Air Force Institute of Technology

AFIT Scholar

9-2012

Guaranteeing Convergence of Iterative Skewed Voting Algorithms for Image Segmentation

Doru C. Balcan
Georgia Institute of Technology

Gowri Srinivasa
PES School of Engineering - Bangalore

Matthew C. Fickus
Air Force Institute of Technology

Jelena Kovacevic
Carnegie Mellon University

Follow this and additional works at: https://scholar.afit.edu/facpub

Part of the Harmonic Analysis and Representation Commons

Recommended Citation

Guaranteeing Convergence of Iterative Skewed Voting Algorithms for Image Segmentation

Doru C. Balcan[a,*], Gowri Srinivasa[b], Matthew Fickus[c], Jelena Kovačević[d]

[a]*School of Interactive Computing, Georgia Institute of Technology, Atlanta, USA*
[b]*Dept. of Information Science and Engineering, and Center for Pattern Recognition, PES School of Engineering, Bangalore, India*
[c]*Dept. of Mathematics and Statistics, Air Force Institute of Technology, Wright-Patterson AFB, USA*
[d]*Dept. of Biomedical Eng., Electrical and Computer Eng. and Center for Bioimage Informatics, Carnegie Mellon University, Pittsburgh, USA*

Abstract

In this paper we provide rigorous proof for the convergence of an iterative voting-based image segmentation algorithm called Active Masks. Active Masks (AM) was proposed to solve the challenging task of delineating punctate patterns of cells from fluorescence microscope images. Each iteration of AM consists of a linear convolution composed with a nonlinear thresholding; what makes this process special in our case is the presence of additive terms whose role is to "skew" the voting when prior information is available. In real-world implementation, the AM algorithm always converges to a fixed point. We study the behavior of AM rigorously and present a proof of this convergence. The key idea is to formulate AM as a generalized (parallel) majority cellular automaton, adapting proof techniques from discrete dynamical systems.

Keywords: active masks, cellular automata, convergence, segmentation.

1. Introduction

Recently, a new algorithm called *Active Masks* (AM) was proposed for the segmentation of biological images [14]. Let the "image" f be any real-valued function over the domain $\Omega := \prod_{d=1}^{D} \mathbb{Z}_{N_d}$ and refer to the $N := N_1 N_2 \ldots N_D$ elements of Ω as *pixels*; here, $\mathbb{Z}_{N_d}$ denotes the finite group of integers modulo N_d. A *segmentation* of f assigns one of M possible *labels* to each of the N pixels in Ω. For the fluorescence microscope image depicted in Figure 1(a), one example of a successful segmentation is to label all of the background pixels as "1," assign label "2" to every pixel in the largest cell, "3" to every pixel in the second largest cell, and so on. Formally, a segmentation is a *label function* $\psi : \Omega \to \{1, 2, \ldots, M\}$, or, equivalently, a collection of M binary *masks* $\mu_m : \Omega \to \{0, 1\}$ where, at any given $n \in \Omega$, we have $\mu_m(n) = 1$ if and only if $\psi(n) = m$. That is, μ_m at any iteration i can be defined as

$$\mu_m^{(i)} := \begin{cases} 1, & \psi_i(n) = m, \\ 0, & \psi_i(n) \neq m, \end{cases}$$

In AM, these masks actively evolve according to a given rule. To understand this evolution, it helps to first discuss *iterative voting*: in each iteration, at any given pixel, one counts how often a given label appears in the neighborhood of that pixel—weighting nearby neighbors more than distant ones—and assigns the most frequent label to that pixel in the next iteration. For example, if a pixel labeled "1" in the current iteration is completely surrounded by pixels labeled "2", its label will likely change to "2" in the next iteration. Formally speaking, iterative voting is the repeated application of the rule:

$$\text{Iterative Voting:} \qquad \psi_i(n) = \underset{1 \leq m \leq M}{\operatorname{argmax}} \left[(\mu_m^{(i-1)} * g)(n) \right], \tag{1}$$

---

*Corresponding author

Email addresses: `dbalcan@cc.gatech.edu` (Doru C. Balcan), `gsrinivasa@pes.edu` (Gowri Srinivasa), `Matthew.Fickus@afit.edu` (Matthew Fickus), `jelenak@cmu.edu` (Jelena Kovačević)

[1]Email: dbalcan@cc.gatech.edu. Mailing address: College of Computing Building, room 218, Georgia Institute of Technology, 801 Atlantic Drive, Atlanta, GA 30332-0280. Phone: (404) 385-8547, Fax: (404)894-0673.

where i is the index of the iteration, $g : \Omega \to \mathbb{R}$ is some arbitrarily chosen fixed weighting function and "$*$" denotes circular convolution over Ω. Iterative voting is referred to as a *convolution-threshold* scheme since it simplifies to rounding the filtered version of $\mu_1^{(i)}$ in the special case $M = 2$. Experimentation reveals that for typical low-pass filters g, repeatedly applying (1) to a given initial ψ_0 results in a progressive smoothing of the contours between distinctly labeled regions of Ω. Despite this nice property, note that taken by itself, iterative voting is useless as a segmentation scheme, as (1) evolves masks in a manner that is independent of any image under consideration.

The AM algorithm is a generalization of (1) that contains additional image-based terms whose purpose is to drive the iteration towards a meaningful segmentation. To be precise, the AM iteration is:

$$\text{Active Masks:} \qquad \psi_i(n) = \underset{1 \le m \le M}{\text{argmax}} \left[(\mu_m^{(i-1)} * g)(n) + R_m(n) \right], \qquad (2)$$

where the region-based distributing functions $\{R_m\}_{m=1}^{M}$ can be any image-dependent real-valued functions over Ω. These will be referred to as *skew functions* in this paper, due to their role to bias the voting. Essentially, at any given pixel n, these additional terms skew the voting towards labels m whose $R_m(n)$ values are large. For good segmentation, one should define the R_m's in terms of features in the image that distinguish regions of interest from each other.

For example, for the fluorescence microscope image given in Figure 1(a), the cells appear noticeably brighter than the background. As such, we choose R_1 to be a soft-thresholded version of the image's local average brightness, and choose the remaining R_m's to be identically zero. When (2) is applied, such a choice in R_m's forces pixels which lie outside the cells towards label "1," while pixels that lie inside a cell can assume any other label. Intuitively, repeated applications of (2) will cause the mask $\mu_1^{(i)}$ to converge to an indicator function of the background, while each of the other masks $\{\mu_m^{(i)}\}_{m=2}^{M}$ converges either to a smooth blob contained within the foreground or to the empty set. Experimentation reveals that the AM algorithm indeed often converges to a ψ which assigns a unique label to each cell provided the scale of the window g is chosen appropriately [14]; see Figure 1 for examples.

