Extending the PowerWatershed framework thanks to Γ-convergence

Laurent Najman

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. Power Watershed, Γ-convergence, image filtering, image segmentation, classification, spectral clustering

AMS subject classifications. 49N99, 68Q25, 68R10, 68U10. 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 $1 \geq \lambda_0 > \lambda_1 > \ldots \lambda_{n-1} > 0$, we set

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

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

$$x^\star \in \lim_{p \to \infty} \arg\min_{x \in \mathbb{R}^m} Q^p(x) \tag{2}$$

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 $p \to +\infty$. Indeed, if $\lambda_0 < 1$, $\sum_k \lambda_k^p Q_k(x)$ tends to 0 uniformly on $\mathbb{R}^m$ when $p \to +\infty$, and any $x \in \mathbb{R}^m$ is a minimizer of 0. Instead, we are interested in the limit $x^\star$ of the minimizers of Q^p as $p \to +\infty$. 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 [17, 7], which has been adapted to the case of space of graphs in [11, 20]. 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 coercivity[1] hypothesis (in our case, that would be on the Q_k) not applicable in our framework. We shall replace

*Submitted to the editors DATE.

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

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

[1] A continuous function $g : \mathbb{R}^m \to \mathbb{R}$ 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 $\text{limit}_{\|\mathbf{x}\| \to +\infty}\, g(\mathbf{x}) = +\infty$.

the coercivity hypothesis by a compactness argument. For the sake of completeness, we shall expose in this paper what is needed to understand the proof, without explicit reference to Γ-convergence.

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

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

Intuitive explanation of the contributions of this paper: as it is clear from the content of the present paper, the proposed framework allows to combine in a specific and rigorous way minimum-spanning-tree based optimization with other optimization algorithms, such as for instance random-walk or spectral clustering. It is well known that minimum spanning tree is an efficient and fast way to do a clustering, although this algorithm is subject to leaks [38] (also known as the chaining effect) and can not impose any constraint on the border of the clusters (such as, for example, regularity). The intuitive idea behind the main results of this paper, is to compute a minimum spanning tree where it is easy to do so, i.e., the centers of the clusters, and to apply a more evolved algorithm on an "extended" boundary of the clusters, in order to impose the constraints. This is somehow similar to what is often done for solving practical problems when the data is large: in such cases, one often starts by reducing the data size by applying a first clustering algorithm, such as the minimum spanning tree. Thus, a different way to look at this paper is the following: we propose a theory and some algorithms to combine data reduction and optimization-based clustering, based on first principles.

Section 2 provides some motivations for solving (2), the main one being the power-watershed [14]. In section 3, the main theorem of this paper is demonstrated, with its associate generic algorithm. In section 4, a specific algorithm dedicated to a particular class of functions is given and proved. Section 5 clarifies the links of the proposed framework with the (union of) maximum spanning trees. Two examples of applications are then developed. Section 6 shows how total-variation is related to watershed-based mosaic images. Section 7 exhibits an application to spectral clustering. Finally, in section 8, we propose some ideas for extension and future work.

2. Motivation.

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

2.1. Notations.

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

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

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

This problem was introduced for segmentation purposes in [14], with $q \geq 1$. In this case, Eq. (3) is a discrete formulation of the many possible variations on total variation denoising. In these formulations, w_{ij} are the *pairwise* weights, which can be interpreted as a weight on the gradient of the target configuration, such that the first term penalizes any unwanted high-frequency content in x and essentially forces x to vary smoothly within an object, while allowing large changes across the object boundaries. The second term enforces fidelity of x to a specified configuration f, w_i being the *unary* weights enforcing that fidelity. If $q > 1$, the function W^p in (3) is usually (depending on the w_{ij}) coercive, proper and strictly convex[2], and a unique minimum $\arg\min_x W^p(x)$ exists for each $p > 0$. The existence and the unicity of the solution $x^\star$ to Eq. (2) thus depends on the convergence of these solutions. The earlier proof provided in [14] is, unfortunately, incomplete and difficult to understand. In [1], there is a proof corresponding to the case $q = 1$, where several solutions $\arg\min_x W^p(x)$ may exist. For a different perspective on the same topic in a different framework, see [10].

