# Energetic-lattice based optimization

Bangalore Ravi Kiran

Université Paris-Est
Ecole doctorale MSTIC

Thèse soumise pour obtenir le grade de
Docteur de l'Université Paris Est en Informatique

**Bangalore Ravi KIRAN**

Directeur de thése: Jean SERRA
Co-Encadrant: Jean COUSTY

# Energetic-Lattice based optimization

## L'optimization par Trellis-Energetique

First version submitted on 30 August 2014

Membres du Jury:

| | |
|---|---|
| Hugues TALBOT | (President) |
| Philippe SALEMBIER | (Rapporteur) |
| Jocelyn CHANUSSOT | (Rapporteur) |
| Jesús ANGULO | (Examinateur) |
| Michael H.F. WILKINSON | (Examinateur) |
| Frank R SCHMIDT | (Examinateur) |
| Jean COUSTY | (Examinateur) |
| Jean SERRA | (Directeur de thése) |

*"Thinking evolves the objective. All the three worlds exist through thinking.*
*The Kosmos melts away on its dissolution. This thinking should carefully be diagnosed."*

- Yogavasishtha (Ramacharka, The spirit of the Upanishads)

Dedicated to my Guru Jean, Marie Françoise
and my Parents Ravi and Kalpana.

தேடிச் சோறுநிதந் தின்று - பல
சின்னஞ் சிறுகதைகள் பேசி - மனம்
வாடித் துன்பமிக உழன்று - பிறர்
வாடப் பலசெயல்கள் செய்து - நரை
கூடிக் கிழப்பருவ மெய்தி - கொடுங்
கூற்றுக் கிரையெனப்பின் மாயும் - பல
வேடிக்கை மனிதரைப் போலே - நான்
வீழ்வே னென்று நினைத் தாயோ?

-மகா கவி பாரதி

"Did you think I too will
Spend my days in mundane search of food,
Telling petty tales and gossips,
Worrying myself with unwanted thoughts,
Hurting others by my selfish acts,
Turn senile old man with grey hair
To end up as fodder to the
relentless march of timeless Death,
As yet another faceless man?"

- Mahakavi Bharathi

# Abstract

Hierarchical segmentation has been a model which both identifies with the construct of extracting a tree structured model of the image, while also interpreting it as an optimization problem of the optimal scale selection. Hierarchical processing is an emerging field of problems in computer vision and hyperspectral image processing community, on account of its ability to structure high-dimensional data.

Chapter 1 discusses two important concepts of *Braids* and *Energetic lattices*. A braid of partitions is a richer hierarchical partition model that provides multiple locally non-nested partitioning, while being globally a hierarchical partitioning of the space. The problem of optimization on hierarchies and further braids are non-tractable due the combinatorial nature of the problem. We provide conditions, of *h-increasingness*, *scale-increasingness* on the energy defined on partitions, to extract unique and monotonically ordered minimal partitions.

Furthermore these conditions are found to be coherent with the Braid structure to perform constrained optimization on hierarchies, and more generally Braids. Chapter 2 demonstrates the Energetic lattice, and how it generalizes the Lagrangian formulation of the constrained optimization problem on hierarchies.

Finally in Chapter 3 we apply the method of optimization using energetic lattices to the problem of extraction of segmentations from a hierarchy, that are proximal to a ground truth set.

Chapter 4 we show how one moves from the energetic lattice on hierarchies and braids, to a numerical lattice of Jordan Curves which define a continuous model of hierarchical segmentation. This model enables also to compose different functions and hierarchies.

Chapter 5 compiles the scale-climbing algorithms by Guigues and Salembier-Garrido, over the hierarchies of partitions, and provides the new dynamic program for the Braids of partitions. Further it discusses a perspective on using intersection graphs to solve the optimal cut problem, and identities "Partition Graphs" to be one of the good graph structures to model partition selection. It finally concludes by formulating the optimal cut problem on hierarchies as a flow-maximization on a tree structure, the case of braids are also discussed.

**Keywords**: Hierarchical segmentation, Lagrangian Multipliers, Lattice optimization, Mathematical Morphology.

# Résumé

La segmentation hiérarchique est une méthode pour produire des partitions qui représentent une même image de manière de moins en moins fine. En même temps, elle sert d'entrée à la recherche d'une partition optimale, qui combine des extraits des diverses partitions en divers endroits. Le traitement hiérarchique des images est un domaine émergent en vision par ordinateur, et en particulier dans la communauté qui étudie les images hyperspectrales et les SIG, du fait de son capacité à structurer des données hyper-dimensionnelles.

Le chapitre 1 porte sur les deux concepts fondamentaux de tresse et de treillis energétique. La tresse est une notion plus riche que celle de hierarchie de partitions, en ce qu'elle incorpore, en plus, des partitions qui ne sont pas emboîtées les unes dans les autres, tout en s'appuyant glolalement sur une hiérarchie. Le treillis énergétique est une structure mixte qui regroupe une tresse avec une énergie, et permet d'y défninr des éléments maximaux et minimaux. Lorsqu'on se donne une énergie, trouver la partition formée de classes de la tresse (ou de la hiérarchie) qui minimise cette énergie est un problème insoluble, de par sa complexité combinatoriale. Nous donnons les deux conditions de h-croissance et de croissance d'échelle, qui garantissent l'existence, l'unicité et la monotonie des solutions, et conduisent à un algorithme qui les détermine en deux passes de lecture des données.

Le chapitre 2 reste dans le cadre précédent, mais étudie plus spécifiquement l'optimisation sous contrainte. Il débouche sur trois généralisations du modèle Lagrangien.

Le chapitre 3 applique l'optimisation par treillis énergétique au cas de figure où l'energie est introduite par une "vérité terrrain", c'est à dire par un jeu de dessins mauel, que les partitions optimales doivent serrer au plus près.

Enfin, le chapitre 4 passe des treills énergétiques à ceux des courbes de Jordan dans le plan euclidien, qui définissent un modèle continu de segmentations hierarchiques. Il permet entre autres de composer les hiérarchies avec diverses fonctions numériques.

Chapitre 5 compile les algorithmes d'escalade par Guigues, sur les hiérarchies de partitions, et en plus fournit le nouveau programme dynamique pour les tresses de partitions. En outre, il décrit une perspective sur des graphes d'intersection qui aide à résoudre le problème de la coupure optimale, et identifie les "Partition Graphs" comme l'une des bonnes structures de graphes pour modéliser la sélection de la partition a partir d'un famille hierarchique. Il conclut enfin par la formulation du problème de coupe optimal sur les hiérarchies, comme un problème de maximisation de flux, sur une structure d'arbre, le cas de tresses est également discuté.

**Mot Clé**: Segmentation hiérarchique, Multiplicateurs de Lagrange, Optimisation dans les treillis, Morphologie mathématique.

# Acknowledgments

The thesis has been an fruitful experience, thanks to my *Maître à penser*, Guru, Jean Serra, and of course, his semi-parental role in these fews years, away from my family. I will always be grateful for the kindness, he and Marie Françoise, have shown me. He has been a demonstrative and overwhelming influential force, not just in my discovery of mathematics, but with long, wide and deep discussions ranging from epistemology, propositional logic to interesting sociological phenomena. The personal and academic freedom I experienced were unparalleled. The variety of projects and ideas embarked upon together will always be a cherished memory, with my colleague and office-mate, in the very familiar, but discrete office 4357.

The thesis by itself is a life experience, aside its professional front. Here I would like to take the opportunity to appreciate the wisdom, friendship and kindness shared by my friend Senthil Kumar, thanks to whose motivation and I find myself pursuing my goal to perform academic research. His patience, kindness and friendship have helped me get past some tough times. In the same spirit, Norbert Bus, provided me with support and, of course exquisite Hungarian sarcasm, bordering usually on pithy wisdom. I would like to thank Frank Schmidt for his advice, support and friendship. Discussions with him not only improved my understanding of various mathematical problems, but also my presentation of the thesis. Working with him has been a great pleasure. I also thank Jesús Angulo, for his friendship, kindness, and his innovative suggestions and professional unbiased opinions on various subjects. I would like acknowledge Hugues Talbot and Yukiko Kenmochi for their kindness, friendship and timely help.

I am grateful to my referees Philippe Salembier and Jocelyn Chanussot for their constructive criticism of thesis, which has greatly helped in improving the quality of the final draft. I would also like to thank Jean Cousty for reviewing and commenting on parts of the manuscript. I would also like to thank Jesús Angulo and Santiago-Velasco for their corrections and detailed comments on the thesis.

I am also grateful to ESIEE and Université Paris-Est to have provided me with the PhD Grant and the opportunity to pursue an interesting research problem. The thesis was part of the KIDICO research program ANR-2010-BLAN-0205-0.

It was a pleasure, to work with Christine Voiron, and Pierre-Alain Mannonni, at UMR CNRS 7300 ESPACE, Sophia-Antipolis, on problems of geo-spatial and demographical analysis. This helped with appreicate the real world problems in GIS applications. The thesis was enriched by application problems motivated in this group.

The A3SI-lab in ESIEE, was an interesting work-place during the three years I spent there. Starting with the student camarades, old and new: Camille Couprie, Benjamin Rayanal, Laszlo Marak, John Chaussard, Nadine Dommanget, Anthony Giroud, Ngo Phuc, Christophe Guentleur, Olivia Miraucourt, Yongchao Xu, Rolland Levillain, Imen Melki, Ania Jezierska, Fiona Zolyniak, Hiroko Mitarai, Ali Kanj, Mathieu Lagarde, Kacper Pluta, Michel Revollo, who in various ways shown their kindness. Among professors and colleagues, I experienced a great work environment with: Michel Couprie, Gilles Bertrand, Laurent Najman, Jean Cousty, Benjamin Perret, Nabil Mustafa, Yukiko Kenmochi, Laurent Tsang, Vicent Nozick, Venceslas Biri, Denis Bureau. I would also like to thank my secretaries of the lab, Sylvie Cache, Martine Elichabe, Christine Auger for their help. I would also like to thank Joëlle Delers, and the team at Access-Paris Est in helping me with various administrative procedures and organizing the stay at the intenational residence.

And the occasional *pause café* with the signal processing lab girls, Feriel Abboud, Mireille El Gheche, Manar Qamhieh, Audrey Rappeti, Mai Quyen Pham, will remain memorable, amongst other antics. The vision group from Ecoles des Ponts, were also a source of vibrant discussions and *sorti*s with Matheusz Kozinki, Raghudeep Gadde, Sergey Zagoruyko, and Maria Vakalopoulou. Also not to forget my small but significant Indian friend circle during these years, with Ajith Sivadasan and Divij Babbar.

During my stay at International Residence, life has been fun thanks to my friends, Norbert, Phuc, Magda and Matheusz, Francesco Dolce, Sonja Hiltunen, Feriel Abboud, Sergiu Ivanov, Matija Hustić, Jacopo Corbetta, Jacopo Mancin, Pavel Heller, Kacper Pluta (thanks for the headphones!), Chi-Wei Chen, Katia Martemianova, as well as Aurelie and Pascal for their friendship and time spent together doing most importantly, nothing. I would also like to mention my *Fontainebleau connection* in the Center of Mathetmatical Morphology(CMM), with some good times with Andres Serna, Santiago Velasco, and Enguerrand Couka. I thank them for their advice, friendship and many a dinner together.

Finally I would like to end by thanking my parents Raja Ravi Sekhar and Kalpana for their support, valuable advice and patience.

# Contents

# Abbreviations

| | |
|---|---|
| **PP** | **P**artial **P**artitions |
| **HOP** | **H**ierarchy **O**f **P**artitions |
| **BOP** | **B**raid **O**f **P**artitions |
| **BFOS** | **B**reiman, **F**riedman, **O**lshen, **S**tone |
| **KKT** | **K**arush **K**uhn **T**ucker |
| **MST** | **M**inimum **S**panning **T**ree |
| **MSF** | **M**inimum **S**panning **F**orest |
| **MIS** | **M**axmimal **I**ndependent **S**set |
| **MWIS** | **M**axmimally **W**eighted **I**ndependent **S**set |
| **SLC** | **S**ingle **L**inkage **C**lustering |
| **CLC** | **C**omplete **L**inkage **C**lustering |
| **HAC** | **H**ierarchical **A**gglomerative **C**lustering |
| **UCM** | **U**ltrametric **C**ontour **M**ap |
| **GT** | **G**round **T**ruth |
| **LME** | **L**agrange **M**inimization by **E**nergy |
| **LMCC** | **L**agrange **M**inimization by **C**ut **C**onstraints |
| **CCM** | **C**lass **C**onstrained **M**inimization |

# Notations

| | | |
|---|---|---|
| $E$ | — | Image domain under study $\mathbb{R}^2$ or $\mathbb{Z}^2$ |
| $x, y$ | — | Points of $E$ |
| $\mathcal{P}(E)$ | — | Set of the all subsets of $E$ |
| $A, S, T_i$ | — | Classes of $E$ |
| $\pi$ | — | Partition of $E$, sometimes written as $\pi(E)$ |
| $\Pi$ or $\Pi(E)$ | — | Set of all partitions of $E$ |
| $\pi(S) : S \in \mathcal{P}(E)$ | — | Partial partition of $E$ of support $S$ |
| $\{S\}$ | — | Partial partition with unique class $S$ |
| $\mathcal{D}(E)$ | — | Set of all partial partitions of $E$ |
| $\sqcup$ | — | Disjoint union: $S = S_1 \sqcup S_2 \Leftrightarrow S_1 \cup S_2 = S$ and $S_1 \cap S_2 = \varnothing$ |
| $\sqcap$ | — | Disjoint intersection: $\pi \sqcap S := \pi(S)$, $\pi$ is a partition, $S \subseteq E$ |
| $H = \{\pi_i, i \in I\}$ | — | Hierarchy, i.e. family of increasing partitions |
| $B$ | — | Braid of partitions, monitor is characterize by partitions in $B$ |
| $\Pi(E, H)$ | — | Set of all cuts of Hierarchy $H$ that partitions $E$ |
| $\Pi(E, B)$ | — | Set of all cuts of Braid $B$ that partitions $E$ |
| $S_0(x)$ | — | Classes of the leaves partition $\pi_0(E)$ containing point $x$ |
| $S_i(x)$ | — | Classes of the partition $\pi_i(E)$ at level $i$, containing point $x$ |
| $\mathcal{S}$ | — | Family of all classes of $H$ or $B$ |
| Cut $\pi(S)$ | — | Partition of $E$ into classes taken in $\mathcal{S}$ |
| $\omega : \mathcal{D} \to \mathbb{R}^+$ | — | energy, i.e.non negative function on $\mathcal{D}(E)$ |
| $\preceq_\omega, \curlywedge_\omega, \curlyvee_\omega$ | — | energetic ordering, infimum, and supremum, w.r.t. energy $\omega$ |
| $\Pi(\omega, H)$ | — | $\omega$-energetic lattices on cuts of Hierarchy $H$ |
| $\Pi(\omega, B)$ | — | $\omega$-energetic lattices on cuts of Braid $B$ |
| $\pi^*$ | — | Minimal cut in an $\omega$-energetic lattice |
| $\lambda, \mu :$ | — | Scalar parameters (e.g. Lagrangian multipliers) |
| $\lambda, \mu :$ | — | vector of parameters (or multipliers) |
| $\pi^*(\lambda)$ | — | minimal cut in the energetic lattice of energy $\omega(\lambda)$ |
| $\pi^*_\varphi, \pi^*_\partial$ | — | minimal cut in the energetic lattice of energy $\omega_\varphi$, and $\omega_\partial$). |

# Publications

## Conferences

1. Jean Serra, B Ravi Kiran, Energetic Order for Optimization on Hierarchies of Partitions: Continuous hierarchy and Lagrange optimization, Internation Conference on Image Processing (**ICIP**) 2014, Paris, Special Session: Variational and Morphological Optimizations: A Tribute to Vicent Caselles.

2. Bangalore Ravi Kiran and Jean Serra. Scale space operators on hierarchies of segmentations. In Arjan Kuijper, Kristian Bredies, Thomas Pock, and Horst Bischof, editors, Scale space and Variational Methods (**SSVM**), volume 7893 of LNCS, pages 331–342. Springer, 2013. ISBN 978-3-642-38266-6. [60]

3. Bangalore Ravi Kiran and Jean Serra. Ground truth energies for hierarchies of segmentations. In Cris L.Luengo Hendriks, Gunilla Borgefors, and Robin Strand, editors, Mathematical Morphology and Its Applications to Signal and Image Processing (**ISMM**), volume 7883 of LNCS, pages 123–134. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-38293-2.[61]

4. Bangalore Ravi Kiran and Jean Serra. Optima on Hierarchies of Partitions, In Cris L.Luengo Hendriks, Gunilla Borgefors, and Robin Strand, editors, Mathematical Morphology and Its Applications to Signal and Image Processing (**ISMM**), volume 7883 of LNCS, pages 147-158. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-38293-2.[103]

5. Bangalore Ravi Kiran, Jean Serra, and Jean Cousty. Climbing: A unified approach for global constraints on hierarchical segmentation. In Andrea Fusiello, Vittorio Murino, and Rita Cucchiara, editors, ECCV 2012. Workshops and Demonstrations, HiPot: **ECCV 2012 Workshop** on Higher-Order Models and Global Constraints in Computer Vision, volume 7585 of LNCS, pages 324–334. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33884-7. [62]

6. Jean Serra, Bangalore Ravi Kiran, and Jean Cousty. Hierarchies and climbing energies. In Luis Alvarez, Marta Mejail, Luis Gomez, and Julio Jacobo, editors, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications (**CIARP**), volume 7441 of LNCS, pages 821–828. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33274-6. [104]

## Journals

7. B Ravi Kiran, Jean Serra, Fusions of Ground Truths and of Hierarchies of Segmentations, Volume 47, 1 October 2014, Pages 63–71, Pattern Recognition Letters (**PRL**) 2014. [58]

8. B Ravi Kiran, Jean Serra, Global-local optimizations by hierarchical cuts and climbing energies, Pattern Recognition(**PR**), Volume 47, Issue 1, Jan 2014, Pages 12–24. ISSN 0031-3203. [59]

## Articles in Preparation/Submission

9. Optimization on Braids of Partitions, B Ravi Kiran, Jean Serra (In preparation).

10. Energetic Lattices, B Ravi Kiran, Jean Serra (In Preparation).

# Chapter 0

# Motivation and Thesis Overview

Hierarchical Clustering has been used extract clusters from data which are assumed to have an underlying tree structure. This has been evident in the field of gene expression array analsysis, phylogenetics, phylolinguistics. It has also been in the domain of image segmentation, with perceptual organization trees [70] where one the segmentation problem was motivated by a inclusion tree of regions/objects.

The basic utlities of hierarchical clustering and the following segmentation/classification is to represent images, summarize non-hierarchical data like images into hierarchical classes for various tasks like objection detection recognition, recover underlying structures of the data, for example in texture images.

Hierarchical clustering is an classical method to group set of datapoints in nested classes by introducing a metric between pairs of points each time. The grouping occurs iteratively with the recalculation of pairwise distances.

There are two approaches to hierarchical clustering: we can go from the "bottom up"(Agglomerative), grouping small clusters into larger ones, or from the "top down" (Bifurcative), splitting big clusters into small ones. These are called *agglomerative* and *divisive* clusterings, respectively. We consider Agglomerative clustering since the bottom-up approach bounds the complexity to polynomial at worst, while divisive clustering requires a combinatorial search of every possible division of cluster, leading to exponential complexity.

Hierarchical segmentations methods have been studied to be able to parameterize the number of segments, and later chose an optimal scale of partitions for the purpose of segmentation. The methods in literature study split-merge heuristics, the feature dissimilarities, to produce segmentations. This thesis will study how to introduce optimality

when the solution space is the hierarchy of partitions, while using its inherent lattice structure, and avoid the combinatorial nature of the problem. [1].

## 0.1 Motivation

This thesis studies the optimality on constrained and unconstrained problems on hierarchies of partitions. The thesis continues on the lines of the work by Salembier et al., Guiges et al., Casselles et al., [13, 47, 100], mainly. They pose the problem of constrained minimization of an energy on the space of the hierarchy of partitions. The goal in these studies and also one of the goals of the thesis is to define conditions on the energy that obtains: unique solutions and in case of parmetrized energy, conditions for achieving refinement ordered solutions with increasing parameters. In said studies, parametrized energies are of the Lagrangian form $\omega_\varphi(\pi) + \lambda \cdot \omega_\delta(\pi S)$, where $\omega_\varphi(\pi), \omega_\delta(\pi)$ represent a objective and a constraint term on partitions $\pi(E, H)$ in the hierarchy.

There are many methods to obtain a hierarchy of segmentations, and this thesis assumes its construction to be an independent step. We will thus consider the hierarchies themselves to be the input of our optimization problem.

### 0.1.1 Convex vs. Lattice based optimization

Most of conventional mathematical optimization depends on conditions ascertaining local optimality that implies global optimality. Consequently the theory has evolved into subject of study of convex sets. In contrast, the thesis uses lattice-theoretic approach and follows in the lines of lattice programming methods such as [11, 112] which are concerned with the *order of optimal solutions* and so are led to a development based on lattices. The important difference between convex analysis and lattice programming methods are seen when one removes one of the following restrictions: differentiability, convexity(or concavity), continuity, local analysis, and adds the following qualities: using order-based properties. This has been demonstrated in the area of pricing in economical applications by the Topkis Theorem [112], and in the area of sub-modular optimization [12].

As an example if one considers integer constraints, e.g., one must order in *multiples* of a given batch size like a case, a box, etc. This destroys convexity, but preserves sublattices. Thus the presence of integral constraints enormously complicates the results

[1] A small square image of 5 by 5 pixels can partitioned in $10^{18}$ of different manners, which is the Bell's Number [15] to count the number of different partition of 25 pixels. This number grows exponentially large with the number of elements.

of convex programming. By contrast, the monotonicity results of lattice programming carry over without change to their integer counterparts [11].

### 0.1.2 A Morphological Approach

In earlier work which performed constrained optimization over the HOP in, Guigues et al. [47], the optima were ordered based on the choice of the scale parameter $\lambda$. This that was left open as choice, thus calculating all the ordered set of partitions corresponding for each feasible value of multiplier $\lambda$. [42, 100] provided a rate-distortion interpretation to the optimization problem following from the work on tree structured vector quantizers [43], and calculate the $\lambda$ that got closest to the constraint.

These methods do not use the inherent lattice structure. On a HOP with partition dependent functions, like perimeter, there is no guarantees of continuity and topological constraints of these functions. For example, one cannot describe the convex composition of partitions, and must accept its discrete combinatorial nature. Furthermore we will see how even when the perimeter is rational or even real valued there might be instances when we cannot achieve the desired constraint value.

The area of Mathematical Morphology [101] is well known for its use of lattice structure in image processing. Transformations, like the dilation, erosion, in morphology are defined on complete lattices which are partially ordered sets, containing a unique infimum and unique supremum. Morphological operators on such lattices frequently lead to scales spaces, these are classical examples of non-linear scale-spaces. In this thesis we will introduce a numerical lattice structure on *partial partitions* from the HOP. The Lagrangian approach to the problem, in the thesis is substituted by a generalized energy based ordering of partial partitions and a subsequent lattice to define the optima. While operating on different energetic lattices corresponding to the objective, constraint and Lagrangian, it generalizes Guigue et al's [47] method to obtain monotonically ordered optimal cuts, for a constrained optimization problem on the HOP. Further the dynamic program first used in classification and regression trees, by Breiman et al. [21] and later generalized for the HOP in Selmbier-Garrido & Guigues [47, 100], are further generalized in this thesis for non-linear energies.

Further the hierarchy of partitions were a result of a multi-scale segmentation step that produced either nested or disjoint classes. This is usually a result of a greedy aggregative step from the application of HAC algorithms. We present in the thesis a more relaxed and richer structure, which is still hierarchical, while allowing non-nested partitions. This is the Braids of partitions, and the advantage is that we have a larger structured

space of solutions where the energetic lattice structure works and hope for a better infimum on the energy.

## 0.2 Thesis Organization

The organization of the thesis is as follows. While one can find the contributions summarized in the conclusion section.

**Chapter 1**

**Definitions**: To begin with, this chapter provides the basic definitions including the partial partition, hierarchies of partitions, energies, classes, and a basic recall of lattices.

**Review on Optimization on Hierarchies**: Following this a review of the constrained optimization problem hierarchies is presented, with a recapitulation of the basic problems.

**Braids and Energetic Lattices**: This is a crucial section and contribution of this chapter introduces the new structure of Braids of partitions(BOP). And also the key definitions of energetic ordering, and energetic lattices, followed by properties of multi-scale energies, namely $h$-increasingness, scale-increasingness and singularity. Examples of different compositions and their optimal cuts are demonstrated with simple examples.

**Chpater 2**

**Rate-distortion Interpretation**: This bibliographic section 2.1 recapitulates the Rate-Distortion theory and its applications in tree based source coding problems. Following which, we lead to the work of Guigues & Salembier-Garrido, [47, 100].

**Lagrangian Relaxation**: Firstly we discuss Lagrangian of the constrained optimization problem on the HOP, already established in [47, 100]. We establish how this is a Lagrangian relaxation, and give important implications with respect to the Lagrangian dual problem. Further we demonstrate how one can use penalty based methods to obtain a better upper bound on the minimization.

**Three models**: We provide three models for the constrained problem on a general Braid of partitions (and thus HOP as well), namely the Lagrangian relaxation using the energetic lattice, A partition or Cut constraint based Lagrangian relaxation, and finally a class based local constraint model.

**Chapter 3**

**Experiments and Demonstrations**: The first part, demonstrates the applications of the energetic lattice, in the problems of, calculating the optimal cuts for different image segmentation models, mainly Mumford-Shah. We introduce here the problem of extraction of partition from a hierarchy that is proximal to a ground truth partition, and its resolution with the energetic lattice based on local Hausdorff distances. We introduce further evaluation measures for hierarchies of segmentations w.r.t ground truth partition.

**Chapter 4** studies the saliency function and how it can be defined by considering the numerical lattice over a finite family of Jordan curves partitioning the Euclidean plane. This is an extension from Chapter 3, where the energetic lattice is no more defined on the partial partitions, but purely on contours of an initial fine partition. Demonstrations include, transforming a hierarchy by performing openings on the lattice of Jordan curves, creating a braid of partition by operating on intersection of Jordan nets.

**Chapter 5**

**Algorithms**: This section compiles the important algorithms, including the dynamic program on hierarchies by Guigues, the generation of braids, and finally the extension of the dynamic program to the braids.

**Partition Graphs**: This section reviews maximal independent sets on intersection graphs for a family partitions and the formulation of image segmentation problem as the calculation of a Maximally weighted independent set of the intersection graph first proposed by Brendel-Todorovic [22]. Further we recall the definition of a *partition graph* which is an intersection graph over a family of subsets, that form a covering of the image domain $E$, where the maximal independent sets of this graph, are partitions of $E$. This basically is mapping between the MIS set and the partitions or cuts from the hierarchy. This correspond to a address partitions from the covering, and helps repair few pathological cases in case of Todorovic et al. [22]. The chapter ends with a Max-flow formulation on trees, which reformulates the optimal cut problem. The case of braids of partitions are studied for both the Partition Graph and Max-flow formulation.

**Chapter 6** provides a conclusion to the thesis. It discusses the main contributions of the thesis and its future perspectives.

# Chapter 1

# Braids and Energetic Lattices

**Publications Associated with Chapter**

- [59] Global-local optimizations by hierarchical cuts and climbing energies, Pattern Recognition(PR) 2014.
- [103] Optima on Hierarchies of Partitions ISMM 2013
- [62] Climbing: A unified approach for global constraints on hierarchical segmentation, ECCV 2012 Workshop.
- [104] Hierarchies and climbing energies, CIARP 2012

After a brief reminder on lattices, partitions and hierarchies, this chapter develops theoretical tools necessary for the rest of the thesis, namely the braid of partitions and energetic lattice. Though several notions are defined to analyze constrained optimization, we will actually deal with the problem only in the second chapter.

## 1.1 Basic notions and Notations

This section presents the basic constructs of the hierarchies of partitions (HOP). The space $E$ under study is arbitrary, continuous or discrete, finite or not, topological or not. $\mathcal{P}(E)$ represents the power set of set $E$. The elements $x, y$ of $E$ are called points, or leaves.

### 1.1.1 Partitions and Partial Partitions

**Definition 1.1.** *A partition $\pi$ of the image domain $E$ is a family of sets $S$:*

$$\pi = \{S \subseteq E\} \tag{1.1}$$

*where $S : E \to \mathcal{P}(E)$, and for each point $x \in E$, we have $x \in S(x)$, and*

$$x, y \in E \quad \Rightarrow \quad S(x) = S(y) \quad or \quad S(x) \cap S(y) = \varnothing \tag{1.2}$$

*These $S$ are called the classes of the partition $\pi$.*

Intuitively, a partition is a division of image domain $E$ into classes which are pair-wise disjoint and whose union restores $E$ in its entirety. We also refer to the partition as $\pi(E)$ in some cases.

First introduced by Ronse in [97], a partial partition is a *local* partitioning of a subset $S \subseteq E$ of the input space.

**Definition 1.2.** *A partial partition $\pi(S)$ of support $S \in \mathcal{P}(E)$ is a set,*

$$\pi(S) = \{A_i | A_i \subseteq S, A_i \cap A_j = \emptyset\} \tag{1.3}$$

*where $S = \cup A_i$, is called the support of partial partition $\pi(S)$.*

An example is demonstrated in figure 1.1. The partial partition of $S$ into the single class $S$ is denoted by $\{S\}$.

The set of all partitions $\pi$ of $E$ forms a complete lattice $\Pi(E)$ for the partial ordering of the refinement, where $\pi_i \leq \pi_j$ when each class $S_i(x)$ of $\pi_i$ is included in the class $S_j(x)$ of $\pi_j$ at the same point $x \in E$:

$$\pi_i \leq \pi_j \quad \Leftrightarrow \quad S_i(x) \subseteq S_j(x). \tag{1.4}$$

This refinement lattice is denoted by $\Pi(E)$. The refinement infimum of a family $\{\pi_i, i \in I \subseteq \mathbb{R}\}$ in $\Pi(E)$ is the partition $\pi$ whose class at point $x$ is $\cap S_i(x)$, and the refinement supremum is the finest partition $\pi'$ such that $S_i(x) \subseteq S'(x)$ for all $i \in I$ and $x \in E$. This is demonstrated in figure 1.1.

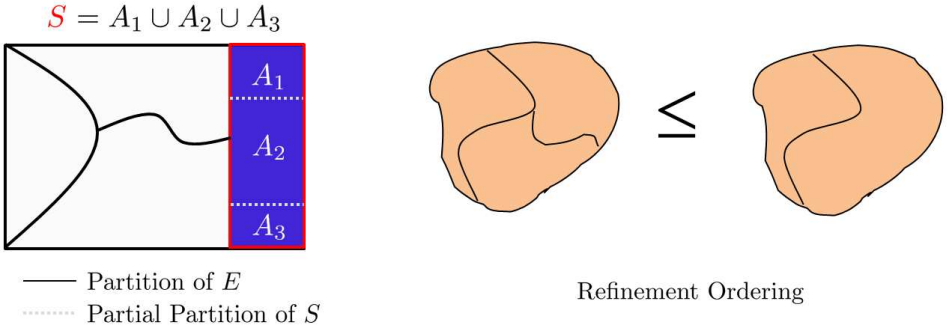



FIGURE 1.1: Left: Partial partition with support $S$ highlighted in red, with local partitioning shown in dotted lines. Right: Partial partition refinement ordering.

### 1.1.2 Hierarchy

Hierarchies of partitions are the matter of an abundant literature (see for example [10], [90], [78]). The definition that we propose here is based on two axioms:

**Definition 1.3.** *(Hierarchy of Partitions(HOP)) A family $\{\pi_i, i \in I \subseteq \overline{\mathbb{Z}}\}$ of partitions of $E$ defines a hierarchy when,*

*(i) The partitions $\pi_i$ are nested, i.e. they form a chain for the refinement ordering:*

$$H = \{\pi_i, i \in I\} \quad with \quad i \leq k \;\Rightarrow\; \pi_i \leq \pi_k, \qquad I \subseteq \overline{\mathbb{Z}}, \tag{1.5}$$

*where the finest partition $\pi_0$ is called the leaves, and the coarsest one, is the root;*

*(ii) The number of leaves is finite in any class of the hierarchy, except possibly, in the class $\{E\}$.*

One often takes the whole space $\{E\}$ for the root. A toy example is given by the following hierarchy of nested partitions of $\overline{\mathbb{Z}}$ where the central class enlarges:

$$\begin{aligned}
&i = 0 \quad \pi_0 = \text{ all points of } \overline{\mathbb{Z}}\\
&i = 1 \quad \pi_1 = \{-\infty\}...\{-3\};\{-2\};[-1,+1];\{+2\};\{+3\}...\{+\infty\}.\\
&i = 2 \quad \pi_2 = \{-\infty\}...\{-4\};\{-3\};[-2,+2];\{+3\};\{+4\}...\{+\infty\}.\\
&i = 3 \quad \pi_3 = \{-\infty\}...\{-5\};\{-4\};[-3,+3];\{+4\};\{+5\}...\{+\infty\}\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots
\end{aligned}$$

Though $|I| = \infty$, the number of leaves at any class $S_i(x)$, $x \in \mathbb{Z}$ remains finite as soon as the label $i < \infty$. Similar situations also occur in the Euclidean spaces. One can think of hierarchies where the leaves are Voronoi polygons, or polyhedra, or more generally of stationary random partitions.

In earth sciences, most of the phenomena are not studied in a finite domain. For example Air-borne and satellites images are of this type. Optimal segmentation of such phenomena must be reached by local, or regional information, and not via a global energy, which would involve the whole space, which needs intricate treatment. Here Axiom *(ii)*, opens the door to a regional approach, a door which leads to the energetic lattices.

### 1.1.3 Classes

A hierarchy can be described from its classes, or nodes. At each point $x \in E$ the family of all classes $S_i(x)$ containing $x$ forms a closed chain of nested elements in $\mathcal{P}(E)$, from the leave $S_0(x)$ to $E$. This chain is called *the cone at point* $x$. Let $\mathcal{S} = \{S_i(x), x \in E, i \in I\}$ be the family of all classes of $H$. One directly extends to $\mathcal{S}$ the characterization (1.2) of a partition by its classes containing any points $x, y \in E$,

$$i \leq j \Rightarrow S_i(x) \subseteq S_j(y), \text{ or } \; S_i(x) \supseteq S_j(y), \text{ or } S_i(x) \cap S_j(y) = \varnothing. \tag{1.6}$$

The classes of the partition $\pi_{i-1}$ included in the class $S_i$ of the partition $\pi_i$ are *the sons* of $S_i$. The symbol $\sqcup$ refers to the disjoint union of classes, i.e.

$$S = S_1 \sqcup S_2 \Leftrightarrow S_1 \cup S_2 = S \text{ and } S_1 \cap S_2 = \varnothing.$$

A *cut* of $H$ is a partition of the space $E$ into classes taken in $\mathcal{S}$. The symbol $\Pi(E, H)$ stands for the set of all cuts of $H$. Clearly, $\Pi(E, H)$ is a sub-lattice of $\Pi(E)$, the lattice of all partitions of $E$. If $S \in \mathcal{S}(H)$, then $\Pi(S, H)$ denotes the family of all partial partitions of $S$ whose classes are in $\mathcal{S}(H)$.

### 1.1.4 Energy

An *energy* $\omega$ is a real valued function over the family of partial partitions $\mathcal{D}(E)$ of space $E$:

$$\omega : \mathcal{D}(E) \to \mathbb{R} \tag{1.7}$$

When the energy $\omega$ of a p.p. $a$ is the sum of the energies of its classes, then $\omega$ is *linear*, or *separable* in terms of Guigues [49, 90, 100], and can be written now using the general definition from equation 1.7 as:

$$\omega(\pi(S)) = \sum_{A_i \in \pi(S)} \omega(A_i) \tag{1.8}$$

Furthermore in equation 1.8 when $\omega(\pi(S)) \geq \omega(S)$, i.e. the sum of energies of classes in the p.p of support $S$ is greater than or equal to the energy on the single class $\{S\}$, then $\omega$ is *sub-additive*. Both assumptions of linearity and sub-additivity will be generalized in our theory, already visible in the definition of the energy in equation (1.7). One can find work on the valuation of partial partitions by Ronse in [98, 99].

### 1.1.5 Recall on Orderings and lattices

A set is (partially) ordered when a binary relation $\leq$ is defined on its elements, with $a \leq a$ (reflexivity), $a \leq b$ and $b \leq a \Rightarrow a = b$ (anti-symmetry), and $a \leq b$ , $b \leq c \Rightarrow a \leq c$ (transitivity). The ordering is total when all pairs of elements of $E$ are ordered. The sets of $\mathbb{Z}^2$ are ordered for the inclusion, the real numbers are totally ordered for the numerical inequality.

A lattice is an ordered set $L$ in which any two elements have a greatest lower bound (g.l.b.) and a least upper bound (l.u.b.). They are denoted by $\wedge$ and $\vee$ respectively. When this property extends to any family of elements, possibly infinite, the lattice is said to be complete. A finite lattice is thus always complete. A lattice contains always two extreme, or universal, elements, namely the least and greatest of all. A basic example of a lattice is provided by the elements $\mathcal{P}(E)$ of an arbitrary $E$. They are ordered by inclusion, the two bounds are the intersection $\cap$ and the union $\cup$, and the universal elements are $(\varnothing, E\}$. An ordering relation does not systematically produce a lattice, for example when a connection is defined on $E$, the connected sets of $E$ do not form a lattice, unlike $\mathcal{P}(E)$, though they are ordered by inclusion.

A sub-lattice of $L$ is a family $L' \subseteq L$ closed under $\wedge$ and $\vee$ (a notion sometimes called pseudo sub-lattice in literature). For example, the family of all sets contained in a zone $Z \subseteq \mathbb{R}^2$ is a sub lattice of $\mathcal{P}(\mathbb{R}^2)$. Finally, a totally ordered sub-lattice is called a *chain*.

The lattice is flexible, because a same family of mathematical objects may be the matter of several different lattices. For example, the braid cuts will be provided with two lattices in this chapter, and with four lattices for constrained optimization, considered in chapter 2. Furthermore the lattice structure assures us that any family has a minimal and a maximal elements, both unique, and which belong to the lattice. A set $\Pi$ of partitions of $E$ (e.g. the braid cuts) cannot be significantly modeled by a topological vector space, so that one cannot speak of the continuity, or the zero gradient, or the convexity, of a real valued function over $\Pi$. Minimization questions must be addressed in another manner, what precisely does the lattice approach.

## 1.2 Optimization on Hierarchies of Partitions

In this section we review the problem of optimization on HOP in literature.

### 1.2.1 Classification and Regression Trees

Classification and Regression Trees (CART) were introduced in the 80's by Breiman et al [21], which creates powerful and simple binary tree based models for classification and regression problems, in statistical learning theory [113]. In both cases, the method consisted in creating rectangular partition of a feature space (high dimensional $\mathbb{R}^n$), either fit a model over each of these rectangles in case of regression, produce a classification. These trees (now called decision trees) described then the estimator for the regression function, or a linear separator for classification tasks. The trees we consider here only address recursive binary partitions as show in in the figure 1.2.

To avoid over-fitting the data, such trees are dealt with commonly in two broad ways: Firstly, prune them (which is what we will consider for the rest of our study). Secondly one can use *aggregations* of different subset of points to average over many such trees, which also includes the classical Random Forests which consists in using a random subspace of features to partition a node and *aggregate* them.

#### 1.2.1.1 Pruning CART trees

First the tree is grown until a minimal number of samples per class is reached. The growth is done by splitting pre-existing classes such that at each step the split point picked minimizes the quadratic deviation from the mean value, in each of the split produced. This is the case for regression, while for classification we have the node impurity and entropy based measures [37].

#### 1.2.1.2 Optimal tree pruning

Deciding the tree size is important parameter since this describes the model's complexity as well the solution space being spanned. Breiman et al. proposes to grow a initial large tree $T_0$, until a minimal number of samples per rectangular class is reached (a minimum number of points of population), following which a pruning is performed to reduce the complexity of the tree for the purpose of classification or fitting a regression model for function.

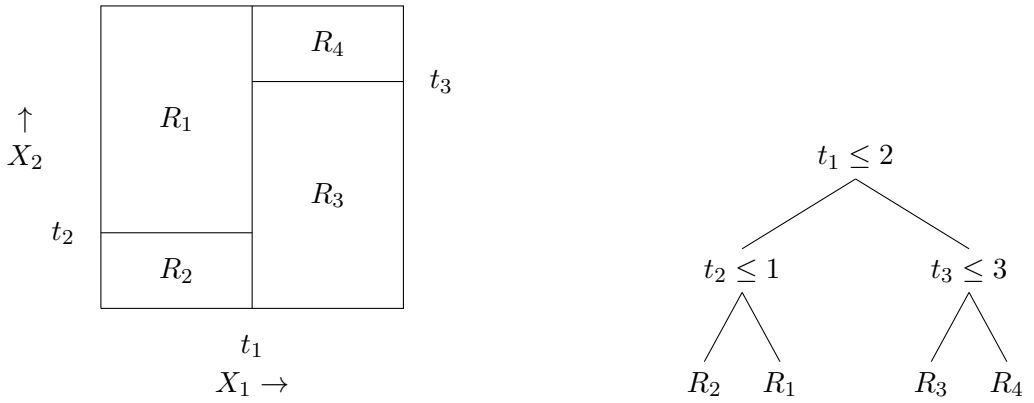



FIGURE 1.2: A two dimensional feature space is recursively partitioned using a binary tree where the partitions consist of purely rectangles. The corresponding binary tree is shown. One continues growing the tree until each rectangular encloses a minimal number sample points.

**Tree Terminology**: Considering Breiman et al.'s [21] notation for the trees just for this section, we denote by $T_0$ the complete non-trivial binary tree that is grown by recursive binary partitioning of the feature space. $T_1$ is said to be a subtree of $T_0$ of the triple $T_1$, $\text{left}_1(\cdot), \text{right}_1(\cdot)$ forms a tree. If $T_1$ is a subtree of $T$ and $T_2$ a subtree of $T_1$ then $T_2$ is subtree of $T_0$. Given any node $t \in T_0$ , the subset $T_t$ consisting of $t$ and all its descendants is called the branch of $t$ stemming from $t$. This is a subtree of $t$. Let $\tilde{T}$ denote the terminal nodes of the tree. A tree $T$ is trivial if the cardinality of set is empty, i.e. $|T| = 1$, or also when the set $T - \tilde{T}$ is empty. A subtree $T_1$ of $T_0$ is called a *pruned* subtree of $T_0$ if $\text{root}(T_1) = \text{root}(T_0)$; this is denoted by $T_1 \preceq T_0$. $\preceq$ is transitive, and so is $\prec$. This provides a top-down definition of a pruned subtree, while Breiman also provides a bottom up definition.

Any arbitrary subset $T_1$ of $T_0$ is a subtree of $T_0$ having a root $t$, if an only if it is a pruned subtree of $T_0$ rooted at $t$, iff, it is a pruned subtree of $T_t$. Given now a real number $\lambda$ ($\alpha$ in Breimans notation), lets define, $R_\lambda(t) = R(t) + \lambda$ for $t \in T_0$. Given a subtree $T$ of $T_0$, let, $R(T) = \sum_{\tilde{T}} R(t)$ and

$$R_\lambda(T) = \sum_{\tilde{T}} R_\lambda(t) + \lambda \left|\tilde{T}\right|.$$

**Cost complexity Pruning**: If $T \preceq T_0$ be a subtree obtained by pruning $T_0$ (Pruning a branch $T_t$ from a tree $T$ consists of deleting from $T$ all descendants of $t$), and $R_m$ represents the $m$ different regions representing the terminal nodes of tree $T$.

The optimal pruned subtree is one which minimizes:

$$R_\lambda(T) = \sum_{m=1}^{|T|} \sum_{x_i \in R_m} (y_i - \mu_{R_m})^2 + \lambda \left|\tilde{T}\right|. \tag{1.9}$$

over each node which results in an optimal pruned subtree $T_\lambda \preceq T_0$ for a given $\lambda$. Following [113], here $\mu_{R_m}$ is the mean value of observed variable $y$ in the region $R_m$. $\lambda$ is a parameter that governs the trade-off between tree size and fidelity to data.

The expected training error is non-monotonic with the subtrees of decreasing size. This shows that as the trees initially decrease in size, the error rate decreases. The error reaches a minimum at a particular subtree size and begins to climb again as the trees get too small. This behavior can be attributed to a trade-off between bias and variance.

The optimal $\hat{\lambda}$ is adaptively determined by cross-validation. For each $\lambda$ the *weakest link pruning*, i.e. successively collapse each internal node, that produces the smallest per node increase in the in an optimal pruned subtree's quadratic term in 1.9, is performed and continued till the root. $\hat{\lambda}$ minimizes the cross-validated sum of squares.

#### 1.2.1.3 Uniqueness and Monotonicity

In defining a Cost-Complexity measure that is a linear combination of misclassification error and tree size, Breiman et al. [21] analyses the following questions:

- **Uniqueness** Is there a unique subtree $T \prec T_0$ which minimizes $R_\lambda(T)$?
- **Monotonicity** In the minimizing sequence $T_1, T_2, ...$ is each subtree gotten by upward pruning from the previous subtree, i.e. does the following nesting relation hold: $T_1 \succeq T_2, ..., \succeq T_0$.

**Optimally pruned subtree**: A pruned subtree $T_1$ of $T_0$, is said to be one of the optimally pruned subtree of $T_0$, with respect to $\lambda$ if:

$$R_\lambda(T_1) = \min_{T' \prec T_0} R_\lambda(T') \tag{1.10}$$

Since there are only finitely many pruned subtrees of $T$, there is clearly an optimal one, but not necessarily a unique one. An optimally pruned subtree $T^*$ is said to be the *smallest optimally* pruned subtree of $T_0$ if $T^* \succeq T'$ for every optimally pruned subtree $T'$ of $T$. There is clearly at most one smallest optimally pruned subtree of $T$ (with respect to $\lambda$); when it exists, it is denoted by $T(\lambda)$.

To avoid the combinatorial problem of choosing an optimal subtree from the exponentially growing number of choices, Breiman et al. suggest the following optimality conditions. We now follow Breiman et al.'s book [21], in chapter 10, Theorem 10.9, and state the following:

**Proposition 1.4.** *(Breiman's Conditions) The smallest minimizing subtree $T(\lambda)$ for complexity parameter $\lambda$ is defined by conditions:*

- $R_\lambda(T(\lambda)) = \min_{T \prec T_0} R_\lambda(T)$,
- *If* $R_\lambda(T) = R_\lambda(T(\lambda))$, *then* $T(\lambda) \leq T$

Breiman et al. states the condition of uniqueness though implicitly. It resolves the choice of minimal cost-complexity across many optimal subtrees, possibly unordered, by picking smallest(sized) subtree, minimizer of $C_\lambda$. This ensured uniqueness for Breiman et al. as well as monotonicity.

In proposition 3.8 Chapter 3 of [21] Breiman shows that for any non-terminal node $t$ of $T$, we have

$$R(t) > R(T_t) \tag{1.11}$$

where $R$ can be for example the error term in 1.9.

This gives us the core idea of the cost-complexity pruning, in that it starts with a pruned subtree, eliminates the weakest sub-branch, to produce the next optimal pruned subtree. This is called *weakest link cutting*.

For a singleton node $\{t\}$, where $t \in T_0$ let consider,

$$R_\lambda(\{t\}) = R(t) + \lambda(1).$$

For any branch $T_t$ rooted in $t$, lets define

$$R_\lambda(T_t) = R(T_t) + \lambda \left|\tilde{T}_t\right|.$$

The value of $\lambda$ at which $T_t$ (child) becomes equally optimal in complexity w.r.t the subtree root node (parent) $t$, we solve the inequality $R_\lambda(T_t) < R_\lambda(\{t\})$, giving us

$$\lambda < \frac{R(t) - R(T_t)}{|T_t| - 1} \tag{1.12}$$

Breiman et al. provides conditions of monotonicity, where for increasing $\lambda$, optimal pruned subtrees monotonically reduce.

Breiman also defines a mapping that spans all nodes and associating the critical parameter value in equation 1.12.

$$\Lambda_1(t) = \begin{cases} \frac{R(t)-R(T_t)}{|\tilde{T}_t|-1}, & t \notin \tilde{T}_1 \\ +\infty, & t \in \tilde{T}_1 \end{cases} \tag{1.13}$$

The weakest link $t_1$ in $T_1$ is a node such that

$$\Lambda_1(t_1) = \min \Lambda(t), t \in T_1$$

and for the next new complexity parameter value put

$$\lambda_2 = \Lambda_1(t_1)$$

After this step we define a new tree

$$T_2 = T_1 - (T_{t_1})$$

We continue performing a the weakest link cutting on this new tree, by recalculating a new function $\Lambda_2$. This done until we reach the root. This yields a monotonically ordered set of optimally pruned subtrees, $T_1 > T_2 > T_3 > ... > t_1$.

#### 1.2.1.4 Departing from Binary Trees

One can observe that the conditions for optimal pruning are not restricted to binary trees. Guigues [47] provides a generalization for the weakest link pruning on a general hierarchy of partitions. Though it is also notable that any general tree can be rewritten as an equivalent binary tree, and such reorganization of tree topology don't actually change the optimal cut, i.e. the leaves set of the optimally pruned subtree.

We shall demonstrate this with a quick numerical example in figure 1.3. As seen in this figure, one can consider any general tree with a monotonically increasing function $R$ in equation 1.11 on the nodes of the tree in the place of the variance term, while just adding a complexity cost of 1. This penalizes subtrees which have too many leaves. We can see

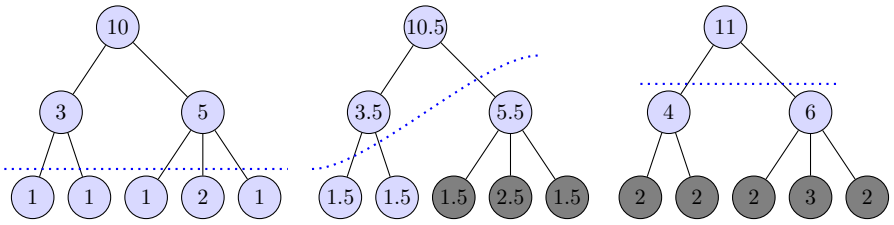



FIGURE 1.3: Pruning example demonstrating Cost-Complexity pruning. Figure demonstrates a tree with classification the cost function given for each node given by equation (1.9) where values for different values of $\lambda = 0$(left), 0.5(center), 1(right), are shown with their corresponding pruned optimal subtrees. The pruned nodes are presented in gray. One can also see the representation of the the *cuts* in blue. They represent the leaves $\left|\tilde{T}\right|$ of the optimally pruned subtree. These cuts represent not only the pruned subtree, but also the partition of the space. It is sufficient to refer to the cut and not the whole subtree.

that the *cut* progressively moves up the tree to fewer number of classes as the value of $\lambda$ increases. This is a very simple version of the constrained optimization problem on the tree. From now on we will simply refer to the optimal subtree as the optimal cut, both of which refer to the same result as demonstrated.

### 1.2.2 Salembier-Garido's Optimal Pruning & Guigue's Scale-set

Salembier-Garrido and Guigues [47, 100] generalized the CART framework for the constrained optimization problem. Salembier-Garrido study binary partition trees, which represent a hierarchy of partitions created by using the max-tree representation, while Guigues considers a hierarchy created from complete linkage on regions of an over-segmentation [48], titled as Cocoons. In both studies the *cost-complexity* pruning, a greedy strategy is used to find a constrained minimum, while replacing the constraint of the size of the subtree $\left|\hat{T}\right|$ by a more general function which is the perimeter of the partition. See Equation (1.14).