The purpose of this paper is to provide a rigorous investigation of the convergence behavior of the AM algorithm. To be precise, we note that in a real-world implementation the AM algorithm occasionally fails to converge to a ψ which is biologically meaningful. However, even in these cases, the algorithm always seems to converge to something. Indeed, when ψ_0 and the R_m's are chosen at random, experimentation reveals that the repeated application of (2) always seems to eventually produce ψ_i such that $\psi_{i+1} = \psi_i$. At the same time, a simple example tempers one's expectations: taking $\Omega = \mathbb{Z}_4$, $M = 2$, $w = \delta_{-1} + \delta_0 + \delta_1$ and $R_1 = R_2 = 0$, we see that AM will not always converge, as repeatedly applying (2) to $\psi_0 = \delta_0 + \delta_2$ produces the endless 2-cycle $\delta_0 + \delta_2 \mapsto \delta_1 + \delta_3 \mapsto \delta_0 + \delta_2$. In summary, even though random experimentation indicates that AM will almost certainly converge, there exist trivial examples which show that it will not always do so. The central question that this paper seeks to address is therefore:

Under what conditions on g and $\{R_m\}_{m=1}^{M}$ will the AM algorithm always converge to a fixed point of (2) ?

We show that when g is an even function, AM will either converge to a fixed point or will get stuck in a 2-cycle; no higher-order cycles are possible. We can further rule out 2-cycles whenever g is taken so that the convolutional operator $f \mapsto f * g$ is positive semidefinite. The following is a compilation of these results:

Theorem 1.1. *Given any $\Omega := \prod_{d=1}^{D} \mathbb{Z}_{N_d}$, initial segmentation $\psi_0 : \Omega \to \{1, \dots, M\}$ and any real-valued functions $\{R_m\}_{m=1}^{M}$ over Ω, the Active Mask algorithm, namely the repeated application of (2), will always converge to a fixed point of (2) provided the discrete Fourier transform of g is nonnegative and even.*

A preliminary version of the results in this paper appears in the conference proceeding [2]. Though the specific AM algorithm was introduced in [14], iterative lowpass filtering has long been a subject of interest in applied harmonic analysis, having deep connections to continuous-domain ideas such as diffusion and the maximum principle [8]. For instance, [9] gives an edge detection application of a discretized version of these ideas. Meanwhile, [6] gives diffusion-inspired conditions under which lowpass filtering is guaranteed to produce a coarse version of a given image. One way to prove the convergence of iterative convolution-thresholding schemes is to show that lowpass filtering decreases the number of zero-crossings in a signal; such a condition is equivalent to a version of the maximum principle [11]. More recently, the continuous-domain version of (1) has been used to model the motion of interfaces between media [12, 13]; in that setting, (1) is known to converge if $M = 2$. Since the AM algorithm is iterative, many of the proof techniques we use here were adapted from *majority cellular automata* (MCA), a well-studied class of
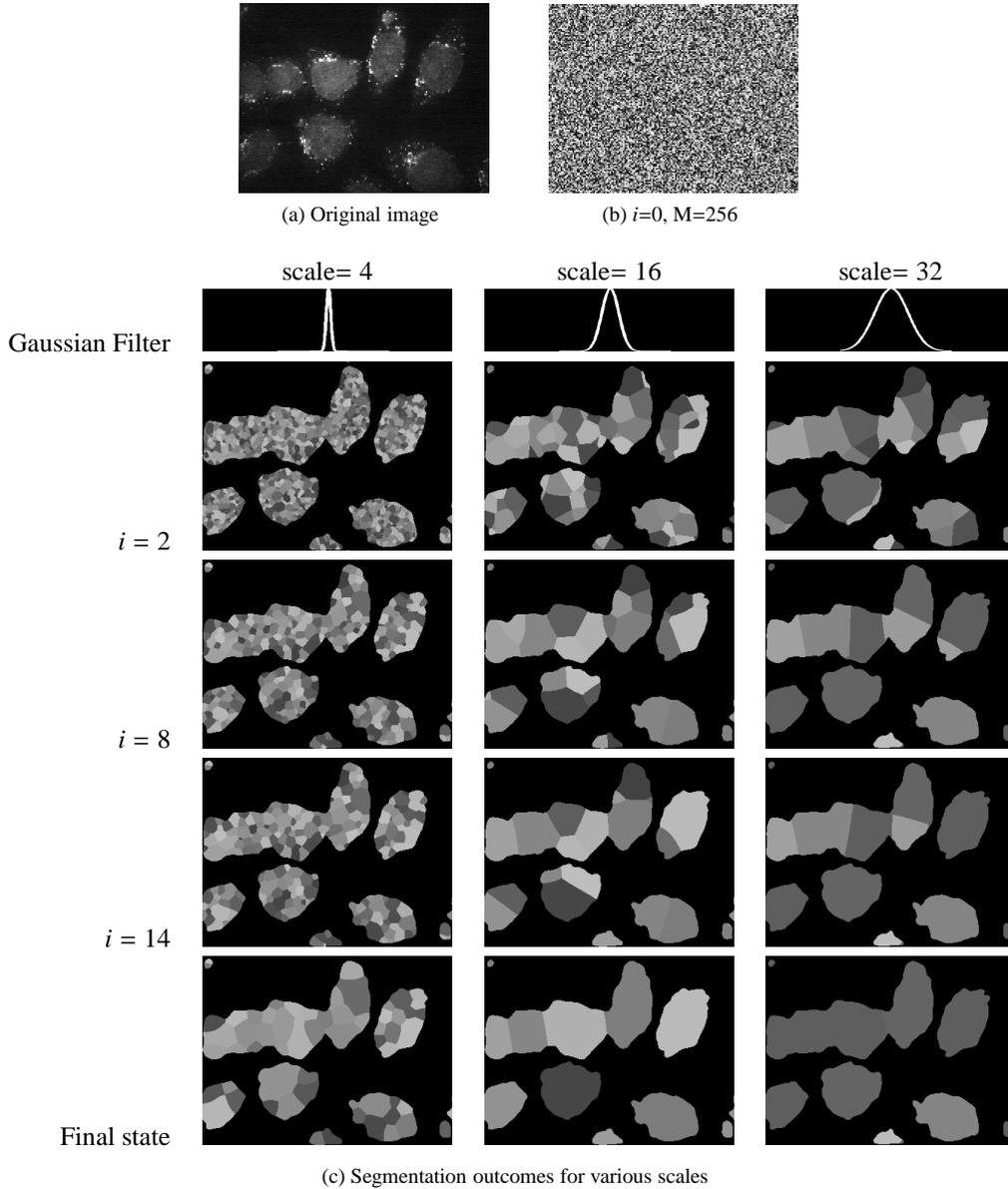
(a) Original image