We can rewrite Eq. (3) as follows:

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

with $\lambda_0 > \lambda_1 > \ldots > \lambda_{n-1}$, where $n \leq |V| + |E|$ is the number of different weights present in the graph G, be they pairwise (i.e., on a edge linking two different x_i and x_j) or unary (i.e., on a edge linking x_i to f_i), E_k is the set of edges with weights equal to λ_k and V_k is the set of vertices with data-fidelity weights equal to λ_k.

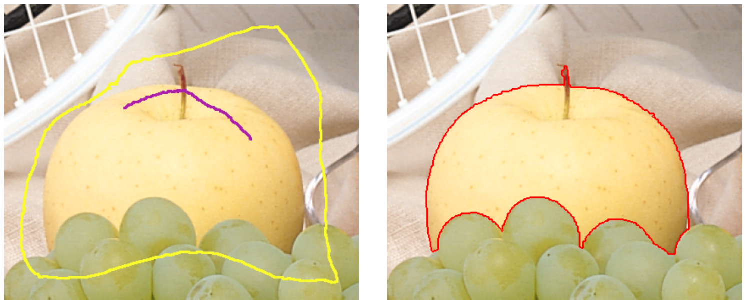
For $0 \leq k < n$, we set

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

We have

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

[2]See article by Combettes & Pesquet [12] for all the necessary hypotheses.



(a) Some seeds (b) Result of the segmentation

Figure 1: From [14], 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)

Theorem 3.3 on page 6 ensures the convergence of the minimizers of W^p. In particular, when $q > 1$ and when the problem is strictly convex, we have unicity of the limit of the minimizers. An example of application to seeded segmentation is shown in Figure 1 (see [14] for more details on this example.)

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

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

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

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

with

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

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

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

$$T_k(x) = \frac{1}{q_k} \sum_{v_i\in\mathcal{V}_k} \left(\sum_{\{v_j|\{v_i,v_j\}\in\mathcal{E}_k\}} a_{ij}|x_i - x_j|^2 \right)^{\frac{q_k}{2}} + \sum_{v_i\in\mathcal{V}_k} \frac{a_i}{2}|x_i - f_i|^2 \tag{10}$$

where $(\mathcal{V}_k, \mathcal{E}_k)$ is any subgraph of G, $a_{ij} \geq 0$ and $a_k \geq 0$. Let us write

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

We recognize in Eq. (10) a discrete Total-Variation-based regularization (more precisely, a discrete weighted p-Dirichlet regularization) of the subgraph $(\mathcal{V}_k, \mathcal{E}_k)$ weighted with the corresponding a_{ij} and a_i [22]. Theorem 3.3 (on page 6) allows us to combine several graph regularizations into one unique formulation: minimizing Eq. (11) can then be thought of as a combination of several scales of Total-Variation regularizations.

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

3.1. A simple example.

For any $\varepsilon > 0$, let $Q_\varepsilon : \mathbb{R}^2 \to \mathbb{R}$ be defined by

$$Q_\varepsilon(x_0, x_1) = \varepsilon(x_0 - x_1)^2 + \varepsilon^2 \left((x_0 - 1)^2 + x_1^2\right) \tag{12}$$

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

3.1.1. Direct approach.

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

$$Q_\varepsilon(\lambda) = 2\varepsilon^2(\lambda + \frac{1}{2})^2 + 4\varepsilon\lambda^2 = \lambda^2(4\varepsilon + 2\varepsilon^2) + 2\lambda\varepsilon^2 + \frac{\varepsilon^2}{2} \tag{13}$$

A derivation with respect to λ leads to

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

A second derivation yields:

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

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

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

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

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

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

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 sequence to $(\frac{1}{2}, \frac{1}{2})$ when ε tends to 0.