One must note now that the difference in these methods w.r.t CART is the interpretation of the constrained minimum. Salembier-Garido and Guigues provide a Lagrangian multiplier interpretation of the optimization [39]. This is discussed later in section 2.1. Constraints are now based on the number of regions and length of perimeter, thus providing a larger range of constraint values, while not requiring a large initial tree to start pruning like in the case of CART. Salembier-Garrido [100] provided a Rate-Distortion interpretation of the constraint problem of choosing a segmentation from a hierarchy with the constrained perimeter.

In Guigues thesis [47] he maps the family of cuts in an input hierarchy $H$[1] to an ordered set of cuts forming a new hierarchy. He views this step as a precursor to image segmentation, where the all choices of scale parameter values $\lambda$ are indexes of the new hierarchy of segmentations $H_\lambda$. The optimal $\lambda$ for image segmentation is not really considered. He refers to the monotonicity in $\lambda$, as *causality* using terminology of scale-spaces, since the scale-sets are image descriptors defined by a hierarchy, energy and a numerical function generating the energy.

Furthermore Guigues provides conditions of monotonicity akin to Breiman et al, by describing multi-scale energies on the nodes in the hierarchy of partitions. The constraint function (the number of leaves in a sub-branch in case of CART) is defined to be sub-additive[2], i.e. $f(A \cup B) \leq f(A) + f(B)$. While the objective function(variance in the case of CART) was defined to be super-additive, i.e. $f(A \cup B) \geq f(A) + f(B)$. One must note that both conditions of super-additivity and sub-additivity are sufficient for monotonicity, but not uniqueness. As we have already seen there can be many optimal cuts, but there can be only one smallest/largest (in terms of refinement) optimal cut. This is how uniqueness is assured. Please refer to the summary in table 2.1 for the different conditions on uniqueness, monotonicity and constraint sense.

We interject here to note that the pruning of trees has been a rich area of study, with recent theses on using them in a variety of problems. In the domain of morphological filtering Yongchao Xu's thesis [123], studies pruning strategies on *attribute*(of gray scale components) trees, instead of directly on the max-tree. One can also cite the optimal pruning on Binary Partition Tree(BPT), for region based classification map generation on hyper-spectral images, by Valero et al. [119] as well as in Valero's thesis [118].

### 1.2.3 Problem review on hierarchies

Now we are in a position to formally state the problem of constrained optimization on the HOP as formulated by Guigues [47]. This generalizes of the complexity of the model in CART as formalized in equation (1.9).

$$\begin{aligned} &\underset{\pi \in \Pi(E,H)}{\text{minimize}} && \sum_{S \in \pi} \omega_\varphi(S) \\ &\text{subject to} && \sum_{S \in \pi} \omega_\partial(S) \leq C \end{aligned} \tag{1.14}$$

[1] and thus the family of subtrees $T \subseteq T_0$ in terms of CART

[2] which in case of Breiman is implicitly set by the size of a subtree constraint $\left|\widetilde{T}\right|$.

where the objective and constraint functions $\omega_\varphi(S), \omega_\partial(S)$ is defined over every partial partition following the separability condition[3]. The problem above is solved using a dynamic program which follows from the CART weakest link pruning algorithm. The value $C$ is a constant on bounding the constraint function. Here $\omega_\varphi(S)$ is identified by Guigues as the fidelity or data term, while, $\omega_\partial(S)$ is identified as the model complexity term. Further he states, that for two models with equal data terms, the simpler of the two should be preferred. This is to choose the parent w.r.t the children, when both have the same energies.

Guigues states the following to be a *dual* problem, which is, between two equally complex models, the closer to the data of the two should be preferred, giving the following constrained optimization problem:

$$\begin{aligned} \underset{\pi \in \Pi(E,H)}{\text{minimize}} \quad & \sum_{S \in \pi} \omega_\partial(S) \\ \text{subject to} \quad & \sum_{S \in \pi} \omega_\varphi(S) \leq K \end{aligned} \tag{1.15}$$

This second constraint problem in equation (1.15) has an interpretation in the rate-distortion interpretation introduce by Salembier-Garrido [100]. This will be again discussed in section 2.1.

The constrained optimization problem in equation (1.14) with no conditions on the energies are NP-hard, with an exponentially growing solution space of the lattice of partitions belonging to a HOP.

Guigues does not directly solve the constrained optimization problem in equations (1.14) and (1.15). Instead he provides conditions on $\omega_\varphi, \omega_\partial$ to achieve monotonicity. This is termed as the *scale-set* which is a family of partitions $\pi_\lambda$, which are monotonic mapping of cuts or partitions $\Pi(E, H)$ of hierarchy $H$, to the levels of an *optimal* hierarchy. These monotone cuts of the hierarchy $H$ are a result from a total ordering. We will discuss later why we will not need this additive monotonicity condition and will be replaced by the less strict condition of scale-increasingness 1.22.

**Unconstrained optimization**: The problem in equation (1.14) were solved by Salembier-Garrido and Guigues, by performing unconstrained minimization on the Lagrangian function, to solve the constrained optimization problem in equation (1.14). We will now distinguish two types of optimization problems, firstly performing unconstrained optimization for any general energy $\omega$, discussed subsequently in this chapter, while constrained optimization using the Lagrangian will be detailed in chapter 2.

[3]Though one must note that in further development we will not require additivity condition and we can compose these energies by non-linear operators like the supremum or infimum.

### 1.2.4 Dynamic program for minimal cut

The algorithm achieving the optimally pruned subtree in [21], also known as the BFOS (after the authors, Breiman, Friedman, Olshen, Stone) algorithm in literature, consists of a dynamic program, that performs a bottom-up scan, while pruning off a branch or the node its rooted in.

Dynamic program consists in solving a structured complex problem by decomposing it into smaller simpler subproblems. These subproblems require to be *overlapping* in nature, such that a solution to a subproblem is calculated only once and serves to solve a larger subproblem. Furthermore this partial solution when aggregated with others, produces a global solution. The aggregation is quite often has a successive approximation interpretation. Following Guigues, in view of the Bellmenian characteristic of any dynamic program, in order to find the optimal cut, while one performs a bottom up search starting at the leaves level, the algorithm will implicitly have to solve subproblems appearing at higher scales before some lower scales. This means, some nodes appearing early lower down in the hierarchy might be part of a global solution.

In the current case the subproblem consists in comparing the energies of parents w.r.t to their children. The subproblem can be brought down to over a partial support of the image domain, where one needs to check whether to keep the children or prune it and keep the parent node.

In our terms, it is whether a parent or a child partial partition is retained. This has been used in a more general setting of general partitions from a hierarchy by Salembier-Garrido and Guigues to calculate provisional optimal cut over a given support of the image domain.The optimal structure of the dynamic program is:

$$\omega^*(\pi(S)) = \min\left\{\omega(\{S\}), \sum_{a\in\pi(S)} \omega(a)\right\} \tag{1.16}$$

$$\pi^*(S) = \begin{cases} \{S\}, & \text{if } \omega(S) \leq \sum_{a\in\pi(S)} \omega(a) \\ \pi(S), & \text{otherwise} \end{cases} \tag{1.17}$$

Here we see in equation (1.17) that one either chooses the parent class $\{S\}$ or its child partial partition $\pi(S)$. The advantage of the partial partition structure can be especially seen here. It helps encode any partitioning of a local support $S$, and represent the provisional partition in the dynamic program. Further in subsection 1.6.4 on $h$-increasingness we will see the generalization of this dynamic program with non-linear compositions of energies of classes of child partial partitions, as well as any change in the multi-scale

energy which can start out linear at lower scale and turn non-linear at higher scales, as long as it abides by the $h$-increasingness.

### 1.2.5 Going from Scale-sets To Energetic Lattices

With the following, we connect the work on HOP in literature with our approach:

1. The sequence of studies by Breiman et al., Salembier-Garrido, and Guigues work on the space of HOP. Here it is critical to note that we will work on the expanded space of Braids, introduced further on.
2. All energies in above mentioned studies are always linear. Though there exists non-linear energies on HOP, such as area filtering, Soille's constrained connectivity [108], Ackay-Ackcoy filtering on multi-spectral images [2]. We propose necessary and sufficient conditions where the dynamic program of weakest link pruning still holds, namely $h$-increasingness.
3. Monotonicity of cuts have up until now been assured by sub/super additivity of the constraint function, in conjunction with said linear energies. For the classes of both linear and non-linear energies, we introduce a necessary and sufficient, monotonicity condition, namely scale-increasingness.
4. Uniqueness in the above studies have not been explicated, while leaning on Breiman's condition of the smallest optimal subtree. We achieve uniqueness, by introducing a lattice structure, i.e. energetic lattice.

In the following sections we will introduce the Braids, energetic ordering and energetic lattices, $h$-increasingness, scale-increasingness, and finally inf-modularity, respectively.

## 1.3 Braids of Partitions

The Braid of partitions, provides a hierarchical structure richer than the hierarchy while lending itself to the constrained optimization problem.

In the lattice of all partitions of a set, the hierarchies form chains, i.e. totally ordered sub-lattices. Can it be possible to construct other sub-lattices, which no longer form chains, while they share hierarchical properties, and help improve the search space for the constrained optimization and the inherent dynamic programming structure? The need for such models arises in several situations, in frequency domain analysis, multivariate segmentations, where partition contours are not totally ordered. One can also

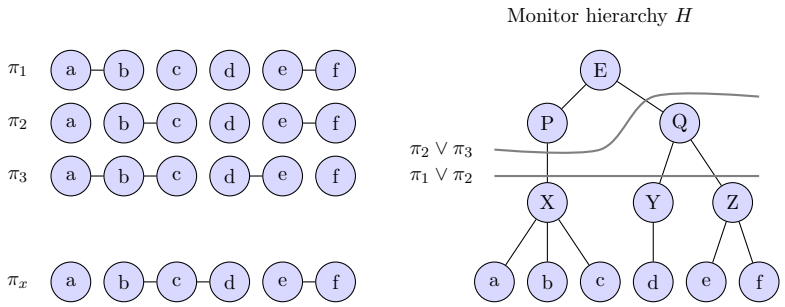



FIGURE 1.4: Toy example of a braid. $E$ is partitioned by leaf nodes $\{a, b, c, d, e, f\}$. The set $B_1 = \{\pi_1, \pi_2, \pi_3\}$ forms a braid whose pairwise supremum is indicated at the right. One can note now that $\pi_1(X), \pi_2(X)$ have a common parent $X$, but $\pi_2(Q), \pi_3(Q)$ a common grand parent $Q$. The partition $\pi_x$ cannot be added to $B_1$. There does not exist a supremum class, in the monitor hierarchy $H$, other than from the whole space $E$, thus not producing a braid structure.

consider situations where hierarchies have to be combined, or enlarged to more adapted structures. One can view such structures as chains of segmentations of an input image where some levels are uncertain, and yield several partially ordered variants. In such cases, it should be preferable to maintain all of variants and to choose among them later, in a dynamic program step.

In figure 1.4 we demonstrate a simple example of a braid with its dendrogram structure. A formal definition is provided in equation 1.18. As we can see the partitions $\pi_1, \pi_2$ are not nested nor disjoint, and basically correspond to different segmentation hypotheses that exist in the stack of segmentations.

### 1.3.1 Definition

To define the braids, we start from the lattice $\Pi(E)$ of the partitions of $E$, of minimal element the leaves partition $\pi_0$. Next we introduce a hierarchy $H$ which serves as a parameter. A braid $B$ is a family of partitions of $E$. The family $B$ is not arbitrary, but monitored by a non-trivial hierarchy $H$, in the sense that the refinement supremum of any two elements of $\Pi(E, B)$ is a cut of $H$. This leads to the more formal definition:

**Definition 1.5.** *(Braid of Partitions) Let $\Pi(E)$ be the complete lattice of all partitions of set $E$; let $H$ be a hierarchy in $\Pi(E)$. A* braid *$B$ of monitor $H$ is a family in $\Pi(E)$ where the refinement supremum of any pair $\pi_1, \pi_2$ in $B$ is a cut of $H$, other than $\{E\}$, and belongs to $\Pi(E, H) \setminus \{E\}$:*

$$\forall \pi_1, \pi_2 \in B \quad \Rightarrow \quad \pi_1 \vee \pi_2 \in \Pi(E, H) \setminus \{E\} \tag{1.18}$$

In other words, the classes of supremums $\pi_1 \vee \pi_2$ are classes of monitor hierarchy $H$, and the monitor by itself is not uniquely defined. Though a unique monitor can be imposed by extracting the largest hierarchy (in terms of refinement) whose classes contain the classes of the partitions in the braid. One thus still has a scale selection to perform in the context of chosing a monitor hierarchy for a given application. Just as the cuts of $H$, which were denoted by $\Pi(E, H)$, we now define the cuts of $B$ as the partitions whose classes are taken in $B$, and denote the class of all these cuts by $\Pi(E, B)$.The hierarchy $H$ may itself belong to the braid, or not. On the other hand, any hierarchy is a braid with itself as monitor. When $H \subseteq B$, we have $\Pi(E, H) \subseteq \Pi(E, B) \subseteq \Pi(E)$, i.e. the braid cuts $\Pi(E, B)$ are in between the cuts of the hierarchy $H$ and the set of all partitions of $E$. A braid cannot be represented by a saliency, except when it reduces to a hierarchy whose classes are connected sets.

**Remark**: The partition with one class $\{E\}$ is not considered in definition 1.18, since this would imply that any family of arbitrary partitions would form a braid with $\{E\}$ as supremum, thus losing any useful structure. We also assume a locally finite number of classes in such cases, like in the case of hierarchies.

The braids of partitions (BOP) provide an alternative hierarchical structure. In case of a hierarchy the cone or family of classes containing a point $x \in E$, can only be nested or disjoint. While the cone of classes in the BOP, that contain a single point, are not necessarily nested, though their suprema are.

The braid structure's definition is a general and provides multiple ways of creating the braid. We state one direction here, which is a composition law on tuples of hierarchies which produces a braid.

**Proposition 1.6.** *Given three hierarchies $H, H_1, H_2$ formed from the same leaves, such that, $H_1 \leq H$, $H_2 \leq H$, then the family of partitions given by $\{H_1 \cup H_2\} \setminus \{E\}$ forms a braid with the monitor $H$.*

Here $H_1 \leq H$ on hierarchies says that each class of hierarchy $H_1$ is contained in the classes of $H$. The union of hierarchies form braids, while braids are not necessarily decomposable into hierarchies.

**Lack of discriminatory ultrametric**: A classical result about hierarchy tells that the set $\mathcal{S}$ of the classes of $H$ is a metric space, where the distance is the absolute difference between the levels of the classes. It is defined as an ultrametric, where the triangle inequality by addition, is replaced by the maximum, which is stricter [68]. This ensures a characterization of each hierarchy, by its ultrametric function. In the case of a braid $B$, because two different classes $S$ and $S'$ of $\pi$ and $\pi'$ respectively may be inserted at

the same level of the monitor hierarchy $H$ (hence have a zero distance), the ultrametric distances between such classes lose characterization, and the ultrametric may not be very discriminative. However, the classes of $B$ can always be described by other distances (e.g. measure of the intersection, Hausdorff metric in case of tessellation of $\mathbb{R}^n$, etc.).

Particular versions of braids have appeared in classification problems, for example Diday [36], demonstrates pseudo-hierarchies, called pyramids, where a child may have two parents.

**1-D Example**: Given a hierarchy $H = \{\pi'_i\}$, and when the level index $i$ is odd, let us associate with $\pi'_i$ a second partition (residue) $\pi''_i \leq \pi'_{i+1}$. This generates the following sequence:

$$\pi'_0, \{\pi'_1, \pi''_1\}, \pi'_2, ..., \pi'_{2i}, \{\pi'_{2i+1}, \pi''_{2i+1}\}, \pi'_{2i+2}, ..., \{E\}, \tag{1.19}$$

which resemble a braid of hair, by successive enlargements and shrinkages. One demonstrates a more flexible braid using equation (1.19). To do this, let the number of supplementary partitions at the odd levels be random. One can also let the hierarchy $H$ monitor the residue, once in every three levels, at level $i$ say, and introduce supplementary partitions at levels $i+1$ and $i+2$ with the condition that all these supplementary partitions are smaller than $\pi'_{i+3}$. Note that the partitions $\pi''_k \in R$ may not correspond to some level of $H$. It is the case for example when $\pi''_1, \pi''_2, \pi''_3 \in R$, with $\pi''_1 \cup \pi''_2 = \pi'_i$ and $\pi''_2 \cup \pi''_3 = \pi'_{i+10}$, where $\pi'_i, \pi'_{i+10} \in H$.

An example of the use of braid is depicted in Figure 1.5 Voronoi partitioning are utilized to remove parasitic grains of small sizes. One later decides during the optimization phase from the constructed braid, which of these partitions have minimal energy. This example also shows that the braid based minimization works also when the input is a single partition, the monitor hierarchy is part of the construction step of the braid.

In figure 1.6 we demonstrate a braid created by composing two hierarchies, which are both created from flooding area and volume attributes of a gray scale image. The partitions of the braid consists of partitions from the two hierarchies. Choosing the attribute and gradient function enables one to create a family of non-trivial braids using the watershed transformation. We demonstrate more examples on color and depth as well as an algorithm to generate braids in chapter 5.

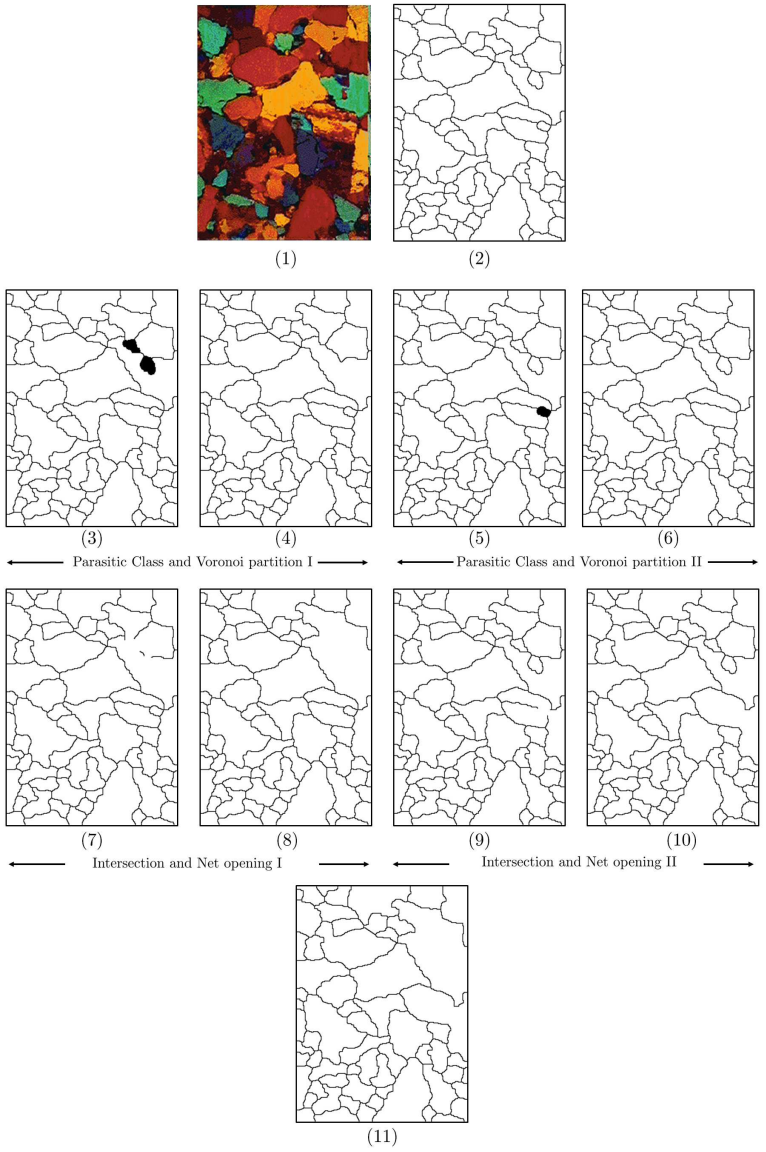



FIGURE 1.5: The intial color image (1) was segmented, giving the partition (2). Two zones are perhaps not correctly segmented demonstrated in (3), that we filled up by a Voronoi partitioning producing partition (4). Further a second parasitic class in (5) is removed and replaced with a Voronoi partitioning giving (6). (7) indicates the intersections of the contours of the initial segmentation (2) and the Voronoi partition in (4), and (8) is the net opening (refer to operator in chapter 4) of (7) and similarly for (9) and (10) with the second parasitic class in (5). (11) depicts the supremum of the two partitions (8) and (10). The three partitions (1, 8, 11), with the whole space form the monitor hierarchy $H$ of the braid made by (1, 4, 5, 8, 10, 11), where the (8) = sup(1, 4) and (11) is the smallest element of $H$ larger than the sup(1,6) and also (11) = sup(8,10).

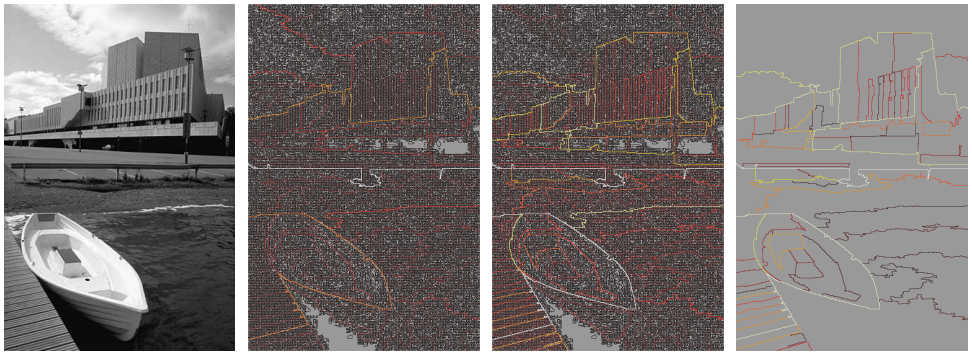


FIGURE 1.6: From left to right: Initial gray-scale image, watershed hierarchy saliency by area attribute flooding, Watershed hierarchy saliency by volume attribute flooding, the monitor hierarchy saliency obtained by net opening over the intersection of Jordan nets, see subsection 4.8.2.

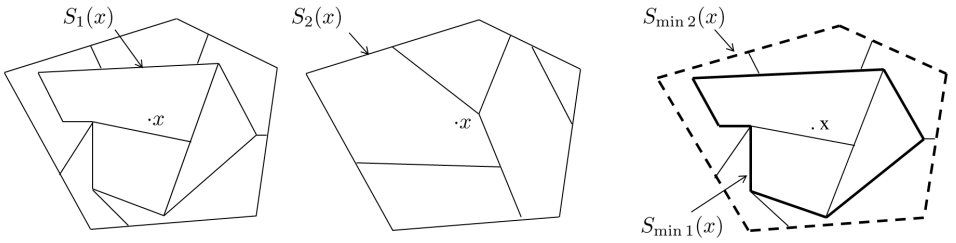
### 1.3.2 Classes of Braids

We now study the classes of any cut $\pi$ from a braid. The family cuts are denoted by $\Pi(E, B)$. Here are few properties of these classes.

- $\forall \pi_1, \pi_2 \in B, \pi_1 \vee \pi_2$ contains a finite number of classes of $\pi_1$ and $\pi_2$. This is a direct consequence of the second axiom of a hierarchy, in definition 1.3.
- The definition of a braid is transitive. Indeed, $\Pi(E, B)$ is a sub-lattice for the refinement, so that if $\pi_1 \vee \pi_2$ and $\pi_2 \vee \pi_3$ belong to $\Pi(E, B)$, then $\pi_1 \vee \pi_2 \vee \pi_3 \in \Pi(E, B)$. Moreover, the supremum in relation (1.18) extends to infinite families. Let $\{S_j(x)\}$ be set of all classes at point $x$ of a possibly infinite family $\{\pi_j, , j \in J\}$ of cuts of the braid $B$. The finite unions of these classes forms a cone of classes in the monitor hierarchy $H$, since they all contain the point $x$. By definition of a hierarchy their union $S(x) = \cup S_j(x)$ also belongs to $H$. As point $x$ spans $E$, the $S(x)$ generate the classes of a cut $\pi$ of $H$, which turns out to be the lowest upper bound (l.u.b.), in $H$, of the family $\{\pi_j, j \in J\}$.
- The next two properties are the concern of minimal covering of braid cuts by hierarchy cuts.

**Proposition 1.7.** *Every braid cut $\pi \in \Pi(E, B)$ admits a lowest upper bound $\pi_{\min}$ among the cuts $\Pi(E, H)$ of the monitor hierarchy $H$. The classes $S_{\min}(x)$ of $\pi_{\min}$ are supports of p.p. of $\pi$.*

*Proof.* Let $V(x)$ be the class at point $x$ of the braid cut $\pi \in \Pi(E, B)$. Given another braid cut $\pi_j$ consider the class $S_j(x)$ of the hierarchy cut $\pi \vee \pi_j$. Make $\pi_j$ span all braid




Partitions $\pi_1$ of Braid $B$ Partitions $\pi_2$ of Braid $B$ Classes $S_{\min 1}(x), S_{\min 2}(x)$ in Monitor $H$

FIGURE 1.7: Left: and middle: details of the partitions $\pi_1$ and $\pi_2$ of the braid $B$, with the two classes $S_1(x)$ and $S_2(x)$ at point $x$. Right: $S_{\min 1}(x)$ and $S_{\min 2}(x)$ are the two lowest upper bounds of $S_1(x)$ and $S_2(x)$ in the monitor hierarchy. $S_{\min 2}(x)$ is the support of a partial partition of $\pi_1$ because $S_{\min 1}(x) \subseteq S_{\min 2}(x)$.

cuts $\Pi(E, B) \backslash \pi$. The intersection

$$S_{\min}(x) = \cap_j \{S_j(x)\} \supseteq V(x) \tag{1.20}$$

is a class of $H$ because the $S_j(x)$ are nested, thus are classes of partitions of the chain $H$ as well as their intersection. Therefore, when the point $x$ describes $E$ then the l.u.b. $\pi_{\min}$ of $\pi$ in $H$ is generated. The inequality $\pi \leq \pi_{\min}$ shows that the class $S_{\min}(x)$ of $\pi_{\min}$ at point $x$ is the support of a p.p. of $\pi$. □

As a consequence, if $B$ is a braid of monitor hierarchy $H$, then $B \cup H$ is a braid of same monitor.

For the braid of Figure 1.7 for example, the cut $\pi_{\min}(\pi_1) = \pi_{\min}(\pi_2) = \pi_1 \vee \pi_2$. Moreover, the case when two $S_{\min}(x)$ are ordered is instructive:

**Proposition 1.8.** *Let $\pi_1, \pi_2 \in \Pi(E, B)$ be two cuts of a braid $B$ of monitor $H$. Let $S_{\min 1}(x)$ and $S_{\min 2}(x)$, the classes at $x$ of the two associated l.u.b.. If $S_{\min 1}(x) \subseteq S_{\min 2}(x)$, then $S_{\min 2}(x)$ is the support of a partial partition of $\pi_1$.*

*Proof.* The proposition is obviously true when $S_{\min 1}(x) = S_{\min 2}(x)$. Suppose that $S_{\min 1}(x) \neq S_{\min 2}(x)$. Let $y \in S_{\min 2}(x) \backslash S_{\min 1}(x)$ and $S_{\min 1}(y)$ be l.u.b. class of $\pi_1$ at point $y$. As $y \in S_{\min 1}(y) \cap S_{\min 2}(x)$, which are both classes of the hierarchy $H$, we have either $S_{\min 1}(y) \supset S_{\min 2}(x)$ or $S_{\min 1}(y) \subseteq S_{\min 2}(x)$. The first case is impossible for $S_{\min 1}(x)$ is not empty, and $S_{\min}(x)$ and $S_{\min}(y)$ are disjoint, thus $S_{\min 1}(y) \subseteq S_{\min 2}(x)$. As this inclusion is satisfied for all points $y \in S_{\min 2}(x) \setminus S_{\min 1}(x)$, we finally obtain that $\pi_1 \sqcap \{S_{\min 2}(x)\}$ is a partition of $S_{\min 2}(x)$. □

Figure 1.7 illustrates the lower bounding $S_{\min 1}, S_{\min 2}$ classes in proposition 1.8.

### 1.3.3 Motivations

The braid structure by itself can be seen in problems of segmentations and evaluation. We state briefly the interest of the braid of partitions.

- This type of nested structure is already well known in the area of super-pixel merging, and the generation of "segmentation-soup" in Malisiewicz et al. [69], where the use of segmentation merging is proven emperically to improve detection support. Furthermore this improvement in detection due to the merging operation is shown to be independent of the segmentation algorithm producing the family of partitions.

- Uncertain partition boundaries produce many possible partial partitions corresponding to the same image values, based on the algorithm used, and the quality measured used to evaluate the segmentation. For further details on the study of segmentation evaluation, please refer to Unnikrishnan et al. [115]. This has been demonstrated in figure 1.8. They introduce segmentation evaluation measure, Normalized Probability Rand (NPR) index, which is a meaningful measure in that it only penalizes fragmentation in regions that are unsupported by the ground-truth images, and allows refinement without penalty if it is consistently reflected in the ground-truth set. We aimed in creating the braid structure to handle similar problems in hierarchical partition structure. This basically enables accommodate boundary ambiguity, which is basically different machine segmentation of regions with contours characterized by varying gray-scale or color gradient values.

- Multivariate segmentations operate on independent functions, like the case of hyper-spectral segmentations or RGB color images. Partition contours here could have a braid structure given certain compositions of functions. Furthermore, one can envisage a composition of segmentations resulting from independent channels, provided they have a non-trivial braid structure.

- The primary motivation is the use of the braid in Breiman's dynamic program that calculated optimum on hierarchies. In this thesis we show that the dynamic program works for a larger class of partitions which is the braids and further more it ensures better optimum than hierarchies, when operating on braids with non-trivial monitors. This is further discussion in detail in section 1.5, and demonstrated in figure 1.14.

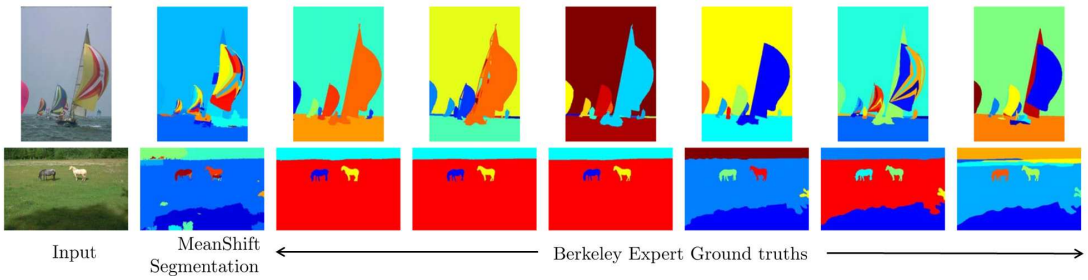



FIGURE 1.8: No single Ground truth segmentation is a refinement of the mean-shift segmentation, but their suprema are. This is a result in the inconsistent partition contours in case of both the human and machine based mean shift segmentation. Image reproduced from Unnikrishnan et al. [114, 115]. This helps handle boundary ambiguity.

## 1.4 Energetic lattices of braids

First of all, what does "minimal cut" mean? A cut of minimal energy? For a finite set of only 25 leaves partitioning $E$ generates $0.5 \times 10^{18}$ different potential partitions, following the bell's number. The set of all energies, which is in practice an interval in the positive integers, risks to be too poor for the purpose, and we may try and act directly on some lattice of cuts, which should of course involve the energy $\omega$ by some modalities. Then the existence and uniqueness of minimal cuts will be ensured by this lattice structure itself.

In figure 1.9 we demonstrate how one can obtain multiple optimal cuts. One requires singular energies to ensure the existence of a unique optimum.

**Definition 1.9.** *Let $\omega$ be an energy on the partial partitions $\mathcal{D}(E)$, and $B$ be a braid $B$ of monitor hierarchy $H$. Energy $\omega$ is singular when*

*(i) the energy $\omega(\{S\})$ of every class $S$ of $H$ is either strictly smaller, or strictly greater, than the energies of all partial partitions of $B$ of support $S$:*

$$\forall\ \pi(S) \in \Pi(S), \omega(\{S\}) < \omega(\pi(S))\}\ \text{or}\ \omega(\{S\}) > \omega(\pi(S))\}, \tag{1.21}$$

*(ii) if $\forall \pi_1, \pi_2 \in B$ and $\pi_1 \vee \pi_2 = \{S\} \in \Pi(E, H)$, then $\omega(\pi_1) \neq \omega(\pi_2)$.*

This is demonstrated in figure 1.9. Definition 1.9 extends the cone structure at point $x$ defined for hierarchies (i.e. $i \leq j \Rightarrow S_i(x) \leq S_j(x)$). In a braid several different classes $S_i^1(x)$, $S_i^2(x)$ etc., may coexist at the same level.

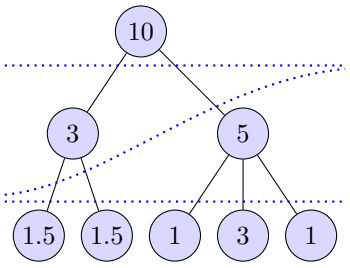



FIGURE 1.9: A hierarchy with multiple optimal cuts with the same energy. One requires the general condition of singularity to ensure the existence of a unique optimum. In a hierarchy, an arbitrary energy $\omega$ becomes singular when the parent picked w.r.t its children in case they have equal energies.

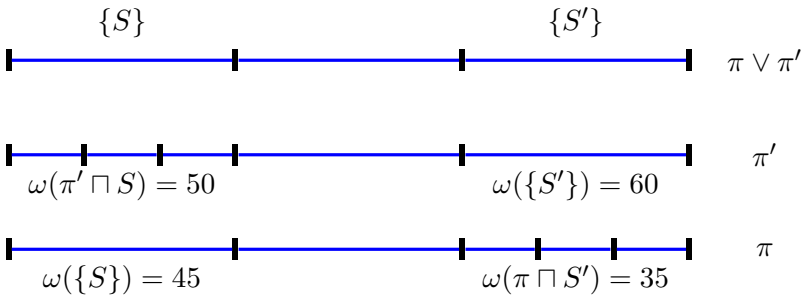



FIGURE 1.10: An example of energetic ordering: We have $\pi \preceq_\omega \pi'$ since in each class of $\pi \vee \pi'$, the energy $\omega$ of $\pi$ is lesser than or equal to that of $\pi'$.

### 1.4.1 Energetic ordering

Consider two partitions $\pi_1, \pi_2$ of $E$, and the class $S$ of the refinement supremum $\pi_1 \vee \pi_2$ at point $x \in E$. $S$ is the support of two p.p. $a_1$ of $\pi_1$ and $a_2$ of $\pi_2$ (see Figure 1.10). Intuitively, one may assess that, in some sense, $\pi_1$ is less energetic than $\pi_2$ for an energy $\omega$ when $\omega[\pi_1 \sqcap \{S\}] \leq \omega[\pi_2 \sqcap \{S\}]$ in each class of $\pi_1 \vee \pi_2$. This intuition is true and has the meaning of an ordering relation when $\omega$ is singular and $B$ is a braid.

**Theorem 1.10.** *Let $\Pi$ be a family of partitions of $E$, and let $\pi_1, \pi_2 \in \Pi$. Given an energy $\omega$, the partition $\pi_1$ is said to be less energetic than $\pi_2$, and one writes $\pi_1 \preceq_\omega \pi_2$ when in each class of $\pi_1 \vee \pi_2$ the energy of the partial partition of $\pi_1$ is smaller or equal to that of $\pi_{2:}$ :*

$$\pi_1 \preceq_\omega \pi_2 \quad \Leftrightarrow \quad \{S \in \pi_1 \vee \pi_2 \ \Rightarrow \ \omega(\pi_1 \sqcap \{S\}) \leq \ \omega(\pi_2 \sqcap \{S\})\} \tag{1.22}$$

*The relation $\preceq_\omega$ is an ordering relation for all singular energies $\omega$, if and only if the family $\Pi(E)$ is the set of cuts of a braid.*

*Proof.* The reflexivity is obvious. For the anti-symmetry, we observe that $\pi_1 \preceq_\omega \pi_2$ and $\pi_2 \preceq_\omega \pi_1$ involve the same refinement supremum $\pi_1 \vee \pi_2$. Therefore in any $S \in \pi_1 \vee \pi_2$

the energies of $\pi_1 \sqcap \{S\}$ and $\pi_2 \sqcap \{S\}$ are the same (Rel.(1.22)). Then the singularity imposes that these two partial partitions are identical.

As for transitivity, consider three elements $\pi_1, \pi_2, \pi_3 \in B$, with

$$\pi_1 \preceq_\omega \pi_2 \text{ (a) and } \pi_2 \preceq_\omega \pi_3 \text{ (b)} \tag{1.23}$$

We must prove that $\pi_1 \preceq_\omega \pi_3$. Let $V_1$ (resp. $V_2$, resp. $V_3$) be the class of $\pi_1$ at point $x$, and $S_1 = S_{\min}(V_1)$ (resp. $S_2$, resp. $S_3$). The classes $S_1, S_2, S_3$ of $H$, which contain the point $x$, are thus ordered in one of the possible three manners $S_1 \cup S_3 \subseteq S_2$, $S_1 \cup S_2 \subseteq S_3$, or $S_1 \cup S_2 \subseteq S_3$. Suppose firstly that in the first case $S_2 = S_3$ thus $S_1 \subseteq S_3$. The proposition 1.8 tells us that $S_3$ is the support of a p.p. of $\pi_1$, and according to Rel.(1.23) the restrictions of $\pi_1, \pi_2$, and $\pi_3$ to this support satisfy $\omega(\pi_1 \sqcap S_3) \leq \omega(\pi_2 \sqcap S_3)$ and $\omega(\pi_2 \sqcap S_3) \leq \omega(\pi_3 \sqcap S_3)$, so that the transitivity is satisfied at point $x$. The same proof still applies for $S_1 = S_3$, and extends to the two other cases by circular permutation.

If $S_1 \cup S_3 \subset S_2$, then there exists a partial partition $a_3$ of $\pi_3$ such that $\{S_3\} \sqcup a_3$ has $S_2$ for support, and Rel.(1.23b) implies that $\omega(\{S_2\}) \leq \omega(\{S_3\} \sqcup a_3)$, thus by singularity that $\omega(\{S_2\}) < \omega(\{S_3\} \sqcup a_3)$. Similarly, the inclusion $S_1 \subseteq S_2$ leads to $\omega(\{S_1\} \sqcup a_1) < \omega(\{S_2\})$ (by applying Rel. (1.23a)), which contradicts the singularity axiom. Therefore, the two possible orders are $S_1 \cup S_2 \subseteq S_3$ and $S_2 \cup S_3 \subseteq S_1$.

If $S_1 \cup S_2 \subset S_3$, there exist two p.p. $a'_1$ and $a'_2$ with $\{S_1\} \sqcup a'_1 = \{S_3\}$ and $\{S_2\} \sqcup a'_2 = \{S_3\}$. By Rel. (1.23b), we find $\omega(\{S_2\} \sqcup a'_2) \leq \omega(\{S_3\})$. Then by singularity, all p.p. of $S_3$, thus $\{S_1\} \sqcup a'_1$, have an energy $\leq \omega(\{S_3\})$, so that $\omega(\{S_1\} \sqcup a'_1) \leq \omega(\{S_3\})$, which shows the transitivity at point $x$. If $S_2 \cup S_3 \subseteq S_1$, the same proof yields the same conclusion. Since local transitivity is true for all point $x \in E$, the relation $\pi_1 \preceq_\omega \pi_2$ itself is transitive.

For the "only if" statement, we have to prove that the ordering vanishes either when $\omega$ is not singular, or when $B$ is not a braid. Consider first an ordering $\preceq_\omega$ whose energy is not singular, and two cuts $\pi$ and $\pi'$ identical everywhere except in the class $S'(x)$ of $\pi'$, where $\pi$ is locally the p.p. $a$. Suppose that $\omega(a) = \omega(S'(x))$. This implies $\pi \preceq_\omega \pi'$ and also $\pi' \preceq_\omega \pi$. However we do not have $\pi' = \pi$ since $a \neq S'(x)$. Thus singularity is needed. Suppose now $\omega$ singular, and applied to the three partitions $\pi_1, \pi_2$, and $\pi_3$ as indicated in Figure 1.11. These three partition do not belong to a braid, because the three classes of $\pi_1 \vee \pi_2$, $\pi_2 \vee \pi_3$, and $\pi_3 \vee \pi_1$ at point $x$ are not nested. We have $\pi_1 \preceq_\omega \pi_2, \pi_2 \preceq_\omega \pi_3$ but not $\pi_1 \preceq_\omega \pi_3$, which achieves the proof. □

The "only if" part of the theorem means that the braids, and their cuts, have the exact level of generality to work. As a consequence, all the downstream results which involve the theorem, including energetic lattices and constraint minimizations, are valid for braid

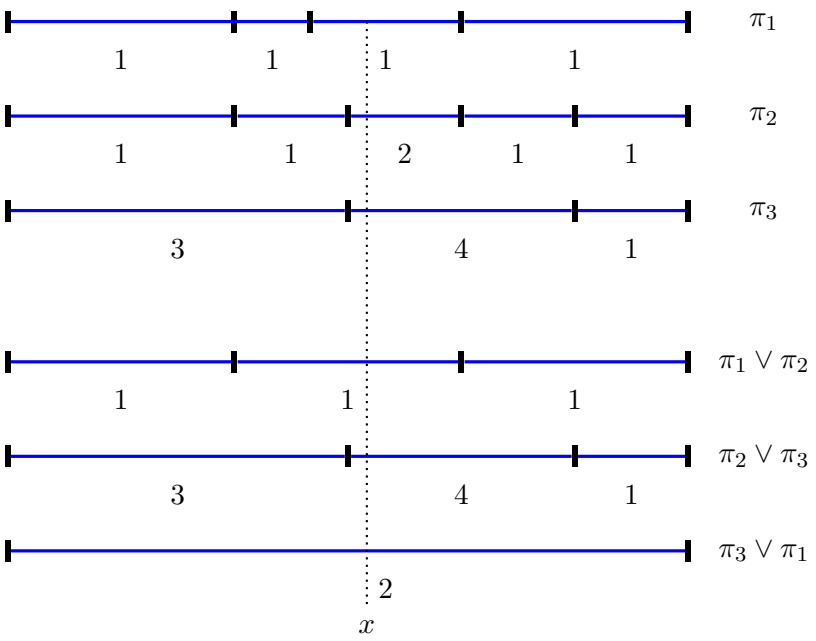



FIGURE 1.11: The three partitions $\pi_1, \pi_2$,and $\pi_3$ cannot come from a braid, because the three classes of $\pi_1 \vee \pi_2, \pi_2 \vee \pi_3$ and $\pi_3 \vee \pi_1$ at point $x$ are not nested. The values of energy $\omega$ for the classes are indicated above them, and the energy of a p.p. is the sum of its classes. The transitivity of the relation $\preceq_\omega$ is not satisfied.

cuts only. If one wants to build up energetic orderings on other families of partitions, the energy $\omega$ must be more specified.

### 1.4.2 Energetic lattice

Given the energy $\omega$, the cuts of any braid $B$ form a complete lattice structure w.r.t. their energetic ordering $\preceq_\omega$. We will prove it in two steps, by beginning with the finite families of cuts

**Lemma 1.11.** *Any finite family of cuts $\{\pi_j, j \in J\}$ in $\Pi(E, B)$ admits a greatest lower bound $\curlywedge_\omega \pi_j$ and a lowest upper bound $\curlyvee_\omega \pi_j$.*

*Proof.* As the family is finite, it suffices to prove the results for the pairs of partitions $\pi_1$, $\pi_2 \in \{\pi_j, j \in J\}$. In each class $S$ of the refinement supremum $\pi_1 \vee \pi_2$ the three p.p. $\{S\}, \{S\} \sqcap \pi_1, \{S\} \sqcap \pi_2$ have three energies which are different by singularity. One can always choose the less (resp. most) energetic one. By doing the same for all classes of $\pi_1 \vee \pi_2$ we obtain the unique largest lower-bound $\pi_1 \curlywedge_\omega \pi_2$ (resp. smallest upper bound $\pi_1 \curlyvee_\omega \pi_2$) of $\pi_1$ and $\pi_2$, which achieves the proof. □

**Theorem 1.12.** *Let $B$ be a family of partitions of $E$, and $\omega$ be a singular energy. The set of all cuts of $B$ forms a complete lattice $\Pi(\omega, E, B)$ for the energetic ordering $\preceq_\omega$*

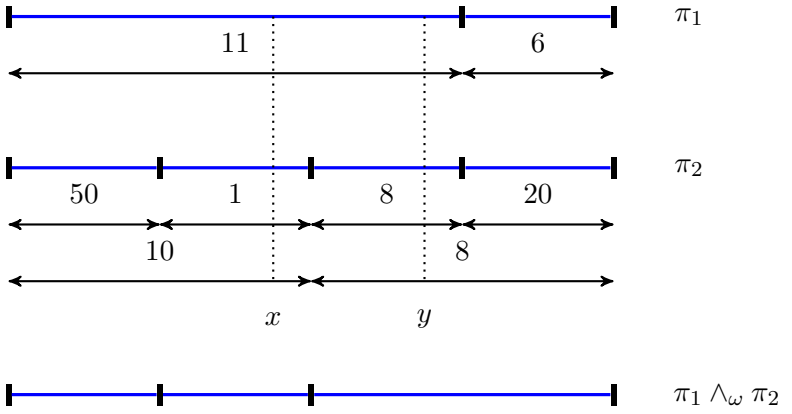



FIGURE 1.12: Energetic infimum of three partitions: At point $x$ we look for the largest class to be less energetic than the constituting internal p.p., this is $\pi_2$ and at point $y$ it is the class of $\pi_1$. The energetic infimum $\curlywedge_\omega$ is the partition drawn at the bottom.

*if and only if $B$ is a braid. Given a family $\{\pi_j, 1 \leq j \leq p\}$ of cuts in $\Pi(\omega, B)$, the infimum $\curlywedge_\omega \pi_j$ (resp. supremum $\curlyvee_\omega \pi_j$) is obtained by taking the p.p. of lowest energy (resp. highest energy) in each class of the refinement supremum $\vee \pi_j$.*

*Proof.* If $B$ is not a braid, the relation $\preceq_\omega$ does not define an ordering. Suppose $B$ to be a braid, and let $\{S_j(x)\}$ be set of all classes at point $x$ for a family $\{\pi_j, j \in J \subseteq I\}$ of cuts of $B$. We saw that these classes form a cone, and that their union $S_M(x) = \cup S_j(x)$, which belongs to $H$, has a finite number of leaves. Therefore the number of possible partitions of these leaves is finite, as well as the number of different partitions $\pi_j \sqcap \{S_M(x)$. Lemma 1.11 applies and leads to the local infimum $\wedge[\pi_j \sqcap \{S_M(x)\}]$. The global infimum is obtained by making $x$ vary, i.e. $\wedge \pi_j = \sqcup\{\wedge[\pi_j \sqcap \{S_M(x)\}], x \in E\}$. By duality, we have also $\vee \pi_j = \sqcup\{\vee[\pi_j \sqcap \{S_M(x)\}], x \in E\}$, which achieves the proof. □

The universal infimum of the lattice $\Pi(E, B)$ is denoted by $\pi^* = \curlywedge_\omega\{\pi, \pi \in \Pi(E, B)\}$. It is the unique cut of $B$ smaller than all the other cuts of $\Pi(E, H)$ for the ordering $\preceq_\omega$. Remarkably, the theorem was established without assuming any linearity, or $h$-increasingness, or sub-modularity, of the energy $\omega$. The theorem is prior to these notions, though they will be useful later.

Figure 1.12 depicts a toy example of the energetic infimum $\curlywedge_\omega$ for a HOP.

### 1.4.3 The three lattices

The assumption of singularity is crucial. If we drop it, we lose all theorems of this paper which involve energetic lattices: the scale increasingness structure is no longer valid, the Lagrange model is undefined, etc.. Fortunately, the singularity hypothesis is not very

restricting in practice, where most of the energies admit a singular version, up to minor changes.

In the notation, the three symbols $\leq, \vee$, and $\wedge$ (without $\omega$ subscript) are allocated to the refinement lattice $\Pi(E, B)$, and $\preceq_\omega, \curlyvee_\omega$ and $\curlywedge_\omega$ to the energetic lattice $\Pi(\omega, E, B)$. The expression "minimal cuts" always refers to energy infima $\curlywedge_\omega$, the only ones for which the expression makes sense (the refinement universal infimum is the leaves partition). Moreover, the meaning of the energy infimum is twofold:

**Proposition 1.13.** *The minimal cut $\pi^*$ of $\Pi(\omega, E, B)$ is not only the $\curlywedge_\omega$ infimum of the family of all cuts, but it is also less energetic than every cut $\pi$ in each class of $\pi \vee \pi^*$. It thus turns out to be both* local *and* global.

Three lattices (and orders) interact on the family of cuts of a braid $B$:

1. Numerical lattice ( $\leq, \vee$, and $\wedge$) for the energies $\omega$,
2. Refinement lattice ( $\leq, \vee$, and $\wedge$) for the cuts of $B$,
3. Energetic lattice $\Pi(\omega, E, B)$, again for the cuts $(\preceq_\omega, \curlyvee_\omega, \curlywedge_\omega)$

$h$-increasingness in section 1.5, studies the relations between the energetic and numerical order of energies as will be demonstrated in relation (1.25).

*Remark* 1.14. (Finite energies) Energetic lattices allow for infinite image domain on account of its local nature. Consider a finite window $Z$ centered at the origin, and $Z_h$ translated by $h$. Suppose now that all classes encountered within $Z_h$ have finite energies, as do their partial partitions. When not the energies these classes and are infinite. Then, for any point in the space, the cone of classes containing the point, given that the energies of these classes in the cone are finite, of course except $E$, one can determine now the class forming the minimal cut, containing said point.

## 1.5 $h$-increasing energies

This section is devoted to the links between an energetic ordering $\preceq_\omega$ on the cuts $\Pi(E, B)$ and the numerical ordering of the energies of these cuts. The theorem 1.12 says nothing about the energy of a minimal cut, and does not tell whether the energetic ordering $\pi \preceq_\omega \pi'$ between two cuts implies the same sense of variation for the energies themselves, i.e. $\omega(\pi) \leq \omega(\pi')$. Indeed, one easily sees that it is not always the case. For example, take for singular energy $\omega(\pi) = 0$ (resp. 1) when the number of classes of the p.p. $\pi$

is odd (resp. even), and greater than 1, the energy of the one class partitions being 2. Then, in Figure 1.10, $\pi \preceq_\omega \pi'$ whereas $\omega(\pi) = 1$ and $\omega(\pi') = 0$. A new condition is needed, namely that of $h$-increasingness:

**Definition 1.15.** *($h$-increasingness) Let $(a_i, a'_i)$ be elements of two different p.p. of the same support $S_i$, and $\{S_i, S_i \in E, i \in I\}$ a family of disjoint supports. A finite singular energy $\omega$ on the partial partitions $\mathcal{D}(E)$ is $h$-increasing when for every triplet $\{a_i, a'_i, S_i \in E, i \in I\}$ one has, $\forall i \in I$:*

$$\omega(a_i) \leq \omega(a'_i) \;\Rightarrow\; \omega(\sqcup a_i) \leq \omega(\sqcup a'_i) \tag{1.24}$$

*When in addition one has $\omega(a_i) < \omega(a'_i)$ for one $i$ at least, and when this leads to $\omega(\sqcup a_i) < \omega(\sqcup a'_i)$, then the energy $\omega$ is strictly $h$-increasing.*

For example, a linear energy, i.e. an energy where $\omega(\sqcup a_i)$ is the sum of the $\omega(a_i)$ is $h$-increasing, an even *strictly* $h$-increasing since

$$\omega(a_i) < \omega(a'_i) \text{ for all } i \in I \quad \Rightarrow \quad \omega(\sqcup a_i) < \omega(\sqcup a'_i).$$

Unlike, the $h$-increasing energy $\omega(\sqcup a_i) = \sum \omega(a_i)$ when $\sum \omega(a_i) < K$ and $= K$ when not, is not strictly $h$-increasing. Figure 1.13 shows the geometrical meaning of the $h$-increasingness.