(b) i=0, M=256

scale= 4   scale= 16   scale= 32

Gaussian Filter

$i = 2$

$i = 8$

$i = 14$

Final state

(c) Segmentation outcomes for various scales

Figure 1: Active mask segmentation of punctate patterns of proteins [14]. (a) Original image (courtesy of Linstedt Lab [1]) (b) Initialization using M=256 random masks. After one iteration of (2), the background is separated from the foreground by the region-based skew function R_1. (c) Segmentation results using various scales of the voting filter at iterations 2, 8, 14 and at convergence. The first row shows a cross section of three Gaussian filters (scale=4, 16 and 32 respectively). The second row shows the segmentation result after two iterations of the algorithm. When scale=4, we observe a large number of small regions in the foreground. This is contrasted by the fewer number of regions when scale=16 and scale=32. Subsequent rows represent the state of the system at various stages of evolution. The last row represents the convergence states. Note that the algorithm converges regardless of the scale chosen, but the segmentation is only biologically meaningful at the proper scale of 16; at scales=4, 32 the cells are oversegmented or undersegmented, respectively.

3

discrete dynamical systems. Indeed, theoretical guarantees on the convergence of a symmetric class of MCA have been known for several decades; see [3, 10], and references therein. Such results were recently generalized to a quasi-symmetric class via the use of Lyapunov functionals [7]. Whereas much of traditional MCA theory focuses on the convergence of repeated applications of (1), our work differs due to the presence of the additive R_m terms in (2).

The paper is organized as follows. In the next section, we use an MCA formulation of AM to prove our main convergence results. In Section 3, we then briefly discuss the generalization of our main results to a less elegant yet more realistic version of (2) involving noncircular convolution. We conclude in Section 4 with some examples illustrating our main results, as well as some experimental results indicating the AM algorithm's rate of convergence.

2. Active Masks as a Majority Cellular Automaton

Cellular automata are self-evolving discrete dynamical systems [3]. They have been applied in various fields such as statistical physics, computational biology, and the social sciences. A tremendous amount of work in this area has focused on studying the convergence behavior of various types of automata. In this section, we formulate the AM algorithm (2) as an MCA in order to facilitate our understanding of its convergence behavior. To be precise, we consider a generalization of (2) in which the convolutional operator $f \mapsto f * g$ may more broadly be taken to be any linear operator A from $\ell(\Omega) := \{f | f : \Omega \to \mathbb{R}\}$ into itself:

$$\psi_i(n) := \min\Big(\underset{m}{\operatorname{argmax}}[(A\mu_m^{(i-1)})(n) + R_m(n)]\Big), \qquad \mu_m^{(i-1)} := \begin{cases} 1, & \psi_{i-1}(n) = m, \\ 0, & \psi_{i-1}(n) \neq m. \end{cases} \tag{3}$$

Here, the contribution of mask m in deciding the outcome at location n at iteration i is $(A\mu_m^{(i-1)})(n)$, and any ties are broken by choosing the smallest m corresponding to a maximal element. Note that given any initial segmentation ψ_0, applying (3) *ad infinitum* produces a sequence $\{\psi_i\}_{i=0}^{\infty}$. However, as there are only M^N distinct possible configurations for $\psi : \Omega \to \{1, \ldots, M\}$, this sequence must eventually repeat itself. Indeed, taking minimal indices i_0 and $K > 0$ such that $\psi_{i_0+K} = \psi_{i_0}$, the deterministic nature of (3) implies that $\psi_{i+K} = \psi_i$ for all $i \geq i_0$. The finite sequence $\{\psi_i\}_{i=i_0}^{i_0+K-1}$ is called a *cycle* of (3) of *length K*. Note that $\{\psi_i\}_{i=0}^{\infty}$ converges if and only if $K = 1$, which happens precisely when ψ_{i_0} is a fixed point of (3).

Thus, from this perspective, proving that (3) always converges is equivalent to proving that $K = 1$ regardless of one's choice of ψ_0. The following result goes a long way towards this goal, showing that if A is self-adjoint, then for any ψ_0 we have that the resulting K is necessarily 1 or 2. That is, if A is self-adjoint, then for any ψ_0, the sequence $\{\psi_i\}_{i=0}^{\infty}$ will either converge in a finite number of iterations, or it will eventually come to a point where it forever oscillates between two distinct configurations ψ_{i_0} and ψ_{i_0+1}.

Theorem 2.1. *If A is self-adjoint, then for any ψ_0, the cycle length K of* (3) *is either 1 or 2.*

Proof. As we are not presently concerned with the rate of convergence of (3), but rather the question of whether it does converge, we may assume without loss of generality that $\{\psi_i\}_{i=0}^{\infty}$ has already entered its cycle. That is, we reindex so that ψ_0 is the beginning of the K-cycle, and heretofore regard all iteration indices as members of the cyclic group $\mathbb{Z}_K$. We argue by contrapositive, assuming $K > 2$ and concluding that A is not self-adjoint. For any $i = 1, \ldots, K$, (3) is equivalent to the system of inequalities:

$$(A\mu_{\psi_i(n)}^{(i-1)})(n) + R_{\psi_i(n)}(n) > (A\mu_m^{(i-1)})(n) + R_m(n) \qquad \text{if } 1 \leq m < \psi_i(n), \tag{4a}$$

$$(A\mu_{\psi_i(n)}^{(i-1)})(n) + R_{\psi_i(n)}(n) \geq (A\mu_m^{(i-1)})(n) + R_m(n) \qquad \text{if } \psi_i(n) \leq m \leq M. \tag{4b}$$

Here, (4b) follows from the fact that $\psi_i(n)$ is a value of m that maximizes $(A\mu_m^{(i-1)})(n) + R_m(n)$. Moreover, in the event of a tie, $\psi_i(n)$ is chosen to be the least of all such maximizing m, yielding the strict inequality in (4a). For any i and n, picking $m = \psi_{i-2}(n)$ in (4a) and (4b) leads to the subsystem of inequalities:

$$(A\mu_{\psi_i(n)}^{(i-1)})(n) - (A\mu_{\psi_{i-2}(n)}^{(i-1)})(n) + R_{\psi_i(n)}(n) - R_{\psi_{i-2}(n)}(n) > 0 \qquad \text{if } \psi_{i-2}(n) < \psi_i(n), \tag{5a}$$

$$(A\mu_{\psi_i(n)}^{(i-1)})(n) - (A\mu_{\psi_{i-2}(n)}^{(i-1)})(n) + R_{\psi_i(n)}(n) - R_{\psi_{i-2}(n)}(n) \geq 0 \qquad \text{if } \psi_{i-2}(n) \geq \psi_i(n). \tag{5b}$$