3.1.2. Scale-based approach. We first note that

$$\frac{Q_\varepsilon(x)}{\varepsilon} = (x_0 - x_1)^2 + \varepsilon\left((x_0-1)^2 + x_1^2\right) \tag{19}$$

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

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

Remark 3.1. Using results from [20], Γ-theory allows to prove that

$$(1,0) = \lim_{\varepsilon\to\infty} \arg\min_x Q_\varepsilon(x). \tag{20}$$

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 $Q^p := \sum_{0\le k<n} \lambda_k^p Q_k$, where $(\lambda_k)_{0\le i<n} \in \mathbb{R}^n$ is such that $1 \ge \lambda_0 > \lambda_1 > \ldots > \lambda_{n-1} > 0$, and $(Q_k)_{0\le k<n}$ 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:

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

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

*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 $p > 0$, $\|x_p\|_\infty \le C$), 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:*

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

where $m_k = \min_{x\in M_k} Q_k(x)$ and where $o(\lambda_{n-1}^p)$ is the Landau notation for negligibility[3].

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.

[3] In our case, this means that $\lim_{p\to\infty} \frac{Q^p(x) - \sum_{0\le k<n}\lambda_k^p m_k}{\lambda_{n-1}^p} = 0$

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 $p \to Q^p(.)$, we can also remark that if for all $p > 0$, Q^p has only one minimizer, then there exists a unique $\overline{x} \in \mathbb{R}^m$ such that $\{\overline{x}\} \subseteq M_{n-1}$.

We are now ready to prove Theorem 3.3.

3.2.1. Scale λ_0. We write

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

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

$$\min_x Q^p(x) = \lambda_0^p m_0 + o(\lambda_0^p) \tag{25}$$

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

$$R_0^p(x) = Q^p(x) - \lambda_0^p Q_0(x) \tag{26}$$

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 we have

$$\min_x Q^p(x) = \lambda_0^p m_0 + \lambda_1^p m_1 + o(\lambda_1^p) \tag{27}$$

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

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

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

and the whole process is repeated for all k.

Repeating the process for all scales λ_k, we obtain

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

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

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

Hence the Theorem 3.3.

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

3.3. A generic algorithm.

The algorithm 1 is a direct application of Theorem 3.3. However, in its generic form, it is not easy to implement. In the next section, we are going to particularize the function Q_k, so that we can provide a more specific implementation and so that we can link this implementation with minimum spanning tree algorithms.

4. Algorithm for the Power Watershed.

4.1. Notations and preliminary concepts.

We are going to detail how the limit of the minimizers in Eq. (2) is obtained for a specific class of functions Q^p that extends the power-watershed formulation given by Eq. (9).

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

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

Let $V^f \subseteq V$, $V^f \neq \emptyset$. The set V^f is the set of boundary conditions for which the value $x_i = f_i$ 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 one can show that the minimum of Q^p is bounded.

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

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

Then Q^p can be written as

$$Q^p(x) = \sum \lambda_k^p Q_k(x) \tag{32}$$

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

In the rest of the paper, we suppose that the graph G is connected. For any $t \geq 0$, the *level set* $[w]_t$ is the graph induced by $\{e_{ij} \in E \mid w_{ij} \geq t\}$. We remark that any two connected components of $([w]_t)_{t>0}$ are either disjoint or nested. Thus, these connected components can be organized in a tree structure[4]. $[w_t]$ is a subgraph of G. If a connected component C of $CC([w]_t)$ 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 connected component of one the level sets is seeded.

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 if $x_i = x_j$ for any $v_i, v_j \in V(C)$.*

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 achieves the minimum is constant on C.

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

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

4.2. Algorithm for the power watershed.. To provide an algorithm computing Eq. (2) where Q^p is given by Eq. (30), we need to define the contract operation. We say that we *contract* two connected vertices v_1 and v_2 when we remove the edge $\{v_1, v_2\}$ linking v_1 to v_2 while simultaneously merging the two vertices into a unique vertex. We say that we *contract* a connected subgraph C when we repeat the contraction until there is no edge in C, *i.e.* when C is reduced to a unique vertex. We remark that the order of the sequence of edge removal has no consequence on the result. We also remark that when we contract a connected subgraph C of a simple graph G, the resulting graph G' might no longer be a simple graph, but G' can be a *multi-graph*, *i.e.*, a graph in which an edge is repeated several times[5].