Here we demonstrate in figure 1.13 the $h$-increasingness condition generalized to the BOP(bottom), along with the HOP where we necessarily have an ordering of partial partitions on a given support as seen in figure 1.27.

When operating on parametrized energies, across various parameter values, please refer to scale increasingness.

### 1.5.1 The two orderings $\preceq$ and $\leq$

$h$-increasingness bridges the gap between the energetic ordering $\preceq_\omega$ for partitions and the numerical ordering of their energies. Consider two cuts $\pi$ and $\pi'$ of a braid $B$, and denote by $\{S_i, i \in I\}$ the set of all classes of $\pi \vee \pi'$. If $a_i$ and $a'_i$ stand for the p.p. of support $S_i$ of $\pi$ and $\pi'$ respectively, and $\omega$ for a $h$-increasing energy, then the left member of (1.24) means that $\pi \preceq_\omega \pi'$ and the right one that $\omega(\pi) \leq \omega(\pi')$, hence:

$$\pi \preceq_\omega \pi' \quad \Rightarrow \quad \omega(\pi) \leq \omega(\pi'). \tag{1.25}$$

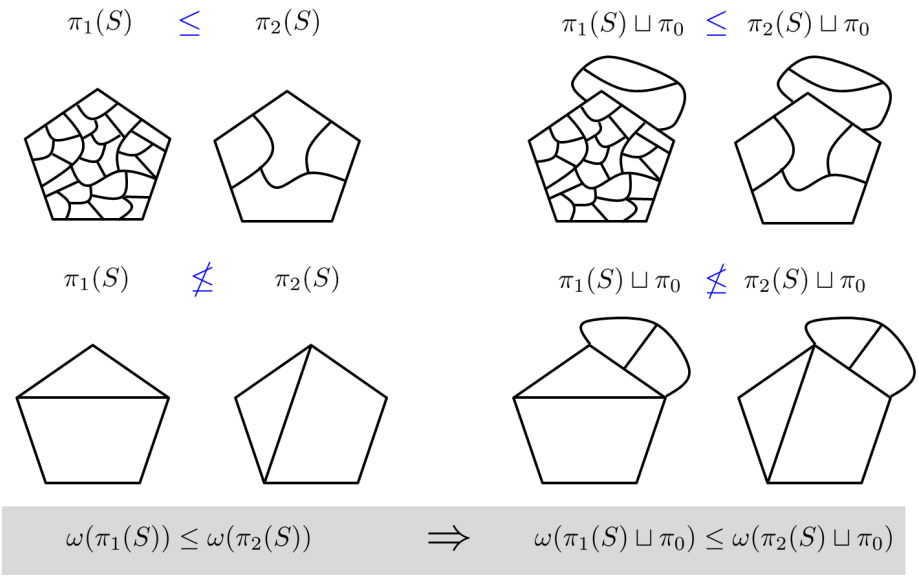



FIGURE 1.13: An example of hierarchical increasingness on HOP(top) and BOP(bottom). We see that the condition of $h$-increasingness holds generally for any family of partial partitions.

with in particular

$$\pi^* = \curlywedge_\omega\{\pi \in \Pi(E,B)\} \quad \Rightarrow \quad \omega(\pi^*) = \wedge\{\omega(\pi), \pi \in \Pi(E,B)\} \tag{1.26}$$

The converse is false for a general case, since several cuts can share the same energy, as demonstrated in figure 1.9. However in case of strict $h$-increasingness the minimal cut is unique, and is infimum of the energetic lattice.

**Proposition 1.16.** *When energy $\omega$ is strictly $h$-increasing, and the set $\Pi(E,B)$ is finite, then implication (1.26) becomes an equivalence.*

*Proof.* By uniqueness of the minimum in the $\omega$-lattice, $\pi^* \prec \pi$ for $\pi \in \Pi(E,B)\backslash\pi^*$. It means that there is a class $S$ of $\pi^* \vee \pi$ such that $\omega(\pi^* \sqcap \{S\}) < \omega(\pi \sqcap \{S\})$. By strict $h$-increasingness, this gives $\omega(\pi^*) < \omega(\pi)$, and by finiteness $\omega(\pi^*) < \wedge\{\omega(\pi), \pi \in \Pi(E,B)\backslash\pi^*\}$. Therefore, if a cut $\pi \in \Pi(E,B)$ has $\omega$ for energy, it can only be $\pi^*$. □

The axiom of *$h$-increasingness* has already been introduced in [59] for the case of a finite number of classes by the Rel.(1.27) below. The above definition 1.15 generalizes it to infinite situations:

**Proposition 1.17.** *When the family $\{a_i, a_i' \in \mathcal{D}(E), i \in I\}$ of Definition 1.15 is finite, then the $h$-increasingness is equivalent to:*

$$\omega(a) \leq \omega(a') \quad \Rightarrow \quad \omega(a \sqcup a_0) \leq \omega(a' \sqcup a_0), \quad a, a' \in \Pi(S, B) \tag{1.27}$$

*where $a$ and $a'$ are two p.p. of same support $S$, and where $a_0$ is a p.p. of support $S_0$ disjoint of $S$.*

*Proof.* The implication (1.24) $\Rightarrow$ (1.27) is obvious. For the reverse sense, consider the two pairs $(a_1, a_1')$ and $(a_2, a_2')$. The relation (1.27) allows us to write

$$\begin{aligned} \omega(a_1) \leq \omega(a_1') \quad &\Rightarrow \quad \omega(a_1 \sqcup a_2) \leq \omega(a_1' \sqcup a_2) \\ \omega(a_2) \leq \omega(a_2') \quad &\Rightarrow \quad \omega(a_1' \sqcup a_2) \leq \omega(a_1' \sqcup a_2') \end{aligned}$$

hence $\omega(a_1 \sqcup a_2) < \omega(a_1' \sqcup a_2')$. Under iteration, this inequality extends to any finite family $\{\omega(a_i), \omega(a_i'),\, i \in I\}$, i.e. to Relation (1.24). □

### 1.5.2 Minimal cut and $h$-increasingness

The finite definition (1.27) yields a dynamic algorithm for scanning the classes of $H$ or $B$ only once :

**Proposition 1.18.** *Let $H$ be a hierarchy, and $\omega$ be a singular energy on $\mathcal{D}(E)$. Consider a node $S$ of $H$ with $p$ sons $T_1..T_p$ of optimal cuts $\pi_1^*, ..\pi_p^*$. The cut of optimal energy of $S$ is either the cut*

$$\pi_1^* \sqcup \pi_2^* .. \sqcup \pi_p^*, \tag{1.28}$$

*or the one class partition $\{S\}$ itself, if and only if $\omega$ is $h$-increasing.*

*Proof.* We firstly prove that the condition in (1.27) is sufficient. The $h$-increasingness of the energy implies that the cut in (1.28) has the lowest energy among all the cuts in the family $\Pi'(S) = \sqcup\{\pi(T_k); 1 \leq k \leq p\}$, and this cut is unique by singularity. Now, every cut of $S$ is either an element of $\Pi'(S)$, or $S$ itself. Therefore, the set formed by the cut (1.28) and $S$ contains the optimal cut of $S$.

Conversely, suppose that $\pi_1^* \sqcup \pi_2^* .. \sqcup \pi_p^*$ is a cut of optimal energy for the partial hierarchy $H(S)$. It means that when we replace $\pi_1^*$ by another cut $\pi_1$ of $T_1$, i.e. such that $\omega(\pi_1) \geq \omega(\pi_1^*)$, then $\pi_1 \sqcup \pi_2^* .. \sqcup \pi_p^*$ has an energy $\geq$ than that of $\pi_1^* \sqcup \pi_2^* .. \sqcup \pi_p^*$. As this is true for all partial partitions of $E$, the energy $\omega$ is therefore $h$-increasing. □

There is an obvious extension to the family of Braids.

To compare the energy of the one class partition $\{S\}$ to the energies of all its descendants, it suffices to compare $S$ to its sons. The lower descendants do not intervene. Moreover, if $\omega$ is not singular, one can always decide to choose $\omega(\{S\})$ when $\omega(\{S\}) = \omega(\pi), \pi \in \Pi(S, B)$. This choice makes $\omega$ singular and preserves its $h$-increasingness (Proposition 4.4 of [59]).

What happens if we drop the singularity axiom in proposition 1.18? We risk to meet a node $S$ which has the same energy as the p.p. of its sons. This event introduces two solutions which are then carried over the whole induction. And since such a doublet can occur regarding any node $S \in H$, the number of minimal cuts may become huge. However at each node $S$, there is always, among the solutions, a larger partition (for the refinement). By ordering the solutions, we thus structure them in a complete lattice. uniqueness reappears, but the question "find the cut that minimizes the energy" has been replaced by "find the largest (or the smallest) cut that minimizes the energy".

Instead of using the refinement, we can, alternatively, introduce a second optimization. For example, for color images, $\omega$ can hold on the luminance, and the criterion for choosing between the optimal cuts can derive from the product saturation$\times$ hue.

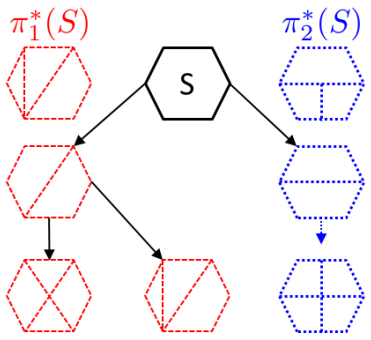



FIGURE 1.14: An elementary step of the dynamic program in a braid structure over a support $S$. The partial optimal cut in each sub-branch is shown. The final step is to compares energies $\omega(S), \omega(\pi_1^*(S)), \omega(\pi_2^*(S))$, where one picks the partial partition with the least energy. Furthermore one needs to implement a consistent rule to obtain a unique solution, in other words, one needs to implement a singular energy.

**Dynamic Program over Braid**:

As demonstrated in figure 1.14, the dynamic program substructure would now consist in making a choice between the parent supremum (if it is a class of the braid), and the partial partitions that it monitors. We consider in the figure a braid composed of two hierarchies (this is to be able to index the partial partitions.). We can now write the dynamic program step first shown for HOP in equations 1.16 and 1.17, now for the BOP:

$$\omega^*(\pi(S)) = \min\left\{\omega(\{S\}), \sum_{a\in\pi_1(S)} \omega(a), \sum_{b\in\pi_2(S)} \omega(b)\right\} \tag{1.29}$$

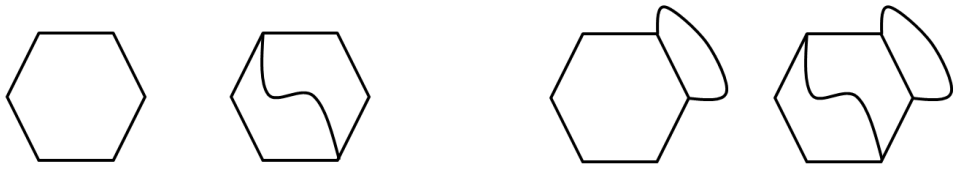
$$\pi^*(S) = \begin{cases} \{S\}, & \text{if } \omega(\{S\}) \leq \min(\sum_{a\in\pi_1(S)} \omega(a), \sum_{b\in\pi_2(S)} \omega(a)) \\ \pi_1(S), & \text{if } \omega(\pi_1(S)) < \min(\omega(\{S\}), \sum_{b\in\pi_2(S)} \omega(a)) \\ \pi_2(S), & \text{if } \omega(\pi_2(S)) < \min(\omega(\{S\}), \sum_{b\in\pi_2(S)} \omega(a)) \end{cases} \tag{1.30}$$

Equation (1.30) demonstrates a subtructure very similar to the hierarchies except now they are applied to the classes of the BOP $B$. When the energies $\omega(\pi_a(S)) = \omega(\pi_b(S))$, and $\omega(\pi_a(S)) < \omega(\{S\}))$, we can either pick randomly, as long as we pick one of the partial partitions, so that in a strict sense to keep the energies remain singular.

When a composition of multiple hierarchies $H_i, i \in \{1, 2, ..., n\}$, leads to a braid of partitions $B$ with a monitor $H'$, the dynamic program on a braid consists of either:

- the (n+1)-ary choice between the partial partitions from the $n$ hierarchies, and the monitoring supremum (if considered part of the braid), or
- Independent dynamic programs in the $n$ hierarchies, when the monitoring hierarchies is trivial, i.e. $H'$ consists of the complete space as a single class $\{E\}$, for all compositions of partitions from the $n$-hierarchies. In such a case the braid structure does not improve the minimum energy of the optimal cut achieved globally.

### 1.5.3 Simple example for a non $h$-increasing energy



$\omega(\pi_1(S)) \leq \omega(\pi_2(S))$ $\omega(\pi_1(S) \sqcup \pi_0) > \omega(\pi_2(S) \sqcup \pi_0)$

| Number of classes $N(\pi)$ | Energy $\omega(\pi)$ |
|---|---|
| 1 | 1 |
| 2 | 2 |
| $> 2$ | 0 |

FIGURE 1.15: For the example energy demonstrated in the table, the energy of a partial partition depends on the number of its classes, by a non h-increasing rule.

The energy $\omega$ in table given in figure 1.15 is a singular energy, which implies that it endows the solution space with a unique minimal cut. But this $\omega$, is not $h$-increasing! We show by a quick demonstration of the lack of a dynamic program to reach the optimum on account of combinatorial explosion.

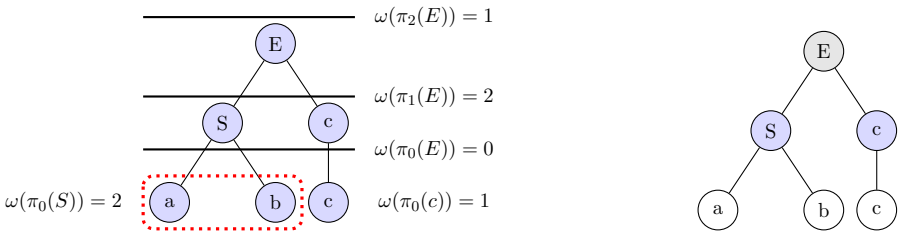



FIGURE 1.16: Figure demonstrating, initial hierarchy with cuts and their energies(left), the minimal cut by dynamic program(center) is $\pi_2(E)$, and the true minimal cut by observing the minimum directly is $\pi_0(E)$, the leaves(in white). The dynamic program fails to extract the minimal cut, and produces $E$ as the minimal cut (in gray). This as well implies that we can not use the global-local property of the energy's optimum, even if the energy is singular.

## 1.6 $h$-increasing compositions and Minkowski norms

$h$-increasingness is a property of energies, which preserves the optimal substructure in extracting the minimal cut problem so that one can use a dynamic program to solve it. As one can see, linear compositions is not the only way to ensure that the optimal cut remains, in the provisional optima of bottom up scan in the hierarchy.

The two common modes of composition are by addition and supremum. The additive mode was studied by Guigues under the name of *separable energies* [47, 49] a context in which he established the Rel.(1.44) below. Denote by $\{T_u, 1 \leq u \leq q\}$ the $q$ sons which partition the node $S$, i.e. $\pi(S) = T_1 \sqcup ..T_u.. \sqcup T_q$. Provide the simply connected sets of $\mathcal{P}(E)$ with an arbitrary energy $\omega$, and extend it from $\mathcal{P}(E)$ to the set $\mathcal{D}(E)$ of all partial partitions by using the sums

$$\omega(\pi(S)) = \omega(T_1 \sqcup ..T_u.. \sqcup T_q) = \sum_1^q \omega(T_u). \tag{1.31}$$

Just as the sum-generated ones, the $\vee$-generated energies on the partial partitions are defined from an energy $\omega$ on $\mathcal{P}(E)$ followed by a law of composition, which is now the supremum.

$$\omega(\pi) = \omega(T_1 \sqcup ... \sqcup T_n) = \vee\{\omega(T_i)\}. \tag{1.32}$$

Both laws are indeed particular cases of the classical Minkowski expression

$$\omega(\pi(S)) = \left[ \sum_{u\in[1,q]} \omega(T_u)^\alpha \right]^{\frac{1}{\alpha}} \tag{1.33}$$

which is a norm in $\mathbb{R}^n$ for $\alpha > 0$. Even though over partial partitions $\mathcal{D}(E)$, it is no longer a norm, it yields $h$-increasing energies for all $\alpha \in [-\infty, +\infty]$:

**Proposition 1.19.** *Let $E \in \mathcal{P}(E)$, let $\omega : P(E) \to \mathbb{R}^+$ be a positive or negative energy defined on $\mathcal{P}(E)$. Then the extension of $\omega$ to the partial partitions $\mathcal{D}(E)$ by means of Relation (1.33) is $h$-increasing.*

*Proof.* We have to prove the relation (1.27), for any two partial partitions $\pi$ and $\pi'$ of $S$. When $\alpha \geq 0$, the mapping $\sqrt[\alpha]{*}$ on $\mathbb{R}^+$ is increasing and, according to Relation (1.33), the inequality $\omega(\pi) \leq \omega(\pi')$ implies

$$\sum_1^q [\omega(T_u)]^\alpha \leq \sum_1^{q'} [\omega(T'_u)]^\alpha \tag{1.34}$$

which in turn implies, for the same reason

$$\sum_1^q [\omega(T_u)]^\alpha + \omega(\pi_0) \leq \sum_1^{q'} [\omega(T'_u)]^\alpha + \omega(\pi_0) \tag{1.35}$$

hence $\omega(\pi_1 \sqcup \pi_0) \leq \omega(\pi_2 \sqcup \pi_0)$.

When $\alpha \leq 0$, the sense of the inequality changes in relations (1.34) and (1.35) but changes again when taking the $\sqrt[\alpha]{*}$ in (1.35), which again lead to $\omega(\pi_1 \sqcup \pi_0) \leq \omega(\pi_2 \sqcup \pi_0)$, and achieves the proof. □

Note that the relation 1.33 preserves order: the optimal cut does not change when the energies of the classes are multiplied by the same constant. One can easily check that the proposition remains true when $\omega : P(E) \to \mathbb{R}^-$ is a negative energy. Some particular cases of $\alpha$ are of interest, namely

| $\alpha$ | $\omega(T_i)$ Composition Law | Applications |
|---|---|---|
| $-\infty$ | infimum | Ground truth energies [61] |
| $-1$ | harmonic sum | - |
| 0 | number of classes | CART [21] |
| $+1$ | sum | Salembier-Garrido, Guigues [47, 100] |
| $+2$ | quadratic sum | - |
| $+\infty$ | supremum | Valero, Veganzones, Soille [107, 118, 121] |

which all provide $h$-increasing energies.

**Corollary 1.20.** *If $\{\alpha_j, j \in J\}$ stands for a family of non negative weights, then the weighed sum $\sum \alpha_j \omega_j$ and supremum $\bigvee \alpha_j \omega_j$ of $h$-increasing energies $\omega_j$ turn out to be $h-$increasing.*

Unconstrained optimization of the Lagrangian function corresponding to a constrained optimization problem on the BOP is a case corresponding to this corollary demonstrated in chapter 2. A number of other laws are compatible with $h$-increasingness, such as multiplication.

**Minkowski's norm or power mean**: Minkowski's norm was used by Allene et al. [3], to relate the Maximum Spanning Forest(MaxSF), to the graph-cut on edge weighted graphs, with source and sink labels . Cousty et al. related the Minimum spanning forests and Watershed-Cuts [31] again on edge weighted graphs. This transition between various algorithms are done by using the q-th power on the gradient weighted edges of the graph. This preserved the ordering required for the MaxSF and the watershed-cut, for the limiting value of $q \to \infty$, while producing the graph-cut case for $q = 1$. This was further generalized by the seminal work of C.Couprie et al. by introducing the Powerwatershed framework [28], further regrouping the Random walker algorithm by Sinop-Grady[106], into a compact energy minimization framework, parameterized by exponents on the weights on the edges, and its coefficients. In particular when $q = 2$, the power watershed leads to a multi-label, scale and contrast invariant, unique global optimum obtained in practice in quasi-linear time [29].

In our case though, we use the Minkowski-norm to generalize the $h$-increasing composition laws, and provide a way to parametrize this choice, and explore this a bit further in the next chapter, for penalty based constrained optimization.

**Significance of exponential parameter** $\alpha$: The value $\alpha$ basically provides a way to determine a scale of a partition by fixing the energy values of the parents and child classes. This parameter $\alpha$ has nothing to do with singularity, while it acts in a complementary way to choose the child or parent classes based on their energies.

A number of other laws are compatible with $h$-increasingness, outside the Minkowski norm generalization, such as the alternating sum-sup composition laws, demonstrated further down the chapter. Application involving the inf and sum compositions are demonstrated for in Section 3.

### 1.6.1 Soille's Constrained Connectivity and Hierarchies

In this subsection we demonstrate the different ways of enforcing uniqueness while using different non-linear compositions especially the case of $\alpha$-flat zones by Soille et al. [108] and Akcay-Aksoy [2]. Both of these methods choose an optimal cut which ensure uniqueness and monotonicity conditions.

We provide a quick recall of $\alpha$-connected components or the quasi flat zones [107, 108]. Following the minimum dissimilarity metric to define a single linkage, Soille et al. define the $\alpha$-connected component to be connected sets within which there is at least one path with a difference in function bounded by $\alpha$ along each pair of points along the path. Such a path based gradient definition can be defined in $\mathbb{R}^2$ while one defines this on a more accessible pixel-graph with 4-adjacency. This min-metric dissimilarity yields an ultrametric distance and thus a hierarchy of partitions.

$$\alpha\text{-CC}(x) = \{x\} \cup \{y | \exists P(x \to y) : \quad \forall x_i \in P(x \to y)\} \wedge x_i \neq y, \quad d(x_i, x_{i+1}) \leq \alpha. \tag{1.36}$$

The $\alpha$-connected component for a hierarchy with increasing $\alpha$. In other words,

$$\alpha\text{-CC}(x) \subseteq \alpha'\text{-CC}(X), \qquad \forall \alpha \leq \alpha'$$

The problem with such connected components, like with single linkage is the chaining effect which produces very long chains with small path-wise differences, even when the global contrast (max-min) maybe we large. For such cases Soille et al defines the $(\alpha, \omega$-connected component containing a point $x$ is the largest $\alpha_i$-connceted component containing $x$ with its global range $(\sup f(y) - \inf f(y), y \in \alpha_i\text{-CC}(x)) \leq \omega$.

$$(\alpha, \omega)\text{-CC}(x) \subseteq (\alpha', \omega')\text{-CC}(X), \qquad \forall \alpha \leq \alpha', \omega \leq \omega'$$

#### 1.6.1.1 $(\alpha, \omega)$-components composed by supremum

The simplest $\vee$-composition of energies consists like comparing the supremum of the energies of the classes in the child w.r.t the parent energy. Here we will see that in the $(\alpha, \omega)$-connected component hierarchy, the maximal $(\alpha, \omega)$-component is a choice between parent component and child components that binarized global range parameter $\omega$, and picks the largest. We are lucky here that the $\omega$ corresponding to our energy over partial partitions and the $\omega$ for the global contrast term in $\alpha, \omega$-components are the same!

Consider a binary $\vee$-energy $\omega$ such that for all $\pi, \pi_0, \pi_1, \pi_2 \in \mathcal{D}(E)$ we have

$$\omega(\pi(S)) = 1 \quad \Rightarrow \quad \omega(\pi(S) \sqcup \pi_x) = 1, \tag{1.37}$$

$$\omega(\pi_1(S)) = \omega(\pi_2(S)) = 0 \quad \Rightarrow \quad \omega(\pi_1(S) \sqcup \pi_x) = \omega(\pi_2 \sqcup \pi_x). \tag{1.38}$$

This binary $\vee$-energy is obviously $h$-increasing. A numerical function $f$ is now associated with hierarchy $H$. Consider the range of variation $\delta(S) = \max\{f(x), x \in S\} - \min\{f(x), x \in S\}$ of $f$ inside set $S$, and the $h$-increasing binary energy $\omega^k(\langle S \rangle) = 0$ when $\delta(S) \leq k$, and $\omega^k(\langle S \rangle) = 1$ when not. Compose $\omega$ according the law of the supremum, i.e. $\pi = \sqcup \langle S_i \rangle \Rightarrow \omega^k(\pi) = \bigvee_i \omega^k(\langle S_i \rangle)$. Then the class of the optimal cut at point $x \in E$ is the larger class of $H$ whose range of variation is $\leq j$. When the energy $\omega^k$ of a father equals that of its sons, one keeps the father when $\omega^k = 0$, and the sons when not. As $k$ varies a climbing family is generated.

### 1.6.2 Dominant ancestor by supremum

Here is an example of ordered energy due to H.G.Akcay and S. Aksoy [2] who study airborne multi-bands images and introduce (up to a small change) $\mu(S)$ =Area $(S) \times$ (mean of all standard deviations of all bands in $S$). They work with energy maximization. Allocate a non negative measure $\mu(S)$ to each node of a hierarchy $H$, where $\mu$ takes its values in a partially ordered set $M$, such as a color space. The energy $\omega$ is ordered by the two conditions

$$\omega(S) \leq \omega(S') \;\Leftrightarrow\; S \supseteq S' \;\&\; \mu(S) \geq \mu(S') \quad S\,,\, S' \in \mathcal{P}(E),\ \mu \in M. \tag{1.39}$$

The node $S^*$ of the optimal cut at point $x$ is the highest more energetic than all its descendants. The optimal cut $\pi^*$ is obtained in one pass, by Guigues' algorithm [49].

The dominant ancestor pruning is also seen in the maximum decision rule by Valero et al. [118]. The pruning decision consists of a maximum decision rule which considers that a node is removed if and only if all its descendant nodes can be removed.

### 1.6.3 Composition of $\vee$-generated energies

Though the weighted supremum of $\vee$-generated energies is $h$-increasing (Prop. **??**), the infimum is not. In practice, this half-result is nevertheless useful, since the $\vee$, paradoxically, expresses the intersection of criteria. For example, when the function $f$ to optimize is color, one can take for energies:

- $\omega_1(S) = 0$ when range of luminance in $S < k_1$, and $\omega_1(S) = 1$ when not,
- $\omega_2(S) = 0$ when range of saturation in $S < k_2$, and $\omega_2(S) = 1$ when not.

Then the $h$-increasing energy $\omega_1(S) \vee \omega_2(S) = 0$ when $S$ is constant enough for both luminance and saturation.

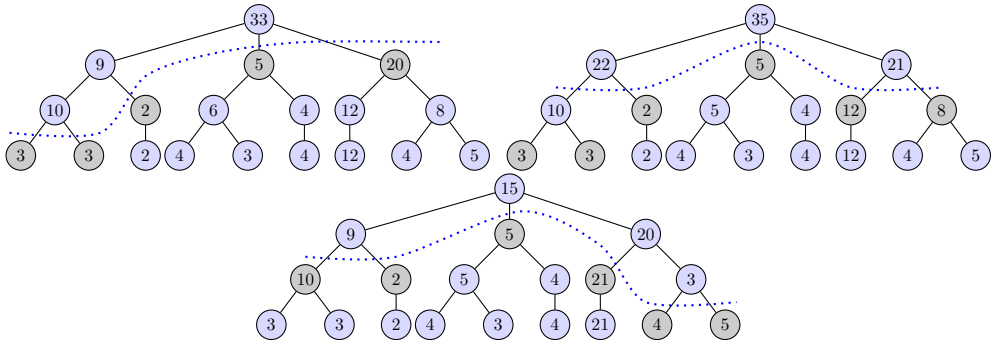



FIGURE 1.17: Optimal cuts for composition laws: addition, $(\alpha, \omega)$-component supremum and Akcay's refinement ordering. For the composition by addition, one compares the energies of parent and sum of energies of child nodes. For the binary energy based supremum, in the current example each node consists of $\alpha_i$-components and the value withing each component is the global contrast parameter $\omega$. Here we demonstrate a cut for which the $\omega \leq 20$. Though this is not a direct composition by supremum. Finally in the Akcay's refinement ordering example the optimal cut picks nodes whose descendants are all smaller that itself. This requires a two pass algorithm. According to the application other laws may be used e.g. both supremum and infimum for the proximity of ground truth with Hausdorff distances [61]. It is interesting to note that the uniqueness conditions in all the 3 cases have been assured by choosing the smallest/largest amongst the optimal cuts in the different cases.

Figure 1.17 shows optimal cuts for three different laws of composition. In a) the additive mode chooses the father $S$, when $\omega(S) \leq \sum \omega(T_j)$. In b) the mode by supremum chooses the $S$, when $\omega(S) \leq \vee\ \omega(T_j)$. Finally, in c) one takes the largest node which is more energetic than all its descendants (maximization of $\omega$).

### 1.6.4 $h$-increasingness and generalizing DP

Now we describe how the $h$-increasingness from equation (1.24) provides a non-linear and multi-scale generalization of the dynamic program (DP) in equation (1.17).

The *separability* condition defined by Guigues is generalized for non-linear compositions while still maintaining the monotonicity of cuts. $h$-increasingness preserves the order of energies across partial partitions during the intermediary steps DP. This enables one to interleave different composition of energies, thus creating a richer way to describe multi-scale energies.

For example consider the following multi-scale energies:

$$\omega(\pi(S)) = \begin{cases} \sum_{a\in\pi(S)} \omega(a), & \text{if Level}(\pi) \leq \frac{N}{2} \\ \sup_{a\in\pi(S)} \omega(a), & \text{otherwise.} \end{cases} \tag{1.40}$$

This energy of partial partition $\pi(S)$ is linearly composed for all partial partitions that are contained in partitions whose level in the hierarchy is below $N/2$, where $N$ corresponds to the total number of levels. Above $N/2$ all partial partitions energies are composed supremum.

### 1.6.5 Composition by alternating sum-supremum

We see in figure 1.18 that the optimal cut in such a case is obtained using $h$-increasingness condition even when we alternate with compositions of addition and supremum. This is very similar to max-pooling in neural networks. One important observation is that the sum-supremum optimal cut is lower-bounded by the optimal cut by supremum, while upper-bounded by the optimal cut by addition. This is the case since

$$\sum_{T_i\in\pi(S)} \omega(T_i) \geq \bigwedge_{T_i\in\pi(S)} \omega(T_i) \tag{1.41}$$

One method to be able to obtain set of cuts between the those of addition and supremum is to alternate the composition rules, but this would still have the problem of choosing parents when supremum is smaller but the sum is not, and vice versa. To this effect we can also consider the $h$-increasing $\alpha$-cuts from equation 1.33, that has no effect on the supremum, but produces cuts that are finer than the composition by sum. By alternating the composition laws one can span finer cuts between the composition by sum and supremum. One can refer to [71] where the Masci et al. proposes a neural

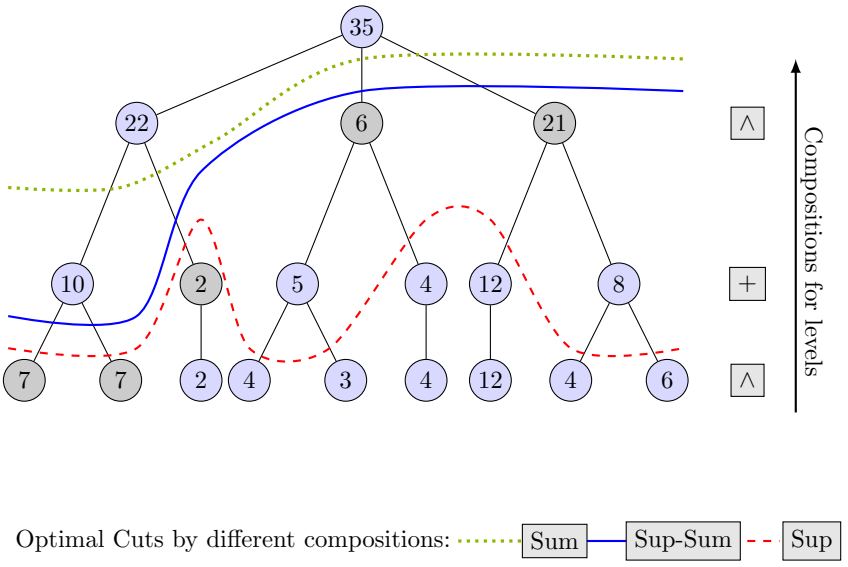



FIGURE 1.18: This demonstration shows an alternating composition: at each odd level we compose the energies by addition while at each even level we compose the energies by supremum. It's notable that the supremum of child energies is always smaller than or equal to their parent energies.

network structure that learns the morphological structuring element and the composition of operators. The pooling step in such operations are by supremum.

In summary $h$-increasingness condition applies to composition by addition, supremum as well as infimum. It helps generalize the optimal substructure in the dynamic program that chooses between parent or child based on their energies. $h$-increasingness applies in case of braids too, which requires a choice now between the monitoring supremum class and the set of child partial partitions being *monitored*. When the monitor hierarchy does not belong to the braid, the choice is just between the child partial partitions.

One can see an example of an alternating composition in Veganzones-Channusot et als. [121] work on the obtaining an optimally pruned binary partition tree. Here they propose multiple models to perform spectral-unmixing on the BPT, using the variants on the Lagrangian model. One can also find work on optimal pruning on the BPT for region based hyper-spectral image segmentation, by Valero [118]. This thesis demonstrate how one can extend compositions to a multidimensional setting.

We shall see an example of infimum in the applications involving partition extraction based on local Hausdorff measures.

## 1.7 Scale increasing families of energies

We first begin by comparing the energetic ordering $\preceq_\omega$ with that $\leq$ of the refinement, when both apply on the partitions $\Pi(E, B)$. At a first glance, combining these two structures is not very intuitive. This is because they do not hold on the same features of the partitions $\Pi(E, B)$. But we can enlarge the scope and consider a family $\{\omega(\lambda), \lambda \in \overline{\mathbb{R}}\}$ of singular energies which act on $\Pi(E, B)$. Each energy $\omega(\lambda)$ induces a minimal cut $\pi^*(\lambda)$.

*Remark* 1.21. How are the minimal cuts structured as scale parameter $\lambda$ varies?

In other words what conditions of monotonicity can we establish in the given energetic lattice structure.

### 1.7.1 Scale increasingness

We will define now a generalized monotonicity property, namely scale increasingness of the family $\{\omega_\lambda\}$:

**Definition 1.22.** *A family $\{\omega(\lambda), \lambda \in \overline{\mathbb{R}}\}$ of energies on the partial partitions $\mathcal{D}(E)$ of $E$ is* scale increasing *when*

$$\lambda \leq \mu \quad \textit{and} \quad \omega(\{S\}, \lambda) \leq \omega(a, \lambda) \Rightarrow \quad \omega(\{S\}, \mu) \leq \omega(a, \mu), \quad S \in \mathcal{P}(E), \ a \sqsubseteq \{S\} \tag{1.42}$$

These inequalities become strict when the scale increasing $\omega(\lambda)$ are singular energies.

**Theorem 1.23.** *Let $\{\omega(\lambda), \lambda \in \overline{\mathbb{R}}\}$ be a family of singular energies acting on a braid $B = H \cup R$, and let $\{\pi^*(\lambda), \lambda \in \overline{\mathbb{R}}\}$ the minimal cuts of the energetic lattices $\{\Pi(\omega(\lambda), B), \lambda \in \overline{\mathbb{R}}\}$. When the family $\{\omega(\lambda), \lambda \in \overline{\mathbb{R}}\}$ is scale increasing, then the least upper bound $S_{\min}(x \mid \lambda)$ of each class $S^*_\lambda(x)$ of $\pi^*(\lambda)$ increases with $\lambda$ at all points $x \in E$ i.e.*

$$\lambda \leq \mu \quad \Rightarrow \quad S_{\min}(x \mid \lambda) \subseteq S_{\min}(x \mid \mu) \qquad x \in E \quad \lambda, \mu \in \overline{\mathbb{R}}. \tag{1.43}$$

*and the partitions $\pi_{\min}(\lambda) = \sqcup\{\{S_{\min}(x \mid \lambda)\}, x \in E\}$ form the hierarchy $H_{\min} = \{\pi_{\min}(\lambda), \lambda \in \overline{\mathbb{R}}\}$*

*Proof.* Put $S_{\min}(x \mid \lambda) = S_\lambda(x)$ and $S_{\min}(x \mid \mu) = S_\mu(x)$. As $H$ is a hierarchy, we have either $S_\lambda(x) \subseteq S_\mu(x)$, or $S_\mu(x) \subset S_\mu(x)$. If $S_\mu(x) \subset S_\mu(x)$ Proposition 1.8 shows that there exists a partial partition $a_\mu$ of $\pi^*_\mu$ of support $S_\lambda(x)$. The $\omega(\lambda)$-energy of $a_\mu$ is $\geq \omega(\{S_\lambda(x), \lambda\})$, because $S_\lambda(x)$ is a class of the minimal cut $\pi^*(\lambda)$ and $a_\mu \sqsubseteq \{S_\lambda(x)\}$. Then we have by scale increasingness $\omega(\{S_\lambda(x)\}, \mu) \leq \omega(a_\mu, \mu)$. On the other hand, as $\pi^*_\mu$ is the

minimal cut for the $\omega(\mu)$-energetic lattice $\Pi(\omega(\mu), B)$, the p.p. $a_\mu$ is less$\omega(\mu)$-energetic than its support $\{S_\lambda\}$, i.e. $\omega(a_\mu, \mu) \leq \omega(\{S_\lambda(x), \mu\})$, thus $\omega(a_\mu, \mu) = \omega(\{S_\lambda(x), \mu\})$. But such an equality contradicts the singularity of $\omega(\mu)$. Therefore we have $S_\lambda(x) \subseteq S_\mu(x)$. As $\lambda$ increases, this inclusion characterizes the cone at point $x$ of a hierarchy, namely that of the partitions $\pi_{\min}(\lambda) = \sqcup\{\{S_\lambda(x)\}, x \in E\}$, which achieves the proof. □

The *if* part of the theorems already appears in [49] for linear energies. Note that we also have $\pi^*(\lambda) \preceq_{\omega(\lambda)} \pi^*_\mu$ by scale increasingness (1.42).

**Corollary 1.24.** *The family $B = H_{\min} \cup \{\pi^*(\lambda), \lambda \in \overline{\mathbb{R}}\}$ is a braid.*

**Corollary 1.25.** *When a braid reduces to a hierarchy $H$, then the minimal cuts generated on $H$ by the family $\{\omega(\lambda), \lambda \in \overline{\mathbb{R}}\}$ form a hierarchy, i.e.*

$$\lambda \leq \mu \quad \Rightarrow \quad \pi^*(\lambda) \leq \pi^*_\mu \qquad \lambda, \mu \geq 0. \tag{1.44}$$

*if and only if the energies $\{\omega(\lambda), \lambda \in \overline{\mathbb{R}}\}$ are singular and scale increasing.*

*Proof.* For the *if* part we observe that Rel. (1.43) becomes $S(x \mid \lambda) \subseteq S(x \mid \mu)$, thus $\pi^*(\lambda) \leq \pi^*_\mu$. For the *only if* part, suppose that there exists a set $S \subseteq E$ for which $\omega(\{S\}, \lambda) \leq \omega(\pi, \lambda), \pi \in \Pi(S, H)$ does not imply $\omega(\{S\}, \mu) \leq \omega(\pi, \mu)$, hence implies $\omega(\{S\}, \mu) > \omega(\pi, \mu)$. It means that $\{S\}$, which is a class of $\pi^*(\lambda)$, is replaced by $\pi$ in $\pi^*_\mu$, so that $\pi^*(\lambda) \not\leqslant \pi^*_\mu$, which achieves the proof. □

The theorem 1.23 was stated for a scalar parameter $\lambda$. It extends, however, to the vector case. Let $\boldsymbol{\lambda}$ (in bold) be a positive vector in $\mathbb{R}^p$, and $\{\lambda_1, \lambda_2, ...\lambda_p\}$ its $p$ positive coordinates. The relation

$$\boldsymbol{\lambda} \leq \boldsymbol{\lambda}' \quad \Leftrightarrow \quad \lambda_i \leq \lambda'_i \qquad 1, ...i...p$$

defines a partial ordering. We can go from $\lambda$ to $\lambda'$ in $p$ "scalar" steps, by firstly changing only $\lambda_1$ into $\lambda'_1$, then $\lambda_2$ into $\lambda'_2$,, so on. Between two steps of the sequence we have,

$$\{\lambda_1, \lambda_2, \lambda_3 ...\lambda_p\} \leq \{\lambda'_1, \lambda_2, \lambda_3, ..., \lambda_p\} \leq \{\lambda'_1, \lambda_2, \lambda_3, ..., \lambda_p\} \leq \{\lambda'_1, \lambda'_2, \lambda_3, ...\lambda_p\} ... \leq \{\lambda'_1, \lambda'_2, \lambda'_3, ..., \lambda'_p\},$$

the vector variation reduces to a scalar one and the theorem 1.23 applies, so that we we can state: $\boldsymbol{\mu}$

**Corollary 1.26.** *The relations (1.43) and (1.44) extend to the case when* $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, ...\lambda_p\}$ *and* $\mu = \{\mu_1, \mu_2, ...\mu_p\}$ *are positive vectors to the Euclidean space* $\mathbb{R}^p$.

The following proposition shows how to easily construct scale increasing families:

**Proposition 1.27.** *When the map* $\lambda \to \omega(\lambda)$ *is increasing, then the family* $\{\omega(\lambda)\}$ *is scale increasing.*

*Proof.* For $\lambda \leq \mu$ and $a \in \Pi(S, B)$, we have $\omega(\lambda, S) \leq \omega(\mu, S)$ and $\omega(\lambda, a) \leq \omega(\mu, a)$. By difference, it comes $\omega(\lambda, a) - \omega(\lambda, S) \leq \omega(\mu, a) - \omega(\mu, S)$. Hence, when $\omega(\lambda, a) - \omega(\mu, S) \geq 0$, then $\omega(\mu, a) - \omega(\mu, S) \geq 0$, i.e. the axiom (1.42). □

Usual energies, like $\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial$, $\omega(\lambda) = \omega_\varphi \vee \lambda\omega_\partial$, or $\omega(\lambda) = \omega_\varphi \wedge \lambda\omega_\partial$ lead thus to hierarchies of minimal cuts. This nice property can be used to compress a hierarchy by reducing the number levels in a significant manner.

### 1.7.2 Scale space properties

The theorem 1.23 indicates, indirectly, that the scale increasing families might build up causal scale-spaces, in Alvarez et Al. sense [4], i.e. induce semi-groups of operators.

Now a semi-group can only be defined when the starting and arrival spaces are the same. It is not the case with the mapping $B_0 \to \pi^*(B_0)$. For generating a convenient space, we have to start from the classes $\mathcal{S}$ of $B_0$, and to introduce the set $\mathcal{B}(\mathcal{S}) = \mathcal{B}$ of all braids whose classes belong to $\mathcal{S}$. Next, we need to provide $\mathcal{B}$ with an ordering relation which extends to braids the set-wise refinement ordering. We can state that $B_1 = (H_1, R_1)$ is smaller than $B_2 = (H_2, R_2)$ when $H_1 \leq H_2$ and $R_1 \subseteq R_2$. The first inequality means that at each level $i \in I$ the partition $\pi_1(i) \in H_1$ is smaller than its homologue $\pi_2(i) \in H_2$; the second inequality means that there are less partitions in the supplement family $R_1$ than in $R_2$. This relation is clearly the matter of an ordering $\dashv$ which generates a lattice on $\mathcal{B}$, where the supremum of $B_1$ and $B_2$ is $B = [(H_1 \vee H_2), (R_1 \cup R_2)]$, (dual relation for the infimum).

Let $\omega$ be a singular energy acting on $B_0$, and giving the $\omega$-energetic minimal cut $\pi^*$. We observe that $\pi^*$ is the same for all braids $\mathcal{B}$, as it depends only on the classes $\mathcal{S}$ of $B_0$. Transform the braid $B \in \mathcal{B}$, of monitor $H$, in the following manner: if at level $i$ the section $\pi(i)$ of the monitor $H$ is $\leq \pi^*$ (for the refinement) replace it by $\pi^*$; if $\pi(i) \not\leq \pi^*(\lambda)$ don't change it, and in all cases keep unchanged the additional partitions $R$ of $B$. We denote this operation by $\varphi_\omega(B) = B \otimes B^*$.

**Proposition 1.28.** *When the energy $\omega$ is singular, the operation $\varphi_\omega(B) = B \otimes B^*$ from $\mathcal{B}$ into itself is a closing.*

*Proof.* The map $\varphi_\omega$ is obviously extensive. It is also idempotent because all braids of $\mathcal{B}$ admit the same minimal cut $\pi^*$, so that $\varphi_\omega(B) = B \otimes B^* \otimes B^* = B \otimes B^*$. Finally, it is increasing. Let $i_1$ and $i_2$ be the first levels of $H_1$ and $H_2$ where the section is not replaced by $\pi^*$. If $B_1 \subseteq B_2$, then we have $i_1 \geq i_2$. Above $i_1$, the sections are those of $H_1$ and $H_2$; between $i_1$ and $i_2$ the sections of $H_1$ are $\pi^*$ and those of $H_2$ are $\geq \pi^*$, and below $i_2$ the sections of both $H_1$ and $H_2$ are $\pi^*$. In all cases $\varphi(B_1) \subseteq \varphi(B_2)$, which achieves the proof. □

**Corollary 1.29.** *Let $\{\omega(\lambda), \lambda \in \overline{\mathbb{R}}\}$ be a scale increasing family of singular energies acting on the braids $B \in \mathcal{B}$ . The closings $\{\varphi_\lambda, \lambda \in \overline{\mathbb{R}}\}$ of the proposition form the Matheron semi-group[4], i.e.$\lambda$,*

$$\varphi_\lambda \varphi_\mu = \varphi_\mu \varphi_\lambda = \varphi_{\max(\lambda,\mu)}. \tag{1.45}$$

*Proof.* The semi-group from equation (1.45) is equivalent to the implication $\lambda \leq \mu \Rightarrow \varphi_\lambda \leq \varphi_\mu$, which is a direct consequence of Rel.(1.44). □

The meaning of the semi group is the following: $\varphi_\lambda$ replaces the lower sections of $H$ by a cylinder of section $\pi^*_\lambda$. As the parameter $\lambda$ increases, the cylindric part develops upwards, and one can start form any "half cylindric" transform $\varphi_\lambda(B)$ to get the transform $\varphi_\mu(B)$, as soon as $\lambda \leq \mu$.

### 1.7.3 The Scale Function $\Lambda$

We call Scale function the scale of appearance of each class $S$ in a singular and scale increasing family $\{\omega_\lambda, \lambda \geq 0\}$. As the parameter $\lambda$ increases, there happens a first $\lambda$ such that $\omega_\lambda(\{S\}) < \omega_\lambda(\pi)$, $\pi \in \Pi(S, B)$. Following Guigues [49], we denote by $\lambda^+(S)$ this first scale of appearance of $S$. The leaves, which have no descendants, are given $\lambda(S) = 0$.

Now each point $x \in E$ labels nested sets $\{S_i(x)\}$ in the hierarchy. The axiom of scale increasingness shows that the class $S_i(x)$ is a candidate to participate in a minimal cut $\pi^*_\mu$ if its scale of appearance is $\lambda^+(S_i(x)) \leq \mu$ and if the scale of appearance of each $S_j(x), j > i$ is $\nu^+ > \mu$. Then $S_i(x)$ will effectively belong to the minimal cuts for all scales $[\lambda^+, \wedge\nu[$ since in this interval it is not covered by a larger class $S_j(x), j > i$.

[4]There are two main types of scale spaces semi-groups used in scale space studies. First is the linear semi-group, based on a vector space. Second is the semi-group of Matheron's granulometries [72] which uses an underlying lattice for analysis, and where the most active transformation imposes its law.

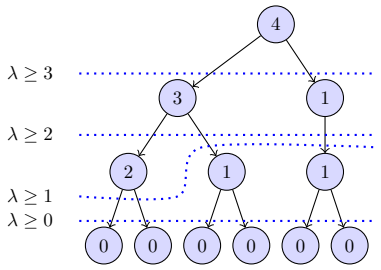



FIGURE 1.19: An example of a scale function. For each $\lambda$ the classes which are hold are just above the dotted line. Note that the class on the right branch with value 1, repeated twice correspond the same same class, and a cut having one or the other are the same cuts.

## 1.8 Inf-modularity

In this intermediary section we describe a property of energies defined on the family of partial partitions akin to sub-modularity for set based functions, which have the property of diminishing returns. Inf-Modularity will be later useful in describing the constraint functions. We will later continue with the resolution of the Lagrangian constrained optimization problem resolution with energetic lattices in section 2.7.

Here we detail the property of the energies on any family of p.p., and does not necessary imply those of braids or hierarchies. We saw the interest in defining scale increasing families $\{\omega_\lambda\}$. Can we define a monotonicity property on a more general class of partial partitions, akin to the discrete concavity sense of sub-modularity [12]?

**Definition 1.30.** *An energy $\omega_\partial : \mathcal{D}(E) \to \mathbb{R}^+$ is said inf-modular when for each p.p. $\pi$ of support $S \in \mathcal{P}(E)$ we have*

$$\omega_\partial(\{S\}) \leq \bigwedge\{\omega_\partial(a), a \sqsubseteq \{S\}\}. \tag{1.46}$$

The inf-modularity resembles the singularity axiom a lot (1.9), but at the same time is different. Firstly, the inequality in equation (1.46) is not strict, and secondly it is a monotonic property all supports $S \in \mathcal{P}(E)$, unlike singularity which just requires distinct energies. Note that if we replace the infimum of (1.46) by a sum, the condition becomes less severe, which is that of sub-additivity.

### 1.8.1 Inf-modularity and scale increasingness

For the Lagrange type energies given by equation (2.22), the two notions of scale increasingness and of inf-modularity coincide, but the latter applies to $\omega_\partial$ only. More precisely, we can state:

**Proposition 1.31.** *The family* $\{\omega_\lambda = \omega_\varphi + \lambda\omega_\partial, \lambda \geq 0\}$ *is scale increasing if and only if* $\omega_\partial$ *is inf-modular.*

*Proof.* If $\omega_\partial$ is inf-modular, and $\mu > \lambda$, we have $(\mu - \lambda)\omega_\partial(\{S\}) \leq (\mu - \lambda)\omega_\partial(\pi)$, If in addition $\omega_\lambda(\{S\}) \leq \omega_\lambda(\pi)$, then by summing the two inequalities, we obtain $\omega_\mu(\{S\}) \leq \omega_\mu(\pi)$, and the scale increasingness is satisfied. Conversely, if the implication (1.42) holds, then by taking the difference between $\omega_\varphi(\{S\}) + \mu\omega_\partial(\{S\}) < \omega_\varphi(\pi) + \mu\omega_\partial(\pi)$ and $\omega_\varphi(\{S\}) + \lambda\omega_\partial(\{S\}) < \omega_\varphi(\pi) + \lambda\omega_\partial(\pi)$, we find $(\mu - \lambda)\omega_\partial(\{S\}) \leq (\mu - \lambda)\omega_\partial(\pi)$, i.e. Rel.(1.46), which achieves the proof. □

The "only if" part of Proposition 1.31 is specific to the $\omega_\varphi(\pi) + \lambda\omega_\partial(pi)$, type energies. In short, inf-modularity $\Leftrightarrow$ Causality (or in our words, additive scale-increasingness), while inf-modularity $\Longrightarrow$ scale-increasingness. For a family such as $\{\omega_\lambda = \omega_\varphi \vee \lambda\omega_\partial\}$ for example, the inf-modularity of $\omega_\partial$ implies the scale increasingness of the $\{\omega_\lambda\}$, but the converse is false.

**An example of inf-modularity**: The example refers to the $\alpha - \omega$-hierarchy by Soille et al. [107, 108]. They have indicated several variants, which all rest on a same idea. A family of previous segmentations of a $2-D$ function $f$ led to hierarchy $H$. One requires the largest classes where function's global contrast is bounded, i.e. at each node $S$ we have the energy $\omega(S) = \sup\{f(S)\} - \inf\{f(S)\}$. The values of $f(S)$ obviously increase monotonically on the hierarchy. A node $S$ is kept when $\omega(S) \leq 20$). The minimal cut is then the union of the largest remaining nodes. see figure 1.17 where this supremum composition over $\omega \leq 20$, is demonstrated. One stops, climbing the hierarchy locally when the global contrast reaches the constraint 20.