4

Now since $K > 2$, there exists a pixel n for which $\{\psi_0(n), \psi_1(n), \ldots, \psi_{K-1}(n)\}$ is not of the form $\{a, a, \ldots, a\}$ nor of the form $\{a, b, a, b, \ldots a, b\}$. At such an n, there must exist an i such that $\psi_{i-2}(n) < \psi_i(n)$. Consequently, at least one inequality in (5) is strict. Thus, summing (5) over all pixels n and all cycle indices i yields:

$$0 < \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} (A\mu_{\psi_i(n)}^{(i-1)})(n) - \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} (A\mu_{\psi_{i-2}(n)}^{(i-1)})(n) + \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} R_{\psi_i(n)}(n) - \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} R_{\psi_{i-2}(n)}(n).$$

Since $\mathbb{Z}_K$ is shift-invariant, $\sum_{i \in \mathbb{Z}_K} R_{\psi_i(n)}(n) = \sum_{i \in \mathbb{Z}_K} R_{\psi_{i-2}(n)}(n)$ for any $n \in \Omega$, reducing the previous equation to:

$$0 < \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} (A\mu_{\psi_i(n)}^{(i-1)})(n) - \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} (A\mu_{\psi_{i-2}(n)}^{(i-1)})(n) = \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} (A\mu_{\psi_i(n)}^{(i-1)})(n) - \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} (A\mu_{\psi_{i-1}(n)}^{(i)})(n), \qquad (6)$$

where the final equality also follows from the shift-invariance of $\mathbb{Z}_K$. To continue, note that for any $i, j \in \mathbb{Z}_K$ we have $\mu_m^{(j)} = 1$ if and only if $\psi_j(n) = m$ and so:

$$\sum_{n \in \Omega} (A\mu_{\psi_j(n)}^{(i)})(n) = \sum_{n \in \Omega} \sum_{m=1}^{M} (A\mu_m^{(i)})(n)\mu_m^{(j)}(n) = \sum_{m=1}^{M} \langle A\mu_m^{(i)}, \mu_m^{(j)} \rangle, \qquad (7)$$

where $\langle f, g \rangle := \sum_{n \in \Omega} f(n)g(n)$ is the standard real inner product over Ω. Using (7) in (6) gives:

$$0 < \sum_{i \in \mathbb{Z}_K} \sum_{m=1}^{M} \langle A\mu_m^{(i-1)}, \mu_m^{(i)} \rangle - \sum_{i \in \mathbb{Z}_K} \sum_{m=1}^{M} \langle A\mu_m^{(i)}, \mu_m^{(i-1)} \rangle = \sum_{i \in \mathbb{Z}_K} \sum_{m=1}^{M} \langle (A - A^*)\mu_m^{(i-1)}, \mu_m^{(i)} \rangle,$$

implying $A - A^* \neq 0$, and so A is not self-adjoint. $\qquad \square$

Theorem 2.1 has strong implications for the AM algorithm (2). Indeed, it is well-known that if g is real-valued, then the adjoint of the convolutional operator $Af = f * g$ is $A^*f = f * \tilde{g}$ where $\tilde{g}(n) = g(-n)$ is the *reversal* of g. As such, if g is an even function, Theorem 2.1 guarantees that AM will always either converge or enter a 2-cycle. We now build on the techniques of the previous proof to find additional restrictions on A which suffice to guarantee convergence:

Theorem 2.2. *If A is self-adjoint and $\langle Af, f \rangle \geq 0$ for all $f : \Omega \to \{0, \pm 1\}$, then (3) always converges.*

Proof. In light of Theorem 2.1, our goal is to rule out cycles of length $K = 2$. We argue by contrapositive. That is, we assume that there exist two distinct configurations ψ_0 and ψ_1 which are successors of each other, and will use this fact to produce $f : \Omega \to \{0, \pm 1\}$ such that $\langle Af, f \rangle < 0$. Substituting $i = 0$ and $m = \psi_1(n)$ into (4a) and (4b) yields:

$$(A\mu_{\psi_0(n)}^{(1)})(n) - (A\mu_{\psi_1(n)}^{(1)})(n) + R_{\psi_0(n)}(n) - R_{\psi_1(n)}(n) > 0 \qquad \text{if } \psi_1(n) < \psi_0(n), \qquad (8a)$$

$$(A\mu_{\psi_0(n)}^{(1)})(n) - (A\mu_{\psi_1(n)}^{(1)})(n) + R_{\psi_0(n)}(n) - R_{\psi_1(n)}(n) \geq 0 \qquad \text{if } \psi_1(n) \geq \psi_0(n). \qquad (8b)$$

Similarly, letting $i = 1$ and $m = \psi_0(n)$ into (4a) and (4b) yields:

$$(A\mu_{\psi_1(n)}^{(0)})(n) - (A\mu_{\psi_0(n)}^{(0)})(n) + R_{\psi_1(n)}(n) - R_{\psi_0(n)}(n) > 0 \qquad \text{if } \psi_0(n) < \psi_1(n), \qquad (9a)$$

$$(A\mu_{\psi_1(n)}^{(0)})(n) - (A\mu_{\psi_0(n)}^{(0)})(n) + R_{\psi_1(n)}(n) - R_{\psi_0(n)}(n) \geq 0 \qquad \text{if } \psi_0(n) \geq \psi_1(n). \qquad (9b)$$

Since ψ_0 and ψ_1 are distinct, there exists $n_0 \in \Omega$ such that $\psi_0(n_0) \neq \psi_1(n_0)$. If $\psi_0(n_0) < \psi_1(n_0)$, we sum (8b) and (9a) over all $n \in \Omega$. If on the other hand $\psi_0(n_0) > \psi_1(n_0)$, we sum (8a) and (9b) over all $n \in \Omega$. Either way, we obtain:

$$0 < \sum_{n=1}^{N} [(A\mu_{\psi_0(n)}^{(1)})(n) - (A\mu_{\psi_1(n)}^{(1)})(n) + (A\mu_{\psi_1(n)}^{(0)})(n) - (A\mu_{\psi_0(n)}^{(0)})(n)].$$