We are now ready to state algorithm 2. This algorithm is essentially the same as the one proposed in [14], although the presentation is simplified by the formalism we introduce in this paper. As stated previously, the various λ_k act as a notion of scale for the problem. Roughly speaking, at a given scale λ_k, this algorithm considers the level set of the graph corresponding to λ_k, and find the minimum of Q_k on this part of the graph. In the subsequent scales, the potentials fixed at previous scales are used as initial conditions.

We have the following theorem.

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 bounded. Then algorithm 2 computes the solution to Eq. (2).*

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

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

[5]As 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.

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)

The rest of the section is dedicated to proving Th. 4.2. We will proceed by successively considering the various scales λ_k.

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

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

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

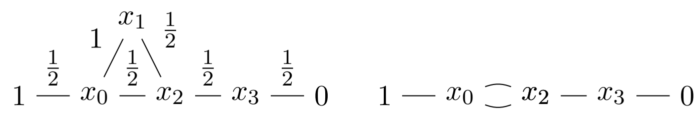
For the sake of simplicity, in the sequel, we still denote by G the resulting contracted multi-graph.

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

As in the previous step, the restriction of Q_1 to any connected component of $[w]_{\lambda_1}$ being positive, we can independently optimize Q_1 on each element $C \in CC([w]_{\lambda_1})$. If C is seeded, then the data attachment allows us to fix the potential x_i of any $v_i \in [w]_{\lambda_1}$ not already fixed at the previous scale. Otherwise, if C is not seeded, (*i.e.*, there is no data attachment for C), 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 vertex. As before, we continue to denote by G the resulting contracted multi-graph.

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

This concludes the proof.



(a) Graph corresponding to Eq. (33)

(b) After the contraction

Figure 2: Applying algorithm 2 on Eq. (33)

4.3. An illustration of the application of Theorem 4.2. Let us consider the following equation.

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

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

$$2(x_0 - x_2)^2 + (x_2 - x_3)^2 + (x_0 - 1)^2 + x_3^2. \tag{34}$$

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

5. Power-watershed and the union of all maximum spanning trees. Algorithm 2 has the structure of Kruskal's algorithm for the maximum spanning tree [25], and thus this invites us to look precisely at the relation between Eq. (2) and the maximum spanning tree problem.

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

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

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

The problem of finding a maximum spanning tree is the oldest problem in combinatorial optimization [30].

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

The following lemma, illustrated on Fig. 3, is easily deduced from the proof of Kruskal's algorithm for maximum spanning tree, and thus its proof is left for the reader. To ease the writing of the lemma, we note by convention $[w]_{\lambda_{-1}} = \emptyset$, where $\lambda_{-1} > \lambda_0$.

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 two situations holds:*

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

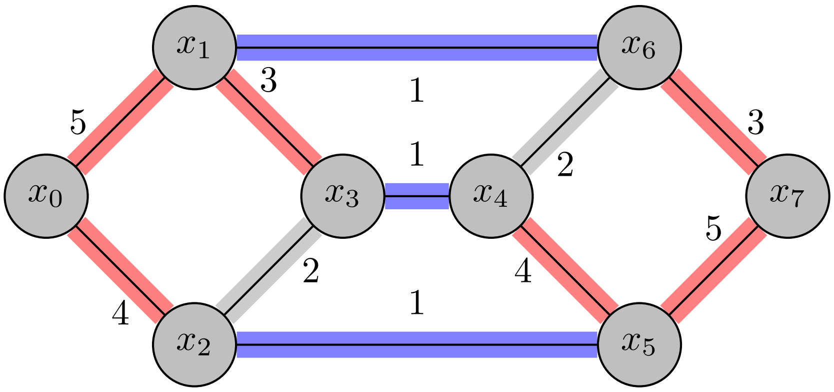


Figure 3: A graph illustrating lemma 5.1. The edges in red and blue form the union of all the maximum spanning trees of the graph. The grey edges $\{x_2, x_3\}$ and $\{x_4, x_6\}$ do not belong to any maximum spanning tree. One can check that both x_2 and x_3 (resp. x_4 and x_6) belong to the same red component of $[w]_3$. Each of the blue edges $\{x_1, x_6\}$, $\{x_3, x_4\}$ and $\{x_2, x_5\}$ belongs to different maximum spanning trees of the graph. One can check that each one of these edges links the two red components of $[w]_2$.