### 1.8.2 Inf-modularity Vs Sub-additivity

The concept of inf-modularity we just introduced is to be comparable with that of sub-additivity. Remember that an energy $\omega$ is sub-additive when, energies over parent class is always smaller than or equal to additive composition of energies on classes of child partial partition. This inequality when $\omega$ is not restricted to positive functions, gives sub-modularity [12] which serves as substitute for the convexity when dealing with the subsets of $E$. As inf-modularity is applied to the partial partitions of $E$, we firstly need to introduce some energy $\omega'_\partial$ on sets that corresponds to $\omega_\partial$, by putting

$$\omega'_\partial(S) = \omega_\partial(\{S\}), \qquad S \in \mathcal{P}(E), \quad \{S\} \in \mathcal{D}(E)$$

with $\omega'_\partial(\varnothing) = 0$. Then we must match sets and partial partitions in some sense. To do this, lets consider the comparison of a p.p. $\pi$ with its classes $T_j, 1 \leq j \leq p$. If we take

$$\omega_\partial(\pi) \leq \sum_{j=1}^{p} \omega_\partial(\{T_j\}) = \sum_{j=1}^{p} \omega'_\partial(T_j), \tag{1.47}$$

then the inf-modularity of $\omega_\partial$ yields inequality

$$\omega'_\partial(S) = \omega_\partial(\{S\}) \leq \omega_\partial(\pi) \leq \sum_{j=1}^{p} \omega_\partial(\{T_j\}) = \sum_{j=1}^{p} \omega'_\partial(T_j),$$

with $\pi \in \Pi(S, B)$, which is nothing but the sub-additivity of $\omega'_\partial$ (i.e. the relation $\omega'_\partial(A \cup B) + \omega'_\partial(A \cap B) \leq \omega'_\partial(A) + \omega'_\partial(B)$ with here $A \cap B = \varnothing$). Now we can state:

**Proposition 1.32.** *Let $\omega'_\partial : \mathcal{P}(E) \to \mathbb{R}^+$ be a sub-additive energy on the sets of $E$. Any extension $\omega_\partial$ of $\omega'_\partial$ to the partial partitions of $E$ which satisfies the inequality (1.47) is inf-modular. Conversely, the restriction $\omega'_\partial$ to sets of an inf-modular energy $\omega_\partial$ is sub-additive.*

For example, in a partition of the plane $\mathbb{R}^2$ the perimeters $\omega'_\partial$ of the classes generate an inf-modular energy $\omega_\partial$ on the partial partitions. We passed from partial partitions to sets, and vice versa, by the relation (1.47) which restricts the set approach. This formulation, defined over partial partitions (and no longer on sets) using inf-modularity, frees ourselves from this limitation.

## 1.9 Summary

**Chapter contribution summary**

- We introduced a new hierarchical structure, namely, the braid of partitions (BOP), that is any general family of partitions whose, where the supremum of any two partial partitions pairs are disjoint or nested. The corresponding hierarchy is called the monitor hierarchy of the braid.
- We introduced the Energetic ordering on partial partitions, and the associated Energetic Lattice, which helps define an infimum in the space of partitions.
- $h$-increasingness was used to formally generalize the dynamic program to a very wide variety of linear and non-linear compositions of energies of partial partitions. Furthermore dynamic program for the new BOP structure, remains the same, while providing a larger search space.
- Unconstrained optimization over the space of partitions using the energetic lattice was the main structure one studies in this chapter, which generalizes various optimization criteria on HOP [2, 49, 100, 108], while being generally applicable to linear and non-linear energies. Chapter 2 will discuss constrained optimization using the energetic lattice.
- The scale-increasing family of energies provide a general multi-scale monotonicity conditions, which generalize Guigue's multi-scale conditions [49]. The energies are defined on the larger family of partial partitions belonging to a braid.
- The four axioms of singularity, $h$-increasingness, scale-increasingness, and inf-modularity are conditions on energies only, and not on the image domain $E$. This space $E$ maybe topological space, vector space, discrete, or neither of these. Therefore, the notions of convexity, connectivity, are not needed for the main results of the chapter.
- The axioms introduced in this chapter serve the purposes of: Singularity $\rightarrow$ existence of energetic lattices and thus uniqueness, $h$-increasingness $\rightarrow$ local-global solution by dynamic programming, scale-increasingness $\rightarrow$ monotonically ordered optimal cuts, and finally inf-modularity is non-linear version of sub-additivity property, which will be used to describe constraint functions in constrained optimization problems.

# Chapter 2

# Constrained optimization

This chapter provides two different contributions[1], to the problem of constrained optimization on hierarchies:

**Chapter contribution summary**

- Firstly in section 2.2, we quickly recap Lagrangian multiplier methods, and the conditions of for optimality for the primal and dual Lagrangian. We show that the constrained optimization problem on HOP in literature are solved by applying Everett's theorem. We discuss how one can improve the lower bound of globally optimal solutions of $\lambda$-cuts by Guigues and Salembier-Garrido [13, 47, 100], by perturbing the Lagrangian. Further we demonstrate the use of a different $h$-increasing penalty function for the constrained optimization problem.
- From section 1.8 onwards, we generalize the constrained optimization problem, at two levels, first by introducing the energetic lattice, and second by enlarging the feasible set by introducing the braids of partitions.

## 2.1 Review on Constrained optimization on Trees

As we have already remarked in the previous chapter, in this section 2.1 we study classification and regression trees(CART) by [21] as applied to the problem of constrained optimization on a generalized hierarchy of partitions (and not just rectangular partitions) [13, 42, 47, 100]. There has been a sequence of studies following CART in the domain of information theory to solve constrained optimization problem of rate-distortion [26]. As Gray et. al [26] explain, pattern recognition used distortion-rate bounds on classification

[1] Articles in prepartion

trees while, information theory use them to solve variable-rate source coding problems. Further, they explain that classification trees can be put into a data compression framework by thinking of the unknown class as a clean probabilistic source corrupted by observation noise and modifying the distortion measure accordingly. The use of tree based models for source coding has been critical in demonstrating how one can achieve rate constraints by alternating across multiple optimal subtrees, even though such a constraint might not exist statically for given distortion level or bandwidth constraint [95, 105].

The goal of this section is to study the Lagrangian optimization framework used in the different papers, and clarify the importance of the optima. Further we will see the connection with the Lagrangian dual and its importance as in the context of the rate-distortion theoretic [95] framework and the Scale-set framework [47].

### 2.1.1 Rate Distortion Theory

To give a quick introduction to rate-distortion theory, we will aim at defining the distortion-rate function, which is the distortion incurred for a given bandwidth or rate constraint on the communication channel.

$$D(R) = \inf_{P_{Y|X}} \{E[\rho(X,Y)] | I(X,Y) \leq R\} \tag{2.1}$$

where $X$ and $Y$ are random variables, where $Y$ represents the coded signal and $X$ the input signal. Here $\rho(X,Y)$ gives error measure, and $P_{Y|X}$ is the conditional probability source distribution. Then for a given channel one can define the rate $I(X,Y)$ and distortion $E[\rho(X,Y)]$ which are achieved by some coding scheme (not necessarily tree structured) [26]. The rate-distortion function is the solution to the constrained minimization problem in equation 2.1. In a tree structured coding scheme, Gray et al. [26] prunes a complete code tree $T$ to achieve pruned subtrees $S$ such that we have:

$$\hat{D}(R) = \min_{S \subset T} \{\delta(S) | l(S) \leq R\} \tag{2.2}$$

gives rate-distortion function for some pruned subtree of a given tree $T$.

### 2.1.2 Image segmentation formulation using Rate Distortion

The rate-distortion function was reformulated in the context of image segmentation where the constraint optimization problem now was defined on the hierarchy of partitions. This was studied first by [100] and in more detail in the thesis [42] over a binary partition tree structure. Further the thesis of Guigues [47] provides different conditions additivity and separability on energies that can be minimized in the above constraint optimization problem, the solution of which can be achieved by solving a dynamic program. This study has later been extended to the minimization of the Mumford-Shah energy on hierarchy of partitions generated from tree of shapes in [13]. We will review the optimization problem in the new formulation to understand their components:

Caselles et al. [13] view the distortion measure as mean square error between the original image and recovered image, $D = E[(f(x) - g(x))^2]$. Here for a maximum allowed distortion $D^*$ one can achieve a lower bound on the bit-rate given by $R(D^*)$. Conversely when bounding the rate we have achievable distortion given by $D(R^*)$. This gives us the following two constraint problems which are equivalent:

$$\min R(D), \text{subject to} \quad D \leq D^* \tag{2.3}$$

and while minimizing the distortion

$$\min_{B \in \mathcal{S}_b} D(B), \text{subject to} \quad R(B) \leq R(B^*(\lambda)) \tag{2.4}$$

where the rate $R(B)$ is measured in terms of the cost of encoding the curves, plus, the cost of encoding gray level values of the regions.

Caselles et al. [13] following Shoham-Gersho [105] use the Everett's theorem [39] to obtain the optimal solution $B^*(\lambda)$ by solving the unconstrained problem:

$$\min_{B \in \mathcal{S}_b} (D(B) + \lambda R(B)) \tag{2.5}$$

Shoham-Gersho [105] clearly suggest that if one can calculate the global minimum to unconstrained Lagrangian in equation 2.5 one has a solution for the constrained problem in 2.4. We will discuss this further showing that this is called the Everett's main theorem [105], and is used to solve constrained optimization problems where the objective and constraint functions are not necessarily continuous or derivable.

Shoham-Gersho and Garrido [42, 105] suggest the use of successive approximation methods. Garrido repeats a bottom up scan analysis while searching the Lagrangian parameter values in a secant iterative search to approximate the rate function $R$. While they achieve the desired rate with the smallest distortion possible, they in reality are achieving perturbed and thus approximate solution of the primal problem. In such cases the optimum corresponding to optimal $\lambda$ corresponds to one that achieves nearly exact rate.

For an understanding of the dual and primal problems and their relations in channel capacity-distortion optimization problems, we refer the reader to Chiang-Boyd [25, 34]. One can note here that duality is not always present inherently and can be overridden by achieving a *mapping* i.e by some simple mappings of signs, variables, constant parameters, and mathematical operations.

### 2.1.3 Tree structured Vector Quantization(TSVQ)

The BFOS algorithm [21] has been used to optimally prune tree structured vector quantizers(TSVQ). The reader is highly consulted to read the book on the subject by Gersho-Gray [43]. This formulation is one of the important motivations for the rate-distortion theoretical framework for constrained optimization and its application in image analysis, with a whole domain developing different ways to quantize signals using the tree structure. The resulting sequence of pruned subtrees would lie on the convex hull of the operational rate-distortion points.

In this section we remark that the *tree functionals* or energy being minimized are the expected distortion and expected length of codeword. That is the energy to minimize is given by [43]:

$$\bar{d} = E(d(\boldsymbol{X}, \hat{\boldsymbol{X}}) + \lambda \text{len}(i(\hat{\boldsymbol{X}})) = E(\rho(\boldsymbol{X}, \hat{\boldsymbol{X}})) \tag{2.6}$$

where len() represents the length of the code word, d represents the distortion measure defined on the random variables $\boldsymbol{X}$ and $\hat{\boldsymbol{X}}$.

The expected error has a quadratic form, which consequently is also a sub-additive function. The crucial difference to be noted here is that the energy functionals are defined on trees, pertain to a distribution of input signal/image values. The tree structure encodes the signal up to a required constraint, but over a distribution of values, thus allowing a tractable assumption of convexity the subsequent validity of application of the Lagrangian multipliers method.

The existence of a multiplier value for a given constraint function value, has been a problem in the domain of continuous optimization. But in case of TSVQ's applied using wavelet bases one achieves the convex hull of the Rate-distortion Curve by time sharing across two valid constraints to achieve the expected or average constraint value on the R-D curve. In most cases the multiplier method assures only an upper bound.

The purpose of this section on rate-distortion theory was to remind the reader that the constraint and objective functions are defined for distribution $\boldsymbol{X}, \hat{\boldsymbol{X}}$. One needs to be careful when one evaluates minimum using the multiplier method for a constraint function such as perimeter or number of classes, since these is no probabilistic model, and furthermore no assurance of achieving a constraint exactly.

## 2.2 Lagrangian Multipliers and Everett's Theorem

The Lagrange multipliers are associated with an optimization problem which is referred to as the Lagrangian dual[2], or simply dual, problem. The role of the dual problem is to define a largest lower bound on the primal value $d^*$ of the primal (original) problem. The important property of the dual is that its concave.

### 2.2.1 Reminder on Lagrange Multipliers

In continuous constrained optimization, with the space of solutions are n-dimensional points $\boldsymbol{x} \in \mathbb{R}^n$ one poses the constrained minimization problem as:

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{minimize}} \quad & f_0(\boldsymbol{x}) \\ \text{subject to} \quad & f_i(\boldsymbol{x}) \leq 0,\ i = 1, \ldots, m. \\ & h_i(\boldsymbol{x}) = 0,\ i = 1, \ldots, p. \end{aligned} \tag{2.7}$$

where, $f_0$ is called the objective function, while $f_i, h_i$ are constraint functions. The Lagrangian multipliers method consists in obtaining the unconstrained minima of a Lagrangian function. In the processing describing the minima of such a function, one has to have the gradient of the objective function, and the gradient constraint functions, scaled differently by a scalar multiplier. Following [20] Lagrangian is classically written as:

[2]The dual problem was discovered during the 1920s by John Von Neumann in matrix games, but had for a long time implicitly been used also for nonlinear optimization problems before it was properly stated and studied by Arrow, Hurwicz, Uzawa, Everett [39], Falk, Rockafellar, others, starting in earnest in the 1950s. The original problem referred to as the primal problem, was a name given by George Dantzig's father, a Greek scholar. [80]

$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \nu_i h_i(\boldsymbol{x})$$

The solution of the the problem in equation (2.7) requires derivatives

$$\nabla f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=p}^{m} \nu_i \nabla h_i(\boldsymbol{x}) = \mathbf{0} \tag{2.8}$$

We obtain the classical Lagrangian case when we have purely equality constraints, giving us $\nabla f_0(\boldsymbol{x}) + \sum_{i=p}^{m} \nu_i \nabla h_i(\boldsymbol{x}) = \mathbf{0}$. This is important to remark in the case of the scale-sets framework, since there can exist a constraint for which one cannot find a multiplier, but one can only find a lower bound.

In our expression for the energy we will use $\omega_\varphi + \lambda\omega_\partial + \nu\omega_\epsilon$ to represent the Lagrangian, with $\lambda$ the Lagrange multiplier, where $\omega_\varphi$ is objective function, to be minimized, and $\omega_\partial$ and $\omega_\epsilon$ are constraint functions.

### 2.2.2 The Relaxation Theorem

Given a constraint optimization problem of the form

$$f^* := \inf_{\boldsymbol{x}} f(\boldsymbol{x}), \text{subject to } \ \boldsymbol{x} \in S, \tag{2.9}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $S \subseteq \mathbb{R}^n$. One defines relaxation by the following:

$$f_R^* := \inf_{\boldsymbol{x}} f_R(\boldsymbol{x}), \text{subject to } \ \boldsymbol{x} \in S_R, \tag{2.10}$$

where $f_R : \mathbb{R}^n \to \mathbb{R}$ such that $f_R(\boldsymbol{x}) \leq f(\boldsymbol{x}), \forall \boldsymbol{x} \in S$, and $S_R \supseteq S$. For problem pair in equations (2.9, 2.10) we state the relaxation theorem [80] as:

**Theorem 2.1.** *(Relaxation Theorem)*

*(i) Relaxation:* $f_R(\boldsymbol{x}) \leq f(\boldsymbol{x})$

*(ii) In-feasibility: If the relaxed problem is infeasible so is the original problem.*

*(iii) Optimal relaxation: If 2.10 has an optimal solution* $\boldsymbol{x}_R^*$*, then*

$$\boldsymbol{x}_{\boldsymbol{R}}^* \in S, \text{and } \ f_R(\boldsymbol{x}_{\boldsymbol{R}}^*) = f(\boldsymbol{x}_{\boldsymbol{R}}^*), \tag{2.11}$$

*then* $\boldsymbol{x}_{\boldsymbol{R}}^*$ *is an optimal solution to original problem in equation (2.9).*

**Definition 2.2.** *We call a vector* $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ *a Lagrange multiplier if it is non-negative and if* $f^* = \inf_{\boldsymbol{x}\in X} L(\boldsymbol{x}, \boldsymbol{\lambda}^*)$*, for a given Lagrangian function* $L(\boldsymbol{x}, \boldsymbol{\lambda})$*.*

We can thus state the conditions on the existence of Lagrange multipliers from [80]:

(i) If there is no duality gap, then the set of Lagrange multipliers equals the set of optimal dual solutions (which however may be empty).

(ii) If there is a duality gap, then there are no Lagrange multipliers.

The relaxation theorem is stated here to understand the Everett's theorem. Lagrangian relaxation is employed in optimization problems to find approximate solutions, by decomposing the constraints to produce "easier" subproblems.

#### 2.2.2.1 Everett's Main, $\lambda$ and $\epsilon$ Theorems

Everett's seminal paper studies resource allocation problem as evident from its use by Shoham-Gersho [105] to study variable rate set quantizers. For the proofs of the $\lambda, \epsilon$ theorems refer the reader to the original paper by Everett [39].

**Theorem 2.3.** *(Everett's Main Theorem) Given the multiplier vector* $\boldsymbol{\lambda} \in \mathbb{R}^m \times \mathbb{R}^p$*, and the lagrangian function,*

$$\min_{\boldsymbol{x}\in X}\{f(\boldsymbol{x}) + \boldsymbol{\mu}^T g(\boldsymbol{x}) + \boldsymbol{\lambda}^T h(\boldsymbol{x})\}$$

*The solution* $\bar{x}(\boldsymbol{\lambda})$ *to this unconstrained minimization problem, is also an optimal solution to perturbed primal problem given by*

$$\begin{aligned} \underset{\boldsymbol{x}\in X}{\textit{minimize}} \quad & f(\boldsymbol{x}) \\ \textit{subject to} \quad & g_i(\boldsymbol{x}) \leq g_i(\bar{\boldsymbol{x}}(\boldsymbol{\lambda})); i = 1, \ldots, m. \\ & h_j(\boldsymbol{x}) = h_j(\bar{\boldsymbol{x}}(\boldsymbol{\lambda})); j = 1, \ldots, p. \end{aligned}$$

The Everett's main theorem states the following: for any non-negative $\boldsymbol{\lambda}$ if an unconstrained minimum of the Lagrangian function can be found, with solution $\bar{\boldsymbol{x}}(\boldsymbol{\lambda})$, then this solution is also the solution to the constrained problem whose constraints are, in fact, the amount of each resource expended in achieving the unconstrained solution. This

implies that the constraints are set by choosing the $\boldsymbol{\lambda}$ vector. Any arbitrary set of non-negative $\lambda$'s works here, notably causing the original constraint optimization problem to be unknown, and is only to be defined once the Lagrangian's solutions is determined.

Its important to recall here that there are no conditions of continuity nor derivability on the objective and constraint functions, other than being real-valued.

**Theorem 2.4.** *($\lambda$-Theorem) Given $\lambda_1^k, \lambda_2^k$ be two multiplier values that produce solution $x_1^*, x_2^*$, respectively, and only one of the constraint values are not met, i.e. $g_k(x_1^*) = g_k(x_2^*), k \neq j$ and that $g_j(x_1^*) > g_j(x_2^*)$, then,*

$$\lambda_2^j \geq \frac{f(x_1^*) - f(x_2^*)}{g_i(x_1^*) - g_i(x_2^*)} \geq \lambda_1^j \tag{2.12}$$

This theorem states that, given that we have two optimal solutions produced by two Lagrange multipliers corresponding to two different constraint functions, for which only one of the constraints are met, the ratio of the change in optimal objective value to that of the optimal constraint is bounded between the two multiplier values that correspond to the value change. In the case where the objective function values are derivable w.r.t those of the constraints, the partial derivative w.r.t single constraints, gives the corresponding multiplier $\lambda_i$.

$$\lambda_i = \left[\frac{df(x^*)}{dg_j(x^*)}\right]_{g_k \text{constant}}$$

The $\lambda$-theorem is useful in multiple constraint case, to demonstrate that the multiplier values corresponding to a particular constraint, produce changes in the constraint, where the multiplier measures how far one goes from the objective's optimal value. When applying Lagrangian multiplier methods it is of interest to know the distance from the approximate upper bound to the global minimum, and its variation with the constraint function. This helps in identifying if one is lower bound whose resource-payoff variation is not drastically varying. This following theorem is for case where one can only guarantee an objective value which is $\epsilon$ away from the true optimum possible.

**Theorem 2.5.** *($\epsilon$-Theorem)*

1. *$\bar{x}$ comes within $\epsilon$ of maximizing the Lagrangian, i.e. for all feasible points,*

$$f(\bar{x}) - \sum \lambda^k g_i(\bar{x}) > f(\bar{x}) - \sum \lambda^k g_i(\bar{x}) - \epsilon$$

2. *$\bar{x}$ is a solution of the constrained problem with constraint function value $g_i(\bar{x})$ that itself being within $\epsilon$ of the maximum for those constraints.*

### 2.2.3 Lagrangian Dual and KKT conditions

We state the KKT conditions to view both the primal and dual problems on the HOP and helps discuss, the feasibility in both domains. The Karush-Kuhn-Tucker(KKT) conditions generalizes the method of Lagrange multipliers to the situation where one has inequalities. We review shortly here, following the [20], the primal problem and its Lagrangian from equation (2.8), without the equality conditions, since we are in such a case:

**Problem 1.** ***Lagrange Primal problem:***

$$\begin{aligned} &\underset{\boldsymbol{x}\in\mathbb{R}^n}{\textit{minimize}} && f_0(\boldsymbol{x}) \\ &\textit{subject to} && f_i(\boldsymbol{x}) \leq 0,\ i = 1, \ldots, m. \end{aligned} \tag{2.13}$$

The Lagrangian for the primal problem is given by

$$L(x, \lambda_i) = f_0(x) + \sum_{1}^{p} \lambda_i f_i(x).$$

The unconstrained minimization of the Lagrangian $L$ replaces the initial constrained minimization of $f_0$. The $\lambda_i$ are the Lagrange multipliers. Interpreted as the coordinates of a vector $\boldsymbol{\lambda}$ in $\mathbb{R}^p$, they yield the dual Lagrange function $g(\boldsymbol{\lambda})$ by the relation:

$$g(\boldsymbol{\lambda}) = \inf \left\{ f_0(\boldsymbol{x}) + \sum_{1}^{p} \lambda_i f_i(\boldsymbol{x}), \boldsymbol{x} \in \boldsymbol{dom}\text{-}f_0 \right\} \tag{2.14}$$

**Problem 2.** ***Lagrange dual problem:*** *find multiplier* $\lambda \succeq 0$ [3] *which maximizes* inf-*Lagrangian function:*

$$\begin{aligned} &\textit{maximize} && g(\boldsymbol{\lambda}) \\ &\textit{subject to} && \boldsymbol{\lambda} \succeq \mathbf{0} \end{aligned} \tag{2.15}$$

The solution to the dual problem in 2.15 for a feasible dual parameter, is only a lower bound to the solution to the primal in general, and are only equal in case of strong duality. This is written as:

$$g(\boldsymbol{\lambda}) \leq f_0(\boldsymbol{x}^*). \tag{2.16}$$

which is termed as the lower bound property.

[3] $\succeq$ here refers to the positive semi-definiteness of matrix $\boldsymbol{\lambda}$.

In the domains **dom**-$f_0$ and **dom**-$g$, of possible solutions in $f_0$ and $g$ respectively, we always have *weak duality* when,

$$g(\boldsymbol{\lambda}^*) \leq f_0(\boldsymbol{x}^*). \tag{2.17}$$

We have *strong duality* when there is no gap, that is the primal's optimum value equals the dual's optimum value,

$$\sup_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda}) = \inf_{\boldsymbol{x}} f_0(x), \tag{2.18}$$

which corresponds to local extrema of both $\boldsymbol{x}^*$ and $\boldsymbol{\lambda}^*$. To summarize, for any optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions[4] [20].

**Theorem 2.6.** *(KKT conditions) Let $f_0$ and $\{f_i, 1 \leq i \leq p\}$ be $p+1$ continuously differentiable functions: $\mathbb{R}^n \to \mathbb{R}$, and let*

*(i)* $f_i(x) \leq 0 \quad 1 \leq i \leq p$ *(primal feasibility)*

*(ii)* $\lambda_i \geq 0, \quad 1 \leq i \leq p$ *(dual feasibility),*

*(iii)* $\lambda_i f_i(x) = 0, \quad 1 \leq i \leq p$ *(complementary slackness),*

*(iv)* $\nabla f_0(x) + \sum_1^p \lambda_i \nabla f_i(x) = 0$, *(minima condition),*

*be the four so called KKT conditions. If $f_0$ and $\{f_i, 1 \leq i \leq p\}$ are convex, and $x^*$ and $\boldsymbol{\lambda}^*$ satisfy the four KKT conditions, then they are optimal. Conversely, if strong duality holds, and if $x^*$ and $\boldsymbol{\lambda}^*$ are optimal, then they satisfy the four KKT conditions .*

The first two conditions are the constraints set by the two problems 2.13 and 2.15. The third condition of complementary slackness is obtained when we assume Strong duality.

The condition of complementary slackness basically requires that, at the optimal solution, every constraint that is does not apply, have a zero valued multiplier. This is because, if this constraint does not bind, then we could just as well have solved the problem without that constraint, and setting the corresponding multiplier equal to 0 effectively deletes the constraint from the Lagrangian.

[4]The KKT theory is more general. It may involve additional constraints $h_j = 0$ (as in Lagrange' initial work) and extends to the pseudo-convex case (a function $f$ on $\mathbb{R}^n$ is pseudo-convex if $\nabla f(x).(y-x) \geq 0$ implies $f(y) \geq (f(x))$; a convex function is pseudo-convex). Besides, the continuous derivability can be replaced by a weaker Lipschitz condition. Moreover, this is very wide field of convex analysis. The demonstrations chosen here highlight the comparison with the braid framework required for the next section.

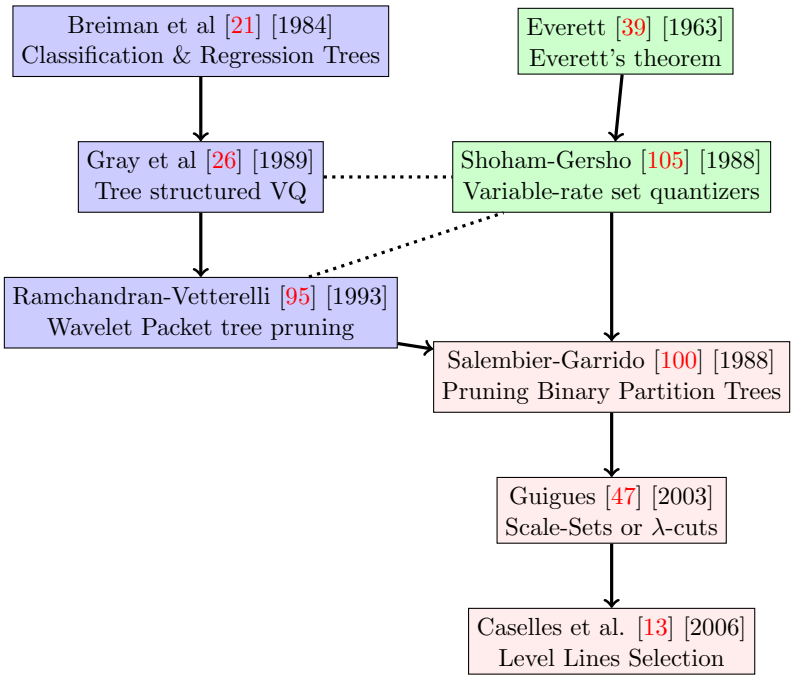



FIGURE 2.1: Blue: Breiman Chain, Green: Everett Chain, Pink: Salembier Chain

In a result similar to Theorem 2.6 for constrained minimization on braids, we present conditions in the following section.

### 2.2.4 Reviewing constrained optimization on hierarchies

The block diagram 2.1 demonstrates the two chains of ideas, converging on a third one.

- Firstly in blue: Breiman's dynamic program to prune classification and regression trees(CART), CART's usage in information theory starting with Grey et al. for source coding, further on CART's DP applied to wavelet tree pruning by Ramchandran-Vetterelli, finally ending in Salembier-Garrido for pruning Binary partition trees.
- Second chain develops the lagrange multiplier based constraint, starting with Everett's theorem, its use by Shoham-Gersho in calculating a optimal source coding schemes, leading to Salembier et al.
- The third chain consists of Salembier-Garrido try to solve the rate-distortion minimization problem on binary partition trees by approximating the constraint rate by searching for an near optimal multiplier, while Guigues established the scale-sets image descriptor for a given HOP and parameterized energy $\omega(\pi, \omega)$. Caselles et al develops Salembier-Garridos model for level line selection in the tree of shapes.

**Multiplier's Role**: In the Breiman chain one uses scale-increasingness w.r.t the multiplier to have a monotonic pruning of HOP, or increasing $\lambda$-cuts, while in the Everett chain one is trying to approximate the constraint by choosing a "good" multiplier.

## 2.3 Guigue's $\lambda$-cuts are upper bounds

We demonstrate how the $\lambda$-cuts calculated by Guigues et al. in [47] are upper bounds the optimal objective function value, achievable in constrained optimization problem on the HOP. One quick intuition to begin with is what happens when we vary the constraint function instead of the Lagrangian multiplier $\lambda$ ?

We will see with a simple counter-example, how the $\lambda$-cuts do not correspond to the global minimum under the conditions for sub-additive constraint and super-additive objective, for a particular cost or constraint function value.

### 2.3.1 Counter-example

The following counter-example considers the Guigues framework of sub-additive and *separable* energies. A dendrogram is depicted in Figure 2.2. The two trees shown $\omega_\varphi$-tree and the $\omega_\partial$-tree represent the energies $(\omega_\varphi(S), \omega_\partial(S))$, associated with each node $S$ in the tree. The dendrogram with node is shown separately, at bottom left of the figure.

To recall, by separability condition $\omega_\varphi$ (and resp. $\omega_\partial$) of a partial partition is the sum of the $\omega_\varphi$ (resp. $\omega_\partial$) of its classes, and when parent and child have the same energy, one chooses the parent. Guigues considers $\omega_\varphi$ to be super-additive, i.e $\omega_\varphi(S) \leq \sum_{a\in\pi(S)} \omega_\varphi(a)$, while the constraint function $\omega_\partial$ to be sub-additive, i.e. $\omega_\partial(S) \geq \sum_{a\in\pi(S)} \omega_\partial(a)$.

Figure 2.2 (bottom right) indicates the lambda function, which gives the value $\lambda$ for which the energy $\omega(\lambda)(S) = \omega_\varphi(S) + \lambda\omega_\partial(S)$ of the class $S$ equals the energy of the partial partition of the children of $S$. The minimal cuts of $H$ w.r.t. $\omega(\lambda)$ are thus the level sets of the lambda function. The three minimal cuts $\pi^*(\lambda)$, for $\lambda = 1, 2, 3$ are shown and denoted by $\pi_1, \pi_2, \pi_3$ for quick reading.

As $\omega(\lambda)$ is scale-increasing, the minimal cuts $\pi^*(\lambda)$ span bottom-up the hierarchy $H$. But this does not mean that they meet all classes of $H$. A class $S$ is met iff it belongs to an ascending path in the $\lambda$-tree. This is not the case, for example, for class $j$ show in Figure 2.2. Following Guigues, we say that these classes are anti-causal. They do not disturb the computation of the $\lambda$-cuts or $\pi^*(\lambda)$.

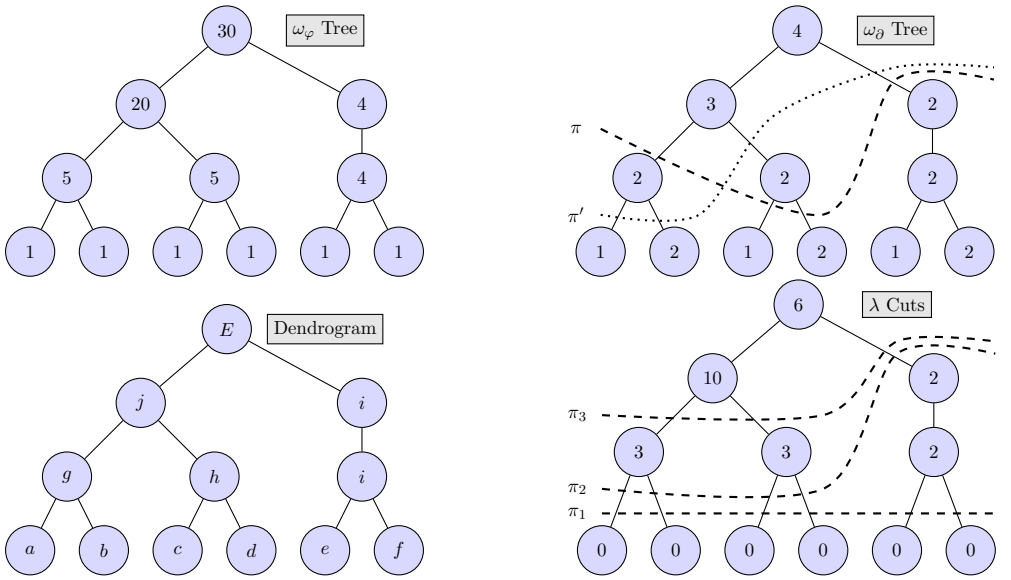



FIGURE 2.2: Bottom Left, a hierarchy $H$ with classes. the pairs of trees in the top row, indicate the two energies $(\omega_\varphi, \omega_\partial)$ associated with the corresponding classes. $\pi$ and $\pi'$ are two cuts of $H$. Bottom right, in the nodes, we depict the lambda values by equating parent and child energies, whose level sets give the minimal cuts w.r.t. the $\omega_\lambda$. They are depicted in the $\lambda$-tree for $\lambda = 2, 3, 4$ as $\pi_2, \pi_3, \pi_4$. The $\lambda$ values for the leaves are assumed to be 0, though in case of Breiman et al. [21] $\lambda$ for the leaf classes are set to $\infty$ to avoid over-fitting.

Though the parameter $\lambda \in \overline{\mathbb{R}}$ varies continuously, both $\omega_\varphi$ and $\omega_\partial$ are piecewise constant functions of $\lambda$. This is due to the finite nature the hierarchy. One observes that $\omega_\varphi$ increases with $\lambda$, while $\omega_\partial$ decreases with it.

Let us consider now consider constrain function value, say $C = 7.5$ and the minimal cut it may correspond to. In the range of the values around 7.5, the only change which may occur in the minimal cut $\pi^*(\lambda)$ is the replacement of $(a, b, c, d)$ by $(g, h, i)$ as $\lambda$ increases. More precisely,

$$\omega(\lambda)(g, h) \leq \omega(\lambda)(a, b, c, d) \quad \Leftrightarrow \quad 10 + 4\lambda \leq 4 + 6\lambda$$

i.e. $\lambda \geq 3$. For $\lambda = 3$ the minimal cut is $(g, h, i)$ by singularity, thus:

$$\begin{aligned} \omega(\lambda)(a, b, c, d, i) &= 8 + 8\lambda \quad \Rightarrow \quad (a, b, c, d, i) = \pi^*(\lambda) \quad \text{for } 2 \leq \lambda < 3 \\ \omega(\lambda)(g, h, i) &= 14 + 6\lambda \quad \Rightarrow \quad (g, h, i) = \pi^*(\lambda) \quad \text{for } 3 \leq \lambda < 6 \end{aligned}$$

For the smallest $\lambda$ such that $\omega_\partial(\pi^*(\lambda)) \leq 7.5$, namely $\lambda = 3$, the corresponding $\omega_\varphi$ value is 14. Compare that with the non optimal cut $\pi = (g, c, d, i)$ of energy $\omega(\lambda) = 11 + 7\lambda$ (Figure 2.2 top left). Cut $\pi$ obviously provides a better result than the minimal cut

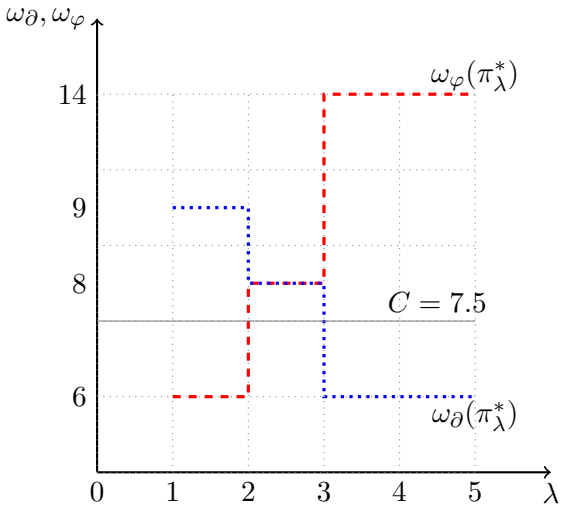



FIGURE 2.3: For $2 < \lambda < 3$ the minimal cut is $(a, b, c, d, i)$ and $\omega_\partial = 8$, for $\lambda \geq 3$ the minimal cut is $(g, h, i)$ and $\omega_\partial = 6$, i.e. $\omega_\partial$ is never equal to the cost $C = 7.5$ at any time.

$(g, h, k)$ since $\omega_\partial(\pi) = 7$ (hence below the cost $C = 7.5$), for an energy $\omega_\varphi(\pi) = 11.5$ (hence smaller than $\omega_\varphi(\pi^*(3)) = 15$).

What's worse is that the two different non-$\lambda$ cuts, $\pi = (g, c, d, i)$ and $\pi' = (a, b, h, i)$ have the same $\omega_\partial = 7$ and $\omega_\varphi = 11.5$. Thus there are many such constrained minimal cuts for the energy $\omega_\varphi$, and none of them have a corresponding value of $\lambda$ that achieves the constraint. And we cannot take their infimum $(a, b, c, d, i)$ because its $\omega_\partial$ energy equals 8, above the cost 7.5.

What happened is as we will see further, is that we do not have a feasible $\lambda$ for the dual problem. We have $\lambda^* = \inf\{\lambda \mid \omega_\partial(\pi^*(\lambda)) \leq 7.5\} = 3$, but $\omega_\partial(\pi^*_3) = 6$ and not 7.5. The is more formally seen in the energetic lattice formulation in Theorem 2.11 cannot give constrained minimal cuts.

### 2.3.2 Lessons from the Counter-Example

We discuss a few important implications of the counter example:

- Lack of Cost→Multiplier mapping: For a given cost $\omega_\partial \leq C$ one is not assured a corresponding multipler $\lambda$, the collection $\{\pi^*(\lambda), \lambda \in \overline{\mathbb{R}}\}$ is not informative enough. A cut that minimizes $\omega_\varphi$ can perfectly not belong to the $\{\pi^*(\lambda)\}$.
- The dual problem is still a combinatorial problem.
- Uniqueness is lost, even when $\omega_\varphi$ is strictly $h$-increasing.

- $\pi^*(\lambda^*)$ is only the upper-bound of the constrained minimal cuts.
- The values of $\lambda$ are discrete due to the discrete nature of the cuts of the hierarchy, furthermore these values of $\lambda$, can be integral or rational, while still lacking a $C \to \lambda$ map.
- One can always reach any one of these minimal cuts (as there may be many). It suffices to begin from any arbitrary class of $\pi^*(\lambda^*)$ and to replace it by its descendants until we reach $\omega_\partial > C$. If one still has $\omega_\partial(\pi) \leq C$ when one arrives to the leaves, one repeats the descent from another class of $\pi^*(\lambda^*)$. This helps improve the upper-bound on the $\lambda$-cuts, for singular $h$-increasing functions.
- The error $\mid \omega_\partial(\pi^*(\lambda^*)) - C \mid$ gives no information about the error $\mid \omega_\varphi(\pi^*(\lambda^*)) - \omega_\varphi(\pi) \mid$ where $\pi$ is a constrained minimal cut.
- One important structure of the constraint problem we see is, for a given cost on the $\omega_\partial$-tree the structure of the solution space forms a lattice. The choice of the partial partition structure, and the energetic lattice to solve the constraint problem, can be seen more clearly in a simple example. The constrained optimization problem will be formulated in a more general framework of the energetic lattice in section 2.7.
- Finally, the counter-example suggests a way to advocate costs independent of $\lambda$. For example it is sufficient to allocate a $\omega_\partial$ to each class and not, globally, to the cuts, which is demonstrated in section 2.10 on class based constraints. This is no more Lagrangian, and depends on a class based constraint.

## 2.4 Improving the upper-bound $\lambda$-cuts

Let us consider again the constrained optimization problem presented earlier in equation 1.14:

$$\begin{aligned} \underset{\pi \in \Pi(E,H)}{\text{minimize}} \quad & \sum_{S \in \pi} \omega_\varphi(S) \\ \text{subject to} \quad & \sum_{S \in \pi} \omega_\partial(S) \leq C \end{aligned}$$

Guigues and Salembier-Garrido's optimization perspective consisted in achieving a $\lambda$ value that provides a cut with minimal objective and at the same time approaches an given constraint function value over the cut is $\omega_\partial(\pi) \leq C$. This basically due to Everett's theorem 2.3. In this section we discuss two approaches which try to improve the upper bounding $\lambda$-cuts by perturbing the original problem, so as to achieve a *better* cut.

### 2.4.1 Perturbing the Scale Function

Guigues's λ-cuts have addressed the problems of monotonicity and constraint satisfaction in just one sense: multiplier-constraint. The constraint function values, correspond to the result of a solution to a perturbed problem where the constraint value is set by the choice of the multiplier. Here we will now perturb the values of the constraint function(or equivalently the objective) so as to achieve a better solution.

Given that we have any general cost $\omega_\partial(\pi^*(\lambda_i)) < C < \omega_\partial(\pi^*(\lambda_j))$ bounded by between any two causal λ-cuts, $\lambda_i, \lambda_j$, more generally between the level sets of the scale function $\Lambda$.

Let us consider now the perturbed scale function $\Lambda(S)$ for each parent class $S$ in the hierarchy or Braid:

$$\Lambda(S,\epsilon) = \frac{\omega_\varphi(S) - \sum_{a\in\pi(S)} \omega_\varphi(a)}{\sum_{a\in\pi(S)} \omega_\partial(a) + \epsilon - \omega_\partial(S)} = \frac{\Delta\omega_\varphi}{\Delta\omega_\partial + \epsilon} \tag{2.19}$$

The perturbation $\epsilon$ corresponds to a small slack variable that is added to the perimeter function $\omega_\partial$. One should note here that the perimeter constraint function has a higher range of values possible compared to the number of classes as constraint, i.e. $1, ...N$, $N$ being the number of leaves. The perturbed scale function $\Lambda_\epsilon(S)$ in such a case improves the constraint value $\rightarrow$ Multiplier mapping ($C \rightarrow \lambda$), by finding cuts that are between a feasible upper-bounding λ-cut and a limiting infeasible λ-cut. The algorithm and results will presented in detail in the applications chapter 3.

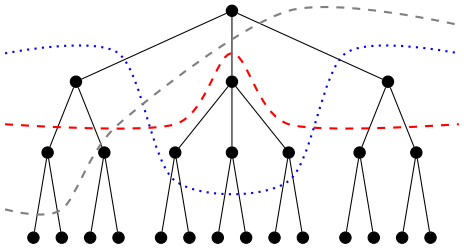


FIGURE 2.4: Demonstrating the feasible space in the hierarchy with the family of $C$-cuts which are the set of partitions that satisfy a constraint of $\omega_\partial(\pi) \leq C$, which here for demonstration are cuts with number of classes no greater than 6.

Given λ-cuts we may have nodes for example $g, h$ in figure 2.2 with the same value of $\lambda$ which are parents of nodes with same $\lambda$, here we have set them to be 1 for nodes $a, b, c, d$. In such a circumstance we have a way to reach a partial partition and thus a cut with a constraint function value closer to $C$, resulting in a cut which is not a λ-cut.

The basic idea is borrowed from Penalty methods associated with Augmented Lagrangian methods [20]. These methods have two multipliers, one that penalizes deviation

from a constraint and the other the regular Lagrangian multiplier. The advantage of these methods are a low span of search values for the penalty multiplier, since the Lagrange multipliers bound the optimal solution. Now we present two approaches for the problem of obtaining a *better* $\lambda$-cut.

**$\lambda$-Perturbation**: For partial partition in the $\lambda$-cut, we perturb the values of one of the partial partitions by a very small value $\lambda' = \lambda + \Delta\lambda$. The effect is that we have no introduced a new family of $\lambda'$-cuts that reaches cost values between the initial $\omega_\partial(\pi^*(\lambda_i)), \omega_\partial(\pi^*(\lambda_j))$. For example in figure 2.2 between $\pi_2, \pi_3$ we have $\pi$ or $\pi'$ as a result of moving the $\lambda$ values of either $g$ or $h$ by $\Delta\lambda$.

### 2.4.2 Penalty Methods

Lets consider the cost $C = 5$, which corresponds to a cut $(j, i)$ which in the scale-set framework is avoided, since this produces does not result in monotonically ordered cuts, and Guigues removes them by scale-climbing, which uses a similar dynamic program but on the scale function. Classes with non-monotonically ordered scale function values are termed by Guigues as *Anti-causal Classes*. Though this problem purely depends on the relation between the objective and constraint function. One can now resolve this ordering problem by changing the constraint function to a penalty function, such that:

$$\omega(\pi(S), \lambda) = \sum_{a \in \pi(S)} (\omega_\varphi(a))^\alpha + \lambda \sum_{a \in \pi(S)} \omega_\partial(a) \tag{2.20}$$

where $\alpha$ is a global parameter for all classes. For the toy example in figure 2.2 it produces the following $\lambda$'s for a range of values of $\alpha$, which are calculated by the scale function equation:

$$\Lambda(S, \alpha) = \frac{\omega_\varphi(S) - (\sum_{a \in \pi(S)} \omega_\varphi(a)^\alpha)^{\frac{1}{\alpha}}}{\sum_{a \in \pi(S)} \omega_\partial(a) - \omega_\partial(S)} \tag{2.21}$$

| | $\lambda$-values for classes (E, j, g/h, i) | Causality |
|---|---|---|
| $\alpha = 1$ | 6, 10, 3, 2 (linear, Lagrangian case), | anti-causal |
| $\alpha = 1.5$ | 66, 67, 9, 6, (floored), | anti-causal |
| $\alpha = 1.525$ | 74, 73, 9, 6, (floored), | causal |
| $\alpha = 2$ | 484, 350, 23, 14, | causal |

A point on the refinement of $\lambda$-cuts is the parameter $\alpha$ in equation (2.20). As already seen in the study of composition of $h$-increasing energies in section 1.6. This $\alpha$ parameterized

penalty energy is $h$-increasing as already shown in proposition 1.19, to enable one to control the choice of parent/children more finely, than just by supremum, infimum, sum or number of classes.

**Increasing Scale-Function range**: In overview from the two methods above we can conclude that the greater the range of values of perturbed $\lambda$, the higher the chance of converging at a better optimal cut meeting constraining $\omega_\partial(\pi) \leq C$. The $\lambda$-perturbation and the penalty function are ways of inspecting $\lambda$-cuts that are finer. While one can note here that the penalty function method provides cuts that are in between the ones address by scale-openings and closings in subsection 1.7.3.

## 2.5 The energies $\omega_\lambda = \omega_\varphi + \lambda\omega_\partial$

Here we present the critical result of constraint minimization over braids by their characteristic energetic lattices. The first chapter dealt with unconstrained minimization of any model of energy on partial partitions. We now will focus on the Lagrangian energy $\omega_\lambda = \omega_\varphi + \lambda\omega_\partial$, whose unconstrained minimization leads to the solution of the original constrained minimization.

**Lagrange Multipliers for constrained optimization on HOP**: When dealing with the cuts $\Pi(E, B)$ of a braid $B$, can we find a cut which minimizes $\omega_\varphi$ on $\Pi(E, B)$ under some energy constraint $\omega_\partial$?

Since Lagrange's starting points seems unrealistic, let's consider his arrival point, which is the Lagrangian like in equation (2.8). For convenience, we begin with a single constraint. We have the Lagrangian $\omega_\lambda$,

$$\omega_\lambda(\pi) = \omega_\varphi(\pi) + \lambda\omega_\partial(\pi) \quad \pi \in \Pi(E, B). \tag{2.22}$$

The formalism (2.22) is classical in image segmentation, as well as the Lagrangian in [100], [49], when the braid reduces to a hierarchy, and when the energies $\omega_\varphi$ and $\omega_\partial$ are linear, i.e. when the energy of a p.p. is the sum of the energies of its classes. These results strongly rest on this linearity assumption, which is in fact a particular case, one can find in the literature various energies which involve other operations, like suprema or infima, in the $\alpha - \omega$-trees [107], as well as a refinement based ordering in [2]. We will now present a more comprehensive approach, which works across these cases.

## 2.6 Discussion on Everett's theorem

### 2.6.1 Gap's and lower bounds

The Lagrangian relaxation used by Everett generates a mapping from the multiplier vectors $\lambda$ into the space of constraint vectors. There is an a priori guarantee that this mapping is onto for a given problem, and one may have inaccessible "gaps" consisting of constraint vectors, for whom $\lambda$ vectors don't exist always (examples can be easily constructed). Optimum in such cases are not guaranteed by the Lagrangian relaxation method and we would need other means to achieve them. The basic cause of the gap's are the non-concavity in the objective function w.r.t the constraints, i.e convexities in the envelope of the set of achievable objective points in the space of feasible set. [39]

The method by itself does not guarantee solutions for a general constraint problem, but ensures a lower bound. It is interesting to note that with the $\lambda$-mapping we ensure to find a cut that is at least $\epsilon$ far from the optimal cut.

Another important point to note is that the existence of a global minima for a given function on a general set $X$ using the Weierstrass theorem [5], which provides the existence of a minimum by necessitating a continuity of the said function [17].

**Discontinuous objective and Constraints**: For function's defined on partial partitions, such as $\omega_\partial$ and $\omega\varphi$, the continuity is mostly never true or even defined, these functions are akin to abstract set value functions.

Lagrangian and KKT multiplier methods depend on the ability to describe the local nature of a minima (or extrema), such as the first order or second order conditions, leading to strict definitions of convexity. In case of non-convex functions with global minima ensure, we find the work that uses the "convexified" epigraph of such functions to obtain the said global minima. In the set valued case the possibility of achieving saddle points are also studied in literature, but we restrict ourselves here to the case of partitions.

**Retrospective on Optimization by Pruning**: In this section we briefly remark in the table 2.1 the development of the pruning based optimization on trees starting with Breiman et al. [21] while comparing with the energetic lattices. We also study the different characteristics of monotonicity, uniqueness, and the fact that if there has been a Lagrangian interpretation used.

[5] Weierstrass Theorem states that if $\mathcal{K} \subset \mathbb{R}^n$ is compact and $f : \mathcal{K} \to \mathbb{R}$ is a continuous function, then $f$ has a maximum and a minimum on $\mathcal{K}$, i.e. there exists $k', k^*$ such that $f(k') \geq f(k) \geq f(k^*), \forall k \in \mathcal{K}$

| Authors | Lagrangian Multiplier | Unique | Monotonic | Cited |
|---|---|---|---|---|
| Breiman et al. (CART) [21] 1984 | No | Yes | Yes | None[(a)] |
| Everett [39] 1963 | Yes | No | No | None[(b)] |
| Shoham-Gersho [105] 1988 | Yes | No | No | [39] |
| Gray et al. [26] 1989 | No | No | No | [21] |
| Ramchandran-Vettereli [95] 1993 | No | No | No | [21] |
| Salembier-Garrido [100] 2000 | Yes | No | No | [21, 95, 105] |
| Guigues [47] 2003 | Yes | Largest Class | Sub/Super additivity[(c)] | [21, 100] |
| Casselles et al. [13] 2007 | Largest Class | Largest Class | Sub/Super additivity | [21, 26, 47] |
| Energetic Lattice [59] 2014 | Yes/No[(d)] | Yes | Scale-increasingness | [47, 100] |

TABLE 2.1: This table summarizes the different components of the algorithms of different authors work, related to the study of extract of optimal cuts, the monotonicity of ordering of the cuts based on $\lambda$, the uniqueness of these solutions of a given constraint, as well as whether the multiplier is chosen or not. (a) First formulation of Dynamic Program for optimal pruning on trees (b) First formulation of multiplier dependent perturbed primal problem for any arbitrary set. (c) Guigues interprets scale-space semi-group structure as in [4] by introducing the causality condition in the scale-sets framework. (d) The energetic lattice can be used to solve the linear Lagrangian case, while also non-linear energies which do not entail a Lagrangian multiplier method and application of Everett's theorem.