Applying (7) four times then gives:

$$0 < \sum_{m=1}^{M} [\langle A\mu_m^{(1)}, \mu_m^{(0)} \rangle - \langle A\mu_m^{(1)}, \mu_m^{(1)} \rangle + \langle A\mu_m^{(0)}, \mu_m^{(1)} \rangle - \langle A\mu_m^{(0)}, \mu_m^{(0)} \rangle] = -\sum_{m=1}^{M} \langle A(\mu_m^{(1)} - \mu_m^{(0)}), (\mu_m^{(1)} - \mu_m^{(0)}) \rangle.$$

As such, there exists at least one index m_0 such that $0 > \langle A(\mu_{m_0}^{(1)} - \mu_{m_0}^{(0)}), (\mu_{m_0}^{(1)} - \mu_{m_0}^{(0)}) \rangle$; choose f to be $\mu_{m_0}^{(1)} - \mu_{m_0}^{(0)}$. $\qquad \square$

The most obvious way to ensure that $\langle Af, f \rangle \geq 0$ for all $f : \Omega \to \{0, \pm 1\}$ is for A to be positive semidefinite, that is, $\langle Af, f \rangle \geq 0$ for all $f : \Omega \to \mathbb{R}$. This in turn can be guaranteed by taking A to be diagonally dominant with nonnegative diagonal entries, via the Gershgorin circle Theorem [5]. Note that in fact strict diagonal dominance guarantees that iterative voting (1) always converges in one iteration. More interesting examples can be found in the special case where A is a convolutional operator $Af = f * g$. Indeed, letting F be the standard non-normalized discrete Fourier transform (DFT) over Ω, we have:

$$\langle Af, f \rangle = \langle f * g, f \rangle = \frac{1}{N} \langle \mathrm{F}(f * g), \mathrm{F}f \rangle = \frac{1}{N} \langle (\mathrm{F}f)(\mathrm{F}g), \mathrm{F}f \rangle = \frac{1}{N} \sum_{n \in \Omega} (\mathrm{F}g)(n) \big| (\mathrm{F}f)(n) \big|^2. \tag{10}$$

As such, if $g : \Omega \to \mathbb{R}$ is even and $(\mathrm{F}g)(n) \geq 0$ for all $n \in \Omega$, then A is self-adjoint positive semidefinite. Moreover, it is well-known that g is real-valued and even if and only if Fg is also real-valued and even. Thus, A is self-adjoint positive semidefinite provided Fg is nonnegative and even. For such g, Theorem 2.2 guarantees that the AM algorithm (2) will always converge. These facts are summarized in Theorem 1.1, which is stated in the introduction. Examples of g that satisfy these hypotheses are given in Section 4.

We emphasize that Theorem 2.2 does not require A to be positive semidefinite, but rather only that $\langle Af, f \rangle \geq 0$ for all $f : \Omega \to \{0, \pm 1\}$. In the case of convolutional operators, this means we truly only need (10) to hold for such f's. As such, it may be overly harsh to require that $(Fg)(n) \geq 0$ for all $n \in \Omega$. Unfortunately, the problem of characterizing such g's appears difficult, as we could find no useful frequency-domain characterizations of $\{0, \pm 1\}$-valued functions. A spatial domain approach is more encouraging: when $\Omega = \mathbb{Z}_N$, writing $f : \Omega \to \{0, \pm 1\}$ as the difference of two characteristic functions $\chi_{I_1}, \chi_{I_2} : \Omega \to \{0, 1\}$, we have:

$$\langle Af, f \rangle = \langle A(\chi_{I_1} - \chi_{I_2}), (\chi_{I_1} - \chi_{I_2}) \rangle = \mathrm{sum}(A_{1,1}) + \mathrm{sum}(A_{2,2}) - \mathrm{sum}(A_{1,2}) - \mathrm{sum}(A_{2,1}),$$

where $\mathrm{sum}(A_{i,j})$ denotes the sum of all entries of the submatrix of A consisting of rows from I_i and columns from I_j. As such, the condition of Theorem 2.2 reduces to showing that $0 \leq \mathrm{sum}(A_{1,1}) + \mathrm{sum}(A_{2,2}) - \mathrm{sum}(A_{1,2}) - \mathrm{sum}(A_{2,1})$ for all choices of subsets I_i and I_j of $\mathbb{Z}_N$.

We conclude this section by noting that (3) is similar to *threshold cellular automata* (TCA) [3, 4]. In fact, (3) is equivalent to TCA in the special case of $M = 2$; in this case, $\mu_0^{(i-1)}(n) = 1 - \mu_1^{(i-1)}(n)$ for all $n \in \Omega$, implying:

$$
\begin{aligned}
(A\mu_1^{(i-1)})(n) + R_1(n) > (A\mu_0^{(i-1)})(n) + R_0(n) \quad &\Longleftrightarrow \quad [A(\mu_1^{(i-1)} - \mu_0^{(i-1)})](n) + (R_1 - R_0)(n) > 0 \\
&\Longleftrightarrow \quad \big\{ A[\mu_1^{(i-1)} - (1 - \mu_1^{(i-1)})] \big\}(n) + (R_1 - R_0)(n) > 0 \\
&\Longleftrightarrow \quad (A\mu_1^{(i-1)})(n) + \tfrac{1}{2}(R_1 - R_0 - A1)(n) > 0 \\
&\Longleftrightarrow \quad (A\mu_1^{(i-1)})(n) + b(n) > 0,
\end{aligned}
$$

where $b(n) := \frac{1}{2}(R_1 - R_0 - A1)(n)$. That is, when $M = 2$, the AM algorithm is equivalent to a threshold-like decision. But whereas the traditional method for proving the convergence of TCA involves associated quadratic Lyapunov functionals [4], our method for proving the convergence of AM is more direct, being closer in spirit to that of [10].

3. Beyond symmetry

Up to this point, we have focused on the convergence of (3) in the special case where A is self-adjoint. In this section, we discuss how Theorems 2.1 and 2.2 generalize to the case of *quasi-self-adjoint* operators, which arise in real-world implementation of the AM algorithm. To clarify, up to this point, we have let the image f and weights g be functions over the finite abelian group $\Omega = \prod_{d=1}^{D} \mathbb{Z}_{N_d}$ and have taken the convolutions in (2) and (3) to be circular. In real-world implementation, the use of such circular convolutions can result in poor segmentation, as values at one edge of the image are used to influence the segmentation at the unrelated opposite edge.