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

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

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

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

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

Proof. This theorem is also a straightforward consequence of Theorem 4.2, Theorem 5.2, and Lemma 5.1. ∎

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

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

In this section, we study the dependence of the solution $x^\star$ of Eq. (2) with respect to data f when $Q^p = TV_q^p$ is given by Eq. (36). This corresponds to a discrete version of the Total Variation problem [31, 19]. More precisely, depending on the value of q, it is a discrete Total Variation model with a data attachment term in L_0 (with the convention that $0^0 = 0$), L_1 (TV-L1) or L_2 (the ROF model) norm.

Let T_q be the operator that maps f to a solution $x^\star$ of Eq. (2).

6.1. Inheriting some properties of TV-L1. The following properties are inherited from the properties of the classical TV-L1 model, see for example [21]. Formal proofs are omitted for brevity.

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

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

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

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

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

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

It is easy to see, thanks to a simple computation that, for $e_{ij} \in E$, as soon as $\mu > |f_i - f_j|$, 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 given by

$$CC_q^\mu(v_i) = \{v_i\} \cup \Big\{ v_j \mid \text{there exists a path } \langle v_i = v^1, \ldots, v^n = v_j \rangle, \text{ such that } |f_k - f_{k+1}| < \mu \quad \forall 1 \leq k < n \Big\} \tag{37}$$

It is the quasi-flat zones hierarchy, introduced by Nagao *et al.* in 1979 [27], recently revived in [33] and shown to be equivalent to an edge-based topological watershed of the gradient in [28].

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 hierarchical in the sense that:

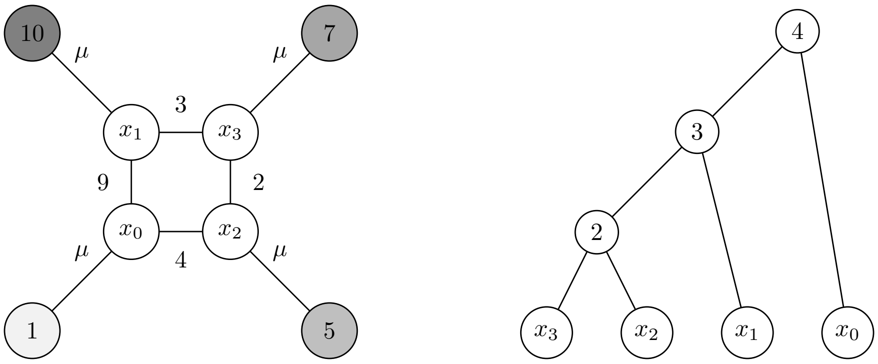
- For any $q \in \{0,1,2\}$, x_q^1 is totally flat. Indeed, we have x_0^1 is the mode of the vector f, x_1^1 is the median of f and $x_2^1 = \frac{1}{N}\sum_{v_i} f_i$ is the mean of f. Hence the whole space is an element of H,
- x_q^0 is exactly f, hence all the points are elements of H,
- two elements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

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

As an example of application, let us consider the graph depicted in Figure 4(a), that corresponds to the following equation

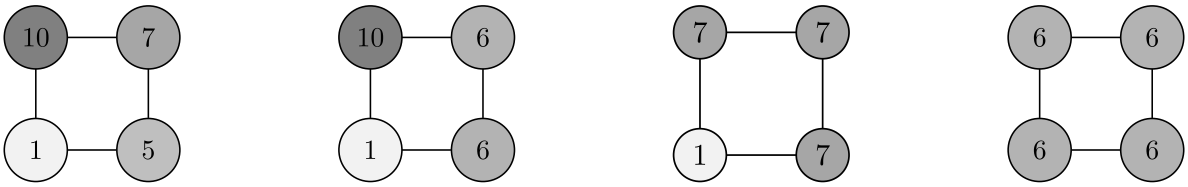
$$\begin{aligned} TVL_1^p(x) = &e^{-9p}|x_0 - x_1| + e^{-4p}|x_0 - x_2| + e^{-3p}|x_1 - x_3| + e^{-2p}|x_2 - x_3| \\ &+ e^{-\mu}\left[|x_0 - 1| + |x_1 - 10| + |x_2 - 5| + |x_3 - 7|\right]. \end{aligned} \tag{38}$$

