: Efficient structure-preserving*
 448 *smoothing with a minimum spanning tree*, Image Processing, IEEE Transactions on, 23 (2014),
 449 pp. 555–569.
 450 [4] C. BERGE, *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and con-*
 451 *convexity*, Courier Dover Publications, 1963.
 452 [5] S. BEUCHER, *Watershed, hierarchical segmentation and waterfall algorithm*, in Mathematical morphology
 453 and its applications to image processing, Springer, 1994, pp. 69–76.
 454 [6] S. BEUCHER AND F. MEYER, *The morphological approach to segmentation: the watershed transforma-*
 455 *tion*, OPTICAL ENGINEERING-NEW YORK-MARCEL DEKKER INCORPORATED-, 34 (1992),
 456 pp. 433–433.
 457 [7] A. BRAIDES, *A handbook of γ -convergence*, Handbook of Differential Equations: stationary partial differ-
 458 ential equations, 3 (2006), pp. 101–213.
 459 [8] E. CARLINET AND T. GÉRAUD, *A comparative review of component tree computation algorithms*, IEEE
 460 Transactions on Image Processing, 23 (2014), pp. 3885–3895.
 461 [9] A. CHALLA, S. DANDA, B. DAYA SAGAR, AND L. NAJMAN, *Gamma limit for spectral clustering*, in
 462 submitted, 2017. <https://hal.archives-ouvertes.fr/view/index/docid/1427957>.
 463 [10] Y. COLIN DE VERDIÈRE, Y. PAN, AND B. YCART, *Singular limits of schrödinger operators and markov*
 464 *processes*, J. Operator Theory, 41 (1999), pp. 151–173.
 465 [11] P. L. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing*, in Fixed-Point
 466 Algorithms for Inverse Problems in Science and Engineering, H. H. Bauschke, R. Burachik, P. L.
 467 Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, eds., Springer-Verlag, New York, 2010.
 468 [12] C. COUPRIE, L. GRADY, L. NAJMAN, AND H. TALBOT, *Anisotropic Diffusion Using Power Watersheds*,
 469 in International Conference on Image Processing (ICIP'10), Sept. 2010, pp. 4153–4156.
 470 [13] C. COUPRIE, L. GRADY, L. NAJMAN, AND H. TALBOT, *Power watershed: A unifying graph-based op-*
 471 *timization framework*, IEEE Trans. Pattern Anal. Mach. Intell., 33 (2011), pp. 1384–1399, <https://doi.org/10.1109/TPAMI.2010.200>, <http://dx.doi.org/10.1109/TPAMI.2010.200>.
 472
 473 [14] J. COUSTY, G. BERTRAND, L. NAJMAN, AND M. COUPRIE, *Watershed cuts: Minimum spanning forests*
 474 *and the drop of water principle*, IEEE Trans. Pattern Anal. Mach. Intell., 31 (2009), pp. 1362–1374.
 475 [15] J. COUSTY AND L. NAJMAN, *Incremental algorithm for hierarchical minimum spanning forests*
 476 *and saliency of watershed cuts*, in 10th International Symposium on Mathematical Morphology
 477 (ISMM'11), vol. 6671 of Lecture Notes in Computer Science, 2011, pp. 272–283, [http://www.esiee.](http://www.esiee.fr/~info/sm/)
 478 [fr/~info/sm/](http://www.esiee.fr/~info/sm/).
 479 [16] G. DAL MASO, *An Introduction to Γ -convergence*, vol. 8, Springer, 1993.
 480 [17] S. DANDA, A. CHALLA, B. DAYA SAGAR, AND L. NAJMAN, *Power tree filter: A theoretical framework*
 481 *explaining and extending the minimum spanning tree filter*, in submitted, 2017.
 482 [18] J. DARBON AND M. SIGELLE, *Image restoration with discrete constrained total variation part i: Fast and*
 483 *exact optimization*, Journal of Mathematical Imaging and Vision, 26 (2006), pp. 261–276.