One solution to this problem—implemented in [14]—is to redefine the set of pixels as a subset $\Omega := \prod_{d=1}^{D} [0, N_d)$ of the D-dimensional integer lattice $\mathbb{Z}^D$, and regard our image f as a member of $\ell(\Omega) := \{f : \mathbb{Z}^D \to \mathbb{R} \mid f(n) = 0 \ \forall n \notin \Omega\}$. Here, the label function ψ and masks μ_m are regarded as $\{1, \ldots, M\}$- and $\{0, 1\}$-valued members of $\ell(\Omega)$, respectively, and the (noncommutative) convolution of any $f, g \in \ell(\Omega)$ with $g \in \ell^2(\mathbb{Z}^D)$ is defined as $f \star g \in \ell(\Omega)$,

$$(f \star g)(n) := \frac{(f * g)(n)}{(\chi_\Omega * g)(n)}, \quad \forall n \in \Omega, \tag{11}$$

where χ_Ω is the characteristic function of Ω, and $*$ denotes standard (noncircular) convolution in $\ell^2(\mathbb{Z}^D)$. For the theory below, we need to place additional restrictions on g, namely that it belongs to the class:

$$\mathcal{G}(\Omega) := \{g \in \ell^2(\mathbb{Z}^D) : (\chi_\Omega * g)(n) > 0 \ \forall n \in \Omega\}.$$

In this setting, for a given $g \in \mathcal{G}(\Omega)$, the AM algorithm (2) becomes:

$$\text{Noncircular Active Masks:} \qquad \psi_i(n) = \operatorname*{argmax}_{1 \leq m \leq M} [(\mu_m^{(i-1)} \star g)(n) + R_m(n)], \qquad \mu_m^{(i-1)} := \begin{cases} 1, & \psi_{i-1}(n) = m, \\ 0, & \psi_{i-1}(n) \neq m. \end{cases} \quad (12)$$

Note that the use of the $\star$-convolution in (12) ensures that any "missing votes" are not counted in favor of any label m. Moreover, the denominator of (11) ensures that when n is close to an edge of Ω, the weights in the g-neighborhood of n are rescaled so as to always sum to one. This rescaling ensures that $\sum_{m=1}^M (\mu_m^{(i)} \star g)(n) = 1$ for all $n \in \Omega$, avoiding any need to modify the skew functions R_m near the boundary.

 We then ask the question: for what g will (12) always converge? The key to answering this question is to realize that the $\star$-filtering operation $Af = f \star g$ can be factored as $A = DB$, where B is the standard filtering operator $Bf = f * g$ and $(Df)(n) = \lambda_n f(n)$, where $\lambda_n = [(\chi_\Omega * g)(n)]^{-1}$. Here, A, B and D are all regarded as linear operators from $\ell(\Omega)$ into itself. More generally, we inquire into the convergence of:

$$\psi_i(n) = \operatorname*{argmax}_{1 \leq m \leq M} [(A\mu_m^{(i-1)})(n) + R_m(n)], \qquad \mu_m^{(i-1)} := \begin{cases} 1, & \psi_{i-1}(n) = m, \\ 0, & \psi_{i-1}(n) \neq m, \end{cases} \quad (13)$$

where $A = DB$ and D is *positive-multiplicative*, that is, $(Df)(n) = \lambda_n f(n)$ where $\lambda_n > 0$ for all $n \in \Omega$. In particular, we follow [7] in saying that A is *quasi-self-adjoint* if there exists a positive-multiplicative operator D and a self-adjoint operator B such that $A = DB$. This definition in hand, we have the following generalization of Theorems 2.1 and 2.2:

Theorem 3.1. *Let A be quasi-self-adjoint: $A = DB$ where D is positive-multiplicative and B is self-adjoint. Then for any ψ_0, the cycle length K of (13) is either 1 or 2. Moreover, if B is positive-semidefinite, then (13) always converges.*

Proof. We only outline the proof, as it closely follows those of Theorems 2.1 and 2.2. Let $(Df)(n) = \lambda_n f(n)$ with $\lambda_n > 0$ for all $n \in \Omega$. We prove the first conclusion by contrapositive, assuming $K > 2$. Rather than summing (5) over all n and i directly, we instead first divide each instance of (5) by the corresponding λ_n, and then sum. The resulting quantity is analogous to (6):

$$0 < \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} \frac{1}{\lambda_n} (A\mu_{\psi_i(n)}^{(i-1)})(n) - \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} \frac{1}{\lambda_n} (A\mu_{\psi_{i-1}(n)}^{(i)})(n) = \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} (B\mu_{\psi_i(n)}^{(i-1)})(n) - \sum_{i \in \mathbb{Z}_K} \sum_{n \in \Omega} (B\mu_{\psi_{i-1}(n)}^{(i)})(n). \quad (14)$$

Simplifying the right-hand side of (14) with (7) quickly reveals that B cannot be self-adjoint, completing this part of the proof. For the second conclusion, we again prove by contrapositive, assuming $K = 2$. Dividing (8a), (8b), (9a) and (9b) by λ_n and then summing either (8a) and (9b) over all n or (8b) and (9a) over all n gives:

$$0 < \sum_{n=1}^N \frac{1}{\lambda_n} [(A\mu_{\psi_0(n)}^{(1)})(n) - (A\mu_{\psi_1(n)}^{(1)})(n) + (A\mu_{\psi_1(n)}^{(0)})(n) - (A\mu_{\psi_0(n)}^{(0)})(n)] = -\sum_{m=1}^M \langle B(\mu_m^{(1)} - \mu_m^{(0)}), (\mu_m^{(1)} - \mu_m^{(0)}) \rangle,$$

implying B is not positive semidefinite. $\qquad\square$

 For a result about the convergence of (12), we apply Theorem 3.1 to $A = DB$ where $\lambda_n = [(\chi_\Omega * g)(n)]^{-1}$ and $Bf = f * g$. Note that we must have $g \in \mathcal{G}(\Omega)$ in order to guarantee that D is positive. Moreover, B is self-adjoint if $g \in \ell^2(\mathbb{Z}^D)$ is even; since g is real-valued, this is equivalent to having its classical Fourier series $\hat{g} \in L^2(\mathbb{T}^D)$ be real-valued and even. Meanwhile, since:

$$\langle Bf, f \rangle = \langle f * g, f \rangle = \langle \hat{f}\hat{g}, \hat{f} \rangle = \int_{\mathbb{T}^d} \hat{g}(x) |\hat{f}(x)|^2 \, dx,$$

then B is positive semidefinite if $\hat{g}(x) \geq 0$ for almost every $x \in \mathbb{T}^D$. To summarize, we have:

Corollary 3.2. *If the Fourier series of $g \in \mathcal{G}(\Omega)$ is nonnegative and even, then (12) will always converge.*

 In the next section, we discuss how to construct such windows g, along with other implementation-related issues.