76

## 2.7 Minimal $\lambda$-cuts and energetic lattices

We develop in this section an energetic-lattice based approach to constrained optimization. Three different models of constraints are proposed, firstly numerical constraint on constraint function on partitions, secondly partition based constraint, and finally a class based constraint. The last two reinforce the model of constrained optimization where the energetic lattice is an optimal framework. From now on, $\omega(\lambda, \pi) = \omega_\varphi(\pi) + \lambda\omega_\partial(\pi)$ is used to denote scalar Lagrangian energy, and $\omega(\boldsymbol{\lambda})$ its vectorial version (2.22). The Lagrangian energy of cut $\pi$ is thus

$$\omega(\pi, \lambda) = \omega_\varphi(\pi) + \lambda\omega_\partial(\pi)$$

A number of lattices are involved in the formalism, we delineate between them, while keeping our notation minimal. The first trivial lattice is of course $\overline{\mathbb{R}}$, which serves in comparing energies. Moreover the same set $\Pi$ of all cuts of $B$, is the matter of the four different partitions lattices:

- $\Pi$ of the refinement ordering, with the leaves as minimal cut and $E$ as maximal one (the symbol $E$ of the previous notation $\Pi(E)$ is dropped for the convenience);
- $\Pi(\omega(\lambda))$ of the energetic ordering $\preceq_{\omega(\lambda)}$ w.r.t the Lagrangian $\omega(\lambda)$, of order $\preceq_{\omega(\lambda)}$. The minimal cut for this energetic lattice is $\pi^*(\lambda) = \curlywedge_{\omega(\lambda)}\{\pi, \pi \in \Pi\}$. The value of $\omega(\lambda)$ for a cut $\pi \in \Pi$ is denoted by $\omega(\pi, \lambda)$, and that for the minimal cut by $\omega(\pi^*(\lambda))$;
- $\Pi(\omega_\varphi)$ (resp. $\Pi(\omega_\partial)$) of the energetic ordering w.r.t. $\omega_\varphi$, of order $\preceq_\varphi$ (resp. the energetic ordering w.r.t. $\omega_\partial$, of order $\preceq_\partial$). If necessary, one makes the notation more precise, and indicates the hierarchy or the braid under study (e.g. $\Pi(\omega_\varphi, B)$;

In the vector case, the vector $\boldsymbol{\lambda} = \{\lambda_i, 1 \leq i \leq p\}$ replaces $\lambda$ in the notation, i.e. $\Pi(\omega(\lambda))$ $\to \Pi(\omega(\boldsymbol{\lambda}))$, $\pi^*(\lambda) \to \pi^*(\boldsymbol{\lambda})$, $\omega(\pi, \lambda) \to \omega(\pi, \boldsymbol{\lambda})$, and the set of constraint $\omega_{\partial i}, 1 \leq i \leq p$ replaces the constraint $\omega_\partial$. Note that the $\pi^*(\lambda)$ are the *only minimal cuts* used below (the minimal cuts in the lattices $\Pi(\omega_\varphi)$ and $\Pi(\omega_\partial)$ play no role).

**Definition 2.7.** *One calls "scalar Lagrange family" of energies any family* $\{\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial, \lambda \in \overline{\mathbb{R}}\}$ *where* $\omega(\boldsymbol{\lambda})$, $\omega_\varphi$, *and* $\omega_\partial$ *are singular, and where* $\omega_\partial$ *is inf-modular.*

*Similarly a "vector Lagrange family" of energies is a family* $\{\omega(\boldsymbol{\lambda}) = \omega_\varphi + \sum \lambda_i\omega_{\partial i}\}$ *where the* $\lambda_i$ *are scalar, and where* $\omega(\boldsymbol{\lambda})$ $\omega_\varphi$, *and* $\omega_{\partial i}$ *are singular, and the* $\omega_{\partial i}$ *are inf-modular.*

Given a braid $B$, we already know that a scalar Lagrange family provides a unique minimal cut $\pi^*(\lambda)$ of $B$ with each $\lambda$, since $\omega(\lambda)$ is singular. Moreover the inf-modularity of $\omega_\partial$ shows that the lower upper bound of these minimal cuts $\pi^*(\lambda)$ enlarge as $\lambda$ increases (Proposition 1.31 and Theorem 1.23). These first results are improved by the following proposition:

**Proposition 2.8.** *Let $\{\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial\}$, be a scalar Lagrange family of energies on the partial partitions of a space $E$, and suppose $\lambda > 0$. Given a braid $B$ on $E$, let $\Pi(\omega(\lambda))$, $\Pi(\omega_\varphi)$ and $\Pi(\omega_\partial)$ be the energetic lattices of the cuts $\pi$ of $B$ w.r.t. the energies $\omega(\lambda)$, $\omega_\varphi$, and $\omega_\partial$ respectively. The minimal element of $\Pi(\omega(\lambda))$ is denoted by $\pi^*(\lambda)$. For all $\lambda, \mu \geq 0$ the two implications*

$$0 \leq \lambda \leq \mu \quad \Rightarrow \quad \pi^*(\lambda) \succeq_{\omega_\partial} \pi^*(\mu) \quad \text{and} \quad \pi^*(\lambda) \preceq_{\omega_\varphi} \pi^*(\mu) \tag{2.23}$$

*are true, i.e. as $\lambda$ increases, the sequence $\{\pi^*(\lambda), \lambda > 0\}$ of the minimal cuts w.r.t. the $\Pi(\omega(\lambda))$ decreases in the energetic lattice $\Pi(\omega_\partial)$ and increases in the energetic lattice $\Pi(\omega_\varphi)$.*

*Proof.* If $\lambda = \mu$, the relation 2.23 is obviously true. Suppose $\lambda < \mu$. Consider the class $S_\mu$ of the minimal cut $\pi^*(\mu)$ at point $x$ and the associated l.u.b. $S_{\max}(x \mid \mu)$. The restriction $\pi^*(\mu) \sqcap S_{\max}(x \mid \mu)$ of $\pi^*(\mu)$ to $S_{\max}(x \mid \mu)$ is the p.p. $a_\mu$. As $\omega_\partial$ is inf-modular, the energy $\omega(\lambda)$ is scale increasing. Then, according to Proposition 1.8, $S_{\min}(x \mid \mu)$ is the support of a p.p. $a_\lambda$ of $\pi^*(\lambda)$. As $\pi^*(\mu)$ is the minimal cut for the energetic ordering $\preceq_{\omega_\mu}$, we have $\omega_\mu(a_\lambda) \geq \omega_\mu(a_\mu)$, i.e.

$$\omega_\varphi(a_\lambda) + \mu\omega_\partial(a_\lambda) \geq \omega_\varphi(a_\mu) + \mu\omega_\partial(a_\mu). \tag{2.24}$$

On the other hand, as $\pi^*(\lambda)$ is minimal cut for the energetic ordering $\preceq_{\omega(\lambda)}$, we have also $\omega(\lambda, a_\mu) \geq \omega(\lambda, a_\lambda)$, i.e.

$$\omega_\varphi(a_\mu) + \lambda\omega_\partial(a_\mu) \geq \omega_\varphi(a_\lambda) + \lambda\omega_\partial(a_\lambda) \tag{2.25}$$

By adding the two inequalities (2.24) and (2.25) we obtain, as $\lambda, \mu > 0$

$$\omega_\partial(a_\lambda) \geq \omega_\partial(a_\mu). \tag{2.26}$$

The inequality (2.26) is true for all supports $S_{\max}(x \mid \mu)$, $x \in E$, which results in $\pi^*(\lambda) \succeq_\partial \pi^*(\mu)$. Similarly, by subtracting the inequality (2.24) from (2.25) we obtain

$$2[\omega_\varphi(a_\mu) - \omega_\varphi(a_\lambda)] \geq (\lambda + \mu)[\omega_\partial(a_\lambda) - \omega_\partial(a_\mu)] \geq 0,$$

which leads to $\pi^*(\lambda) \preceq_{\omega_\varphi} \pi^*(\mu)$, and achieves the proof. □

**Corollary 2.9.** *If in addition $\omega_\partial$ and $\omega_\varphi$ are h-increasing, then*

$$\lambda \leq \mu \Rightarrow \omega_\partial(\pi^*(\lambda)) \geq \omega_\partial(\pi^*(\mu)), \text{ and } \omega_\varphi(\pi^*(\lambda)) \leq \omega_\varphi(\pi^*(\mu)). \tag{2.27}$$

The $h$-increasingness allows us to apply relation (1.25) to both implications (2.23). Then the two energies $\omega_\partial$ and $\omega_\varphi$ vary in opposite senses over the minimal cuts. The important corollary 2.9 generalizes the $\lambda$-cuts by Salembier-Garrido and Guigues [47, 100]: Firstly to braids from hierarchies, and secondly from linear Lagrange families of energies $\omega_\varphi$ and $\omega_\partial$ to various non-linear compositions. Note that $\omega_\partial$ does not need to be $h$-increasing for obtaining $\lambda \leq \mu \Rightarrow \omega_\varphi(\pi^*(\lambda)) \leq \omega_\varphi(\pi^*(\mu))$, and vice versa.

**Corollary 2.10.** *The proposition 2.8 extends to vector Lagrange families, and the implication (2.23) still holds when the vector inequality $\boldsymbol{\lambda} \leq \boldsymbol{\mu}$ replaces the scalar one in (2.23).*

*Proof.* The proof is similar to that of Corollary 1.26. The vector variation from $\boldsymbol{\lambda}$ to $\boldsymbol{\mu}$ can be decomposed into a succession of scalar variations of each coordinate. As the proposition 2.8 applies for each of these scalar steps, we finally get

$$\mathbf{0} < \boldsymbol{\lambda} \leq \boldsymbol{\mu} \quad \Rightarrow \quad \pi^*(\lambda) \succeq_{\omega_\partial} \pi^*(\mu) \text{ and } \pi^*(\lambda) \preceq_{\omega_\varphi} \pi^*(\mu).$$

i.e. the vector version of Rel.(2.23). □

## 2.8 Lagrangian Minimization by Energy (LME)

This first type of minimization focuses on the energies and performs unconstrained minimization of the Lagrangian to obtain a relaxation of the constraint optimization problem. However, as the set $\Pi$ of partitions replaces the Euclidean space $\mathbb{R}^n$, the notions of continuity, derivability, gradient, and convexity vanish and we work on in the space of lattices. For the sake of pedagogy, we view the case of one constraint first. The primal and dual problems are re-stated over the cuts of $B$, and within an energetic lattice:

**LME Primal problem**:

$$\begin{aligned} &\underset{\pi \in \Pi(E,B)}{\text{minimize}} && \omega_\varphi(\pi) \\ &\text{subject to} && \omega_\partial(\pi) \leq C, \end{aligned} \tag{2.28}$$

Now the domain of the feasible cuts is the subset $\Pi'$ of $\Pi$

$$\Pi' = \{\pi, \pi \in \Pi, \omega_\partial(\pi) \leq C\} \tag{2.29}$$

In the Lagrangian energetic lattice $\Pi_\omega(\lambda)$ corresponding to Lagrangian $\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial$, the minimal cut $\pi^*(\lambda)$ has an energy $\omega(\pi^*(\lambda))$ which is itself minimal ($h$-increasingness of $\omega(\lambda)$). The energy $\omega(\pi^*(\lambda))$ turns out to be the dual Lagrangian, $g(\boldsymbol{\lambda})$. The energy $\omega(\pi^*(\lambda))$ is a function of $\lambda, \omega_\varphi$ and $\omega_\partial$, but not of the cuts $\pi \in \Pi(E, B)$. The dual problem can now be stated:

**LME Dual problem**: Given a braid $B$ find the parameter $\lambda$ which maximizes $\omega(\pi^*(\lambda))$, subject to the constraint $\lambda > 0$.

$$\begin{aligned} &\text{maximize} \quad \omega(\pi^*(\lambda)) \\ &\text{subject to} \quad \lambda > 0 \end{aligned} \tag{2.30}$$

The two problems, (LME) 2.28 and 2.30 will be solved jointly by introducing

$$\lambda^* = \inf\{\lambda \mid \omega_\partial(\pi^*(\lambda)) \leq 0\}. \tag{2.31}$$

The constraint function $\omega_\partial$ being inf-modular and $h$-increasing, corollary 2.9 applies, and

$$0 \leq \lambda^* \leq \lambda \quad \Rightarrow \quad \pi^*(\lambda^*) \succeq_\partial \pi^*(\lambda) \;\Rightarrow\; 0 \geq \omega_\partial(\pi^*(\lambda^*)) \geq \omega_\partial(\pi^*(\lambda)).$$

The domain of the feasible $\lambda$ is therefore $\lambda \geq \lambda^*$.

One must notice an immediate difference between the exact dual of the combinatorial problem and the one proposed here. The space of solutions, in the latter are the $\lambda$-cuts $\pi^*(\lambda)$. We can now set the minimization problem more precisely. In case of the dual, the idea is to span the multplier space.

Three conditions are needed:

1. *Primal constraint qualification:* the set $\Pi'$ is not empty,
2. *Dual constraint qualification:* $\lambda^*$ exists and is $\geq 0$,
3. *Relaxation:* $\omega_\partial(\pi^*(\lambda^*)) = 0$.
4. *First order condition for minima replaced by lattice:* $\inf \Pi_\omega(E, B)$

We observe that the two functionals $\omega(\pi^*(\lambda))$ and $\omega_\varphi(\pi)$ are ordered. Indeed, the $h$-increasingness of $\omega(\lambda)$ implies

$$\pi^*(\lambda) \preceq_{\omega(\lambda)} \pi \quad \Rightarrow \quad \omega(\pi^*(\lambda)) \leq \omega(\pi, \lambda) \qquad \pi \in \Pi,$$

On the other hand, for every doublet $\pi \in \Pi'$ and $\lambda \geq \lambda^*$ the Lagrangian $\omega(\pi, \lambda)$ is smaller or equal to $\omega_\varphi$, since its term $\lambda\, \omega_\partial(\pi)$ is $\leq 0$ (condition *2*):

$$\omega(\pi, \lambda) \leq \omega_\varphi(\pi) \qquad \pi \in \Pi', \lambda \geq \lambda^*. \tag{2.32}$$

Hence $\omega(\pi^*(\lambda)) \leq \omega_\varphi(\pi)$ and

$$\bigvee_{\lambda \geq \lambda^*} \omega(\pi^*(\lambda)) \leq \bigwedge_{\pi \in \Pi'} \omega_\varphi(\pi), \tag{2.33}$$

which is nothing but a transposed version of the weak duality inequality (2.17). As $\pi^*(\lambda^*)$ satisfies condition *3*, the Lagrangian $\omega(\pi^*(\lambda^*))$ is reduced to its term in $\omega_\varphi$, i.e.

$$\omega(\pi^*(\lambda^*)) = \omega_\varphi(\pi^*(\lambda^*)) \tag{2.34}$$

and the inequality (2.32), applied to the doublet $\{\pi^*(\lambda^*)\}$, gives

$$\pi \in \Pi' \;\Rightarrow\; \omega_\varphi(\pi^*(\lambda^*)) = \omega(\pi^*(\lambda^*)) \leq \omega(\pi, \lambda^*) \leq \omega_\varphi(\pi). \tag{2.35}$$

This results in $\omega_\varphi(\pi^*(\lambda^*)) \leq \wedge\{\omega_\varphi(\pi), \pi \in \Pi'\}$. But $\pi^*(\lambda^*)$ is an element of $\Pi'$, hence it belongs to the infimum, and

$$\omega_\varphi(\pi^*(\lambda^*)) = \bigwedge\{\omega_\varphi(\pi), \pi \in \Pi'\}, \tag{2.36}$$

which solves the primal problem 2.28. Concerning the dual problem, we draw from (2.33), (2.34), and (2.36) that $\bigvee_{\lambda \geq \lambda^*} \omega(\pi^*(\lambda)) \leq \omega_\varphi(\pi^*(\lambda^*)) = \omega(\pi^*(\lambda^*))$. The reverse inequality also holds because the right member is an element of the supremum, and finally

$$\bigvee\{\omega(\pi^*(\lambda)), \lambda \geq \lambda^*\} = \omega(\pi^*(\lambda^*)) = \; \omega_\varphi(\pi^*(\lambda^*)) = \bigwedge\{\omega_\varphi(\pi), \pi \in \Pi'\}. \tag{2.37}$$

The weak duality of Rel.(2.33) becomes strong duality for the doublet $(\pi^*(\lambda^*), \lambda^*)$ of arguments. This pair solves both primal and dual problems 2.28 and 2.30. At this stage,

$\lambda^*$ is unique, but not necessarily $\pi^*(\lambda^*)$. However, the uniqueness of the solution $\pi^*(\lambda^*)$ is attained when $\omega_\varphi$ is strictly $h$-increasing, by application of Proposition 1.16. This happens, for example, when the energy $\omega_\varphi$ is linear. In conclusion, we can state:

**Theorem 2.11.** *(Braid minimization by energy, scalar case). Given a braid $B$ on $E$, let $\{\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial, \lambda \in \overline{\mathbb{R}}\}$ be a scalar Lagrange family of energies where in addition $\omega_\varphi, \omega_\partial$ and $\omega(\lambda)$ are $h$-increasing. And let $\lambda^* = \inf\{\lambda \mid \omega_\partial(\pi^*(\lambda)) \leq 0\}$. Now If we have,*

*(i) the set $\Pi'$ is not empty,*

*(ii) $\lambda^*$ exists and is $\geq 0$,*

*(iii) $\omega_\partial(\pi^*(\lambda^*)) = 0$,*

*then $\pi^*(\lambda^*)$ and $\lambda^*$ are solutions of the problems 2.28 and 2.30 respectively. When $\omega_\varphi$ is strictly $h$-increasing, then the solution $\pi^*(\lambda^*)$ is unique.*

*Conversely, if $\Pi'$ is empty, or if there is no $\lambda$ such that $\omega_\partial(\pi^*(\lambda)) \leq 0$, then there is no solution. If these two conditions are satisfied, but not the third one, the dual problem is still solved by $\lambda = \lambda^*$, but not the primal one.*

### 2.8.1 Vector case

Theorem 2.11 easily extends to the vector case, of Lagrangian

$$\omega(\pi, \boldsymbol{\lambda}) = \omega_\varphi(\pi) + \sum_1^p \lambda_i \omega_{\partial i}(\pi). \tag{2.38}$$

The previous set $\Pi'$ becomes the family $\Pi'_i = \{\pi, \pi \in \Pi, \omega_{\partial i}(\pi) \leq 0\}$. Applying the corollary 2.10 in the previous proof leads to the

**Theorem 2.12.** *(Braid vector constraint minimization). Given the vector $\boldsymbol{\lambda} = \lambda_1, \lambda_2 ... \lambda_p$ Let*

$$\omega(\pi, \boldsymbol{\lambda}) = \omega_\varphi(\pi) + \sum_1^p \lambda_i \omega_{\partial i}(\pi)$$

*be a vector Lagrange family of energies. Put $\lambda_i^* = \inf\{\lambda \mid \omega_{\partial i}(\pi^*(\lambda)) \leq 0\}, 1 \leq i \leq p$, and vector $\boldsymbol{\lambda}^* = \lambda_1^*, \lambda_2 ... \lambda_p^*$. If*

*(i) the set intersection $\Pi' = \cap\{\Pi'_i, 1 \leq i \leq p\}$ is not empty,*

*(ii) there exist $p$ values $\lambda_i^* \geq 0$,*

*(iii)* $\omega_{\partial i}(\pi^*(\lambda_i^*)) = 0, \quad 1 \leq i \leq p,$

*then* $\pi^*(\lambda^*)$ *and* $\boldsymbol{\lambda}^*$ *are respectively solutions of the problems 2.28 and 2.30. When* $\omega_\varphi$ *is strictly h-increasing, then the solution* $\pi^*(\lambda^*)$ *is unique.*

*Conversely, if* $\Pi'$ *is empty, or if there is one* $i$ *at least such that* $\omega_{\partial i}(\pi^*(\lambda)$ *is always* $> 0$*, then there is no solution. If these two conditions are satisfied, but not the third one, the dual problem is still solved by* $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$*, but not the primal one.*

### 2.8.2 Costs

In the scalar case, one can interpret the constraint in terms of a cost $C$, by letting $\omega_\partial(\pi) = \omega'_\partial(\pi) - C$, hence $\omega_\partial \leq 0 \Leftrightarrow \omega'_\partial \leq C$. Applied to cut $\pi$, this gives $\omega_\partial(\pi) = \omega'_\partial(\pi) - C$. This amounts to a change of origin on the axis of the energies. If $\omega_\partial$ is inf-modular, of $h$-increasing, then $\omega'_\partial$ also is, and Theorem 2.11 still applies. Figure 2.3 depicts a situation where the constraint is compared to a cost $C$.

In the vector case, similarly, $p$ cost constants $C_i$, $1, 2, ...i, ..., p$ can complete the inputs set, and be interpreted as the coordinates of a vector $\boldsymbol{C}$ in $\mathbb{R}^p$. The vector Lagrangian is now written

$$\omega(\pi, \boldsymbol{\lambda}) = \omega_\varphi(\pi) + \sum_1^p \lambda_i[\omega'_{\partial i}(\pi) - C_i], \tag{2.39}$$

The $\omega'_{\partial i}$ are inf-modular, and the theorem 2.12 is valid for them. In the vector case, the solutions of equation (2.39) are the doublets $\{\pi^*(\boldsymbol{\lambda}^*), \boldsymbol{\lambda}^* - \boldsymbol{C}\}$, and in the scalar case the doublet $\{\pi^*(\lambda^*), \lambda^* - C\}$.

### 2.8.3 Discussion

We shall describe some of the similarities we have seen with the KKT conditions. The feasibility assumptions in the KKT theorem 2.6 have counterparts in 2.11 and 2.12. It is not surprising to find them again. Furthermore, the complementary slackness, leading to strong duality, reappears in braid optimization via the condition *3* of Theorem 2.11. This condition *3* appears because we want that the minimal cut $\pi^*(\lambda^*)$ *in the Lagrangian energetic lattice* $\Pi(\omega(\boldsymbol{\lambda}))$, such that $\omega(\pi^*(\lambda^*)) \leq 0$, be also minimal cut in the energetic lattice $\Pi(\omega_\partial)$. The counter example of section 2.3 shows what happens when this condition is dropped.

Four major differences separate KKT conditions and their counterpart conditions in the optimization problem on braid:

1. The counterpart of the points $x$ of $\mathbb{R}^n$ is now the cuts of $\Pi(B)$ and these new "points" are structured by several orderings and lattices.
2. The assumptions of "continuously derivable" functions $f_0, f_i$ have no counterpart in braid optimization. A topology on the set $\Pi(B)$ would just be cumbersome. The concept which replaces the first order minima condition in 2.6 is the infimum of the energetic lattice $\Pi(\omega(\boldsymbol{\lambda}))$.
3. First order minima conditions are replaced inf-modularity.
4. The KKT conditions (2.6) when interpreted on a BOP, works by ordering the cuts themselves (e.g. the energetic lattices *of cuts* versus the numerical one *of energies*), and goes to energies in a second step, via $h$-increasingness.

## 2.9 Lagrange minimization by Cut-Constraints (LMCC)

In the two theorems 2.11 and 2.12, $h$-increasingness seems to be an artificial construction residual of the dynamic program. We now will reformulate the minimization problems directly in the refinement of partitions or cuts, getting closer to a lattice based approach, in such a way that $h$-increasingness will no longer be required. We begin, by looking at cuts $\Pi(\omega_\partial)$ by a given cut itself $\pi_C \in \Pi(E)$.

**Problem 3.** ***LMCC Primal problem****: Find the minimal cut $\pi_\varphi$ in the Energetic lattice $\Pi(\omega_\varphi)$ which is constrained by a cut $\pi_C$*

$$\begin{array}{ll} \underset{\pi\in\Pi(\omega_\varphi)}{\textit{minimize}} & \omega_\varphi(\pi) \\ \textit{subject to} & \pi \preceq_{\omega_\partial} \pi_C \end{array} \tag{2.40}$$

The set of feasible solutions for $\pi_\varphi$ is clearly

$$\Pi_C = \{\pi \mid \pi \in \Pi(E,B), \quad \pi \preceq_{\omega_\partial} \pi_C\}$$

and $\pi_\varphi = \curlywedge_\varphi\{\pi \mid \pi \in \Pi_C\}$.

**Minimal cut using the energetic-lattice**: Equation 2.40 refers here to a minimization using two energetic lattice structures, and not just the numerical order of energies.

This is a critical change in the method and approach to the minimization problem, as the energetic lattice helps order the solution space and thus also the final optimum reached by enforcing the singularity condition. This consists in finding the infimum in $\Pi(\omega_\varphi)$ constrained by $\omega_\partial$-energetically ordered cuts. For clarity, one must note here that this requires interaction between two energetic-lattices.

As before, we introduce the Lagrangian $\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial, \lambda \in \overline{\mathbb{R}}$, which induces the energetic lattice $\Pi(\omega(\lambda))$ of minimal cut $\pi^*(\lambda)$. To obtain the optimal Lagrangian parameter $\lambda^*$, we proceed as in classical Lagrangian dual, and obtain

$$\lambda^* = \inf\{\lambda \mid \pi^*(\lambda) \preceq_{\omega_\partial} \pi_C\}$$

**Problem 4.** ***LMCC Dual problem****: Find the parameter $\lambda$ which maximizes the cut $\pi^*(\lambda)$ in the energetic lattice $\Pi(\omega_\partial)$:*

$$\begin{aligned} &\textit{maximize} \quad \pi^*(\lambda) \in \Pi(\omega_\partial) \\ &\textit{subject to} \quad \lambda \geq 0 \end{aligned} \tag{2.41}$$

To solve both problems (2.40, 2.41) jointly, the following three conditions are required:

**Theorem 2.13.** *Given a braid $B$ on $E$, let $\{\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial, \lambda \in \overline{\mathbb{R}}\}$ be a scalar Lagrange family of energies where in addition $\omega_\varphi, \omega_\partial$, and $\omega(\lambda)$ are h-increasing. Let $\lambda^* = \inf\{\lambda \mid \pi^*(\lambda) \preceq_{\omega_\partial} \pi_C\}$. If*

*(i) Primal feasibility: the set $\Pi_C = \{\pi \mid \pi \in \Pi(E, B), \pi \preceq_{\omega_\partial} \pi_C\}$ is non-empty,*

*(ii) Dual feasibility: $\lambda \geq 0$*

*(iii) Lattice Assumption: $\pi_\varphi \succeq_\varphi \pi^*(\lambda^*)$*

*then the set of feasible multipliers are $\lambda \geq \lambda^*$, and $\pi^*(\lambda^*)$ and $\lambda^*$ are the unique solutions to the problems 2.40 and 2.41 respectively.*

*Proof.* We first prove that the feasible set of multipliers are $\lambda \geq \lambda^*$. If $\lambda < \lambda^*$, then $\pi^*(\lambda)$ does not belong to the space $\Pi_C$ of solutions $\pi_\varphi$. The class $S(\lambda)$ of $\pi^*(\lambda)$ at leaf $x$, contains one or more classes of $\pi^*(\lambda^*)$ which form a partial partition $a(\lambda^*)$. By proposition 2.8 $\omega_\partial(a(\lambda^*)) \leq \omega_\partial[S]$. Moreover, as $\lambda$ varies there is a finite number of different sets $S$ (axiom *(ii)* of the braid definition 1.18), so that $\omega_\partial[a(\lambda)]$ is necessarily one of the $\omega_\partial(S)$, hence $\pi^*(\lambda^*) \preceq_\partial \pi_C$. Therefore $\pi^*(\lambda^*) \in \Pi_C$ which implies $\pi_\varphi \preceq_\varphi \pi^*(\lambda^*)$, and by assumption *(iii)* $\pi_\varphi = \pi^*(\lambda^*)$. The minimal cut (in the Energetic-Lattice $\Pi(\omega(\lambda^*))$ is the solution of the primal LMCC problem 2.40, and the solution is unique since $\pi_\varphi$ must be the minimal element of a lattice. In the $\omega_\partial$-energetic lattice $\Pi(\omega_\partial)$ we

have $\lambda \geq \lambda^*$ implies $\pi^*(\lambda) \preceq_\partial \pi^*(\lambda^*) = \pi_\varphi$, hence $\curlyvee_\partial\{\pi^*(\lambda), \lambda \geq \lambda^*\} \preceq_\partial \pi^*(\lambda^*)$, and finally $\pi^*(\lambda^*) = \curlyvee_\partial\{\pi^*(\lambda), \lambda \geq \lambda^*\}$, which solves the LMCC dual problem 2.41. □

The comparison between the two LME and LMCC approaches is instructive. It shows how $h$-increasingness is not really essential, and that Lagrangians still work for lattices of partitions $\Pi_C$, and not only for the numerical lattice of the energies $\Pi_\omega$'s. But the most interesting feature is that theorem 2.13 applies to family of partitions of possibly infinite space $E$, as long as the number of classes is locally finite. This situation occurs quite often in the "remote sensing", where the area under imaging is incomparably smaller than the actual span of study. The LMCC provides such an independence (while not the LME) on account of the association an optimum with each leaf.

## 2.10 Class constrained minimization (CCM)

We now study a stricter constraint model of Class-local constraint. This restricts the constraint function to be defined now the classes and no more on the partitions. Further we see how this becomes purely a energetic lattice based solution to the solve the constrained optimization problem.

### 2.10.1 Single constraint

This section treats firstly the case of hierarchies and then that of braids, and develops an alternative method for constrained optimization, which does not resort to Lagrangians.

The hierarchy under study here is considered to be finite, and energies $\omega_\partial, \omega_\varphi : \mathcal{S} \to \mathbb{R}^+$ are defined on classes. Consider the following optimization problem:

**CCM problem**

$$\begin{array}{ll} \underset{\pi \in \Pi(\omega_\varphi)}{\text{minimize}} & \omega_\varphi(\pi) \\ \text{subject to} & \omega_\partial(S) \leq C_\mathcal{S}, \forall S \in \pi \end{array} \tag{2.42}$$

The method consists in generating a new hierarchy $H'$ where the minimization of $\omega_\varphi$ is no longer conditioned. Let $\Pi(C_\mathcal{S})$ stand for the family of the cuts $\pi$ of a hierarchy $H$, where constraint function values for each class $S$ is bounded to $C_\mathcal{S}$.

$$\pi \in \Pi(C_\mathcal{S}) \quad \Leftrightarrow \quad \{S \sqsubseteq \pi \quad \Rightarrow \quad \omega_\partial(S) \leq C_\mathcal{S}\}. \tag{2.43}$$

Obviously, the problem is feasible if and only if $\Pi(C_\mathcal{S})$ is not empty.

**Proposition 2.14.** *If the family* $\Pi(C_{\mathcal{S}})$ *is not empty, it is closed for the refinement infimum and supremum.*

*Proof.* As the number of levels of $H$ is finite, it is sufficient to prove that $\{\pi_1;\pi_2\} \in \Pi(C_{\mathcal{S}})$ implies that $\{\pi_1 \wedge \pi_2; \pi_1 \vee \pi_2\} \in \Pi(C_S)$. Consider $\pi = \pi_1 \wedge \pi_2$. At leaf $x \in E$ let the class of $\pi_1$ be $S_1$ and let that of $\pi_2$ be $S_2$. If $S_2 \subseteq S_1$, the p.p. $a_2$ of $\pi_2$ of support $S_1$ is the minimum of $\pi_1 \wedge \pi_2$ in $S_1$, and *vice versa*. In both cases the two infima have energies bounded by $C_{\mathcal{S}}$. This property remains true as point $x$ spans $E$, the infimum $\pi_1 \wedge \pi_2$ belongs to $\Pi(S_C)$. One continues similarly for $\pi_1 \vee \pi_2$. □

Since the family $\Pi(C_{\mathcal{S}})$ is closed under the refinement infimum, it admits a smallest element $\pi_0$

$$\pi_0 = \wedge\{\pi, \pi \in \Pi(C_{\mathcal{S}})\} \tag{2.44}$$

The classes of $\pi_0$ can be interpreted as the set of leaves of a new hierarchy $H'$, identical to $H$ above and on $\pi_0$, but where all classes below $\pi_0$ are removed (see Figure 2.5). The cuts $\pi$ of new hierarchy $H'$ are exactly those of $H$ that satisfy the class constraint $\omega_\partial(\pi) \leq C_{\mathcal{S}}$. The problem now reduces to find the non-conditional minimal cut of $H'$ w.r.t. $\omega_\varphi$, a question that we already know how to treat. If the minimization is considered in the $\omega_\varphi$-energetic lattice $\Pi(\omega_\varphi, H')$ relative to $H'$, we just have to suppose $\omega_\varphi$ is a singular energy. If we want that the minimal cut $\pi^*_\varphi$ induces a minimal energy, then, according Rel.(1.26), we must take $\omega_\varphi$ $h$-increasing (in addition $\pi^*_\varphi$ is then found in one bottom-up pass). Now we can state:

**Proposition 2.15.** *When* $\omega_\varphi$ *is a singular and* $h$*-increasing energy, then the minimal cut* $\pi^*_\varphi$ *in the* $\omega_\varphi$*-energetic lattice* $\Pi(\omega_\varphi, H')$ *is also a cut of smallest* $\omega_\varphi$ *energy in* $\Pi(\omega_\varphi, H)$ *whose all classes* $S^*$ *satisfy the cost constraint* $\omega_\partial(S^*) \leq C$*.*

The result is important. It grants the existence and the uniqueness of the minimal cut $\pi^*_\varphi$ under very large conditions: no prerequisite is needed for $\omega_\partial$, and uniquely singularity and $h$-increasingness for $\omega_\varphi$. Note that the cost $C$ has not to stay constant. Equivalence (2.43) holds on each class separately. $C$ may vary through the space, or according to the level $i$ in the hierarchy.

When the energy $\omega_\varphi$ is also increasing w.r.t. the refinement of the cuts (e.g. the quadratic deviation term Mumford-Shah energy), i.e. when:

$$\pi_1 \leq \pi_2 \quad \Rightarrow \quad \omega_\varphi(\pi_1) \leq \omega_\varphi(\pi_2), \tag{2.45}$$

then the minimal cut $\pi^*_\varphi$ coincides with $\pi_0$, since

$$\pi_0 \leq \pi \quad \Rightarrow \quad \omega_\varphi(\pi_0) = \wedge\{\omega_\varphi(\pi), \pi \in \Pi(C_\mathcal{S})\} = \omega_\varphi(\pi^*_\varphi). \tag{2.46}$$

### 2.10.2 Implementation for inf-modular $\omega_\partial$

It remains to build up the hierarchy $H'$ i.e. to find the leaves $\pi_0$. The search being combinatorial, the complexity reason drives us to call for inf-modularity for $\omega_\partial$. Put

$$\omega_\partial(S) \leq \wedge\{\omega_\partial(T), T \text{ son of } S\}, \tag{2.47}$$

i.e. the energy $\omega_\partial$ of class $S$ is smaller or equal to the smallest energy of the sons of $S$. Such class inf-modularity acts on classes and no longer on p.p. as Rel.(1.46), but both are equivalent. The inequality (2.47) is preserved indeed when any son $T$ is replaced by its own sons, which allows us to progressively obtain any p.p of the right member of (1.46). Conversely, it suffices to particularize Rel.(1.46) to the sons of $S$ for finding Rel.(2.47).

Fast implementation is then obtained by the following greedy top-down algorithm :

- index the classes of $H$ by a lexicographic ordering from the root $E$ to the leaves;
- starting from $E$, go down;
- when class $S$ has all its sons $T$ such that $\omega_\partial(T) \leq C_\mathcal{S}$, then replace $S$ by its sons;
- otherwise don't perform the replacement (this because, according to Rel.(2.47), every cut of the sub-hierarchy of root $S$ presents at least one class $T$ such that $\omega_\partial(T) > C_\mathcal{S}$, and the replacement would introduce undesired classes). Then keep $S$ and continue;
- iterate until all leaves have been processed.

The partition $\pi_0$ is obtained at the end of the scan, i.e. in one pass. A toy example is given in Figure 2.5.

W.r.t. the $\omega_\partial$-energetic lattice $\Pi(\omega_\partial, H')$, the cut $\pi_0$ turns out to be a maximum:

**Proposition 2.16.** *When $\omega_\partial$ is singular and inf-modular, then the infimum $\pi_0$ of $\Pi(C_\mathcal{S})$ is nothing but the maximal cut $\pi^{**}(\omega_\partial)$ whose $\omega_\partial$ energy is $\leq C$ in the $\omega_\partial$-energetic lattice $\Pi(\omega_\partial, H')$.*

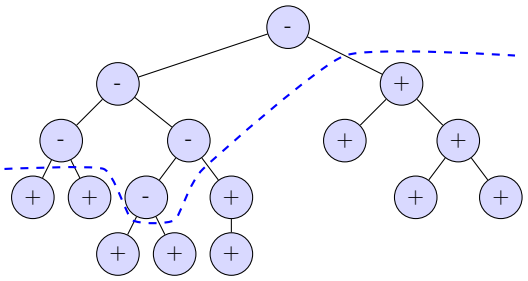


FIGURE 2.5: Minimal CCO cut. The + classes have $\omega_\partial > C$, the − classes have $\omega_\partial \leq C$.

*Proof.* Compare $\pi_0$ with any cut $\pi \in \Pi(C_\mathcal{S})$. The class $S$ of $\pi$ that contains leave $x$ is the support of a p.p. $a_0$ of $\pi_0$, since $\pi_0 \leq \pi$. Then the inf modularity (2.47) gives $\omega_\partial(S) \leq \omega_\partial(a_0)$, which results in $\pi \preceq_{\omega_\partial} \pi_0$ when $x$ spans $E$. As this is true for all cuts $\pi \in \Pi(C_\mathcal{S})$, we obtain $\pi^*(\omega_\partial) = \pi_0$. □

### 2.10.3 Class constraint versus Lagrange minimization by energies

What is the relation between the above results and those we can get by using $\omega_\partial$ as the constraint term in a Lagrange family? Suppose $\omega(\lambda), \omega_\varphi$, and $\omega_\partial$ singular, and extend $\omega_\partial$ to partitions by $\vee$-composition:

$$\omega_\partial(\pi) = \vee\{\omega_\partial(T), T \sqsubseteq \pi\}. \tag{2.48}$$

The energy $\omega_\partial$ inf-modular, since equality (2.48) remains valid when $\pi$ is reduced to the single class $\{S\}$. Therefore, according to Definition 2.7, the family $\{\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial, \lambda \in \overline{\mathbb{R}}\}$ is Lagrange. As above, $\pi^*(\lambda)$ stands for the minimal cut of the Lagrangian energetic lattice $\Pi(\omega(\lambda), H\}$. We saw (Proposition 2.8) that the function $\lambda \to \pi^*(\lambda)$ is increasing for the refinement, and $\pi^*(\lambda) \to \omega_\partial(\pi^*(\lambda))$ decreasing, so that there is a maximal $\omega_\partial(\pi^*(\lambda_0)) \leq C$. Suppose $\omega_\varphi$ strictly increasing for the refinement ordering; it is thus $h$-increasing. If $\pi_0 < \pi^*(\lambda_0)$ then $\omega_\varphi(\pi_0) < \omega_\varphi(\pi^*(\lambda_0))$ which contradicts the meaning of a minimum of $\omega_\varphi(\pi^*(\lambda_0))$ (Rel.(1.26)). On the other hand, we also have Eq.(2.46), which leads to $\pi_0 = \pi^*(\lambda_0)$, and we can state:

**Proposition 2.17.** *Let $\{\omega(\lambda) = \omega_\varphi + \lambda\omega_\partial, \lambda \in \overline{\mathbb{R}}\}$ be a Lagrange family where $\omega_\varphi$ is strictly increasing and $\omega_\partial$ is obtained by $\vee$-composition (Eq.(2.48)). Then the minimal cut $\pi^*(\lambda)$ equals the minimum $\pi_0$ given by Eq.(**??**).*

### 2.10.4 Vector case (multi constraints)

The multi-constraints situations are matter of the same approach. The minimization of $\omega_\varphi$ is now subject to $p$ constraints $\omega_{\partial i}(S) \leq C_i$. Denote by $\boldsymbol{C}$ the cost vector $\boldsymbol{C} = \{C_i, 1 \leq i \leq p \leq\}$ The family $\mathcal{A}(\boldsymbol{C})$ is that of the cuts whose classes $S$ satisfy the $p$ constraints $\omega_{\partial i}(S) \leq C_i$. The proposition 2.14 extends to this vector cases:

**Proposition 2.18.** *The family $\mathcal{A}(\boldsymbol{C})$ is closed for the refinement infimum and supremum.*

*Proof.* We keep the notation used in the proof of proposition 2.14. Let $\pi_1$ and $\pi_2$ be two elements of $\mathcal{A}(\boldsymbol{C})$. If $S_2 \subseteq S_1$, the p.p. $a_2$ of $\pi_2$ of support $S_1$ is the minimum $\pi_1 \wedge \pi_2$ in $S_1$, and all classes $S$ of $a_2$ satisfy the $p$ conditions $\omega_{\partial i}(S) \leq C_i$. The same occurs if $S_1 \subseteq S_2$, which results in that $\pi_1 \wedge \pi_2$ belongs to $\mathcal{A}(\boldsymbol{C})$, and achieves the proof □

We can also view the situation when one constraint at least is satisfied (a case impossible to treat with Lagrangians). The family $\mathcal{A}(\boldsymbol{C})$ is replaced by $\mathcal{B}(\boldsymbol{C})$ such that

$$\pi \in \mathcal{B}(\boldsymbol{C}) \quad \Leftrightarrow \quad \{S \sqsubseteq \pi \quad \Rightarrow \exists i \mid \ \omega_{\partial i}(S) \leq C_i\}.$$

and the previous proposition becomes:

**Proposition 2.19.** *The family $\mathcal{B}(\boldsymbol{C})$ is closed closed for the refinement infimum and supremum.*

(Proof similar to the previous one). Like in case of a single constraint, both $\mathcal{A}(\boldsymbol{C})$ and $\mathcal{B}(\boldsymbol{C})$ lead to hierarchies $H'$ with a unique $\pi_0$. Proposition 2.15 and the greedy algorithm still apply, with minor changes.

**Extension to braids**:

The class constrained minimization that we just developed for hierarchies extends to braids under one more condition. As Proposition 28 involves the refinement minimum of two partitions, i.e. of horizontal of cuts of $H$, we must also assume that the refinement minimum of two partitions of a braid is a partition of the hierarchy. It is clear that when this condition is fulfilled, then all results of the current section remain valid for braids.

### 2.10.5 Overview of the three models

The three models for constrained optimization on Braids and thus also hierarchies, basically introduces three different ways of introducing the constraint: Firstly LME, which

uses a numerical constraint on composition of energies of partial partitions constituting any cut, secondly LLCM, where the constraint is cut belonging to solution space and thus using lattice structure. Finally and thirdly, CCM, which provides a numerical constraint for each class.

**CCM vs LME**: CCM solves the problem of finding the cut of a hierarchy with least energy $\omega_\varphi$ under the constraint $\omega_\partial$; while LME provides an upper bound $\pi^*(\lambda_0)$; One should also note that CCM is valid for hierarchies, LME for the larger class of braids. In CCM, the mapping $\pi \to \omega_\varphi(\pi)$ holds globally on the cuts $\pi$, it is just supposed to be increasing, which is not very demanding; and $\omega_\partial$ is inf-modular; in LME $\omega(\lambda)$ is singular, $\omega_\varphi$ is $h$-increasing, and $\omega_\partial$ is inf-modular and $h$-increasing.

Another difference is that the CCM model is local, in the sense that the constraint $\omega_\partial$ is allocated to every class of the cut under study, and it becomes also global when one take the supremum for law of composition ( i.e. when the energy $\omega_\partial$ of a p.p. equals the supremum of the energies of its classes); for the implementation it is assumed that $\omega_\partial$ is inf-modular. CCM can use non-local costs $C$ which may vary over the space $E$, or remain constant.

Finally one can note that in CCM, the extension to the multi-constrained cases is straightforward, and concerns both logical *and* and *or* of constraints. This latter mode is out of the scope of LME.

## 2.11 Primal Vs Dual: $C$-cuts Vs $\lambda$-cuts

In this section we shortly discuss the scale-sets descriptor of Guigues [47], and multipler approximation by Salembier-Garrido [42], and provide interpret a primal version of the climbing algorithm, and why it is cumbersome. Both Guigues and Salembier-Garrido search values of the Lagrangian multiplier $\lambda$, so as to obtain a cut $\pi$ which reaches or is closest to an input or given constraint function value $\omega_\partial(\pi) \leq C$. We know that the solution to the dual, lower bounds the solution to the primal problem. Furthermore we have also seen that the non-existence of multipliers refers also to the lack of existence a strong duality. Thus spanning the $\lambda$-cuts at best provides us, in this case, with an upper-bounding cut. This is one of the important results in this chapter.

**Dual Domain Composition**: Furthermore a composition by addition of child energies $\omega_\varphi(T_i), \omega_\partial(T_i)$ associates with each braid and thus hierarchy, a $\Lambda$, i.e. scale function. In this dual domain, one can see that the scale-climbing corresponds to a "composition by supremum" of the Lagrangian parameters. This basically correponds to the infimum over the multiplier values in guigues are our result.

We have up until now, seen the following types of optimization on BOP/HOP:

1. unconstrained optimization of $\omega(\pi), \pi \in \Pi(B, E)$ using dynamic program
2. constrained optimization of $\omega_\varphi(\pi)$ subject to $\omega_\partial(\pi)$, for $\pi \in \Pi(B, E)$, using dynamic program on the space multipliers $\lambda$. (DP on Dual Problem)

both using the energetic lattice structure to ensure an optimum.

Now we will consider the problem (2) above in the primal domain, and establish why it is easier consider the dual domain.

These are the family of cuts in the braid whose constraint function has a value smaller than $C$. One thing that can be directly noted about these partitions are that they are in any sense monotonically ordered w.r.t $C$. Furthermore we can see the nature of the energetic lattice unravel itself with this simple parametrized family of cuts. This is demonstrated in figure 2.4.

The $\lambda$-cuts also have the same problem, except that the implicit condition of having the largest partition by Guigues, Breiman and various other authors, provides a unique solution. This has been explicited in two forms, first as scale-increasingness condition on the energies and, second as singularity condition, which ensures a unique minimum.

## 2.12 Summary

**Chapter contribution summary**

- We have shown that the $\lambda$-cuts correspond to an upper-bound and is not a global solution to a constrained optimization problem on HOP and BOP.
- The solutions in fact are shown to be a Lagrangian relaxation of the original constrained optimization problem, who solutions are global only at the cost endowed by the multiplier value.
- We demonstrate three locally constrained optimization models, that use the partition itself as constraints, enforcing further the energetic lattice structure.
- We also demonstrate basic perturbation and penalty methods to obtain a better bounding $\lambda$-cut. An interesting prospect would be to use a energetic lattice based perturbation for quicker convergence rates. A further analysis of the problem would be to consider the global convergence analysis of the constrained optimization problem on HOPs and BOPs. In such a context we propose three different models.
- In using the energetic lattice for constrained optimization, we have three different classes of constrained optimization problems that correspond to the three different ways of enforcing a constraint:
  - Lagrangian Minimization by Energy(LME): (By numerical constraint on engery) Energetic lattice based generalization of the Lagrangian, when one works in the space of partitions from a Braid, instead of $\mathbb{R}^n$.
  - Lagrange minimization by Cut-Constraints (LMCC): (By a partition constraint) We introduce a partition based constraint optimization model, which does not involve any numerical constraint function, but one that is driven by the energetic lattice.
  - Class constrained minimization (CCM): (By numerical constraint on engery on a class) The third model demonstrates how the energetic lattice handles local constraints. There is no Lagrangian formulation here.

# Chapter 3

# Applications of Energetic Lattice

**Publications Associated with Chapter**

- [59] Global-local optimizations by hierarchical cuts and climbing energies, Pattern Recognition(PR) 2014.
- [103] Optima on Hierarchies of Partitions, ISMM 2013.
- [61] Ground truth energies for hierarchies of segmentations, ISMM 2013.

This chapter demonstrates the applications of dynamic program for energetic lattices, involving different compositions of energies. A specfic contribution and application is the formulation of *proximal partition extraction*. This consists in extraction of a partition, given a hierarchy of segmentations and a ground truth partition corresponding to an image, that is *closest* to the ground truth or marker partition.

## 3.1 A few useful $h$-increasing energies

Following the description of $h$-incresingness and the different compositions of energies possible we briefly describe a few common segmentation models that can be formulated as $h$-increasing energies.

### 3.1.1 Mumford and Shah energy

One of popular, notably non-convex image segmentation model is the Mumford-Shah functional [85]. One can find an exhaustive study of this functional in Morel et al.'s book [82], where it is formulated in the Euclidean plane using edges which are composed of rectifiable simple arcs.

$$\omega(\pi(S),\lambda) = \omega_\varphi(\pi(S)) + \lambda\omega_\partial(\pi(S)) = \sum_{1\le u\le q}\int_{x\in T_u} \| f(x) - \mu(T_u) \|^2 + \lambda \sum_{1\le u\le q} (\partial T_u) \tag{3.1}$$

This can be generally written as an affine energy of the form

$$\omega(S,\lambda) = \omega_\varphi(S) + \lambda\omega_\partial(S) \qquad S \in \mathcal{S}. \tag{3.2}$$

The first term, $\omega_\varphi$, is the additive *fidelity term* which sums up the quadratic deviations from the mean value $\mu(T_u)$ over the class $T_u$ over different $u$ producing a partial partitioning of class $S$, and the second term $\omega_\partial = \partial T_u$, the lengths $\partial T_u$ of the frontiers of all $T_u$.

Both increasingness relations, $h$-increasingness and scale-increasingness (1.27) and (1.42), are satisfied by the family of energies $\omega + \varphi, \omega_\partial$ in eqn (3.1). For short these energies are called *climbing*.

When the energy $\omega_\partial$ is sub-additive, i.e.

$$\omega_\partial(\bigcup_{1\le u\le q} T_u) \le \sum_{1\le u\le q} \omega_\partial(T_u), \tag{3.3}$$

then the family is obviously scale increasing, since

$$\omega_\partial(S) = \omega_\partial(\bigcup_{1\le u\le q} T_u) \le \sum_{1\le u\le q} \omega_\partial(T_u) = \omega_\partial(\pi(S)).$$

Conversely, L. Guigues has shown that the condition (3.3) is necessary for scale increasingness [47].

### 3.1.2 Additive energy by convexity

The arc length function is not the only choice. One can also think about another $\omega_\partial(S)$, which reflects the convexity of the class $S$. Consider in $\mathbb{R}^2$ a connected set $S$ without holes and with a non zero curvature everywhere on $\partial S$. Let $d\alpha$ be the elementary rotation of its outward normal along the element $du$ of the frontier $\partial S$. As the curvature $c(u)$ equals $\frac{d\alpha}{du}$, and as the total rotation of the normal around $\partial S$ equals $2\pi$, we have

$$2\pi = \int_{c\ge 0} c(u)du - \int_{c<0} |c(u)|\, du.$$

When dealing with partitions, the distinction between outward and inward vanishes, but the parameter

$$\omega_\partial(X) = \frac{1}{2\pi} \int_{\partial S} |c(u)|\, du \tag{3.4}$$

still makes sense. It reaches the minimum value 1 when set $S$ is convex, and increases with the degree of concavity. For a starfish with five pseudo-podes, it values around 5. Now $\omega_\partial(S)$ is sub-additive for the open parts of contours, therefore it can participate as a regularity term in an additive energy. In digital implementation, the angles between contour arcs must be treated separately (since sub-additivity applies on the open parts).