 484 [19] Y. C. DE VERDIÈRE, *Spectres de graphes*, vol. 4 of Cours Spécialisés, Société Mathématique de France,
 485 1998.
 486 [20] V. DUVAL, J.-F. AUJOL, AND Y. GOUSSEAU, *The tv1 model: a geometric point of view*, Multiscale
 487 Modeling & Simulation, 8 (2009), pp. 154–189.
 488 [21] A. ELMOATAZ, O. LEZORAY, AND S. BOUGLEUX, *Nonlocal discrete regularization on weighted graphs: a*
 489 *framework for image and manifold processing*, Image Processing, IEEE Transactions on, 17 (2008),

- 490 pp. 1047–1060.
- 491 [22] L. GUIGUES, J. P. COCQUEREZ, AND H. L. MEN, *Scale-sets image analysis*, International Journal of
492 Computer Vision, 68 (2006), pp. 289–317, <http://dx.doi.org/10.1007/s11263-005-6299-0>.
- 493 [23] B. R. KIRAN AND J. SERRA, *Global–local optimizations by hierarchical cuts and climbing energies*, Pattern
494 Recognition, 47 (2014), pp. 12–24.
- 495 [24] J. B. KRUSKAL, *On the shortest spanning subtree of a graph and the traveling salesman problem*, Pro-
496 ceedings of the American Mathematical society, 7 (1956), pp. 48–50.
- 497 [25] R. LERALLUT, É. DECENCIÈRE, AND F. MEYER, *Image filtering using morphological amoebas*, Image and
498 Vision Computing, 25 (2007), pp. 395–404.
- 499 [26] M. NAGAO, T. MATSUYAMA, AND Y. IKEDA, *Region extraction and shape analysis in aerial photographs*,
500 Computer Graphics and Image Processing, 10 (1979), pp. 195–223.
- 501 [27] L. NAJMAN, *On the equivalence between hierarchical segmentations and ultrametric watersheds*, J. Math.
502 Imaging Vis., 40 (2011), pp. 231–247, <http://www.laurentnajman.org>. arXiv:1002.1887v2.
- 503 [28] L. NAJMAN, J. COUSTY, AND B. PERRET, *Playing with kruskal: algorithms for morphological trees in*
504 *edge-weighted graphs*, in International Symposium on Mathematical Morphology and Its Applications
505 to Signal and Image Processing, Springer, 2013, pp. 135–146.
- 506 [29] J. NEŠETŘIL, E. MILKOVÁ, AND H. NEŠETŘILOVÁ, *Otakar Borůvka on minimum spanning tree problem:*
507 *translation of both the 1926 papers, comments, history*, Discrete Mathematics, 233 (2001), pp. 3–36.
- 508 [30] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Physica
509 D: Nonlinear Phenomena, 60 (1992), pp. 259–268.
- 510 [31] J. SHI AND J. MALIK, *Normalized cuts and image segmentation*, IEEE Transactions on pattern analysis
511 and machine intelligence, 22 (2000), pp. 888–905.
- 512 [32] P. SOILLE, *Constrained connectivity for hierarchical image decomposition and simplification*, IEEE Trans.
513 Pattern Anal. Mach. Intell., 30 (2008), pp. 1132–1145, <https://doi.org/10.1109/TPAMI.2007.70817>.
- 514 [33] J. STAWIASKI AND F. MEYER, *Minimum spanning tree adaptive image filtering*, in 2009 16th IEEE
515 International Conference on Image Processing (ICIP), IEEE, 2009, pp. 2245–2248.
- 516 [34] G. STRANG, *Maximal flow through a domain*, Mathematical Programming, 26 (1983), pp. 123–143.
- 517 [35] U. VON LUXBURG, *A tutorial on spectral clustering*, Statistics and computing, 17 (2007), pp. 395–416.
- 518 [36] Y.-C. WEI AND C.-K. CHENG, *Towards efficient hierarchical designs by ratio cut partitioning*, in
519 Computer-Aided Design, 1989. ICCAD-89. Digest of Technical Papers., 1989 IEEE International
520 Conference on, IEEE, 1989, pp. 298–301.