The hierarchy H_1 corresponding to all flat zones of the solution x_1^μ to Eq. (2) for all μ is depicted in Figure 4(b). Some x_1^μ are given in Figures 4(c) to 4(f) depending on different values for μ.



(a) Graph corresponding to Eq. subsection 6.2

(b) The corresponding hierarchy



(c) $\mu < 2$

(d) $2 < \mu < 3$

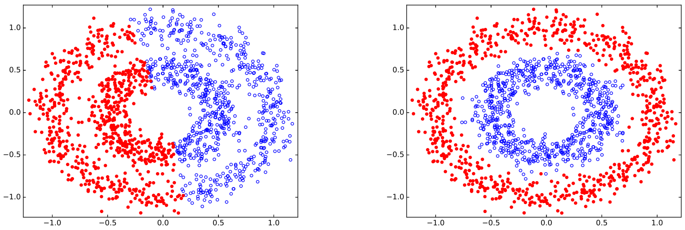
(e) $3 < \mu < 4$

(f) $\mu > 4$

Figure 4: An example of application of the Γ-limit of the TV-L1 scheme corresponding to Eq. subsection 6.2.

For fast algorithms computing the complete hierarchy of mosaic images, we refer to the ultrametric/saliency framework [28, 16, 29, 15]. Such algorithms allows one to compute the whole family of x_q^μ for $\mu \in [0, 1]^m$ in linear or quasi-linear time.

7. An application to spectral clustering: the power-ratio cut. Spectral clustering [36] is a successful approach to clustering and many different variations do exist. One of those variations is called *ratio-cut* [37], and can be described thanks to the following optimization



(a) Ratio-cut result (b) Power-ratio cut result

Figure 5: From [9], an example of comparison between ratio-cut and power-ratio cut spectral clusterings on two nested noisy circular sets of 2D points.

problem. For finding k clusters, compute the solution to

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

where L is the graph-laplacian[6], Tr is the trace operator and $\mathbb{I}$ is the identity matrix.

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

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

By raising the w_k at the power p, we can see that (41) fits into the power-watershed framework. Thus, we can define the *power-ratio cut* as the solution to the γ-limit of (41) when raising w_k at the power p and letting p tends to infinity. Algorithm 2 is not applicable to this situation, but we can apply some variation of algorithm 1. A complete theory, together with an efficient algorithm for solving this problem, as well as illustrations (such as the one of Figure 5) and experiments, can be found in [9].

[6] The 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

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

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

- Exploring the extension of the classical power-watershed, as proposed in Eq. (7), is a must.
- We showed in section 6 how to express the morphological mosaic images as a Γ-limit of some Total-Variation operators. The next steps would be to look at the theory of *scale-set image analysis* [23, 24] and find if it would be possible to obtain optimal hierarchies composed of optimal partitions. This possibly would allows us to think the watershed operator as a projector, and would probably allows to strenghten the ideas developed in [13] regarding anisotropic diffusion.
- Regarding spectral classification, we hint in section 7 at what can be done with the ratio-cut criterion [9]. Other criterions, such as normalized cuts [32], should of course be studied, but this could be more difficult. Such a study could lead to applications in machine learning such as graph matching and non-linear dimensionality reduction techniques.
- Application of the proposed framework for morphological filtering, in order to clarify links and differences between for example MST-based filtering [34, 3] and distance-based filtering (such as the amoeba framework [26]) is also of interest. A first step in this direction can be found in [18].

We list below some other open problems.