#### 3.1.2.1 Additive energies by active contours

The active contours aim to match regular closed curves with the zones of maximum variation in an image, example the Snakes or Chan et al.s Active contours [24, 57]. The energies we view are particular cases of active contours adapted to hierarchies, and derive from the approach proposed by Y. Xu at Al. [124]. The main idea is the following: each node $S \in H$ is dilated and eroded by a disc $B$, and the two crowns $S \oplus B \backslash S$, and $S \backslash S \ominus B$ are compared. This comparison stands for the fidelity term $\omega_\varphi$ in Rel. (3.2), and a function of the curvature (e.g. Rel. (3.4)) stands for the regularity term. One goes from sets to partial partitions by additivity, according to the relation (1.31).

The simplest comparative energy is given by the difference of a given energetic function $f$ on the two crowns:

$$\omega_\varphi(S) = \left| \int_{(S \oplus B \backslash S)} f(x)dx - \int_{(S \backslash S \ominus B)} f(x)dx \right|, \quad S \in \mathcal{P}(E). \tag{3.5}$$

It can be expressed in a dimensionless form by putting:

$$\widetilde{\omega}_\varphi(S) = \left| \frac{\int_{(S \oplus B \backslash S)} f(x)dx - \int_{(S \backslash S \ominus B)} f(x)dx}{a(S)} \right|, \quad S \in \mathcal{P}(E),$$

where $a(S)$ denotes the area of $S$. When the absolute value bars are removed, the both energies $\omega_\varphi$ and $\widetilde{\omega}_\varphi$ become sub-additive. Alternatively, the energy $\omega_\varphi$ proposed in [124] is the sum of the variances of $f$ in the two crowns, divided by the variance of $f$ in the union of these two crowns.

For the regularity term $\omega_\partial$ of the energy (3.2), one classically takes the above function $\nu$ of Rel. (3.4), which is scale increasing and generates the climbing family $\{\omega_\varphi + \lambda\omega_\partial\}$.

### 3.1.3 Mumford-Shah Energy for Color image segmentation

We aim to find an optimal cut which provides the simplified version of a colour image $f$, constrained by compression rate. A hierarchy $H$ has been obtained on input image figure 3.1, by segmentations of the scalar luminance $l = (r + g + b)/3$ based on flooded watersheds [32]. In each class $S \in H$, the simplification consists in replacing the function $f$ by its mean value of colors, i.e. the means of the three channels over $S$. Note that this colour mean does not intervene directly in the three energies (3.6) to (3.8), but rather in the display of the optimal cut. We use the energy $\omega(S)$, as defined in equation (3.2) to demonstrate different optimal cuts.

For the first experiment, as the data fidelity term $\omega_\varphi(S)$ we take the variance of the luminance for each class $S$ of hierarchy $H$ (first term of equation (3.6)). The regularization term $\omega_\partial(\pi)$, is equal to the contour length $\mid \partial S \mid$, plus 24 bits for the average color of $S$. This gives the energy $\omega_{lum}(S)$, whose result is shown in figure 3.1 (left).

$$\omega_{\text{lum}}(S) = \sum_{x \in S} \parallel l(x) - \bar{l}(S) \parallel^2 + \lambda(24 + \mid \partial S \mid), \tag{3.6}$$

In a second experiment depicted in figure 3.1(right), we separate each colour vector $(r, g, b)$ into two components: the vector luminance $\overrightarrow{l} = (l, l, l)$ which gives the gray scale, plus the orthogonal chrominance vector $\overrightarrow{c} = (r - m, g - m, b - m) = (c_1, c_2, c_3)$ whose module is the saturation.The fidelity term of the energy is now the sum of the variances of the components of $\overrightarrow{c}$ over $S$ as shown in (3.7).

$$\omega_{\text{chrom}}(S) = \sum_{x \in S, 1 \leq i \leq 3} \parallel c_i(x) - \bar{c}_i(S) \parallel^2 + \lambda(24 + \mid \partial S \mid), \tag{3.7}$$

The principle idea demonstrated by this experiment is the independence between the function generating the hierarchy and the energy creating the energetic ordering or lattice, and the imminent minimization by a dynamic program. We observe in figure 3.1 (right) that the plant in front of the female duck are now correctly segmented, and that the water in the background has lesser detail.

$$\omega_{\text{Texture}}(S) = \omega_{\text{chrom}}(S) + \sum_{S\prime \in siblings(S)} \frac{\mu}{\sigma^2(Area(S\prime))}, \tag{3.8}$$

This leads to a third experiment, depicted in figure 3.2, based on energy in equation (3.8). This experiment is to extract textures parametrized by regularity. The energy (3.8) in is similar to (3.7), except for the term in which is inversely proportional to the quadratic

FIGURE 3.1: First row: Initial image and Saliency function corresponding to hierarchical watershed flooding [32]. Second Row: Optimal cuts by using variance of luminance(left), chrominance(right).

FIGURE 3.2: Optimal Cuts for texture using variance of chrominance for scale $\lambda = 100$: Left, input Image, middle and right, cuts for parameters at $\mu = 10^{12}$ (low uniformity) and $\mu = 10^{14}$ (high uniformity), in Eq. 3.8.

deviation of sizes (area) of the children from the mean size of the children, which is a trivial $h$-increasing energy. Furthermore the fidely term is a quadratic deviation of the chrominance vector being minimized on the partitions of hierarchy produced from the luminance vector $l$. This experiement basically shows that the optimal partitions that are extracted by using minimal norm of a vector does not remain the same when one obtains when using their components.

Intuitively, texture features are formulated into this multi-scale framework where obtains a two level filtering scale parameter, which combines the effect of minimal deviation of chrominance and structure of texture into one global energy function. $\lambda$ controls the perimeter of the segmentation, while $\mu$ constrols the regularity of classes in the segmentation. The monotonicity of the segmentations in the optimal hierarchy becomes a bit more complicated and is subject to more detailed study, since the pair $\pi(\lambda_i, \mu_k), \pi(\lambda_j, \mu_l)$, for $i < j, k < l$ are cuts that are not predestined to be ordered, and would require a more

general law of scale-increasingness. This demonstration basically shows the flexibility of the multiscale-energy formulation.

We conclude this section by quoting Salembier and L. Garrido [100], and L. Guigues [49], who demonstrate constrained optimization on HOP, by replacing the gradient descent by *climbings* dynamic program for family of $h$-increasing energies.

### 3.1.4 Hierarchical structure based energies

In this experiment we demonstrate how one can use the hierarchical structure itself as a constraint. This is possible by using the integral of the saliency function on support of the contours of the partial partitions (and not purely the length of partial partition).

That is a numerical measure $g(x) = 1 - s(x)$ is introduced on the perimeter $\partial S$

$$\omega_g(S) = \sum_{x \in S, 1 \leq i \leq 3} \parallel c_i(x) - \bar{c}_i(S) \parallel^2 + \lambda \int_{\partial(S)} g(x) \mathrm{d}x, \tag{3.9}$$

This constraint function requires that the saliency value on the partial partitions be the highest, while having a minimal $\omega_\varphi$. A high saliency function value implies higher level in the hierarchy. This enables the use of saliency function itself as a multi-scale constraint function. In our example we use saliency function's value which corresponds to the *connection value* of watershed floodings by volume [32]. This is demonstrated with a set of optimal cuts at different values of $\lambda$.

One can also use a value which is not dependent on the gradient function but a proximity function to a ground truth. This will be demonstrated in the following section.

## 3.2 Ground truth Proximal Energies

In this section we concentrate on a particular application which is supervised image segmentation evaluation. More specifically we will look at ground truth set based segmentation evaluation. For this purpose we work on the Berkeley Dataset [10]. This chapter focusses on the case when the segmentations are part of a hierarchy and different ground truth sets provide different qualitative information on the segmentation. The motivation in ground truth based evaluation of hierarchicy of segmentations lies in the key fact that the ground truth partition can be found at a *finite* scale or cut in the hierarchy of segmentations, above and below which the cuts become non-optimal,

resulting in sub-optimal F-measures. This is refered to in [94] as the upper-bound partition selection, and poses it as a combinatorial optimization problem. Here we will motivate the requirement of an energetic lattice and the $h$-increasing energy that can be formulated using Haussdorf distance to solve the same problem. This provides a way to climb up to the partition *closest* to the ground truth partition.

### 3.2.1 Ground truth evaluation

Here we introduce the problem of ground truth evaluation of hierarchies of segmentations a bit more strictly. First let us determine our inputs and outputs we desire from the problem. Given a hierarchy of segmentations $H$ and a ground truth partition $G$ we would like to determine:

1. The cut $\pi^*$ in hierarchy $H$ that is closest to ground truth partition $G$. Here we will use the *localised* Haussdorf distance to define a maximal distance between two a cut in the hierarchy and a ground truth set. This gives two energies associated to the two sense's of proximities:
   - Hierarchy to Ground truth partition: $\pi \to G, \pi \in H$.
   - Ground truth partition to Hierarchy: $G \to \pi, \pi \in H$.
2. Compare any hierarchy $H$ with multiple ground truth partitions of the same image.
3. Compare any two hierarchies with respect to a common ground truth partition.

We demonstrate in figure 3.3 how the above problem is a scale selection problem by a toy example.

### 3.2.2 Segmentation Versus GT Partitions: Refinements and Overlaps

In this section we consider the different topological possibilites between the classes of the ground truth partition and segmentation being evaluated. This has been first addressed elaborately in [70]. He associates the variablity in human segmentation to different perceptive differences between human annotators: mainly classifying them as varying attention to detail in the scene, thus producing refinements of the same object or region in the image.

To the ends of classifying the different situations possible between a ground truth partition and a Image segmentation, Tuset et al (following Martin's error measures LCE, GCE) [93] classifies them into 4 classes (see figure 3.4):

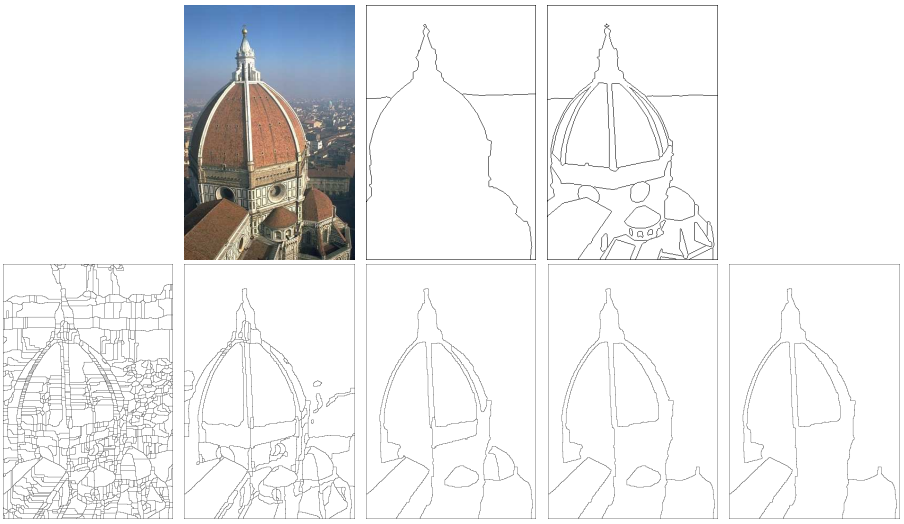


FIGURE 3.3: The first row shows an input image, with two ground truth partitions corresponding to the image, from the Berkeley dataset. The bottom row consists of a sequence of thresholds of the Ultrametric Contour Map(UCM) segmentation hierarchy. The problem now consists in extracting a proximal partition in the hierarchy that is closest to one of these ground truth partitions. Further how do we compose when we have multiple ground truth partitions, and how do we compare hierarchies w.r.t a single ground truth.

1. **Oversegmentation**: When the union of a finite set of classes in the segmentation is a class of the ground truth partition. In other words, the segmentation is locally a refinement of the ground truth partition.

2. **Undersegmentation**: When the union of a finite set of classes in the ground truth partition is a class of the segmentation. In other words, the ground truth partition is locally a refinement of the segmentation.

3. **Overlaps**: When the classes of the segmentation and ground truth overlap but do not produce refinements. This termed as noise in case of [93].

4. **Braids**: When the same support is segmented in a non-nested structure. The supremum of the two partial partitions have the same support, in other words the same monitor. This difference in local segmentation can result, when we have textured regions or smooth zones which are differently segmented by human experts. This has been observed in the evaluation of segmentation algorithms by Unnikrishnan et al. [115]

To be more precise, the cases 1,2,4 are all in the general structure of a Braid of partition.

Now a single or a set of ground truths can demonstrate either one or all of these characteristics. Here we demonstrate how the refinements are handled using the inf-composition. For a similar study on the saliency function, please see 4.7.5.

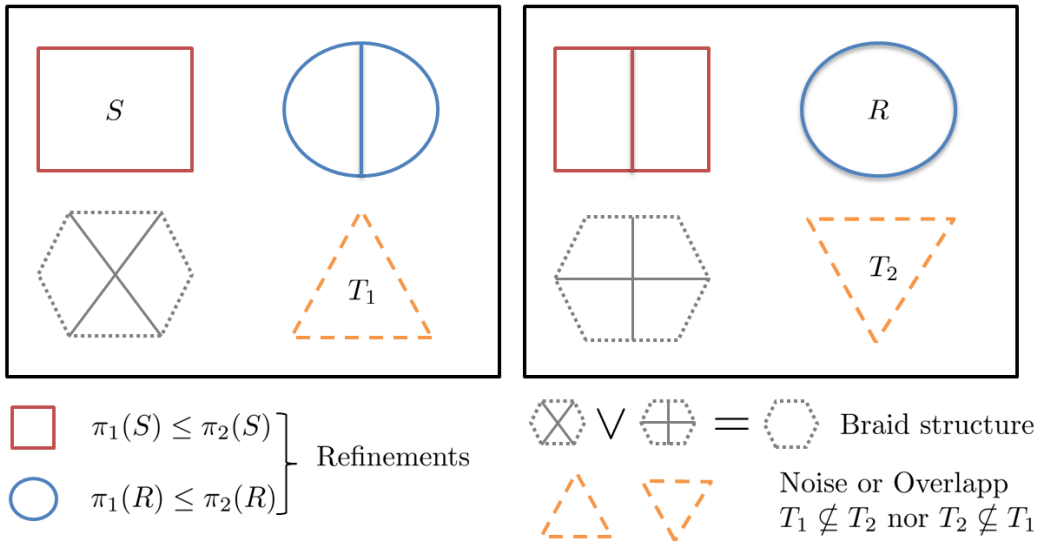



FIGURE 3.4: Example demonstrating two partitions $\pi_1$ and $\pi_2$, where one of them could be a ground truth segmentation while the other being a machine segmentation. Figure demonstrates the different possible configurations of refinement, braid and overlaps of classes.

### 3.2.3 Segmentation Evaluation Measures

The comparison of two segmentations of the same image is not clear, as image segmentation is inherently ill-defined: there is no single ground truth (partition, since when we compare segmentations it necessitates that the ground truth is a partition too!) that can capture the faulty and correct labellings in the two segmentatations [115]. Once this is lack of definition is noted, one can find heuristic assumptions on the correspondence between ground truth set and segmentation, providing us with a variety of evaluation measures in literature. We cite few which are important for our discussion later. The measures evaluating hierarchy of partitions has been studied extensively in the thesis by Tuset et al. [93].

1. Many region based measures have the Jaccard distance between sets as a begining point. The Jaccard index between finite sets $S, S'$ is given by:

$$J(S, S') = \frac{S \cap S'}{S \cup S'} \tag{3.10}$$

The Jaccard distance is given by:

$$d_J(S, S') = 1 - J(S, S') = 1 - \frac{S \cap S'}{S \cup S'} \tag{3.11}$$

2. We can have different refinement between the ground truth partition and the image segmentation. That is if subsets of regions in one segmentation consistently merge into some region in another segmentation the consistency error is low [70]. The refinement-invariant evalation measures include Global Consistency Error(GCE) and Local Consistency Error(LCE). The Global Consistency Error (GCE) assumes that one of the segmentations must be a refinement of the other, and forces all local refinements to be in the same direction. The Local Consistency Error (LCE) allows for refinements to occur in either direction at different locations in the segmentation.

3. Segmentation Covering [10]:

   The overlap between two regions (classes of segmentation $S$ or $S'$) $R$ and $R'$, is defined as:

   $$\mathcal{O}(R, R') = \frac{R \cap R'}{R \cup R'}$$

   and the segmentation covering is defined as:

   $$\mathcal{C}(S \to S') = \frac{1}{N} \sum_{R \in S} |R| \cdot \max_{R' \in S'} \mathcal{O}(R, R') \tag{3.12}$$

4. Boundary based measures: D.Martin thesis [70] provides a variety of Ground truth based evaluation measures. Using the boundary of segmentations, the unitary element: *Edgel*, where he performs edgel correspondence between the segmentations. This is an example of contour/boundary based measure.

It would be interesting now to observe that the contour based measures are sensitive to the placement of class contours w.r.t each other unlike the region based measures. It is important to note that a single human annotated ground truth rarely provides an objective segmentation containing all objects of interest. Figure 3.5 demonstrates the problem across different Ground truth partitions of the same image.

### 3.2.4 Haussdorf Distance

The Haussdorf distance is a tool used in the image processing community to compare images (cites needed). It is well known that the Hausdorff distance is a metric over the set of all closed, bounded sets, while satisfying properties of identity, symmetry and triangle inequality.

Let $E, d$ be a metric space of distace $d$ and let $A, B \subset E$ be non-empty subsets of space $E$. The Haussdorf distance can be defined now as:

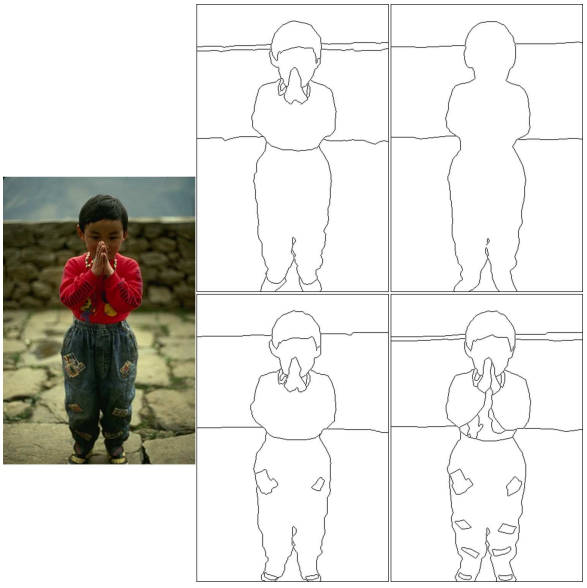


FIGURE 3.5: GT's corresponding to an input image. This demonstrates how the human experts in this case have drawn different scales of details in the scene. All scales are not present in no single GT partition. This as described well across literature is due to the ill-posedness of the segmentation problem. The OIS averages the results of choosing the right scale of partition from the UCM across various GTs to evaluate the segmentation hierarchy. We will use instead an inf-composition to extract the proximal partition. We remark the inherent braid structure in such cases.

$$d_H(A,B) = \max \sup_{a\in A} \inf_{b\in B} \rho(a,b), \sup_{b\in B} \inf_{a\in A} \rho(a,b) \tag{3.13}$$

The Haussdorf distance in equation (3.13) can be calculated by the supremum of minimal radii of dilation by a ball of one set to cover the other set [102]. If $\delta_r(X)$ represents a dilation of set $X$ by a compact ball of radius $r$, then:

$$d_H(A,B) = \inf\{r : B \subseteq \delta_r(A); A \subseteq \delta_r(B)\} \tag{3.14}$$

equivalently for an erosion operation $\epsilon_r(X)$,

$$d_H(A,B) = \inf\{r : \epsilon_r(A) \subseteq B; \epsilon_r(B) \subseteq A\} \tag{3.15}$$

**Problems with the Haussdorf Distance**: The Hausdorff distance is very sensitive to even a single "outlying" point of $A$ or $B$. In the case of classes of partitions this corresponds to a small convexities/concavities in shapes that produces large values of haussdorf distance between the sets or classes. The staggered class determines by its distance alone the Haussdorf distance $H(A,B)$ between the two sets, and subsequently

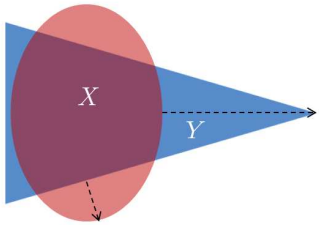



FIGURE 3.6: Haussdorff distance Assymetry: Smallest disc dilation of X that contains Y is drastically larger than theat of X to contain Y, thanks to a the difference in convexities of the shape. The same situations occur when dealing with classes of a partition.

the partitions. This is a well known problem with the Haussdorf distance. This is demonstrated in figure 3.6.

We have to address thus two questions:

- How to define the Haussdorf distance for partial partitions or classes of partition?
- How to handle asymmetric shapes by formulating a composition of energies over the partial partitions ?

Felzenszwalb et al. propose the use of a dynamic program approach using curves of shapes, ordered in a tree [41] for deformable shape matching.

### 3.2.5 Hausdorff distance

Most of the supervised evaluations of hierarchies, including Arbelaez et al., Tuest et al. [7, 10, 94], derive from the intuition of the Hausdorff distance, in various critical manners. Let us briefly recall this background.

In a metric space $E$ of distance $d$ we aim to match the support $S(\pi)$ of a bounded partial partition $\pi$ with a set $G$ of points and lines, considered as a GT drawing. The smallest isotropic dilation of $G$ that covers the contour $S(\pi)$ has a radius

$$\rho_G = \inf\{\rho \mid G \oplus \rho B \supseteq S(\pi)\}, \tag{3.16}$$

where $\rho B$ is the disc of radius $\rho$ centred at the origin. One can interpret $\rho_G$ as the "energy" required for reaching $\partial S$ from the GT $G$. In the same way, the counterpart covering is given by the radius $\rho_A$:

$$\rho_A = \inf\{\rho \mid S(\pi) \oplus \rho B \supseteq G\}. \tag{3.17}$$

By introducing the so called *distance function* $d(x, Z)$ from point $x$ to the fixed set $Z$, i.e.

$$d(x, Z) = \inf\{d(x, z), z \in Z\} \quad x \in E \tag{3.18}$$

we see that

$$\rho_G = \sup\{d(x, G), x \in S(\pi)\} \quad \text{and} \quad \rho_A = \sup\{d(x, S(\pi)), x \in G\}, \tag{3.19}$$

an interpretation which connects the distance function with the partial order on sets by inclusion. In Rel.(3.19) the value $\rho_G$ (resp. $\rho_A$) is the maximal distance from a point of $\partial S$ to $G$ (resp. of $G$ to $\partial S$). The first one, $\rho_G$, indicates how precise is $S$ w.r.t. the GT, the second one, $\rho_A$, how representative is this GT. In indexation, these two numbers are respectively named *precision* and *recall*. The symmetric expression $\rho = \max\{\rho_G, \rho_A\}$ is the well known Hausdorff distance

Hausdorff distance is lacking of finesse because it is a global notion, and of robustness because it uses suprema. If we could define a local equivalent, associated with each class $T$ of $\pi$, and no longer with the whole $S(\pi)$ itself, then the regions with a good fit would be treated separately from the others. And in addition, if this equivalent was $h$-increasing, then it would provide an energy for calculating easily the associated optimal cut [59], i.e. the smallest upper bound of all cuts of the hierarchy, in the wording of [94]).

### 3.2.6 Half Haussdorf distances

**Composition of Ground truths for Segmentation Evaluation**: Peng et al. [92] evaluate segmentations with multiple ground truths as human-labeled ground truths are only a small fraction of all the possible interpretations of an image. Furthermore they describe that the labeled ground truths set by itself might not be optimal to compare with the input segmentation, and conclude that such evaluation leads to a certain biassed evaluation. Their key observation lies in the local structural similarities between the groundtruhths and the segmentation. They create a composite ground truth, which remains similar across all ground truths in the set, created by a labeling minimizing a potts prior. We will use this structural similarity feature in defining our composition of energies across multiple ground truths.

**Region based and Contour Based Measures** The ground truth set can be interpretted as a boundary or a set of segments based on which there can be many measures defined. Movhedi et al. [84] results show that a *Contour Mapping* measure based upon contour bimorphisms between the boundaries of the object segmentations under comparison were most consistent with psychovisual studies involving human evulation. Further more this also suggests that the ground truth set need not always contain partitions locally.

Based on the above motivations we have constructed the half(complementary) haussdorf local energies $\omega_G, \theta_G$ as shown in figure 3.7. These energies are called precision and recall energies, by corruption of classical usage of terms precision and recall refering to the type I and II of errors.

#### 3.2.6.1 Precision energy

We now focus on the classes $\{T_i\}$ whose concatenation $T_i \sqcup T_2 ... \sqcup T_k$ generates the partition $\pi$. The $\{T_i\}$ are said to be the children of parent $S$. Consider the class $T_i$ of the partition $\pi$. The smallest dilate $G \oplus \rho B$ that covers $T_i$ has a radius:

$$\omega_G(T_i) = \inf\{\rho \mid G \oplus \rho B \supseteq T_i\}. \tag{3.20}$$

By taking the supremum of all $\omega_G(S)$ we find the above value $\rho_G$ of Rel.(3.16):

$$\rho_G = \bigvee\{\omega_G(S), S \sqsubseteq \pi_A\}. \tag{3.21}$$

This shows the soundness of $\omega_G$. But a problem arises when we want to extend it from sets to the partial partitions $\mathcal{D}(E)$ of $E$ by some law of composition between the $T_i$. When the chosen energy is $h$-increasing, which will always be the case here, finding optimal cuts in hierarchies amounts to compare the partition energies of parents and children [59]. If we compose the energies of the children by supremum, then we trivially always find $\omega_G(\pi) = \omega_G(S)$, the parent. If we compose by infimum, we have $\omega_G(\pi) = \omega_G(S)$ when the $\omega_G(T_i)$ all identical, and $\omega_G(\pi) < \omega_G(S)$ when not. And if we compose the energies of the sons by averaging, we obtain again $\omega_G(\pi) < \omega_G(S)$. Therefore, in all cases, we arrive to an optimal cut which can only be at the lowest level of hierarchy $H$, i.e. the leaves, or at the highest one, i.e. the space $E$ itself.

For being more informative, we can introduce a trade off based on mutual comparisons of the energies of the sons. An easy way consists in adding a quantizer $\lambda$ in the composition

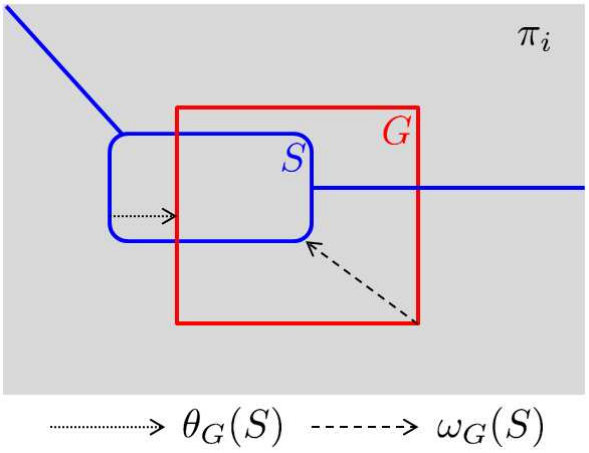



FIGURE 3.7: Energies $\omega_G(S), \theta_G(S)$ for each class $S$ in a segmentation, defined w.r.t class from a ground truth partitionin red. The composition of these energies decid the local distance measure introduced and minimized.

by infimum, so that

$$\omega_G(\pi) = \omega_G(T_i \sqcup T_2 ... \sqcup T_k) = \inf\{\omega_G(T_i)\} + \lambda. \tag{3.22}$$

As this new energy is $h$-increasing, the optimal cut is reached in one pass by comparing the respective energies of sons and fathers [59]. As $\omega_G(S) = \sup\{\omega_G(T_i)\}$, we have $\omega_G(\pi) < \omega_G(S)$ iff $\lambda < \sup\{\omega_G(T_i)\} - \inf\{\omega_G(T_i)\}$.

The parent replaces its children when the latter are sufficiently "identical" , parameterized by complexity parameter $\lambda$. For each value of $\lambda$ we thus obtain the cut which minimizes the distances to the ground truth $G$, i.e. the smallest upper bound of all cuts, as posed by Tuset et al. [94]. To give an idea of the distribution of the energies $\omega_G(S), \Theta_G(S)$ shown in Fig 3.8 For two different ground truths, over different partitions from a hierarchy. As seens there are cases where the parent is as proximal as the child.

#### 3.2.6.2 Recall energy

The number $\omega_G(S)$ informs us about those points of $\partial S$ close enough to $G$, but not on those of $G$ close to $\partial S$. We cannot take, here, the dual form of the $\omega_G(S)$ of Rel.(3.20), as we did before with the global Hausdorff distance. Such a dual energy would be

$$\omega'_G(S) = \inf\{\rho \mid S \oplus \rho B \supseteq G\}, \tag{3.23}$$

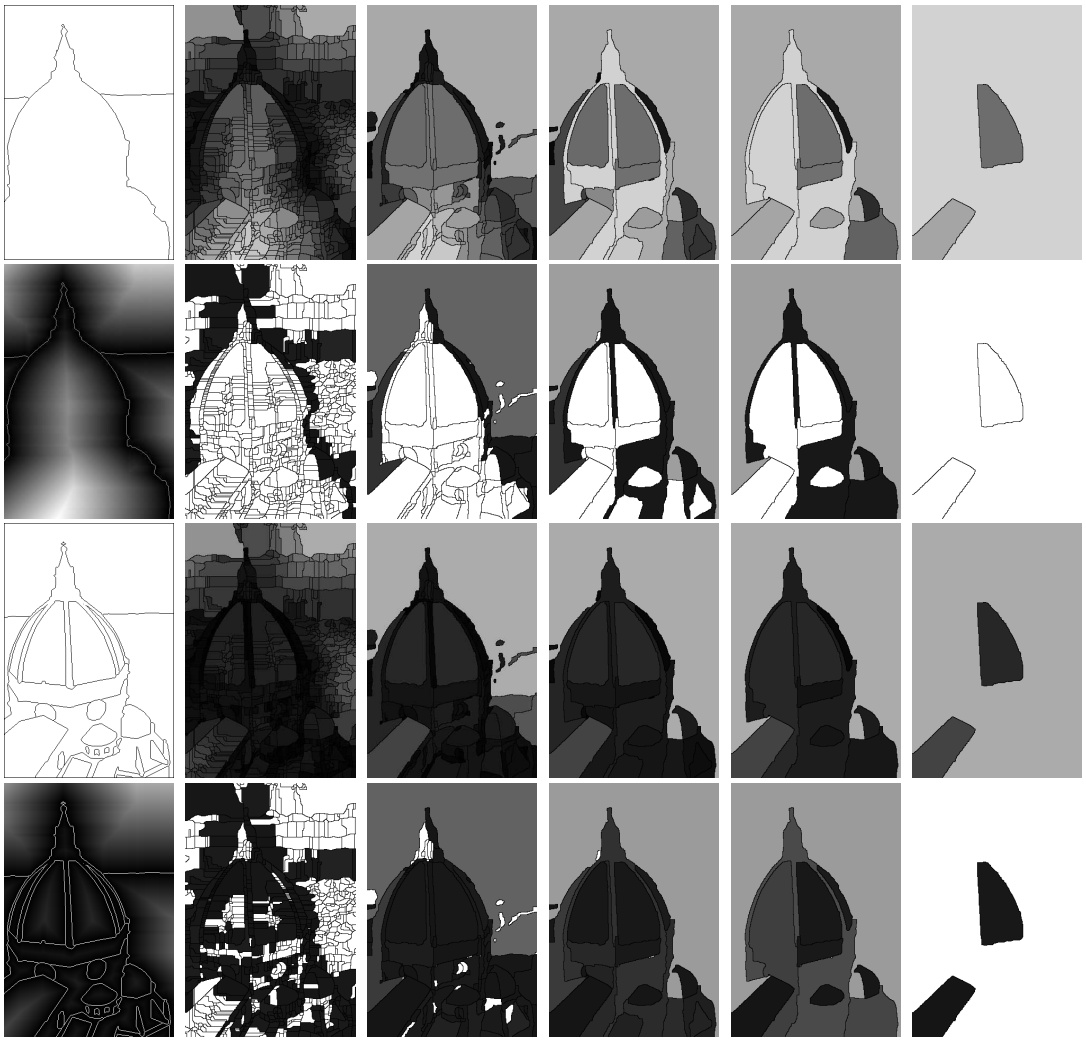


FIGURE 3.8: Row 1: $\omega_{GT2}(S)$, Row 2: $\theta_{GT2}(S)$, Row 3: $\omega_{GT7}(S)$, Row 4: $\theta_{GT7}(S)$. Figure shows the different half haussdorf proximity functions $\omega(S)$ and $\theta(S)$ for each class from different partitions in a hierarchy. The two ground truhs chose are of different scales. The gray scale value 0 corresponds to closest while 255 corresponds to farthest. Ground truth and associated distance function on left, energy values over 6 different partitions from the hierarchy on its right. One can already get a quick idea of what the dynamic program would extract an a minimal cut looking at the individual values. One can see that the scale of the ground truth partition affects the energy associated with classes of the hierarchy of segmentatations. What's left is to obtain a good composition.

a quantity which risks to be extremely large, for the drawing $G$ may spread over the whole space, whereas class $S$ is locally implanted. Fortunately, when dealing with $h$-increasing energies, one is less interested in the actual values of the energies than by their *increments* between fathers and sons. Now, when a point of $G$ is outside class $S$, then its distance to $S$ is the same as the max of the distances to the sons $T_i$ of $S$:

$$x \in G \cap S^c \quad \Rightarrow \quad d(x,S) = d(x,\partial S) = \bigvee d(x,T_i) = \bigvee d(x,\partial T_i), \tag{3.24}$$

so that the part of $G$ exterior to $S$ is not significant. For the sake of comparison, it thus suffices to focus only on the distances involved in the covering of $G \cap S$ by dilations of $\partial S$ on the one hand, and on those of $\partial T_i$ on the other hand. Then the energy $\omega'_G$ of Rel.(3.23) has to be replaced by the more appropriate one

$$\theta_G(S) = \inf\{\rho \mid S \oplus \rho B \supseteq G \cap S\}. \tag{3.25}$$

When $S$ spans all classes of a partition $\pi_A$, then the supremum of all $\theta_G(S)$ gives the value $\rho_A$ of Rel.(3.17)

$$\rho_A = \bigvee\{\theta_G(S), S \sqsubseteq \pi\}, \tag{3.26}$$

and the (global) Hausdorff distance $\rho$ between $\pi$ and $G$ turns out to be the double supremum,

$$\rho = \bigvee\{\{\omega_G(S) \bigvee\{\theta_G(S)\}, S \sqsubseteq \pi\}. \tag{3.27}$$

It remains to verify that $\theta_G$ is $h$-increasing.

**Proposition 3.1.** *Given a ground truth set $G$, the extension of the energy $\theta_G$ of Rel.(3.25) to partial partitions by $\vee$ composition is $h$-increasing.*

*Proof.* Let $\pi(S_1)$ and $\pi'(S_1)$ be two partial partitions of set $S_1$, with

$$\theta_G(\pi(S_1)) = \bigvee\{\theta_G(T_i), T_i \sqsubseteq S\} \leq \theta_G(\pi'(S_1)) = \bigvee\{\theta_G(T'_i), T'_i \sqsubseteq S'_1\} \tag{3.28}$$

Consider a partial partition $\pi(S_2)$, where $S_2 \subseteq S_1^c$. By taking the supremum of each member of inequality (3.28) with $\bigvee\{\theta_G(X_j), X_j \sqsubseteq S_2\}$ one does not change the sense of the inequality, which becomes

$$\theta_G(\pi(S_1) \sqcup \pi(S_2)) \leq \theta_G(\pi'(S_1) \sqcup \pi(S_2)), \tag{3.29}$$

which achieves the proof. □

Note that when $G \cap S = \varnothing$, then $\theta_G(S) = K_{\max}$, which is a large penalty set as a factor of the number of pixels in the input image segmentation.

The energies are demonstrated in figure 3.7. to summarize $\omega_G(S)$ gives the largest radius of dilation of the ground truth set, so as to cover the contours of class $S$, from *any* set of contours of the ground truth, while the energy $\theta_G(S)$ gives the largest radius of dilation of contour of class $S$ so as to cover contours of the ground truth parition covered by the class $S$.

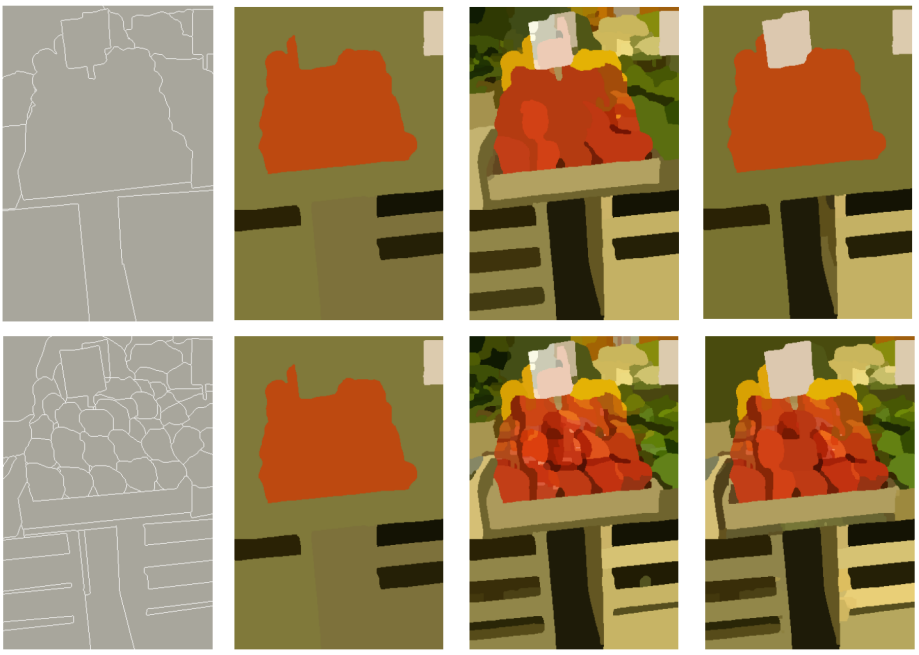



FIGURE 3.9: Ground truth partitions, and corresponding optimal cuts, for energies $\omega_G, \theta_G$ and for composition by sum $\omega_G + \theta_G$. The input hierarchy is the UCM from the Berkeley dataset, consisting of 800 level binary partition tree.

### 3.2.7 Composition of $\omega_G(S)$ and $\theta_G(S)$.

The composition of the energies happens with respect to a single ground truth, or to several ones. In the first case, one can wonder if preferable not to combine $\omega_G$ and $\theta_G$ so that they can provide two separated maps for the precision and for the recall. The two associated overall values may be presented in a 2-D graphic as proposed in [9]. We can also take for final energy either $\max(\omega_G, \theta_G)$, or sum $\omega_G + \theta_G$, they are both $h$-increasing. On the example of the "peppers", and for two different ground truths, one obtains the results depicted in Fig.3.9

### 3.2.8 Composing multiple ground truth sets

In case of multiple ground truths, the usual techniques proposed in literature are additive [9]. Formally speaking, why not? Putting $\omega_G = \sum \omega_{G_i}$ yields an $h$-increasing energy, hence a best cut (which is, of course different from the sum of the best cuts of the various $G_i$). The implicit assumption here is that all ground truths are more or less similar.

But one can also encounter drawings $G_i$ that focus on different regions of the scene. Then if we take the sum, each part of the space risks to be penalized because if is far from

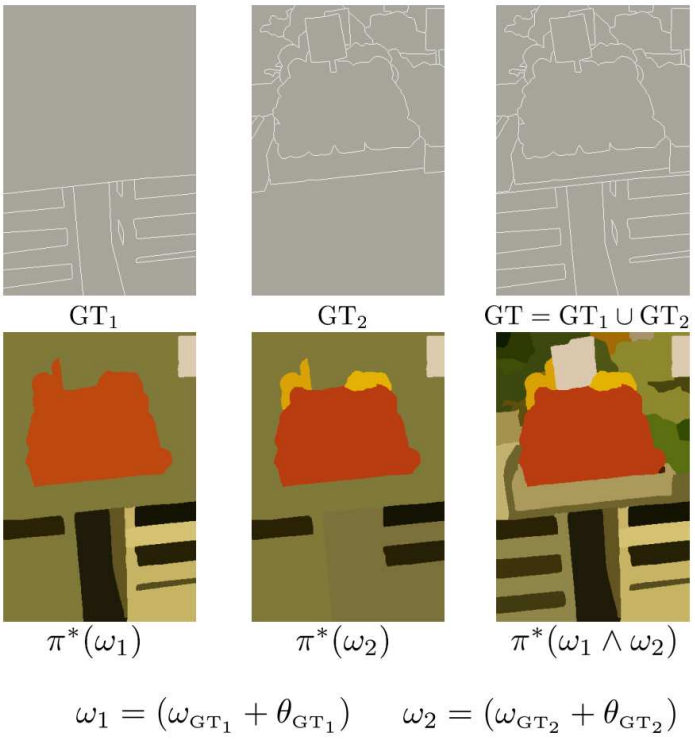



FIGURE 3.10: Two ground truths and their union are shown, with their corresponding optimal cuts, for the energy $\theta_G + \omega_G$. The composition over different ground truth sets is achieved by infimum as shown.

one drawing, at least. This was the case in case of adaptive ground truth compoisition in Peng et al. [92].

For the situation depicted in Fig.3.10, the energies first two best cuts are given by $\sup\{\omega_G, \theta_G\}$ and the third one by taking $\inf\{\sup\{\{\omega_{G_1}, \theta_{G_1}\}, \sup\{\{\omega_{G_2}, \theta_{G_2}\}\}$. When point $x \in E$ is farther from $G_1$ than from $G_2$ then the $G_1$ energy is not taken into account.

### 3.2.9 Number of Classes Constraint

The two energies $\omega_G(S)$ and $\theta_G(S)$ of Rel. (3.25) and (3.22) have been chosen because of their geometrical meanings, but they are far for being the only possible ones. It is iindeed easy to build an energy which fits the features one wants to emphazise. Suppose for example that we decide that the number of classes $n$ of the ground truth is a cruxial feature. Then when applying energy $\omega_G$ we can condition the ascending pass which generates the best cut to stop as soon as the number $n$ of classes is reached. Fig. 3.11 depicts the best cuts w.r.t. $\omega_G$. when the parameter $\lambda$ of Eq.(3.22) equals $0, 10$, and $80$, and when the ground truth is GT7, which has 87 classes. For $\lambda = 0$, we do not obtain

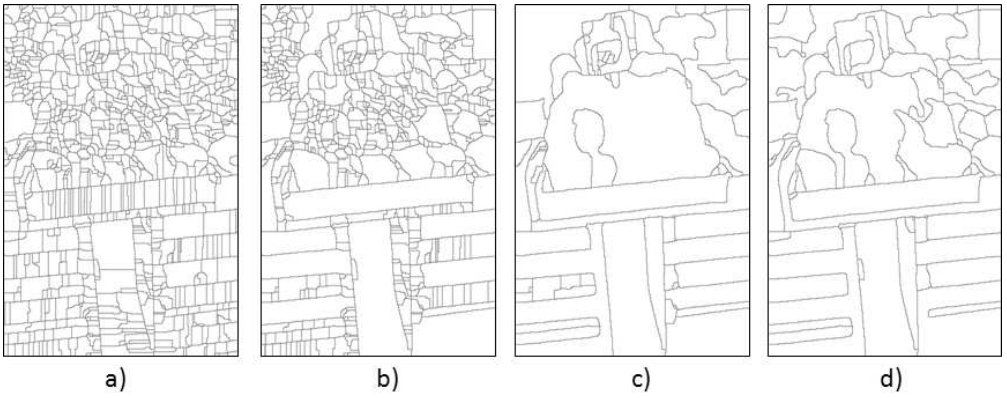



FIGURE 3.11: a) Leaves partition b), c) and d) Conditional $\lambda$-cuts for $\lambda = 0$, 10, & 80.

the leaves partition, because the classes with an equal energy have been clustered, as pointed out previously. In Fig. 3.11c) and d), but not in Fig. 3.11b), one arrives to 87 classes before the end of the climbing algorithm. This explains why the two partitions are not comparable.

### 3.2.10 $h$-increasing Coverage Measures

The measures for evaluating segmentations with ground truths categorize into two types: Region based and Boundary based. The global purpose of these measures, is to be able to evaluate the image segmentation algorithms w.r.t certain metrics. Here we use one of the region based measures to extract an best possible segmentation from a hierarchy of segmentations. One of the common measures is the coverage measure [10]. The coverage criterion is not $h$-increasing due the division by the union of regions. This changes the optimality and we can't ensure a local optimum that is part of the global optima. One can now rewrite it (eqn. (3.12)) in an $h$-increasing energy form as follows:

$$\omega(S) = N - \sum_{j \in [1,n]} |S \cap G_j| + \lambda n \tag{3.30}$$

where $N$ is number of pixels in the image or partition, $G_j$ are the different connected components of the ground truth partition $G$, and $n$ is the total number of connected components in the ground truth partial partition of support $S$.

$$\omega(S) = (N - \sum_{j \in [1,n]} |S \cap G_j|) + \lambda n$$

The sum term in the functional refers to the number of pixels that correspond between the ground truth and the class. This is should be maximum, while the difference $N - \sum_j |S \cap G_j|$ should be minimum.

The second term corresponds to the number of connected components of ground truth in each class $S$. Minimizing this value gives the largest partition that fits $G$.

This equation can be replaced by

$$\omega(S) = N - \sum_{j\in[1,n]} |S \cap G_j| + \lambda \sum_{j\in[1,n]} |\partial G(x)|_{x\in G_j\cap S}$$

### 3.2.11 Local linear dissimilarity

Another variant consists in replacing the supremum that appears in Rel.(3.19) by a $L^p$ sum, which gives less importance to the farthest zones. A similar approach has been successfully used by L. Gorelick et Al. [44] in regional line-search cuts. Among the $L^p$ integrals, the one which weakens the most the weights of the farthest zones is obtained for $p = 1$. Therefore we take for precision energy $\widetilde{\omega}_G(S)$ the integral of distance function $g(x)$ of $G$ along the contour $\partial S$ and for recall energy $\widetilde{\theta}_G(S)$ the integral of the distance function $g(x, \partial S)$ of $S$ on $G \cap S$:

$$\widetilde{\omega}_G(S) = \frac{1}{\partial S}\int_{\partial S} g(x)dx \qquad \widetilde{\theta}_G(S) = \frac{1}{G\cap S}\int_{G\cap S} g(x,\partial S)dx \tag{3.31}$$

The two functionals $\widetilde{\omega}_G$ and $\widetilde{\theta}_G$ are extended from classes to partial partitions by addition, since they both involve integrals, and one easily checks that the two energies are $h$-increasing. The higher $\widetilde{\omega}_G(S)$, (resp.$\widetilde{\theta}_G(S)$), the farther $S$ is from $G$ (resp. $G$ is from $S$). In case of a ground truth given by $k$ drawings, one just sums up the $k$ energies $\widetilde{\omega}_G$ and $\widetilde{\theta}_G$.

### 3.2.12 Global Precision-Recall similarity integrals

Following from the local measures in (3.31) which are integrals of the distance function associated with each class, we propose here a global similarity measures for a hierarchy. Two global measures of precision and recall for a given hierarchy $H$ of segmentations with respect to an input ground truth partition $G$. The measure now is not between 2 partitions any more and deals with the global similarity between hierarchies of partitions $H$ and a single partition $G$. The representative functions we are going to use for the global measures are: $s$ the saliency and $g$ the distance function, the set $S_i$ saliency map

threshold at an index $i$.

$$P = \sum_{i=0}^{1} \frac{i}{N} \frac{\int_{x \in \epsilon(S_i)} (1 - g(x)).S_i(x)\mathrm{d}x}{|S_i|} \quad R = \sum_{i=0}^{1} \frac{i}{N} \frac{\int_{x \in G} (1 - g_{S_i}(x))\mathrm{d}x}{|G|} \tag{3.32}$$

The integral calculates the similarity between partition $S_i$ produced by thresholding the saliency $s$ at $i$ and the ground truth partition $G$ by integrating the inverse distance function $1 - g$ under the binary function $S_i$. Also the sense of the hierarchy is such that $s_{i+1} \subset s_i$ which represents that partition at a higher level in the hierarchy has fewer contours than the one below to respect the inclusion order. Each integral is weighted by the relative rank of the partition within the hierarchy $H$. This is done by weight it by ratio of threshold index $i$ and the total number of levels in the hierarchy $N$ as shown in equation(3.32).

Similarly a global precision value for the contours of the partitions in the hierarchy can be calculated by integrating the distance functions $g_{S_i}$ of partition $S_i$ under the ground truth partition $G$. These integrals are normalized with respect to each image support by dividing by the size of the image.

### 3.2.13 Proximity between hierarchies

The integrals in equation (3.32) is between a partition $G$ (ground truth) and a hierarchy $H$. The same can be extended to measure the proximities between two hierarchies of partitions. Given two hierarchies of partitions, $H_1, H_2$, with N and M number of levels, and partitions indexed by $i$ and $j$ respectively,

$$\phi_{12} = \sum_{j \in [1,M]} \sum_{i \in [1,N]} \frac{\int_{x \in \epsilon(\pi_i)} (1 - g_{\pi_j}(x)).\pi_i(x)\mathrm{d}x}{|\pi_i|} \tag{3.33a}$$

$$\phi_{21} = \sum_{i \in [1,N]} \sum_{j \in [1,M]} \frac{\int_{x \in \epsilon(\pi_j)} (1 - g_{\pi_i}(x)).\pi_j(x)\mathrm{d}x}{|\pi_j|} \tag{3.33b}$$

where $g_{\pi_i}$ is the distance function of the partition $\pi_i$.

The measure lacks the similarity measures across partitions which are not horizontal cuts, but generally cuts from the two hierarchies. This becomes again a combinatorial problem. The refinement of cuts $\pi_i$ from an input hierarchy $H_1$ would have a value of the distance function $g_{\pi_j}$ which decreases on average till the point where the two partitions nearly fit and the integral starts increasing again.

### 3.2.14 Ground truth energies to Saliency functions

Here we shortly describes the motivation to leave the energetic lattice and moves to a lattice defined purely on the finest partition of the space. We have seen in the formulation of the local Hausdorff energies that they consisted in measuring the proximity between the contours of the class of the hierarchy and the ground truth partition.

This requirement of local energies, when changed to an global assignment of a proximity measure over all the contours of the segmentations in HOP, results in a way to transform a HOP, by reordering its contours based on its proximity to a ground truth. This was the starting point for the chapter 4 on saliency functions. Furthermore one can examine such a proximity measure in the two Hausdorff senses, that is, distance from partition in HOP to ground truth, and vice versa. Interestingly here, we deviated from the classical saliency function, [32, 88], which consists in weighting of contours of the minima by their extinction function [116], in other words the result of defining a flooding.

We will see later that a proximity measure, along with a leaves partition is sufficient to define a lattice, and thus produce a saliency function. For a deeper understand we redirect the reader to the chapter 4.

## 3.3 Summary

In summary, one must note that we are demonstrating a framework to perform constrained optimization on hierarchies and braids, using various energies. The image segmentation problem by itself requires another step, which is to determine the optimal scale parameter. This can be done in a variety of methods.

**Chapter contribution summary**

- Demonstration of various $h$-increasing energies to minimize Mumford-Shah functional, texture energies, which enforce different constraints.
- Half Hausdorff energies to calculate proximal cut from hierarchy w.r.t Ground truth, and formulation of the proximity in a constrained optimization framework.
- A global measure which distinguishes between a set of hierarchies of segmentations given a set of GT partitions.