4. Examples of Active Masks in practice

In this section we present a few representative and interesting examples of filter-based cellular automata, and discuss their behavior in relation with the results we proved in the previous sections. We also present some preliminary experimental findings on the rate of convergence of AM. For ease of understanding, let us for the moment restrict ourselves to circulant iterative voting (1), namely the version of AM (2) in which all the skew functions R_m are identically zero. The simplest nonzero filter is $g = \delta_0$. The DFT of δ_0 has constant value 1, and is therefore nonnegative and even. As such, Theorem 1.1 guarantees that (1) will always converge. Of course, we already knew that: since $f * \delta_0 = f$ for all $f \in \ell(\Omega)$, (1) will always converge in one step; as noted above, the same holds true for any g whose convolutional operator is strictly diagonally dominant with a nonnegative diagonal: $g(0) \geq \sum_{n \neq 0} |g(n)|$.

More interesting examples arise from *box filters*: symmetric cubes of Dirac δ's. For instance, fix $N \geq 3$ and consider (1) over $\Omega = \mathbb{Z}_N$ where $g = \delta_{-1} + \delta_0 + \delta_1$. Since g is symmetric, Theorem 2.1 guarantees that (1) will either always converge or will enter a 2-cycle. However, if N is even, then (1) will not always converge, since $\psi_0 = \delta_0 + \delta_2 + \cdots + \delta_{N-2}$ generates a 2-cycle. This phenomenon is depicted in Figure 2(a). This simple example shows that symmetry alone does not suffice to guarantee convergence; one truly needs additional hypotheses on g, such as the requirement in Theorem 1.1 that its DFT is nonnegative. This hypothesis does not hold for $g = \delta_{-1} + \delta_0 + \delta_1$, since $(Fg)(n) = 1 + 2\cos(\frac{2\pi n}{N})$. Similar issues arise in the two-dimensional setting $\Omega = \mathbb{Z}_{N_1} \times \mathbb{Z}_{N_2}$: both the 3×3 box filter (Moore's automaton, see Figure 2(b)) and the "plus" filter (von Neumann's automaton, see Figure 2(c)) are symmetric, meaning their cycle lengths are either 1 or 2, but neither are positive semidefinite, having DFTs of $[1 + 2\cos(\frac{2\pi n_1}{N_1})][1 + 2\cos(\frac{2\pi n_2}{N_2})]$ and $1 + 2\cos(\frac{2\pi n_1}{N_1}) + 2\cos(\frac{2\pi n_2}{N_2})$, respectively. Indeed, when N_1 and N_2 are even, alternating stripes generate a 2-cycle for the box filter, while the checkerboard generates a 2-cycle for the plus filter.

Of course, it is not difficult to find filters g which do satisfy the hypotheses of Theorem 1.1: one may simply let g be the inverse DFT of any nonnegative even function. More concrete examples, such as a discrete Gaussian over $\mathbb{Z}_N$, can be found using the following process. Let $h : \mathbb{R} \to \mathbb{R}$ be an even Schwartz function whose Fourier transform is nonnegative; an example of such a function is a continuous Gaussian. Let g be the N-periodization of the integer samples of h, namely $g(n) := \sum_{n'=-\infty}^{\infty} h(n + Nn')$. Then g is even, and moreover, by the Poisson summation formula:

$$(Fg)(n) = \sum_{n'=0}^{N-1} g(n')e^{-\frac{2\pi i n n'}{N}} = \sum_{n'=0}^{N-1} \sum_{n''=-\infty}^{\infty} h(n' + Nn'')e^{-\frac{2\pi i n n'}{N}} = \sum_{k=-\infty}^{\infty} h(k)e^{-\frac{2\pi i n k}{N}} = \sum_{k=-\infty}^{\infty} \hat{h}(k + \frac{n}{N}) \geq 0.$$

In particular, if g is chosen as a periodized version of the integer samples of any zero-mean Gaussian, then Theorem 1.1 gives that the AM algorithm (2) necessarily converges. This construction method immediately generalizes to higher-dimensional settings where $D > 1$. It also generalizes to the noncircular convolution setting considered in Section 3. There, we further restrict h to be strictly positive, and let g be the integer samples of h. The positivity of h implies $(\chi_\Omega * g)(n) > 0$ for all $n \in \Omega$, implying $g \in \mathcal{G}(\Omega)$ as needed. Moreover, g is even and the Poisson summation formula gives that its Fourier series is nonnegative: $\hat{g}(x) = \sum_{k=-\infty}^{\infty} \hat{h}(k + x) \geq 0$. Any g constructed in this manner satisfies the hypotheses of Corollary 3.2, implying the corresponding noncircular AM (12) necessarily converges.

4.1. The rate of convergence of the AM algorithm

Up to this point, we have focused on the question of whether or not the AM algorithm (2) converges. Having settled that question to some degree, our focus now turns to another question of primary importance in real-world implementation: at what rate does AM converge? Experimentation reveals that this rate highly depends on the configuration of the boundary between two distinctly labeled regions of Ω. This led us to postulate that the number of *boundary crossings* (see Figure 3) should monotonically decrease with each iteration. Experimentation reveals that this number indeed often decreases extremely rapidly, regardless of the scale of g. Figure 4 depicts such an experiment for the fluorescence microscope image shown in Figure 1(a). Starting from a random initial configuration of 64 masks, we used a Gaussian filter under three different scales, with each plot depicting the evolution of 5 independently-initialized runs of the algorithm. We emphasize the algorithm's fast rate of convergence: the vertical axis represents a nested four-fold application of the natural logarithm to the number of boundary crossings. We leave a more rigorous investigation of the AM algorithm's rate of convergence for future work.