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

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

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

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

8.2. Link between the continuous world and the discrete one. When $\varphi_{ij}(x) = \varphi_k(x) = |x|$, Eq. (3) is equivalent to a min-cut/max-flow problem. This problem has a continuous formulation, that was proposed by G. Strang [35]. But the maximum spanning tree problem, being a tree problem, is purely discrete, and we do not see any obvious formulation in continuous terms.

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

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

Acknowledgments. The author would like to thank A. Challa, S. Danda, F. Malgouyre, J.-C. Pesquet and H. Talbot, as well as the anonymous reviewers, for their insightful comments of previous versions of this paper.

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

REFERENCES

[1] C. Allène, J.-Y. Audibert, M. Couprie, and R. Keriven, *Some links between extremum spanning 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.

[2] G. Anzellotti and S. Baldo, *Asymptotic development by γ-convergence*, Applied mathematics and optimization, 27 (1993), pp. 105–123.

[3] L. Bao, Y. Song, Q. Yang, H. Yuan, and G. Wang, *Tree filtering: Efficient structure-preserving smoothing with a minimum spanning tree*, Image Processing, IEEE Transactions on, 23 (2014), pp. 555–569.

[4] C. Berge, *Topological Spaces: including a treatment of multi-valued functions, vector spaces, and convexity*, Courier Dover Publications, 1963.

[5] S. Beucher, *Watershed, hierarchical segmentation and waterfall algorithm*, in Mathematical morphology and its applications to image processing, Springer, 1994, pp. 69–76.

[6] S. Beucher and F. Meyer, *The morphological approach to segmentation: the watershed transformation*, Optical Engineering, 34 (1992), pp. 433–433.

[7] A. Braides, *A handbook of γ-convergence*, Handbook of Differential Equations: stationary partial differential equations, 3 (2006), pp. 101–213.

[8] E. Carlinet and T. Géraud, *A comparative review of component tree computation algorithms*, IEEE Transactions on Image Processing, 23 (2014), pp. 3885–3895.

[9] A. Challa, S. Danda, B. Daya Sagar, and L. Najman, *Gamma limit for spectral clustering*, in Discrete geometry for computer imagery, LNCS, Springer, 2017, https://hal.archives-ouvertes.fr/view/index/docid/1427957. To appear.

[10] K. C. Ciesielski, J. K. Udupa, A. X. Falcão, and P. A. Miranda, *Fuzzy connectedness image segmentation in graph cut formulation: A linear-time algorithm and a comparative analysis*, J. Math. Imaging Vis., 44 (2012), pp. 375–398.

[11] Y. Colin De Verdière, Y. Pan, and B. Ycart, *Singular limits of schrödinger operators and markov processes*, J. Operator Theory, 41 (1999), pp. 151–173.

[12] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, eds., Springer-Verlag, New York, 2010.

[13] C. Couprie, L. Grady, L. Najman, and H. Talbot, *Anisotropic Diffusion Using Power Watersheds*, in International Conference on Image Processing (ICIP'10), Sept. 2010, pp. 4153–4156.

[14] C. Couprie, L. Grady, L. Najman, and H. Talbot, *Power watershed: A unifying graph-based optimization 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.

[15] J. Cousty, G. Bertrand, L. Najman, and M. Couprie, *Watershed cuts: Minimum spanning forests and the drop of water principle*, IEEE Trans. Pattern Anal. Mach. Intell., 31 (2009), pp. 1362–1374.

[16] J. Cousty and L. Najman, *Incremental algorithm for hierarchical minimum spanning forests and saliency of watershed cuts*, in 10th International Symposium on Mathematical Morphology (ISMM'11), vol. 6671 of Lecture Notes in Computer Science, 2011, pp. 272–283, http://www.esiee.fr/~info/sm/.

[17] G. Dal Maso, *An Introduction to Γ-convergence*, vol. 8, Springer, 1993.