# Chapter 4

# Hierarchies and Saliency function

**Publications Associated with Chapter**

- [58] Fusions of ground truths and of hierarchies of segmentations (PRL 2014).
- [60] Scale space operators on hierarchies of segmentations (SSVM 2013).

In the previous chapters, the cuts from a BOP $B$ have been our solution space, and various types of increasing energies were optimized to produce an optimal cuts, and further an optimal hierarchy whose cuts are ordered based on increasing energies. We leave the space braids of partition, and concentrate now on a particular subfamily of the hierarchies, which is the represented by *saliency functions*. Here one can renounce the lattice over the support of classes and purely work on the contours of partitions in the hierarchy. Following the decision to work on contours of the partitions, we will restrict ourselves to hierarchies represented by a saliency functions in $\mathbb{R}^2$.

A saliency function is a numerical representation of a hierarchy, which was first introduced by [88] to represent the scale of disappearance of watershed contours. More generally saliency function is a positive function defined on the frontiers of the classes of a partition $\pi_0$.The function value on a contour represents the altitude at up to which it separates two components, beyond this value, the components fuse into a single component. Thus the saliency characterizes a hierarchy which has the partition $\pi_0$ for leaves. In the Euclidean plane a simple method to model $\pi_0$ is to consider its frontiers as Jordan curves.

With this numerical representation of the hierarchy, we formalize the work in [60], by introducing an order on the contours of the partitions thus avoiding the combinatorial problem of choosing an optimal closest cut. We define the partition of input space to be a finite set of Jordan curves partitioning $\mathbb{R}^2$. Further define a lattice of such Jordan

curves which when associated with a numerical function produces a new ordering and thus a hierarchy. We treat the problem of ground truth evaluation of hierarchies of segmentations. More precisely we use the ground truth distance function to order contours in the hierarchy reflecting the proximity to ground truth. We further demonstrate with other point-wise function to reorder contours, resulting in a new transformed saliency function.

First in section 4.3 we describe the necessary topological framework, and explicit the lattice of Jordan curves. Section 4.3 gives the theory, Section 4.5 and 4.6 describes a fusion between the saliency and external functions, which reorders the initial hierarchy. Section 4.7 demonstrates such a reordering based on the proximity to a ground truth partition. Then the composition of two hierarchies is obtained by using a distance between them. Following which one studies a measure that demonstrates how the distance function produces structural changes in the transformed or reordered hierarchy.

**Saliency function: a numerical representation of partition hierarchy**: We revisit the saliency function, first introduced as a representation of the hierarchy. They have been popularized as Ultrametric Contour Map(UCM) [10] The hierarchy and thus its saliency function, can be generated in many ways, most frequently by watershed of floodings [10, 32, 79], or again they can be regular hierarchies such as, quad-trees, oct-trees, all of which can be represented by the saliency function.

## 4.1 Ground truth Evaluation of Segmentation Hierarchies

In section 3.2.1 we have seen hierarchies of segmentations have been evaluated by expert annotated ground truths by various local, regional and global measures, over classes regions and partitions [10, 94]. This was formulated as an energy using the Hausdorff distance between contours of ground truth and classes in the hierarchy [61], to extracts partitions from a given hierarchy, closest to the ground truth. Until here we have used the energetic lattice from the lattice based optimization framework [59] in chapter 1.

### 4.1.1 Contour proximity

There has been substantial work on comparing the contours of a ground truth set and image segmentation, for the purpose of evaluating the distance between them. One can find the use of mainly three types measures between the classes of ground truth and segmentation: Region measures (Region Intersection Measure/Coverage measure) [84], Boundary-based Measures [55] which use the Hausdorff distance between the contours.

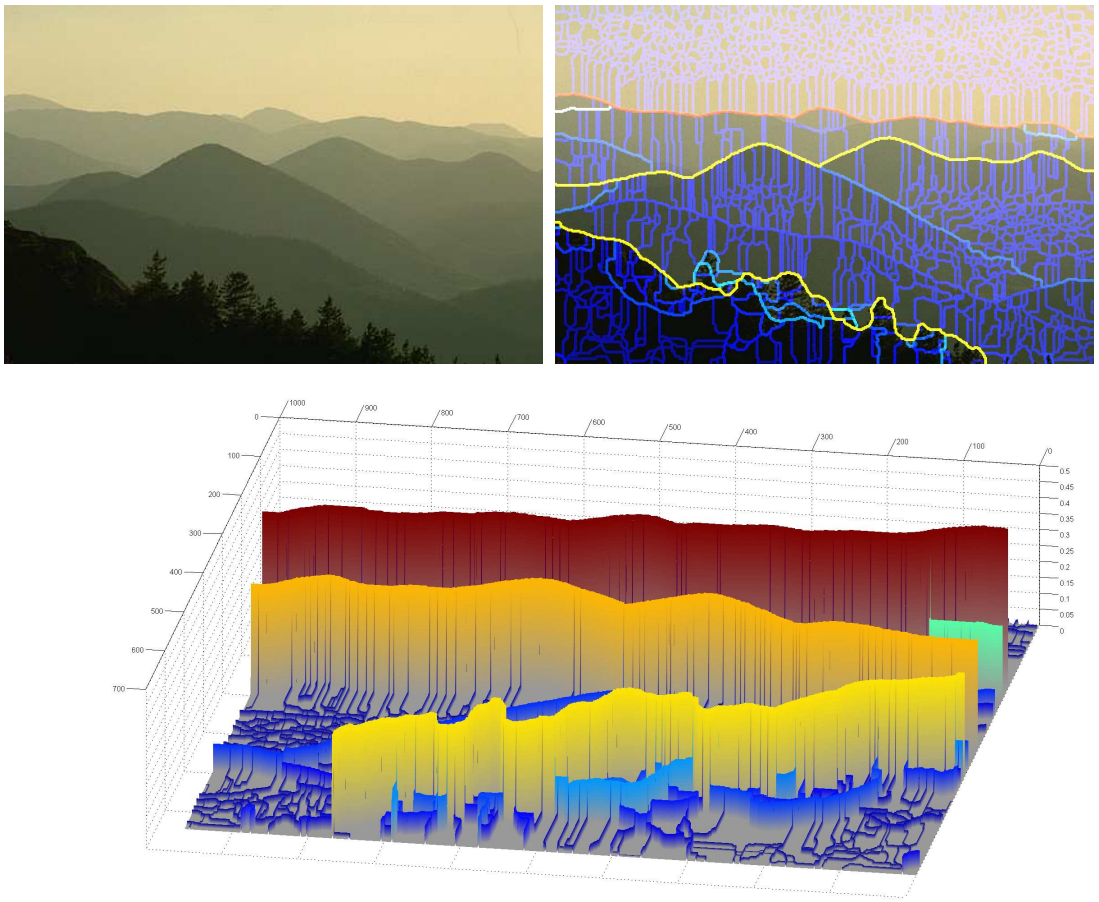
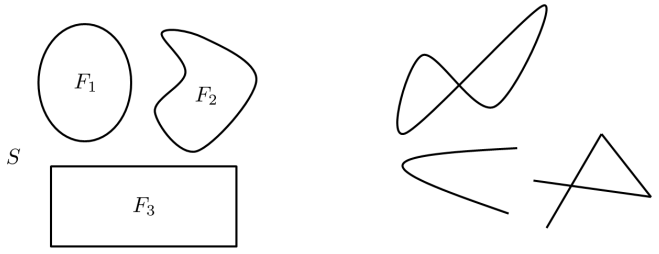


FIGURE 4.1: Saliency Function example: First image is an input image from Berkeley data base. Second image is the corresponding Ultrametric Contour Map(UCM) which represents a hierarchy of partitions. Third image shows the corresponding altitude map for the UCM. Here we see that each arc in the saliency function separates different components at different altitudes. This third image is oriented 180 degrees off to better render the details.

There are also certain mixed regional measures, which weight the false positives and negatives differently. The Hausdorff distance calculates the supremum of distance distribution between two classes, thus for contours of any pair of classes, having the same worst case distance, will be evaluated irrespective of other distances.

In such a context we will consider only contour proximity information. This can be sufficient to extract a set of partitions from a given hierarchy of segmentations. No explicit measure or energy on the classes is required to do so. In this line of work [60] studies how the distance function of the ground truth can be used to order contours of the hierarchy of partitions, so as to be able to pick the partition closest to the ground truth, and the next closest until one completely transforms the hierarchy.




Jordan Curves Not Jordan Curves

FIGURE 4.2: Examples of Jordan Curves. The Jordan curve tessellates the space, with one bounded interior component or face $F$ and an unbounded exterior component $S$

## 4.2 Jordan Curves

We can adopt either the framework of the Euclidean plane $\mathbb{R}^2$, or that of an abstract graph, or that again of a planar graph. We will work on the Euclidean plane in this paper, and further discuss its significance after establishing the main results. The critical feature we need to model is, the distinction between inside and outside of a contour. This is exactly what is achieved by Jordan curves defined in the Euclidean plane $\mathbb{R}^2$ equipped with the topology of Euclidean distance.

Here we present the assumptions of our model:

1. The working space is Euclidean plane $\mathbb{R}^2$,
2. $\mathbb{R}^2$ is partitioned into faces and contours by Jordan curves,
3. There exists a finest partition with a finite number of faces, called leaves.

This last axiom permits to construct a lattice structure and thus hierarchies.

We now define what a Jordan Curve and how it has been used to describe image segmentation or partitions.

**Theorem 4.1.** *A Jordan curve $C$ in $\mathbb{R}^2$ is the image of an injective (i.e. without self-intersection points) continuous map of the unit circle into the plane. According to a famous theorem due to C. Jordan, the complement $\mathbb{R}^2 \setminus C$ consists of exactly two open connected components, the first one, called face $F$ is bounded, and the other, called background $S$, which is not. They are homeomorphic to the inside and the outside of a disk respectively.*

One classically calls a *tessellation* any partitioning of a topological space into open classes, and classes formed by their frontiers [101]. The partition of $\mathbb{R}^2$ into three classes $\{C, F, S\}$ by a Jordan curve is thus a tessellation of $\mathbb{R}^2$.

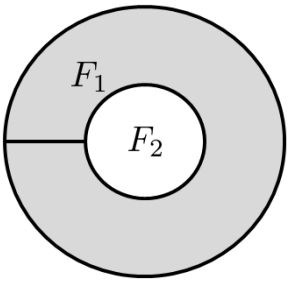



FIGURE 4.3: One can suppress the contour in the interior of region $F_1$, while no increasing the energy $\omega(\pi)$ of the partition by $F_1, F_2$.

### 4.2.1 Normal Segmentations

Here we must interject that the class of hierarchies of segmentations produced here are a subclass of the general hierarchy of partitions with no topological constraint. The constraint requires that two classes in the partition can't be merged if they are not connected an arc is removed from the partition. Thus hierarchies of segmentations which label the classes are a more general class of hierarchies. Here we will use the definition of a Normal Segmentation introduced by Morel et al. in [82] to formalize this:

**Definition 4.2.** *A segmentation $\pi$ is called normal if every sub-segmentation $\pi'$ of $\pi$ verifies $\omega(\pi') > \omega(\pi)$*

Here the *normality* of a segmentation refers to the simplification of a segmentation such that:

- the merging of any two classes necessarily produces a suppression of atleast one Jordan Curve, thus causing a reduction in the energy associated with the partition (in the Mumford-Shah sense)
- the suppression of a Jordan curve produces a merging of 2 classes.

Morel et al. [64] further states to check computationally if the segmentation produced is normal, aside the necessity for finite number of classes, one requires the classes have no internal boundaries. That is if each Jordan Curve separates two different regions. This property ensures finite classes remain at the end of the operation. This is termed by Morel et al as an 1-normal segmentation. An example in figure 4.3 from Morel et al [82] is reproduced.

### 4.2.2 Describing Segmentation with Jordan Curves

The image segmentation process results in a segmentation which is formally defined as a partition of the input space. Many authors have used the concept of Jordan Curves to represent the segmentation of images. To represent the Ultrametric Contour Maps (UCM) [8], Arbelaez refers to a segmentation $K$ as a finite set of rectifiable Jordan curves, which are called the contours of $K$. A finite set of rectifiable Jordan Curves is said form a *normal segmentation* when the removal of any number of curves increases the energy defined the segmentation. One similarly finds the definition of $1-Normal$, $2-Normal$ segmentations by Morel et al. [64, 82]. Furthermore Jordan curves can be used to describe the contours of components(level sets) of a continuous functions [81].

Jordan curves may be extremely irregular (e.g. the fractal Von Koch snowflake), and even have of non measurable lengths. Measurable Jordan curves have been used in image processing by [81, 82] for functionals whose computations involve length and perimeter as seen in the case of Mumford Shah Functional. But this restriction is not pertinent here, since we do not measure lengths.

## 4.3 Jordan Nets

Here we construct a lattice structure that uses a finite set of Jordan curves to create a partition of $\mathbb{R}^2$.

### 4.3.1 Definitions

**Definition 4.3.** *A Jordan net, or J-net $N$, is defined as a set of Jordan curves, which delineate a finite number of open insides. In addition, the empty set $\varnothing$ is also,by definition, a J-net.*

$N$ is thus a set of contours $C_i$ that delimits the bounded faces $F_1$, $F_2, ..., F_p$, $p < \infty$, plus the unbounded background $S$. Both faces and background are open. J-nets may comprise several connected components and faces included in each other. Note that the number of primitive Jordan curves one can extract from $N$ does not reduce to the $C_i$. In Figure 4.4 for example, one can take as primitive J-net curves, the two half circles which share a diameter, while one can also take, just one of them plus the complete disc. The faces we consider are the complement of the Jordan net $N$. One can observe of course that the definition of the Jordan net basically resembles that of a segmentation [82]. According to Jordan theorem, the plane $\mathbb{R}^2$ is partitioned by the union of $N, S$, and $\{F_i\}$.

### 4.3.2 Ordering and lattice of J-nets

Though the space of all J-nets of $\mathbb{R}^2$ is ordered by inclusion, it does not generate a lattice because it is not upper-bounded, and anyway it is broad for our goal: in a circular crown of radius $r > 0$, one can draw an uncountable number of Jordan curves. We will restrict this space by considering only the J-nets included in a finite net $N_0$, whose associated faces define *the leaves*. An example is given by the faces of the tessellation in figure 4.5.

Two families derive from $N_0$: Firstly, the power set $\mathcal{P}(N_0)$, constituted by all sets whose points belong to $N_0$. Secondly, the family $\mathcal{N}(N_0)$ of all J-nets included in $N_0$. We have $\mathcal{N}(N_0) \subseteq \mathcal{P}(N_0)$. Both sets $\mathcal{P}(N_0)$ and $\mathcal{N}(N_0)$ comprise the empty set $\varnothing$ and are ordered by inclusion. $\mathcal{P}(N_0)$, as a power set, is a Boolean lattice. But unlike $\mathcal{P}(N_0)$, the family $\mathcal{N}(N_0)$ is not complemented: if $N \in \mathcal{N}(N_0)$ the complement $N_0 \backslash N$ may have Jordan arcs which are not looped. However, the following property holds:

**Proposition 4.4.** *The set $\mathcal{N}(N_0)$ of all J-nets included in the base $N_0$ forms a lattice with $N_0$ and $\varnothing$ as universal bounds. The supremum of $N, N' \in \mathcal{N}(N_0)$ is the union $N \cup N'$, and the infimum is the union of all Jordan curves common to $N$ and $N'$ (empty set included).*

*Proof.* $\mathcal{N}(N_0)$ admits a greatest element, namely $N_0$. Let $N, N' \in \mathcal{N}(N_0)$. The union $N \cup N'$, composed of Jordan curves belonging to either $N$ or $N'$, is therefore an element of $\mathcal{N}(N_0)$. Concerning the infimum, the largest lower bound of $N$ and $N'$ is obtained by the union of all Jordan curves common to $N$ and $N'$. This family exists, i.e. is not empty, since it contains the empty set. □

This approach focuses on contours, but the duality in $\mathbb{R}^2$ provides the inverse ordering for the faces. If $N \subseteq N'$, then

$$\cup_i \{F_i\} \cup S = \mathbb{R}^2 \backslash N \supseteq \mathbb{R}^2 \backslash N' = \cup_i \{F'_i\} \cup S'$$

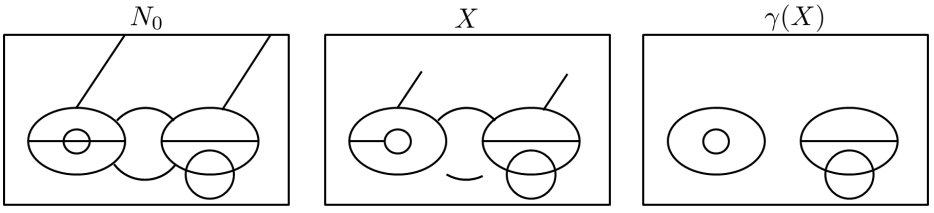



FIGURE 4.4: An elementary Jordan net $N_0$, A set $X$ of arcs and Jordan curves, and their net openings $\gamma(X)$. It is important to note that the number of connected components of white pixels don't change after a net opening $|\gamma(X)|_{CC} = |X|_{CC}$. $\gamma(X)$ removes two types of arcs: open arcs and arcs which are not *normal* [64].

FIGURE 4.5: Initial Image 25098 from Berkeley database, leaves given by lowest(finest partition) threshold of Ultrametric contour map(UCM). The leaves here represents the initial finite net $N_0$.

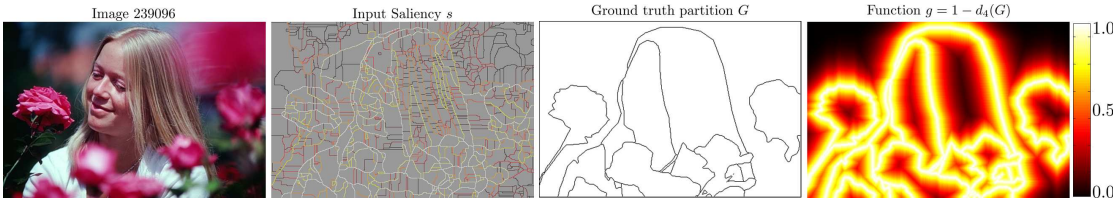



FIGURE 4.6: Initial Image 239096 from Berkeley database, Saliency function $s$: Ultrametric contour map(UCM), Ground truth partition $G$, Inverted distance function $g$. The inputs here we consider are the saliency function $s$ and the inverted distance function $g$.

The faces increase when contours are removed. Proposition 4.4 suggests to associate an opening with the $N_0$-infimum.

**Corollary 4.5.** *Given $X \in \mathcal{P}(N_0)$ the union $\gamma(X)$ of Jordan curves $C$ contained in $X$*

$$\gamma(X) = \cup\{C \subseteq X, C \in \mathcal{N}(N_0)\} \tag{4.1}$$

*is an opening on $\mathcal{P}(N_0)$, called net opening, whose set of invariants is $\mathcal{N}(N_0)$. $\gamma(X)$ provides the largest J-net included in $X$. By duality in $\mathbb{R}^2$, if $Y = \mathbb{R}^2 \setminus X$, the closing $\varphi(Y) = \mathbb{R}^2 \setminus \gamma(X)$ provides the largest classes having $\gamma(X)$ as J-net contours.*

In particular the $N_0$-infimum between $N, N' \in \mathcal{N}(N_0)$ is $\gamma(N \cap N')$. The opening $\gamma(X)$ simplifies $X$ by suppressing all points which are not involved in a Jordan contour included in $X$ [60]. Figure 4.4 illustrates these changes.

### 4.3.3 Net Opening in Literature

The net opening operator $\gamma$ is not a new notion. It appears already in 1982 in [101] under the name of pruning, where it serves indeed to cut branches in digital skeletons. One can also find it in [68] for discrete classifications by ultrametric. A similar line of thoughts is developed for characterizing the types of edges in topological watersheds [16] by Cousty et al. [31], which produces watershed cuts, that partition the vertex set. In [33] a local variant is proposed to model flooding in digital watersheds. In the same context, it reappears in [32], where each pixel is assigned as nodes set, and the edges is obtained as a function of the gradient between two pixels. The dynamic, surface, volume based saliencies are calculated using such edge weight values. The Minimum spanning tree is applied to the edge set to obtain the hierarchies of segmentations on the vertex set.

Further in the framework of graphs, Haximusa et al. [52] perform hierarchical image segmentation by obtaining the infimum of costs over edges separating two components with the costs defined as the largest internal contrast in the fused component. This operator is posed as edge contraction on the Region adjacency graph(RAG), which removes an edge from a graph while merging the two vertices it previously connected. There has also been work in the domain of edge and vertex based labeling studied in the well known Multi-cut problem [56]. The multi-cut is a NP hard combinatorial problem that determines the edges such that the sum of weights of cut-set is minimal. The paper [56] then proposes to optimize over the set of all separating boundaries while minimizing particular energy functions. Applications here are seen in the domain of closed boundary segmentation problems [5]. While the multi-cuts framework is flexible we are looking here for a simpler framework utilizing lattice (sup-inf) based optimization to extract closed contours from a predetermined hierarchy of segmentations.

**Formulating $\gamma$ as an Opening**: By working on Jordan Curves, we can interpret this as an indifference in operating on edges and vertices on graphs, is one of the key motivations which these cited approaches miss. Furthermore the by defining the underlying lattice $\mathcal{N}(N_0)$, whose net opening operator $\gamma$ expresses the infimum, one reaches a unique infimum by nature of the lattice, and is also reason why the net-opening $\gamma$ works so well. Consequently this facilitates the ability to state theorems 4.6 and 4.9, and more importantly to compose several inputs, to introduce the filter $\gamma\varphi$, to provide closest bounds, etc. The opening $\gamma$ results in the largest saliency function under $g$, which corresponds to an optimum. This is in contrast to the multi-cut criterion, where one minimizes the sum of the costs of the edge cutset.

## 4.4 Watershed transformation and Saliency function

This section which is mainly bibliographic revisits the saliency function from its development in the transformation.

### 4.4.1 Watershed

The watershed transformation has been a subject of intense study in literature. The transformation was introduced in the continuous domain by Beucher and Lantuejoul [19], followed by others, including Najman et al. [87], and also in the discrete domain by Vicent et al. [16, 30, 122]. In the continuous domain there have been problems in describing the watershed of a continuous function, since the watershed line produced may be thick, with nonzero area. It may also have so-called *barbs* which are branches of zero area with an end point [87]. There has been an energy minimization, Water-snakes, based calculation of the continuous watershed for functions that are derivable [89].

The watershed transform in its approach, treats the input functions as a relief. The function used to determine the presence of edges, usually considered is the morphological gradient. When the function is seen as a mountain landscape, and one floods topographically the valleys, region boundaries are determined as watershed lines [77]. This analogy is also well described by the drop of water principle: a drop of water falling on a topographic surface follows a descending path and eventually reaches a minimum. The watershed may be thought of as the separating lines of the domain of attraction of drops of water [31]. The above discussion demonstrates that the watershed transformation has many ways of being defined. For example, based on the catchement basins of its minima [77], the watershed contour separating components [31]. In addition, one finds continuous or discrete definitions. There are also many ways of handling cases when one finds a drop on a saddle point (plateaus).

In such cases one requires a distance function, which has been addressed in many ways, in literature, but can be categorized into the following:

1. General Non-Smooth Geodesic Distance: [77], Surfacial mean curvature based watersheds [96].
2. Smooth Geodesic Distance: [89], which uses a geodesic distance weighted by the norm of the gradient of the function to formulate the watershed calculation as a minmization.

3. Morphological gradient weighted distance functions: The waterfall uses a simplified gradient resulting from the iterative mosaicking of the image. [18, 79], A topological watershed of a function, preserves the gradient between its regional minima. The gradient across the regional minima is given by the minimal altitude at which the minima basins recombine [16, 31]. Operations of extending the basins/minima by "raisings" has been suggested in [33] using again the gradient function. One can also find work on using the area, volume and dynamic of connected components of level sets of the function [32].

4. Regularized gradient function: [116] uses flooding by viscous liquid to simulate the regularization of the gradient to avoid over-segmentation problems. [120] uses a diffusion scheme on the RGB gradient vector field, to simplify the gradient.

In all these cases its important to note that one assumes a finite set of local minima of the input function.

#### 4.4.1.1 Saliency functions

Saliency functions have been firstly used to represent the hierarchy of catchment basins, produced when one employs hierarchical watershed methods. This was first introduced in [88] where Najman et al. used the geodesic distance to produce sequence pairwise floodings that recombined minima based on the geodesic distance between pair of points belonging to these minima. The minima of the function are flooded as a function of their extinction values [117], which orders the attributes [32] of the components of the levels sets of a function, for example the area, volume and dynamic, to produce a hierarchy of partitions, and the corresponding saliency function.

Further Najman et al. [86] suggests an equivalence between the set of all saliencies possible and an equivalent ultrametric watershed. It is of interest to note that the watershed depends on the gradient function which is calculated over the original image.

The Watershed clearly combines the two processes: calculating a dissimilarity function (usually a gradient), and calculating the partition separating the minima of the function using its gradient. We distinguish the two processes and define in our axioms:

- A finite set of Jordan nets provides model of segmentation [82]. This states that our starting point is the partition itself. The watershed is a more complicated problem, which involves various questions of thick contours and barbs, that are avoided by having a Jordan curve. The goal in using the Jordan nets is simply to define a hierarchy of partitions bounded under a numerical function.

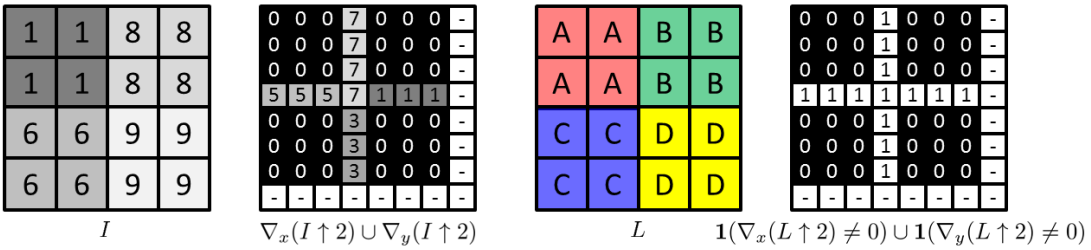



FIGURE 4.7: The figure demonstrates how the image $I$ produces a meaningful gradient function in the xy directions $\nabla_x I \cup \nabla_x I$, using the image values. The label image $L$ on the other hand does not depend on the actual values in the gradient ($\nabla_x L$ depends on the labeling $L$ which can be arbitary) but just the existence of the non-zero gradient value. This can be seen as a indicator function of the gradient function $\mathbf{1}(\nabla L)$.

- In our study, the numerical function that decides the recombination/flooding of the components of the partition is no more dependent on the gradient of a function. This broadens the choice of the gradient based ordering required to generate watershed based saliency functions. It can be any function. In our case, it corresponds to the proximity of segmentation contours to a ground truth set. This is demonstrated in figure 4.7.

A hierarchical partition of the space, one can be determined in multiple ways based on the problem at hand (region merging, graph based hierarchies, waterfall strategies).

The critical difference in the formulation of a saliency function as numerical function on the Jordan net, is the use of a function which is *independent* of the function being partitioned (like Luminance of the image). In this case it depends on an external constraint: the proximity between the contours of a ground truth set and a partition. This is disjunct from the properties/attributes of the components or classes in the hierarchy of partitions.

Finally, the two methods watersheds, net-opening, are close in the type of operations they perform on arcs, but are operations really for two different cases. Watershed are more involved operations that involve gradient calculation which result in a partition of the space. While the net-opening simply tries to define a family of saliency functions that can be defined using a primitive partition under a given function.

## 4.5 Fusion of hierarchies and functions

We now get to the heart of the matter. We would like to distribute weights $g : \mathbb{R}^2 \to \overline{\mathbb{R}}$ on the points of the basic J-net $N_0$ so that one could obtain larger and larger tessellations

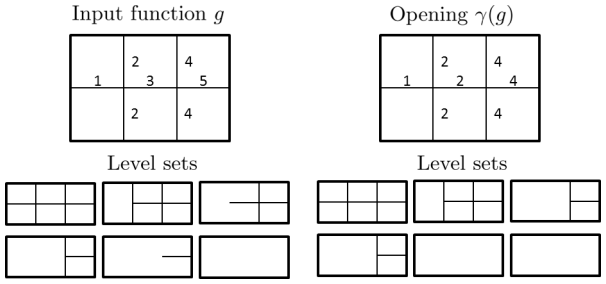



FIGURE 4.8: Input function $g$ on a simple toy Jordan net, The net opening $\gamma(g)$ and their level sets.

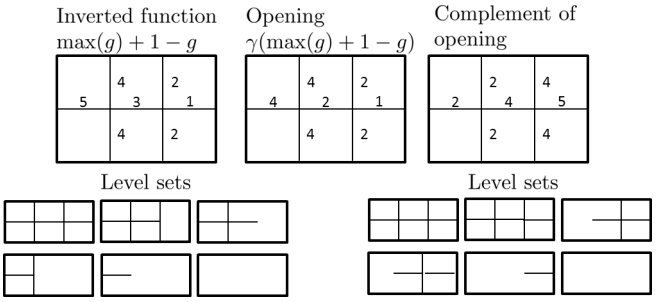



FIGURE 4.9: Complement of input function $g' = \max(g) + 1 - g$, Intermediate result showing the opening of inverted function $\gamma(g')$, and finally for the pseudo closing we use another iteration of the net opening: $\phi(g) = \gamma(\max(\gamma(g') + 1 - \gamma(g')$, with level sets of $g'$ and $\phi(g)$. The opening of the complement is a closing in the space of arcs but not that of the Jordan nets.

as the threshold increases. The cross section $X_t(g)$ at level $t$ is given by

$$X_t(g) = \{x \mid x \in N_0, g(x) \geq t\} \tag{4.2}$$

has no special reason to be a J-net. Breaches can appear in some curves of $N_0$. But $\gamma(X_t)$ has no breach and is the largest J-net smaller than $X_t$. Moreover, as $\gamma$ is increasing, the successive J-nets transforms are nested:

$$t \geq t' \Rightarrow X_t \subseteq X'_t \Rightarrow \gamma(X_t) \subseteq \gamma(X'_t)$$

This orients us towards the unique numerical version of the binary opening $\gamma$ acting on level sets $X_t$, denoted by the same symbol $\gamma$ ($\gamma(X)$ for sets and $\gamma(g)$ for functions). It is given by

$$\gamma(g)(x) = \bigvee \{t > 0 \mid x \in \gamma[X_t(g)]\}$$

or, equivalently by the level sets

$$X(\gamma(g), t) = \bigcap_{s<t} \gamma[X_s(g)] \tag{4.3}$$

The successive thresholds of $\gamma(g)$ delineate increasing tessellations of $\mathbb{R}^2$, i.e. by definition, a hierarchy denoted by $H[\gamma(g)]$. Though $t$ is defined on $\mathbb{R}^+$, the number of level sets of $\gamma(g)$ which are different is finite, just as the number of Jordan curves in $N_0$ is.

The Jordan net opening $\gamma(g)$ generates a saliency on $N_0$, in L. Najman and M. Schmitt's sense [88], i.e. a numerical function whose thresholds are always closed contours. More precisely we can state the following:

**Theorem 4.6.** *Let $\mathcal{F}$ be the family of the positive bounded numerical functions on $\mathbb{R}^2$, let $N_0$ be a basic J-net, and let $\gamma$ be the associated numerical net opening . Given $g \in \mathcal{F}$, $\gamma(g)$ is piecewise constant on $N_0$ and provides the greatest saliency smaller than $g$ on $N_0$. The pair $[N_0, \gamma(g)]$ characterizes a finite hierarchy $H[N_0, \gamma(g)]$. The image $\mathcal{I} = \gamma(\mathcal{F})$ of $\mathcal{F}$ under the opening $\gamma$ is exactly the family of all positive and bounded saliencies on $N_0$.*

*Proof.* Consider a Jordan curve $C_i \subseteq N_0$. Denote by $t_i$ the supremum of the $t$ such that the contour $C_i$ is not damaged, i.e.

$$t_i = \sup\{t \mid C_i = C_i \cap X_t(g)\} \tag{4.4}$$

For $t < t_i$ the opening $\gamma[X_t(g)]$ preserves integrally the internal face $F_i$ of $C_i$, which is thus also preserved for $t = t_i$, according equation (4.3). For $t > t_i$ a breach is made in $C_i \cap X_t(g)$ towards some face $F_j$ adjacent to $F_i$ and $\gamma[X_t(g)]$ makes both faces merge in a larger one, whose contour is included in $N_0$. The set difference $N_0(t_i) = C_i \setminus \cup_{t>t_i} \gamma[X_t(g)]$indicates the points of $C_i$ that vanish above level $t$, or equivalently, the points of $N_0$ where $\gamma(g) = t_i$. The net opening $\gamma(g)$ is thus piecewise constant at values $t_i$ whose cardinal is finite, and its thresholds are closed contours. It is thus a saliency, and the two data of $N_0$ and $\gamma(g)$ characterize the hierarchy $H[N_0, \gamma(g)]$. As each saliency $\gamma(g) \in \mathcal{F}$, the image $\mathcal{I} = \gamma(\mathcal{F})$ provides all saliencies. Besides, the transform $\gamma(g)$ is the greatest saliency smaller than $g$ since, for any saliency $s_1 \leq g \in \mathcal{F}$, we have by increasingness of $\gamma$, that $s_1 = \gamma(s_1) \leq \gamma(g) \leq g$. □

**Corollary 4.7.** *The family $\mathcal{S}$ of all saliency functions $s$ on $N_0$ is a lattice for the pointwise numerical ordering. The supremum of a finite family $\{s_i, 1 \leq i \leq I\}$ is $\vee s_i$, and the infimum is $\gamma(\wedge s_i)$.*

A saliency function $s$ on Jordan net $N_0$ is equivalent to a parametrized hierarchy $H(N_0, s)$ since the threshold

$$N_0(t) = \{C, C \subseteq N_0, s(C) \geq t\}$$

generates a Jordan net at level $t$ and since

$$t' \geq t \Rightarrow N_0(t') \subseteq N_0(t)$$

Moreover this hierarchy $H(N_0, s)$ is parametrized in the sense that the sequence $\{N_0(t), t \geq 0\}$, of successive contours is not only ordered, but labeled by the level $t$ of appearance of each $N_0(t)$. Conversely it is clear that the successive thresholds of a parametrized hierarchy $H(N_0, s)$ induce a saliency function on $N_0$ by Eq(4.4).

The equivalence between saliency functions and parametrized hierarchies leads to the following result:

**Corollary 4.8.** *The lattice $\mathcal{S}$ of the saliency functions provides a lattice structure on the set $\mathcal{H}$ of all parametrized hierarchies on $N_0$, where for $H(s_1), H(s_2) \in \mathcal{H}$ :*

$$s_1 \leq s_2 \Leftrightarrow H(s_1) \leq H(s_2);$$

$$H(s_1 \vee s_2) = H(s_1) \vee H(s_2);$$

$$H(\gamma(s_1 \wedge s_2)) = H(s_1) \wedge H(s_2)$$

## 4.6 Composing hierarchies and numerical functions

The lattice structure of $\mathcal{N}(N_0)$ allows us to combine various bounds by suprema and infima. We can ask the following questions on the compositions of a saliency and any general family of functions, and in what type of problems would different compositions make sense.

1. Given that the base Jordan net $N_0$ is already the support of a predetermined finite hierarchy $H$ with saliency $s$. When a non-negative function $g_1$ on $\mathbb{R}^2$ is introduced, how to compose it with $s$?
2. When in turn a second function, $g_2$, acts on the saliency $s_1$ resulting of $g_1$, how the two effects are composed?

We will successively take up these questions by openings, closings and thickenings.

### 4.6.1 Lower bounds by opening

Theorem 4.6 allows us to combine a saliency $s$ with a function $g$. The following nice properties are indeed direct consequences of this theorem:

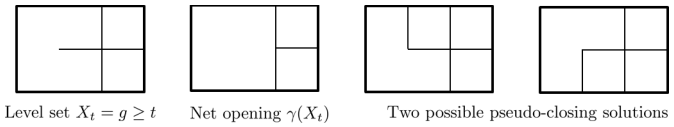



FIGURE 4.10: Lack of upper bounds by closing.

**Theorem 4.9.** *Let $H$ be a finite hierarchy on $N_0$, of saliency $s$, and let $g_1$ and $g_2$ be two non negative functions on $\mathbb{R}^2$, then:*

*(i)* $\gamma(s \vee g_1)$ *(resp.* $\gamma(s \wedge g_1)$*) is the largest saliency smaller than* $s \vee g_1$ *(resp.* $s \wedge g_1$*),*

*(ii) if* $g_1$ *is itself a saliency on* $N_0$*, then the supremum* $g_1 \vee s = \gamma(g_1 \vee s)$ *is a saliency,*

*(iii)* $\gamma(g_1) \vee \gamma(g_2)$ *is the smallest saliency larger than or equal to* $\gamma(g_1)$ *and* $\gamma(g_2)$*,*

*(iv) if* $g_1 \circledast g_2$ *denotes an operation from* $\mathcal{F} \times \mathcal{F} \to \mathcal{F}$*, such as* $+, -, \times, \div, \vee$*,or* $\wedge$*, then* $\gamma(g_1 \circledast g_2)$ *is the largest saliency smaller than* $g_1 \circledast g_2$*, and* $\gamma(g_1 \vee g_2) \leq \gamma(g_1 + g_2)$*.*

*In all cases the resulting saliency is unique.*

### 4.6.2 Lack of upper bounds by closing

Is it possible to reach similar bounds, but from above, by means of closings? We saw that the lattice $\mathcal{N}(N_0)$ is not preserved under complement. Thus it is direct but interesting observation that the complement saliency function $g$ given by $g' = \max(g) + 1 - g$ need not be a saliency function by itself.

Consequently, the closing $\varphi$, dual of the binary net opening $\gamma$ for the complement on $N_0$, maps $\mathcal{P}(N_0)$ on itself, but not on $\mathcal{N}(N_0)$, as shown by Figure 4.10. The asymmetry extends, of course, to the numerical case, and there is no closest closing because the infimum of saliencies may not be a saliency. The closing $\varphi(g)$ in figure 4.9 does not define a hierarchy, but composition product $\gamma\varphi(g)$ in figure 4.12 does.

The saliency $\gamma\varphi(g)$, greater than the largest saliency lower-bound $[\gamma(g) \leq \gamma\varphi(g)]$ and smaller than the upper-bound $\varphi(g)$ is indeed a good interpolation of saliency function from $g$.

### 4.6.3 Upper bound by Geodesic Reconstruction

The net opening $\gamma(g)$ can be interpreted as a geodesic reconstruction of function $g$ under the marker $N_0$, J-net. The dual operation, which consist in a geodesic reconstruction

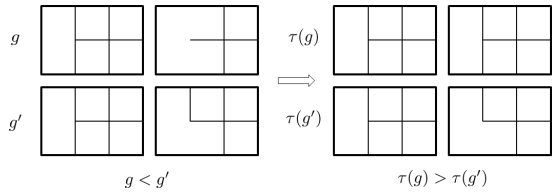



FIGURE 4.11: Dual Closing: The thickening $\tau$ is not increasing. $g$ and $g'$ are binary or two level saliency functions. We have $\tau(g) \geq \tau(g')$ though $g \leq g'$.

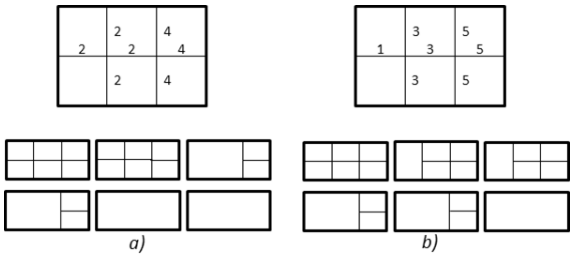



FIGURE 4.12: a) filter $\gamma\varphi(g)$, i.e. the closest saliency below $\varphi(g)$, b) thickening $\tau(g)$.

from above(the complement), provides a saliency function which upper bounds $g$, called netting $\tau$. For constructing it, we start from level set $X_t(g)$ of $N_0$ weighted by a function $g$ as described in equation 4.4:

- if $X_t(g)$, as defined by Eq (4.2), is not a J-net, replace it by the leaves $N_0$, which is a J-net,
- if not, leave $X_t(g)$ unchanged.

In both cases, continue with $X_2(g)$, which is now compared to the transform of $X_1(g)$. When the top of the hierarchy is reached, thus in one pass, then the tessellations at all levels are larger than initially and increase with the level, i.e. $\tau(g)$ is a saliency, and $\tau(g) \geq g$, and this operation is idempotent: $\tau[\tau(g)] = \tau(g)$. However the operation $\tau$ is not increasing (see Figure 4.11), and thus belongs to the class of thickening [101]. Figure 4.12 shows the thickening of function $g$ in figure 4.8. Though $\tau(g)$ is an upper-bound of $g$, but not necessarily the closest one, which may or may not exist.

### 4.6.4 Discussion on Graph based methods

We would like to conclude, firstly, by coming back to three arguments reasons that motivated the Euclidean framework:

1. The only operands used in both theory and experiments are points, Jordan loops, and their interiors, but never edges. The principle of parsimony urged to avoid

the useless distinction between vertices and edges, which would have automatically been introduced by a graph approach. However, our results could also be obtained, probably, in a discrete framework which distinguishes vertices and edges.

2. Jordan nets were weighted with functions $g$ (e.g. distance functions), so that each point, in each arc of the net was given a weight. In case of a finite graph the resolution of function $g$ is restricted to a single value for each edge. But there is no particular reason to impose this quantization to the function $g$, whose resolution can be much finer.

3. Distances in the physical world are isotropic, and we need a mathematical representation to preserve it. Now, an abstract graph ignores isotropy, except one that imposes it with its metric. The planar graphs are better: embedded in $\mathbb{R}^2$, they allow us to build distance functions by emulating the discs of $\mathbb{R}^2$. It is why we used this formulation in the preliminary version of the present study. [45] studies how image segmentation algorithms on graph often have to decide optimal connectivity and topology.

## 4.7 Algorithm and Experiments

### 4.7.1 Net Opening by up-sampling and down-sampling

Figure 4.13a) represents the leaves partition whose components are labeled 2,3, and 4. This partition is up-sampled by 2 according a dilation by a $2 \times 2$ square. This results in Figure 4.13b). The basic J-net separating the leaves is obtained by the support of the gradient of image 4.13b) (still with the same origin), and is depicted in Figure 4.13c). This appears only in the even rows and columns. Introduce now a weighting function $g$. Figure 4.13d) depicts the $g$ weights on the gradient support. The cross section $X_t(g)$ for $t = 2$ is indicated in Figure 4.13e). It opens a breach between the label 2 and 4 which merge in a unique component, of label 4 in Figure 4.13f). By down-sampling, i.e. by removing even rows and columns in Fig. 4.13f), we return to initial space in Figure 4.13g), where the two labels 4 and 2 are clustered into a single label, which is exactly the level $t = 2$ of the hierarchy. This results in the closing of classes as defined in corollary 4.5.

With respect to the up-sampling in the discrete spaces, the discretized image fits correctly with its euclidean version, if we admit two assumptions. Firstly no features is smaller than the elementary square grid, and secondly, there is at most one feature change between two neighbouring pixels. The double resolution is nothing but a mean

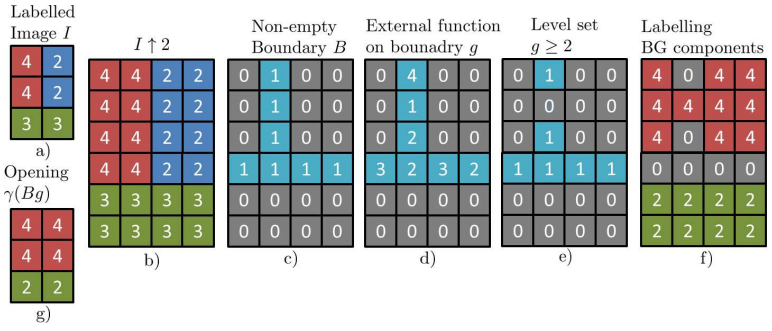



FIGURE 4.13: Toy example showing down-sampling and the different stages of operation to achieve the net opening. One important different w.r.t [86] is that here we have boundary operator that simply calculates the non-zero gradient contours, since the labeling of the components does not produce an ordering dependent on the gradient of the image, and furthermore can be arbitrary. This double resolution to separate cells in discrete topology is the well known Khalimsky topology [76], further one can find the Khalimsky's digital jordan curve theorem in by Kiselman [63].

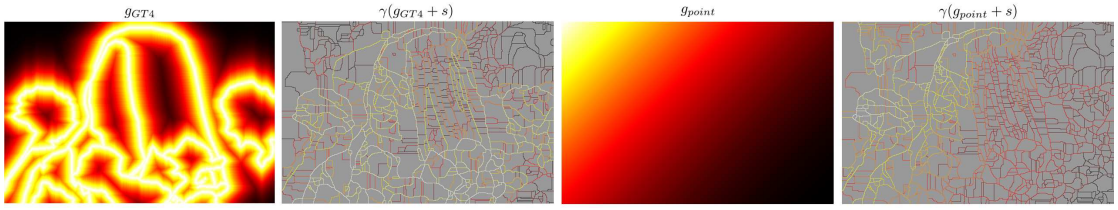



FIGURE 4.14: Inverse distance function $g_{GT4} = 1 - d_4(GT4)$, Transformed saliency $\gamma(s + g_{GT4})$, Point distance function $g_{point}$, Transformed saliency $\gamma(s + g_{point})$.

for exploiting these assumptions. It fits well with the Euclidean Jordan net approach, which divide the space into faces and J-net, since it localizes the J-net in the union of the even rows and columns of the up-sampled image.

### 4.7.2 Fusion of ground truth and hierarchy

Conventionally the ground truth information is intended to assess the quality of a segmentation, here a hierarchy $H$ of segmentations. Here in the place of evaluating the hierarchy, we analyse it with respect to the given ground truth. The saliency transformation by a ground truth is an amelioration of the partitions in the hierarchy by reordering them. They generate new partitions with the same edges ordered by combined effect of proximity to the ground truth, and high value of saliency function(note that this is optional). More clearly, how do we combine a ground truth and a hierarchy?

The inputs given to us are the saliency function $s$ representing the initial hierarchy $H$ and the ground truth set of edges $G$. They are shown in figure 4.15. Here we use the 4-connected distance function of ground truth $d_4$, to define the inverse distance function

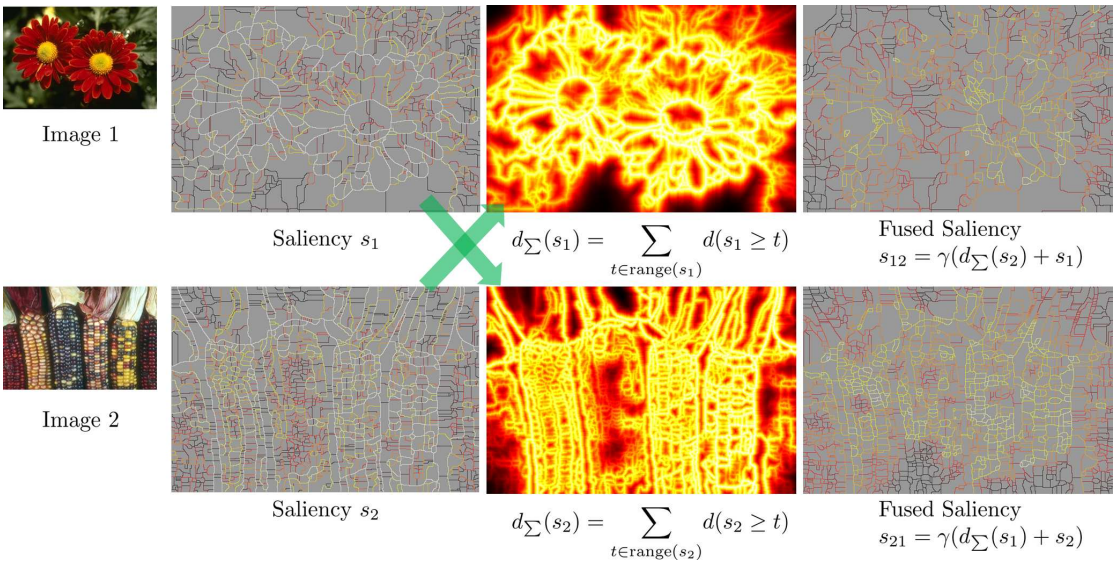



FIGURE 4.15: Hierarchy fusion: Here we fuse two hierarchical structures by introducing a distance function $d_\sum$ which is unique(given its leaves) for every saliency function $s$.

$g = 1 - d_4$. The output is a new saliency $\gamma(s + g)$ and thus a new hierarchy $H_g$ which contains partitions from $H$ that are closest in distance to the ground truth partition $G$ and the initial saliency.

Figure 4.14 summarizes the input inverse distance function $g$ and resulting saliency functions. The input saliency is shown for input image 239096 from the Berkeley database, already shown in figure 4.6. The ground truth $G$ is more or less representative of the image structure in the saliency $s$, and thus the resulting transformed saliency $s_G = \gamma(s+g)$ is not too different, except that in general edges very far from the ground truth are reduced or weakened, while the ones in close proximity are reinforced (see Fig 4.14). An additional example with a point ground truth is used to demonstrate with a function which has nothing to do with the image structure. The choice of the distance function affects the partitions selected as seen in the toy example in Figure 4.16, the diamond is extracted before the circle in case of city-block distance, while the Euclidean distance extracts both at the same scale.

### 4.7.3 Fusions of two hierarchies

In this second example, we will demonstrate a fusion of two hierarchical structures. In this specific example we show how to combine hierarchy of segmentations from images having two very different image structures. In figure 4.15 we present the images of a flower in Image 1 and Corn in Image 2 portraying different textures. The hierarchies are represented by their saliencies $s_1$ and $s_2$.

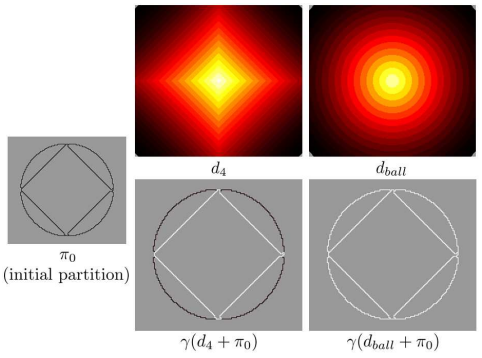



FIGURE 4.16: Scaling space with different distance functions: Here we demonstrate how the inner diamond and the circle get separated when using the (4-connection $d_4$) city-block distance function, while the euclidean ($d_{ball}$) distance function reaches them at the same scale. This produces two different saliency functions and thus hierarchies: $\gamma(d_4 + \pi_0)$ (with 2 levels) and $\gamma(d_{ball} + \pi_0)$ (with single level).

To achieve this fusion one needs a spatial proximity information of the different scales of partitions across the two saliency functions $s_1$ and $s_2$. Thus we convert a saliency function $s$ into an ordered distance function $d_{\sum}(s)$ by summing the distance functions of the different partitions in the thresholds of the saliency function as in equation (4.5).

One of them, $s_1$ say, serves to generate the function $g_1$ by taking the sum $d_\Sigma(s_1)$ of the distance functions of its level sets,

$$d_\Sigma(s_1) = \sum d(X_t), X_t = \{x \mid s_1(x) \geq t\}, \tag{4.5}$$

where $0 \leq t \leq t_{\max}$. The information about $s_1$ is considered as summarized by $d_\Sigma(s_1)$. Thus we use the distance function representing one hierarchy over the leaves (Jordan-net) of another hierarchy to obtain a composition of the two hierarchies. It enters as operand in the net opening $\gamma$, which gives the modified saliency $\gamma(d_\Sigma(s_1) + s_2)$ of the Image 4.15. By inverting the roles of $s_1$ and $s_2$ , we find similarly the modified saliency $\gamma(d_\Sigma(s_2) + s_1)$ Both are depicted in Figure 4.15. Thus here we are able to see in $s_{12}$ the closest set of partitions in the flower saliency function $s_1$ to the texture partitions from the corn saliency $s_2$. We can scan the different scales by weighting by a constant $K$ the distance function $d_{\sum}(s)$, i.e. $s_{12}^K = \gamma(K \cdot d_{\sum}(s_2) + s_1)$. Here we have demonstrated the crossed saliencies for $K = 10$, though to find interesting scalings is another problem.