(a) 3-tap filter automaton

(b) Moore's automaton
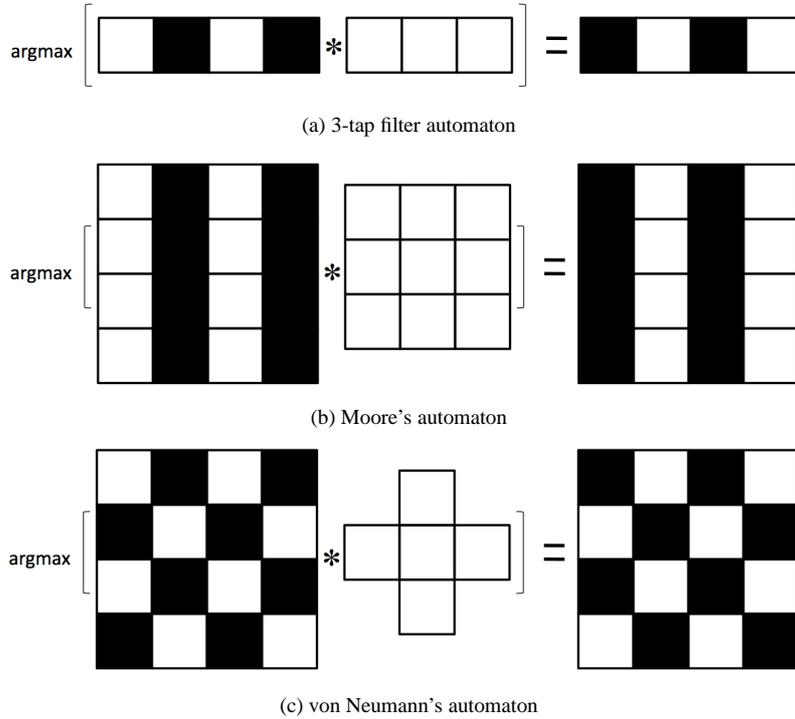

(c) von Neumann's automaton

Figure 2: An illustration of oscillating states produced by various automata: (a) The 3-tap box filter $\delta_{-1} + \delta_0 + \delta_1$ over $\Omega = \mathbb{Z}_4$. Using this g in (1) with $\psi_0 = \delta_0 + \delta_2$ results in the endless 2-cycle $\delta_0 + \delta_2 \mapsto \delta_1 + \delta_3 \mapsto \delta_0 + \delta_2$. This is because at each iteration, each pixel's two neighbors will outvote him in deciding his label in the next iteration. (b) Convergence is also an issue in two dimensions, as illustrated by Moore's automaton—a 3×3 box filter—over $\Omega = \mathbb{Z}_4 \times \mathbb{Z}_4$. (c) Two-cycles persist in two dimensions even when the box filter is replaced by the smoother "plus" filter of von Neumann's automaton. In all three cases, these filters are even and so Theorem 2.1 ensures that the cycle length K of (1) is either 1 or 2. However, none of them are positive semidefinite, as their DFTs attain negative values. As such, the convergence guarantees of Theorem 2.2 do not hold.
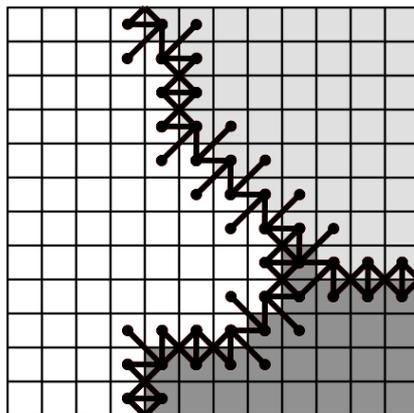

Figure 3: An illustration of the zero-crossings in an image with M=3 masks.
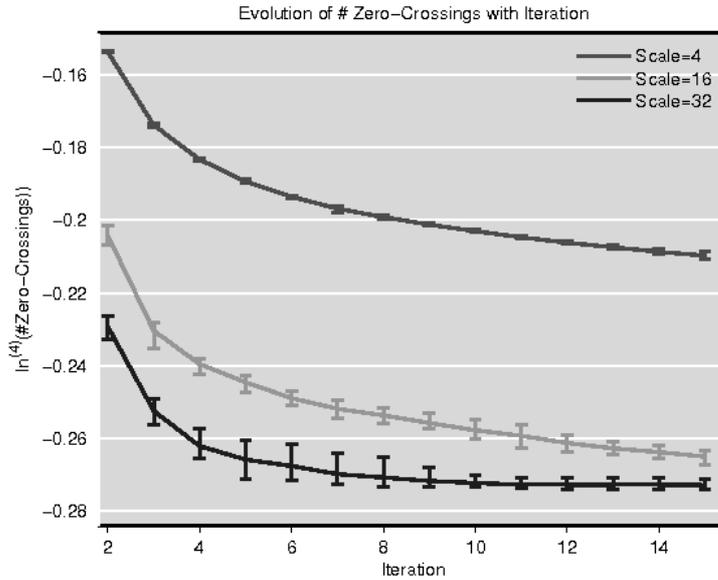
9

Figure 4: The rate of decrease of the AM algorithm in terms of the number of boundary crossings.

Acknowledgments

References

[1] Linstedt Lab at CMU. http://www.cmu.edu/bio/faculty/linstedt.html.

[2] D.C. Balcan, G. Srinivasa, M. Fickus, and J. Kovačević. Convergence behavior of the Active Mask segmentation algorithm. In *IEEE ICASSP*, 2010.

[3] E. Golés and S. Martínez. *Neural and Automata Networks. Dynamical Behavior and Applications*. Kluwer, 1990.

[4] E. Golés-Chacc, Fr. Fogelman-Soulie, and D. Pellegrin. Decreasing energy functions as a tool for studying threshold networks. *Discrete Applied Mathematics*, 12:261–277, 1985.

[5] G. Golub and Ch. L. Van Loan. *Matrix Computations*. Johns Hopkins University Press, London, 3rd edition, 1996.

[6] J.J. Koenderink. The structure of images. *Journ. Bio. Cybern.*, 50(5):363–370, 1984.

[7] P. Melatagia Yonta and R. Ndoundam. Opinion dynamics using majority functions. *Mathematical Social Sciences*, 57:223–244, 2009.

[8] L. Nirenbarg. A strong maximum principle for parabolic equations. *Commun. Pure and Appl. Math.*, 6:167–177, 1953.

[9] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. Patt. Anal. and Mach. Intelligence*, 12(7):629–639, 1990.

[10] S. Poljak and M. Sura. On periodical behaviour in societies with symmetric influences. *Combinatorica*, 3(1):119–121, 1983.

[11] R. A. Hummel. *Readings in computer vision: issues, problems, principles, and paradigms*, chapter Representations based on zero-crossings in scale-space, pages 753–758. Morgan Kaufmann Readings Series archive. Morgan Kaufmann Publishers, Inc., CA, USA, 1987.

[12] S. J. Ruuth. Efficient algorithms for diffusion-generated motion by mean curvature. *Journ. Comp. Phys.*, 144:603–625, 1998.

[13] S. J. Ruuth and B. Merriman. Convolution-thresholding methods for interface motion. *Journ. Comp. Phys.*, 169:678–707, 2001.

[14] G. Srinivasa, M. Fickus, Y. Guo, A. Linstedt, and J. Kovačević. Active mask segmentation of fluorescence microscope images. *IEEE Trans. Imag. Proc.*, 18(8):1817–1829, 2009.