[18] S. Danda, A. Challa, B. Daya Sagar, and L. Najman, *Power tree filter: A theoretical framework explaining and extending the minimum spanning tree filter*, in Mathematical Morphology and Its Applications to Signal and Image Processing: 13th International Symposium, J. An-

gulo, S. Velasco-Forero, and F. Meyer, eds., vol. 10225 of LNCS, Springer, 2017, pp. 199–210, https://hal.archives-ouvertes.fr/hal-01430538/.
[19] J. Darbon and M. Sigelle, *Image restoration with discrete constrained total variation part I: Fast and exact optimization*, Journal of Mathematical Imaging and Vision, 26 (2006), pp. 261–276.
[20] Y. C. de Verdière, *Spectres de graphes*, vol. 4 of Cours Spécialisés, Société Mathématique de France, 1998.
[21] V. Duval, J.-F. Aujol, and Y. Gousseau, *The TVL1 model: a geometric point of view*, Multiscale Modeling & Simulation, 8 (2009), pp. 154–189.
[22] A. Elmoataz, O. Lezoray, and S. Bougleux, *Nonlocal discrete regularization on weighted graphs: a framework for image and manifold processing*, Image Processing, IEEE Transactions on, 17 (2008), pp. 1047–1060.
[23] L. Guigues, J. P. Cocquerez, and H. L. Men, *Scale-sets image analysis*, International Journal of Computer Vision, 68 (2006), pp. 289–317, http://dx.doi.org/10.1007/s11263-005-6299-0.
[24] B. R. Kiran and J. Serra, *Global–local optimizations by hierarchical cuts and climbing energies*, Pattern Recognition, 47 (2014), pp. 12–24.
[25] J. B. Kruskal, *On the shortest spanning subtree of a graph and the traveling salesman problem*, Proceedings of the American Mathematical society, 7 (1956), pp. 48–50.
[26] R. Lerallut, É. Decencière, and F. Meyer, *Image filtering using morphological amoebas*, Image and Vision Computing, 25 (2007), pp. 395–404.
[27] M. Nagao, T. Matsuyama, and Y. Ikeda, *Region extraction and shape analysis in aerial photographs*, Computer Graphics and Image Processing, 10 (1979), pp. 195–223.
[28] L. Najman, *On the equivalence between hierarchical segmentations and ultrametric watersheds*, J. Math. Imaging Vis., 40 (2011), pp. 231–247, http://www.laurentnajman.org. arXiv:1002.1887v2.
[29] L. Najman, J. Cousty, and B. Perret, *Playing with Kruskal: algorithms for morphological trees in edge-weighted graphs*, in International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing, C. L. Luengo Hendriks, G. Borgefors, and R. Strand, eds., vol. 7883 of Lecture Notes in Computer Science, Springer, 2013, pp. 135–146.
[30] J. Nešetřil, E. Milková, and H. Nešetřilová, *Otakar Borůvka on minimum spanning tree problem: translation of both the 1926 papers, comments, history*, Discrete Mathematics, 233 (2001), pp. 3–36.
[31] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D: Nonlinear Phenomena, 60 (1992), pp. 259–268.
[32] J. Shi and J. Malik, *Normalized cuts and image segmentation*, IEEE Transactions on pattern analysis and machine intelligence, 22 (2000), pp. 888–905.
[33] P. Soille, *Constrained connectivity for hierarchical image decomposition and simplification*, IEEE Trans. Pattern Anal. Mach. Intell., 30 (2008), pp. 1132–1145, https://doi.org/10.1109/TPAMI.2007.70817.
[34] J. Stawiaski and F. Meyer, *Minimum spanning tree adaptive image filtering*, in 2009 16th IEEE International Conference on Image Processing (ICIP), IEEE, 2009, pp. 2245–2248.
[35] G. Strang, *Maximal flow through a domain*, Mathematical Programming, 26 (1983), pp. 123–143.
[36] U. Von Luxburg, *A tutorial on spectral clustering*, Statistics and computing, 17 (2007), pp. 395–416.
[37] Y.-C. Wei and C.-K. Cheng, *Towards efficient hierarchical designs by ratio cut partitioning*, in Computer-Aided Design, 1989. ICCAD-89. Digest of Technical Papers., 1989 IEEE International Conference on, IEEE, 1989, pp. 298–301.
[38] C. T. Zahn, *Graph-theoretical methods for detecting and describing gestalt clusters*, IEEE Transactions on computers, 100 (1971), pp. 68–86.