### 4.7.4 Composition by $\wedge$: Hausdorff distance Ordered saliencies

Until this section we have described how the inverted distance function $g = 1 - d$ of Ground truth partition $G$ was used to reorder the Jordan curves in a hierarchy of segmentations $H$ to produce a new hierarchy of partitions ordered by the proximity to the ground truth. Though this operation in the Hausdorff sense has only measured the distance from the ground truth to the partition contour. In this section we briefly demonstrate an example on how to calculate the Hausdorff distance reordering the hierarchy of partitions by going from the hierarchy $H$ to the ground truth partition $G$.

This is a bit more tougher to calculate since in the earlier case the distance function associated with set $G$ gave a point-wise function that could be used on each Jordan net. In this case for each point in the Jordan net, we can now define the complementary distance as the radius dilation at each point so as to cover the ground-truth partition completely. The distance function can now be writing as the infimum of the distance function of a point $x$ on the ground truth set $G$:

$$d_G(x) = \wedge_{x \in G} d(x) \tag{4.6}$$

This basically corresponds to the supremum of the distance function for each point, on the ground truth set.

Thus now one can write the reordering of the Jordan anew now using this new distance function as follows:

$$s_G = \gamma(\pi_0 + (1 - d_G)) \tag{4.7}$$

Now we can produce two saliencies: Going from $G \to H$ and other from $H \to G$. We are now in a position to write the Hausdorff distance ordered saliency:

$$s_{G \leftrightarrow H} = \gamma(N_0 + \frac{1}{d}) \vee \gamma(N_0 + \frac{1}{d_G}) \tag{4.8}$$

There is also another way of writing this saliency function:

$$s_{G \leftrightarrow H} = \gamma(N_0 + \frac{1}{d \wedge d_G}) \tag{4.9}$$

The two methods of writing the Hausdorff distance ordered saliency suggest that it is composed of the supremum of two net openings, which is also an opening and the

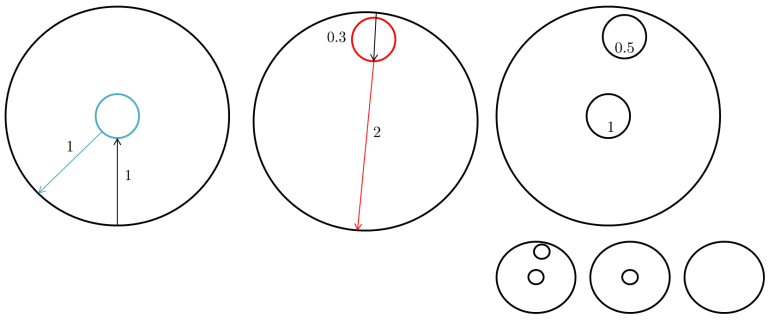



FIGURE 4.17: Toy example demonstrating Hausdorff distance ordering. First case with blue circle demonstrates a symmetrically aligned pair circular contours(black, blue), the second case demonstrates an asymmetrically aligned pair of circular contours(black, red). Aside the arrows we calculate the infimum of radius of dilation for one contour to cover completely, the other, for example, the radius of dilation of set in blue to cover set in black is 1, while radius of dilation for set red to cover set in black is 2. The third figure demonstrates how the two circles are reordered, by associating them with the inverse of the Hausdorff distance between the circles.

resulting function is also a saliency. The inverse of the distance is taken to produce a function which is highly salient when the classes of ground truth and the segmentation are symmetrically placed.

#### 4.7.4.1 Partition Asymmetry and Hausdorff distance

When the ground truth partition $G$ and the base Jordan Net $N_0$, are refinements of each other, i.e. $G \subseteq N_0$ or, $N_0 \subseteq G$, we demonstrate here that the Hausdorff distance between their contours, in this case provide a measure of asymmetry between the partitions.

Consider the simple example of concentric and asymmetrically placed circles, figure 4.17. The Hausdorff distance in case of concentric circles case is the supremum of two equal distances, that correspond to the difference in radii. This is demonstrated by the blue circle. While in case of an asymmetrically placed red circle, the distance from the black$\rightarrow$red $= 0.3$ red$\rightarrow$black $= 2$, is shown. The Hausdorff distance based saliency is calculated using equation 4.9.

The Hausdorff distance function ordered saliency function thus basically produces high saliency value for partitions which are symmetrically placed w.r.t the contours of the ground truth set. Symmetry of object contours itself can be used to extract salient object contours [110]. But here the Hausdorff distance ordering provides different partition contours in base Jordan net $N_0$ that are symmetrically placed w.r.t the contours of the ground truth partition contours. We demonstrate with different refinements of ground truth in figure 4.18

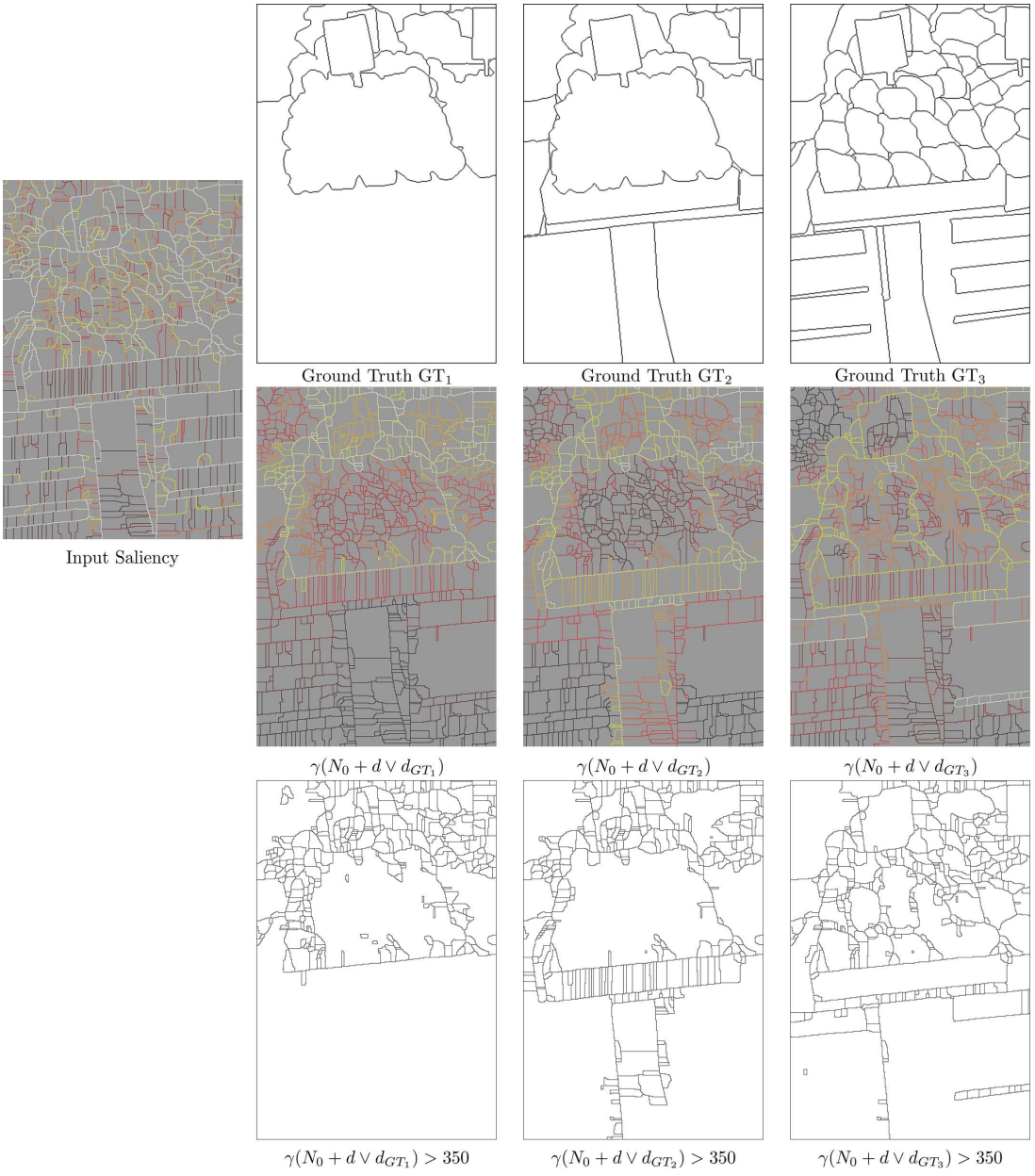



FIGURE 4.18: This example demonstrates 3 scales of ground truths and the corresponding, Hausdorff distance ordered saliency functions. The base Jordan net is extracted from the leaves/finest level of the input saliency function. Partitions corresponding for each ground truth at a threshold(level) of 350 is shown.

### 4.7.5 Combining multiple ground truths with a single hierarchy

As described already in the subsection of refinements and overlaps 3.2.2, the segmentation and ground truth partition, are either refinements of each other locally, or contain overlapping classes. Furthermore different ground truths may not capture the same details (contours). This led us to use an Inf-composition of the set of inverted distance functions so as to combine multiple parts of ground truth partitions for the same scene/image. This is a direct consequence of the fact that the distance function of union of disjoint ground truth contours is the infimum of distance functions of the individual disjoint contours.

Thus in a more formal setting if we have a ground truth $G_1$ with inverted distance function $g_1 = 1 - d(G_1)$, and the elementary Jordan net (leaves) from a hierarchy, we can write the saliency function:

$$s_1 = \gamma(N_0 + g_1) \tag{4.10}$$

Now for a set of disjoint or refined ground truth sets: $G_1, G_2, ...G_n$ with inverted distance functions $d_1, d_2, ...d_n$ we can calculate the saliency on the Jordan $N_0$ as:

$$s_\wedge = \wedge_{i=1}^{n} \gamma(N_0 + (1 - d_i)) \tag{4.11}$$

which can be rewritten using a inf-composition over the functions followed by the net opening:

$$s_\wedge = \gamma(N_0 + (1 - \wedge_{i=1}^{n} d_i)) = \gamma(N_0 + g_\wedge) \tag{4.12}$$

Let's consider the case when we have multiple ground truths $G_i$ drawn by multiple experts, and the problem to now consider is to produce a saliency function weighted by the relative frequency with which the experts draw the same contours as ground truth. In this case the inf-composition of distance functions $d_\wedge = \wedge_i d_i$, it would produce minima's at all points of the ground truth, including the locations where they intersect. In the current problem one would like weighting that produces saliency for contours proportional in value to the number of ground truth contours. This gives us:

$$s_+ = \gamma(N_0 + (1 - \sum_{i=1}^{n}) d_i)) = \gamma(N_0 + g_\sum) \tag{4.13}$$

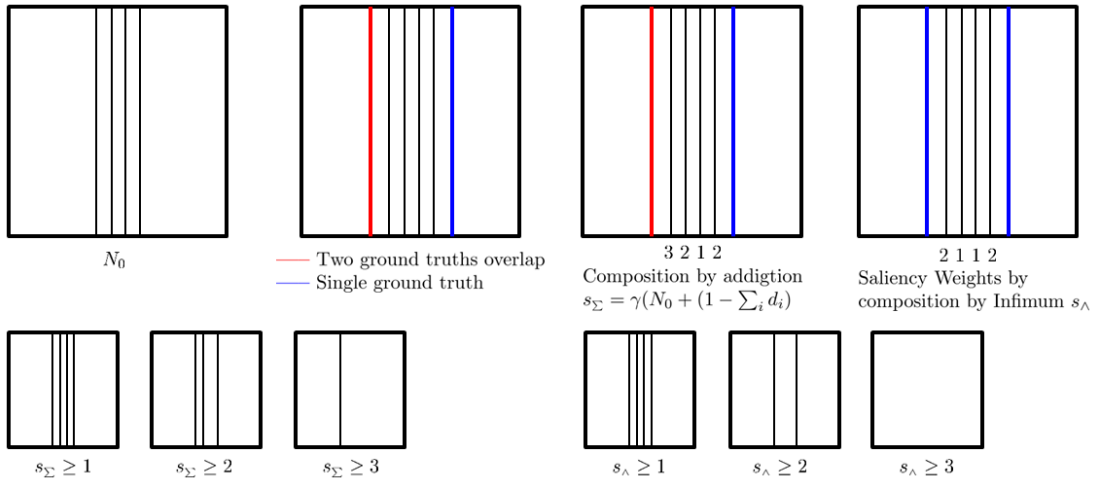



FIGURE 4.19: This toy example demonstrates a partition $N_0$ with three ground truths partitions. Two of the three ground truth lines overlap and is represented in red, while the single ground truth in blue. In such a case one produces the saliency function weights as seen. This is due the composition by addition that weights a partition contour if its proximal to larger number of ground truths (here overlapping, but in general can span space at different locations) than if it is simply close to a single ground truth. This also in contrast produces a different ordering compared to the $\wedge$-compositions.

In the following toy example we demonstrate how co-occurrence of ground truth contours can be used to reorder partition contours. The composition by addition is interesting in cases where the image being segmentation has continuous gradient regions, and there are different possibilities and variations in the human experts segmenting the ground truth. Furthermore in case of the image in study, there are also variations in segmentation possible on account of texture and the segments that the user might interpret as salient [115]. Thus different composition rules serve to extract different partitions from the primitive partition or Jordan Net.

Here we now demonstrate an example (see figure 4.20) over a texture image where the human annotated ground truths vary due the presence of texture, as seen the image is difficult even for a human to segment, since there are a variety of scales and details present. We also show the different inverted distance functions associated with the ground truths, as well as the inverse of the sum of distance funtions $g_{\sum} = 1 - \sum_i d_i$. The different saliencies generated from the different inverted distance functions $g_i = 1 - d_i$ over leaves $N_0$ obtained from origial saliency function (UCM), are in figure 4.21. Too see more clearly the effect of the sum composition, we demonstarte the $M$-measure in equation 4.14 for the different saliencies. The observations we can make, are that the sum composition produces larger range of unique saliency values and thus partitions producing a finer analysis of the hierarchy. This verifies the weighting demonstration in toy example in figure 4.19.

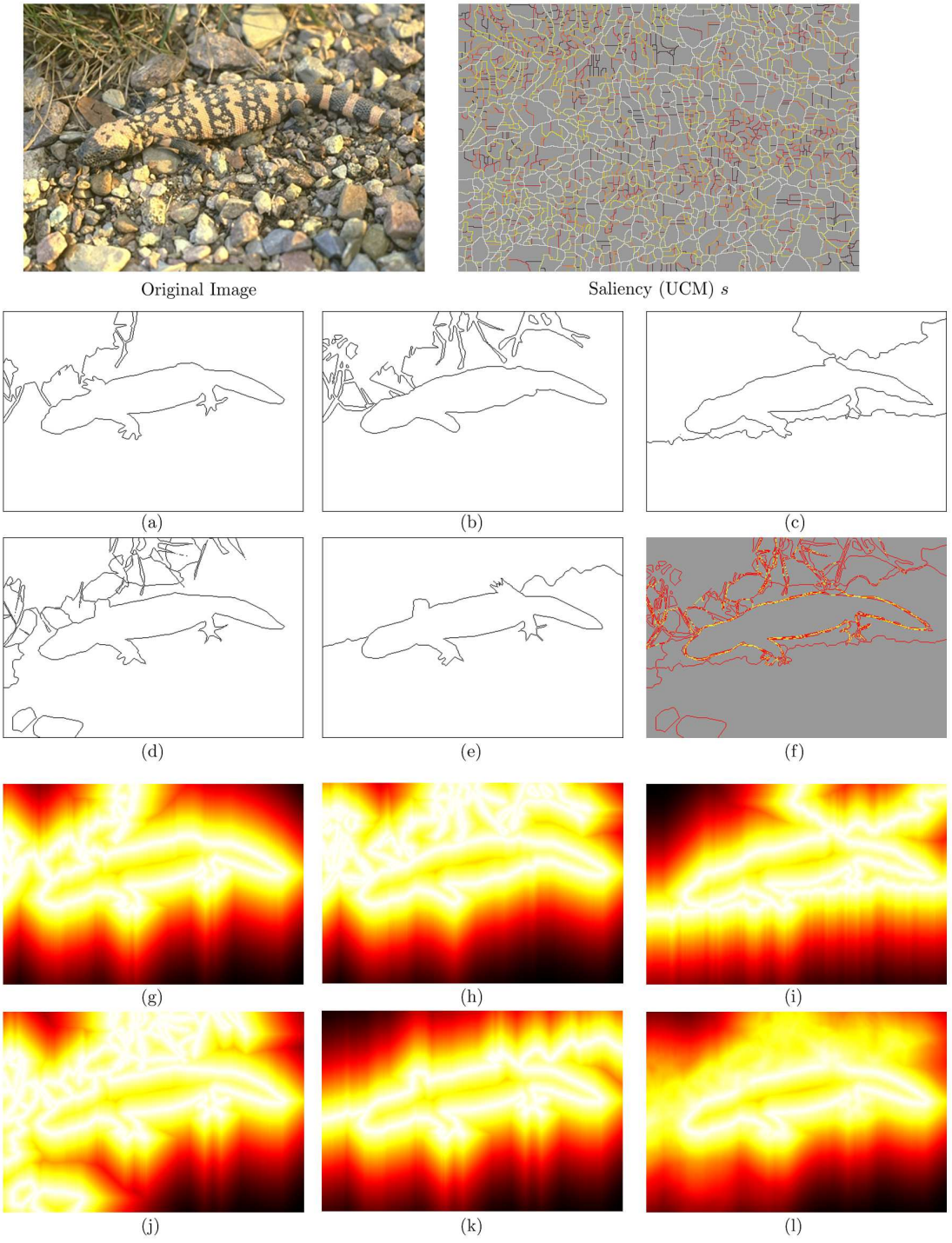



FIGURE 4.20: Initial image with (a,b,c,d,e) representing five different ground truths, with images (g,h,i,j,k) corresponding inverted distance functions of ground truths. While (f) shows the sum of ground truths, and (l) its inverted distance function. We see different contours of the lizard in the image that are reinforced. Further more the ground truth partitions in this case are not simple refinements, and thus validating our use of a composition by addition. Corresponding net openings are demonstrated in 4.21

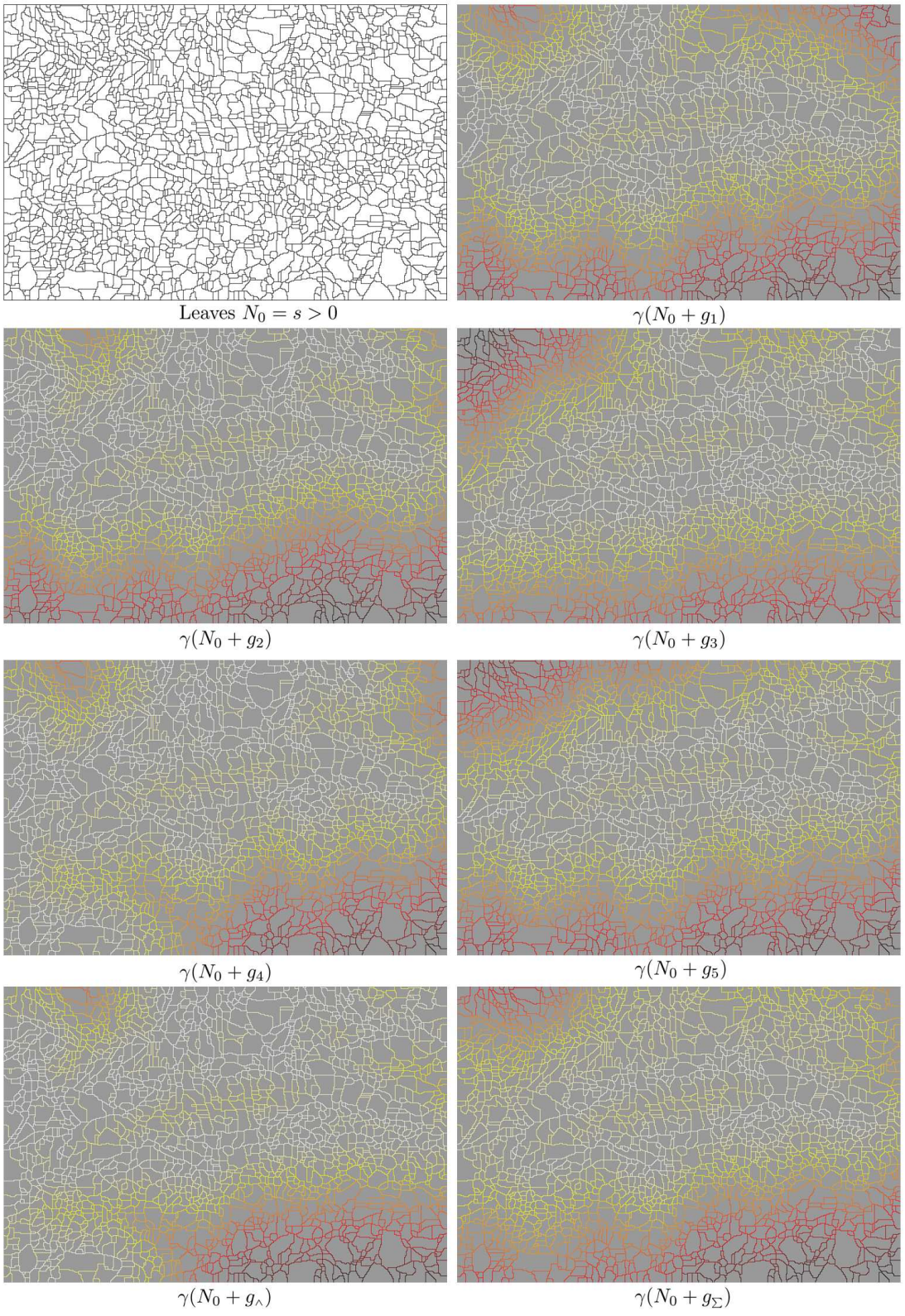



FIGURE 4.21: Figure shows initial leaves partition $N_0$ with three ground truths partitions, and the different net openings possible. The last saliency demonstrates the composition by addition that weights, where higher weight is given to a partition contour, if its proximal to larger number of ground truths (here overlapping, but in general can span space at different locations) than if it is simply close to a single ground truth. Composition by addition (eqn 4.13) also in contrast produces a different ordering compared to the $\wedge$-compositions (eqn) 4.12). Please refer to figure 4.22 to view the different scales that can be extracted using the M-measure from equation(4.14).

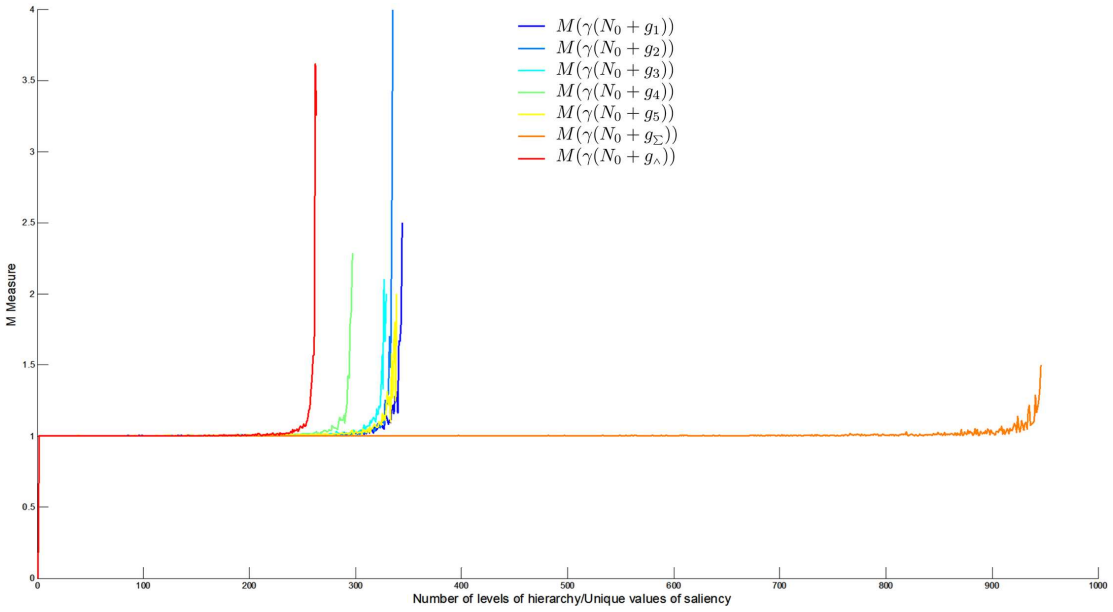



FIGURE 4.22: The plot demonstrates the $M$ measure for the saliencies in figure 4.21 generated with different inverse distance functions of ground truth(in figure 4.20), saliency by Inf-composition and composition by addition. As we can see the inf and sum composition form the bounds of variation of the $M$-measure which not only provides a structural measure of how many children are regrouped by the parent level in each hierarchy but also the number of levels in the hierarchy.

### 4.7.6 Measuring structural changes after transformations

In this section we provide a way to analyse the structural changes in a hierarchy of partitions, resulting from a net opening using a ground truth distance function. This provides one of the possible measures to evaluate the hierarchy, while various studies are already available on the subject [10, 54, 94]. We introduce an evaluation measure:

$$M = \frac{\|\pi_i\|_L}{\|\pi_{i+1}\|_L} = \frac{\#labels \ \ in \ \ childLevel}{\#labels \ \ in \ \ parentLevel} \tag{4.14}$$

where, the quantity $\|\pi_i\|_L$ refers to the number of different labels at a particular level of the hierarchy. The number of labels in the case of saliency function is the number of labeled connected components, in the partition $\pi_i$. The measure is not new and has been widely used in studying dendrograms and hierarchies. Classically in computer science, the branching factor for tree data structures, is the number of children at each node.

Measure $M$ is plotted for different distance functions corresponding to different ground truths and a point distance function, in figure 4.23. The x-axis here corresponds not only to the threshold of the saliency function,i.e. level in hierarchy, but also to the proximity of contours at this level to the respective ground truths. Though this measure $M$ without

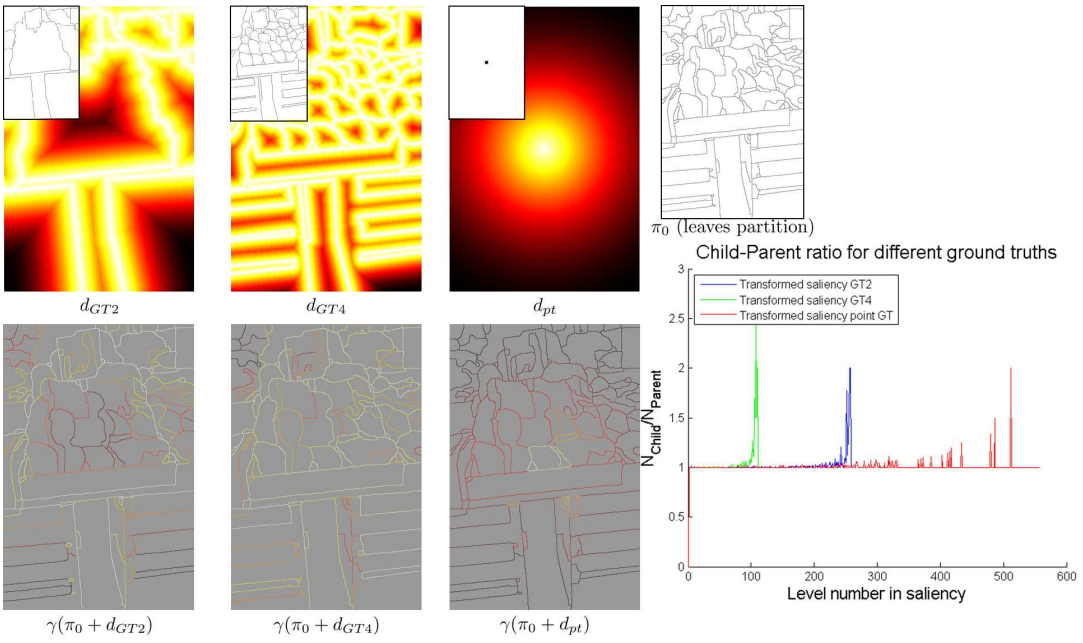



FIGURE 4.23: Evaluating with $M$: The first row shows 3 different distance functions, $d_{GT2}, d_{GT4}, d_{pt}$ with their corresponding ground truths on the top right corner of the images. $\pi_0$ is the leaves partition on which the distance functions will reorder the arcs. In the bottom row we have the transformations $\gamma(\pi + d_{GT2}), \gamma(\pi + d_{GT4}), \gamma(\pi + d_{pt})$ representing new the hierarchies(saliency functions). The plot on the bottom right displays the $M$ measure at different levels of the hierarchy. The maximum number of levels in these hierarchies is bounded by the maximum value of the distance function producing a partition.

the proximity information, is non-informative since it can not distinguish between binary trees. Neither can it distinguish between balanced quad-trees. Its insensitive to uniform branching.

In this experiment the measure is used to compare the scales of classes in the initial saliency and the transformed saliency. The discontinuities in the plot where $M > 1$ correspond to scales of the distance function at which merging/reordering of child partitions occur. With the distance function value on the $x$-axis and branching measure $M$ on $y$-axis, each point on the graph provides at a give radius of dilation, how many number of classes are recombined for the given distance function. This provides a granulometric analysis of the partitions w.r.t the distance function used.

This is another measure extends the Global Precision-Recall similarity integrals described in [61], that measured the integral of product of saliency function and the distance function of ground truth partition. Though in this case, instead of a global measure, we have a value $M$ for each child with a valid proximity value assigned to partition/level in the hierarchy.

### 4.7.7 Geometric and Intrinsic Net openings

The net opening until now involved the combination of an external function and an elementary finite Jordan net. In this section we will show how we not only use an external function, but also the intrinsic properties of partition itself to perform the opening. Intrinsic properties here in this particular case refer to curvature of the partition contours. We use a combination of a corner opening and net opening to basically transform an input weighted quad-tree into a weighted K-d tree saliency function. Such geometrical features have been studied to extract corner points [67]. For a study of geometric corner point openings, or L-openings please refer to our paper [58].

When the geometry of the faces of the Jordan net are well defined as in cases of regular polygons, one can also use this in applications involving polygonal mesh simplifications [35]. In extending the study and relaxing geometric constraints, one can think of performing curvature based net openings, which envisage the use a global point-wise curvature function like in [27]. In the place of corner based L-opening, one can transform the saliency to produce directional net openings that remove contours by ordering them based on the supremum of local radius of curvature. In literature we also find global contour extraction methods that use symmetry of object shape as a prior. This helps extract certain object classes having a certain axis of mirror symmetry, which is analytically measured [110], thus providing us a way to perform a symmetry based net opening.

## 4.8 Braids from net opening

### 4.8.1 Braids from multiple functions on single J-net

In this section we show that given two numerical functions $g_1, g_2 : \mathbb{R}^2 \to \mathbb{R}$, and a input Jordan net $N_0$, we have:

**Proposition 4.10.** *The union any two hierarchies obtained from the individual net opening with any two numerical functions, forms a braid of partitions, where the monitor hierarchy is given by the hierarchy obtained from the net opening using the supremum of the two input functions. That is:*

$$B = \Big\{ H(\gamma(g_1)) \cup H(\gamma(g_1)) \Big\} \setminus \{E\} \quad \textit{on net} \quad N_0 \quad \textit{with} \quad H_{monitor} = \gamma(g_1 \vee g_2) \tag{4.15}$$

This is a direct application of theorem 4.9. We can see that given any two numerical functions $g_1, g_2$ the numerical net openings $\gamma(g_1), \gamma(g_2)$ on a common net $N_0$, are the

highest lower bounds of $g_1, g_2$. Furthermore $\gamma(g_1 \vee g_2)$ is the bounding saliency function for $g_1 \vee g_2$. Also we have, $\gamma(g_1) \leq \gamma(g_1 \vee g_2)$ and $\gamma(g_2) \leq \gamma(g_1 \vee g_2)$. It is now direct that the braid structure results when we are assured of the existence of $\gamma(g_1 \vee g_2)$.

Furthermore one can observe that the any transformed saliency functions $\gamma(s+g)$ of an input saliency function $s$, by any external function $g$ always produces functions that will include the original saliency function $s$ forming a braid.

### 4.8.2 Intersection of multiple Jordan nets

We recall here briefly the proposition 4.11, first proposed and proved in [58] , to further study the net-opening operator in the context of calcuating supremums of partitions which are not hierarchical. Further the same operators helps create braids of partitions.

Consider all J-nets included in a basic one $N_0$, whose associated faces are called *the leaves*. Two families derive from $N_0$. Firstly, the power set $\mathcal{P}(N_0)$, constituted by all sets whose points belong to $N_0$. Secondly, the family $\mathcal{N}(N_0)$ of all J-nets included in $N_0$. We have $\mathcal{N}(N_0) \subseteq \mathcal{P}(N_0)$. Both sets $\mathcal{P}(N_0)$ and $\mathcal{N}(N_0)$ comprise the empty set $\varnothing$ and are ordered by inclusion. $\mathcal{P}(N_0)$, as a power set, is a Boolean lattice. But unlike $\mathcal{P}(N_0)$, the family $\mathcal{N}(N_0)$ is not complemented: if $N \in \mathcal{N}(N_0)$ the complement $N_0 \backslash N$ may have Jordan arcs which are not looped. However, the following property holds:

**Proposition 4.11.** *The set $\mathcal{N}(N_0)$ of all J-nets included in the base $N_0$ forms a lattice with $N_0$ and $\varnothing$ as universal bounds. The supremum of $N, M \in \mathcal{N}(N_0)$ is the union $N \cup M$, and the infimum is the union of all Jordan curves common to $N$ and $M$ (empty set included).*

Proposition 4.11 suggests to associate an opening with the $N_0$-infimum.

**Corollary 4.12.** *Given $X \in \mathcal{P}(N_0)$ the union $\gamma(X)$ of Jordan curves $C$ contained in $X$*

$$\gamma(X) = \cup\{C \subseteq X, C \in \mathcal{N}(N_0)\} \tag{4.16}$$

*is an opening on $\mathcal{P}(N_0)$, called net opening, whose set of invariants is $\mathcal{N}(N_0)$. In particular the $N_0$-infimum between $N, M \in \mathcal{N}(N_0)$ is $\gamma(N \cap M)$.*

To conclude this discussion on Jordan net openings operator, in figure 4.24, we demonstrate the net opening while composing over the intersection of two elementary finite Jordan nets.

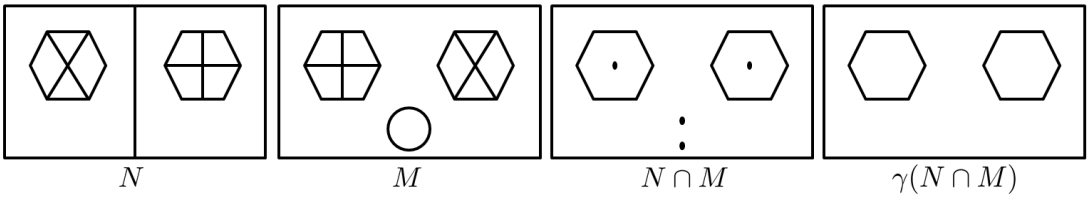



FIGURE 4.24: Net opening over the intersection of two finite Jordan nets $N, M$. This example demonstrates the use of the net opening operator to extract from the intersection of two Jordan nets (which is not necessarily a set of Jordan curves) the largest Jordan net in the base net $N_0$. Since we work on pixel domains, one can consider the finest net $N_0$ to be the one separating all pixels. This operator will be used to create braids while recomposing partitions from different hierarchies. This a very simple operator, and it helps in combining partitions which don't share a common leaf partition or Jordan net.

**Note:** The union of finite countable set of Jordan curves gives a finite set of points, over which a further net opening gives a finite Jordan net. Though this may not be the case with the union of Jordan nets, which might produce Jordan curves with infinite number of faces. Thoght it should be noted that on the base Jordan net representing the pixel grid, the union will always, provide a finite Jordan net, and numerical net opening consists of a union of openings.

### 4.8.3 Braids over multiple hierarchies

In this section we use the elementary binary net opening described in the corollary 4.12, to create the monitor hierarchy. Demonstrated in the figure 4.25, the net opening is used to extract the largest partition whose classes contain the classes of two input partitions. By iterating this binary operation top-down on both input hierarchies, over every pair of partitions extracted for the threshold of the saliency function, one generates the monitor saliency function, and thus hierarchy. Please refer to algorithm 5 in the next chapter, for a better idea of how this is done in a top-down manner. Figure 4.26, shows the finest partition in the monitoring hierarchy. When we stagger two hierarchies of different number of levels, and perform the top-down algorith 5 we can obtain different scales of $\pi_{\min}$ partitions, and thus different monitor hierarchies, but the braid generated by starting top down on all hierarchies produces the largest monitor.

This composition of hierarchies enables a multi-variable hierarchies to be governed by a monitor hierarchy, enabling us now to calculate optimal cuts on the braids. One needs to note that the monitor hierarchy can turn out to be trivial or of less importance if it very few monitoring supremum classes.

There are a variety of experiments we can perform in case of multi-variable hierarchies:

- Hierarchies for color images, for example over the R,G,B images.
- Hierarchies for Hyperspectral images, similar to the RGB case, except larger in dimension
- Hierarchies for Stereo images, to characterize hierarchical structures that remain invariant across the stereo-pair iamges.

In these different applications one has different interpretations for the monitor hierarchy.

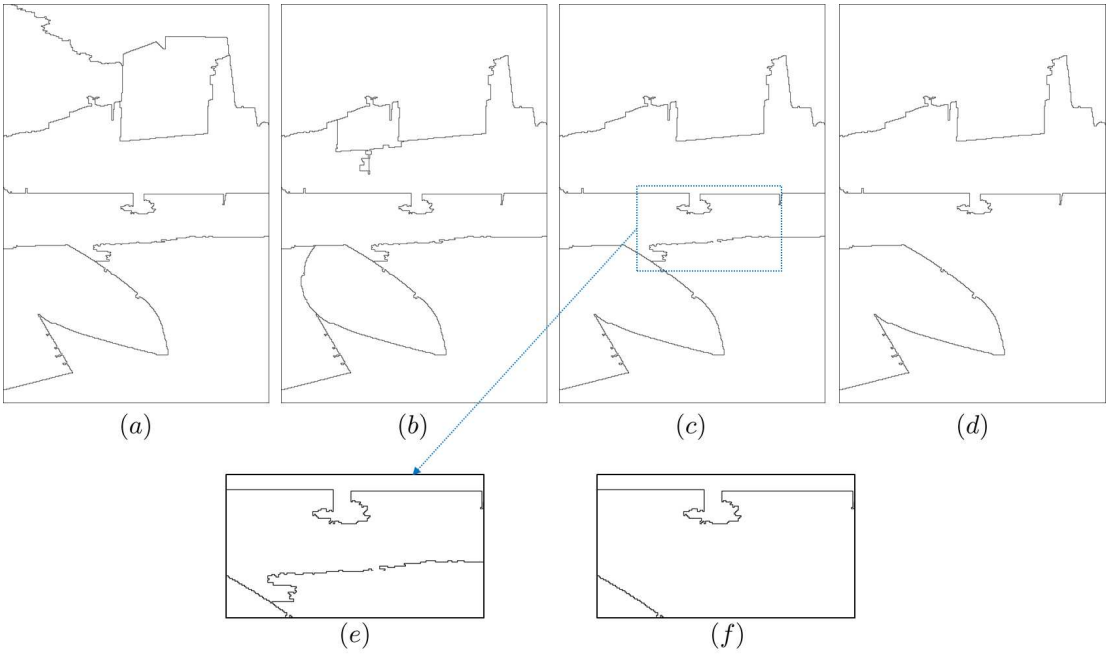
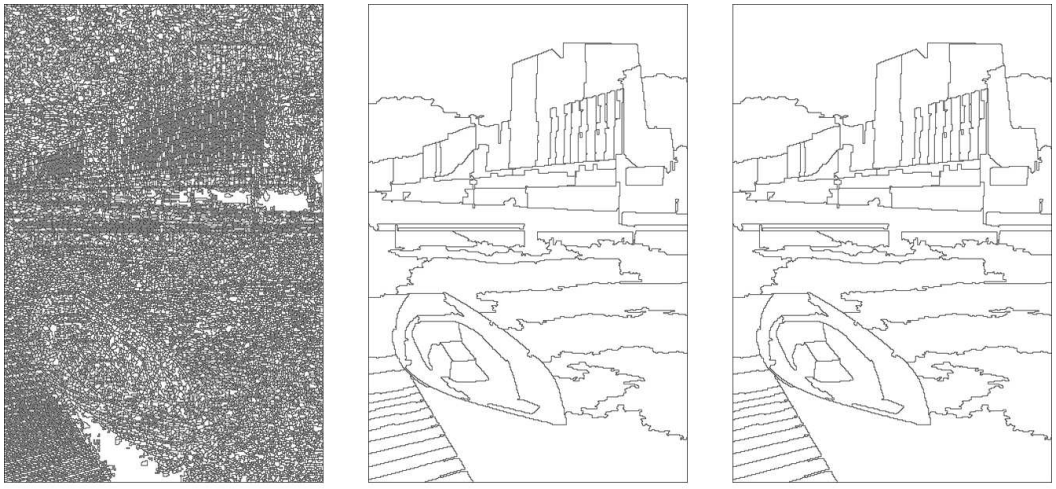



FIGURE 4.25: Calcuating the monitor partition of two partitions using the net opening. The partitions (a), (b) are extracted from watershed flooding by attributes of area and volume respectively [32]. One can note that the watershed by different attributes are not hierarchical. (c) is the intersection of contours between the two partitions, (d) gives the net opening of the intersection, resulting in the monitor partition. (e) shows a magnified view of the contour in the intersection set, where the area and volume floodings have small difference, resulting in a fissure in the intersection set. This leads to a loss of a large class, in the corresponding net opening. We calcuate the monitor of the braid formed in such an event demonstrated in figure 1.6.



$\pi_{\min 1}$ $\pi_{\min 2}$ $\pi_{\min} = \pi_{\min 2}$

FIGURE 4.26: Figure shows, Leaves partition $\pi_{\min 1}$ in watershed saliency by area, $\pi_{\min 2}$, corresponding partition from watershed saliency by volume. And finally $\pi_{\min}$ as described in figure 1.7 in chapter 1, is the net opening of the intersection $\pi_{\min 1} \cap \pi_{\min 2}$. The $\pi_{\min}$ of the braid and finest partition picked from watershed saliency $\pi_{\min 2}$ by volume are the same, thought this might not be the case generally.

## 4.9 Summary

**Chapter contribution summary**

- We worked purely on the contours of a finest Jordan net and defined a net-opening which either removes or preserves the arcs in the net, producing a new hierarchy.
- The net opening was used to find the hierarchy whose saliency is the closest to a given ground truth, represented by its distance function $g$.
- This is the first time the operator has been interpreted as an opening $\gamma$ producing the largest Jordan net that includes any input contour set $X$. The dual closing does not exist since the lattice $\mathcal{N}_0$ is not complemented.
- Geometrical net openings were discussed which help simplify regular hierarchies like Quad-trees and mesh hierarchies. We distinguish here the difference between reordering the Jordan nets based on an external function and based on the intrinsic properties of the partition created by the Jordan net (curvature, junctions, symmetry).
- We demonstrate rules of composition of functions, yielding the largest saliency function, bounded by the composition.
- In Arbelaez et al, Cousty et al. [10, 32], the saliency function is used solely as a representation of the hierarchy of partitions. They provide a visual interpretation of the segmentations at different scales of attributes(dynamic, area and volume) of the gray scale components. In contrast, our study the saliency is used an input numerical function to manipulate the hierarchical structure.
- Braids can be created by composing multiple hierarchies, by performing net opening over the tuples of partitions from the different hierarchies. Braids can also be created by composing net openings using multiple external functions on the same Jordan net.

# Chapter 5

# Algorithms and Graphs

In this short chapter we provide a quick review of super-pixel segmentation methods, demonstrating the different problems encountered in region merging methods, a posteriori. Further we also continue from the work by Brendel-Todorovic [22] who poses the image segmentation problem as an extraction of maximally weighted independent set (MWIS), from the intersection graph over a family of segments generated from any low level segmentation algorithm. To handle them more formally we re-introduce the definition of partition graphs. This class of graphs enables a bijection between the MWIS of the intersection graph and the subsets that provide a partitions of the space. We end the chapter with a max-flow optimization problem on a tree and a graph structure adapted for BOPs.

## 5.1 Region merging methods review

Here we briefly analyse the emergence of *super-pixel/region merging/hierarchical segmentation methods*. In this bibliographic section we will overview a small family of hierarchical segmentation methods. Before this we will shortly provide a motivation on evolution of clustering methods in the domain of image segmentation.

There has been over few decades of study on the problem of image segmentation, and on using region based methods. For computationally tractable segmentation algorithms one often operates on pre-segmentation images. Thus the area of study is now the choice of pre-segmentation to produce a *good* leaves partition or super-pixel segmentation, the set of super-pixel features, to choose a discriminatory dissimilarity metric, as well as statistical measures and others, to recombine the regions. These properties of a good super-pixel algorithm in a way reflect the distinguishing properties of hierarchical agglomerative clustering(HAC) algorithms: *Metric* and *Linkage*.

Here we provide brief over of studies which compare now how different segmentation algorithms generating super-pixel and the subsequent merging orders or linkage criteria. The dissimilarity is usually represented as the weight on the edge set of a Region Adjacency Graph(RAG) associated with the super-pixel segmentation.

To begin with the super-pixel methods are generated not just for reduced complexity but also to the ends of using grouping super-pixels to produce a good segmentation of the image. One classical example one can site is the Normalized Cut based super-pixel segmentation approach to build limb and torso detectors [83], the outputs of which are assembled to classify human posture. One can find a good but brief summary of image segmentation by region merging methods in [111].

Image segmentation has been formulated as a *undersegmentation-agglomerative merge* step. [50] studies Volume based watershed, Mean-shift, and FH algorithm (Felzenszwalb and Huttenlocher) [40] pre-segmentations, using a hierararhical agglomerative merging based on the euclidean distance between the coordinates of colors in CIELAB space. The difficulty evident from the study are the different parameterizations to extract a presegmentation from each of the 3 methods. [48] uses Priority queue based optimized agglomerative clustering algorithm, using the Hotelling $T^2$ statistics on region, again in the CIELAB color space. [23] again study statistical measures of the regions, the Kullback-Leibler Merging Criterion, that maximizes the probability of regions being generated from the same distribution. An interesting scale based measure is the image size dependent scale threshold given by $T = \alpha \cdot \|I\|/n$, where $n$ is the number of regions, and $\|I\|$ of the number of pixels in the image. This threshold is used with the KL merging criterion to control the scale of the partition produced, by varying parameter $\alpha$. One also notes the work in producing hierarchical floodings based on the area, volume and dynamic of gray scale components, producing a hierarchical segmentations by Cousty et al.[32].

The watershed-cuts as already remarked in [86, 109] with Soille's constrained connectivity are calculated rapidly thanks their use of single linkage clustering (SLC), in the sense of linkages in clustering algorithms, while that of Guigues Cocoons [48] provides a complete linkage clustering (CLC).

## 5.2 Algorithms

In this section we compile the algorithms for calculating the minimal $\lambda$-cuts given a HOP or BOP, a Lagrangian parameter $\lambda$, the scale function $\Lambda$ for each parent in a HOP or BOP. Finally given the $\Lambda$-function, one can calculate the family of optimal $\lambda$-cuts.

First we will present the original dynamic program used by Guigues, over the HOP, and later continue with version for the Braids.

### 5.2.1 Optimal Cut DP on HOP

**Algorithm 1:** DP on HOP: Optimal-Cut$(H, \lambda, \omega_\varphi(\mathcal{S}), \omega_\partial(\mathcal{S})$

**Data**: $H$, $\lambda$, $\omega_\phi(\mathcal{S})$, $\omega_\partial(\mathcal{S})$.
**Result**: Optimal cut $\pi^*$, $\omega_\varphi(\pi^*)$, $\omega_\partial(\pi^*)$
**begin**
  $N \longleftarrow |H|$ # levels $F(S) \longleftarrow 0, S \in H$ Set all flags to zero **for** $level \in [2, N]$ **do**
    **for** $S \in H(level)]$ **do**
      $\omega(S) \longleftarrow \omega_\varphi(S) + \lambda\omega_\partial(S)$
      $\omega(\pi(S)) \longleftarrow ComposeFunc(\omega_\varphi(\pi(S)), \omega_\partial(\pi(S)), \lambda)$
      **if** $(\omega(S) \leq \omega(\pi(S)))$ **then**
        $F(S) = 1$ Optimal Parent
        **else**
          $\forall T_i \in \pi(S), F(T_i) \longleftarrow 1$ Optimal Child
        **end**
      **end**
    **end**
  **end**
  **for** $level \in [2, N]$ **do**
    **for** $S \in H(level)]$ **do**
      **if** $F(S)$ **then**
        $\pi^* \longleftarrow \pi^* \cup S$ add optimal classes to cut.
        $\omega_\varphi(\pi^*) = \omega_\varphi(\pi^*) + \omega_\varphi(S)$
        $\omega_\partial(\pi^*) = \omega_\partial(\pi^*) + \omega_\partial(S)$
      **end**
    **end**
  **end**
**end**

Algorithm 1 is the same as BottomUpAnalysis from Garrido's thesis [42] and calculates the optimal cut given a scale parameter $\lambda$. It also calculates the $R, D$ functions in [42] which here are $\omega_\varphi(\pi^*), \omega_\partial(\pi^*)$. The energies $\omega_\varphi, \omega_\partial$ are are available for all classes in the hierarchy $S \in \mathcal{S}$. The energy in case of Salembier-Garrido and Guigues correspond to the Lagrangian $\omega_\varphi(\pi) + \lambda\omega_\partial(\pi)$. While the composition of energies of child classes can be obtained by a general function denoted here by $ComposeFunc(\omega_\phi(\pi(S)), \omega_\partial(\pi(S)), \lambda)$ or $ComposeFunc(\omega(S))$, which represent:

- Addition: $\sum_{T_i \in \pi(S)} \omega_\phi(T_i) + \lambda\omega_\partial(T_i)$
- Supremum: $\bigvee_{T_i \in \pi(S)} \omega(T_i)$
- Infimum: $\bigwedge_{T_i \in \pi(S)} \omega(T_i)$

- Also $\left[\sum_{u\in[1,q]}\omega(T_u)^\alpha\right]^{\frac{1}{\alpha}}$ Minkowski-functional from equation(1.33).

all of which are $h$-increasing energies with compositions that are $h$-increasing. The algorithm 1 can also be used for a single energy, while not involving a constrained optimization problem.

**Algorithm 2:** $\Lambda$-Func$(H, \omega_\phi(S), \omega_\partial(S))$

**Data**: $H$, $\omega_\phi(S)$, $\omega_\partial(S)$.

**Result**: Scale function $\Lambda(S)$.

**begin**

  $N \longleftarrow |H|$ # Levels

  $\forall S \in H, AC(S) \longleftarrow 0$ Set all Anti-causal flags to 0

  $\forall S \in H(1), \Lambda(S), \longleftarrow 0$ Set $\Lambda(S)$ to 0 for leaves level

  **for** *level* $\in [2, N]$ **do**

    **for** $S \in H$*(level)* **do**

      **if** $|\pi(S)| > 1$ **then**

        $\Lambda(S) \longleftarrow \left(\frac{\sum_{T\in\pi(S)}\omega_\varphi(T)-\omega_\varphi(S)}{\omega_\partial(S)-\sum_{T\in\pi(S)}\omega_\partial(T)}\right)$

        $AC(S) = \left(\Lambda(S) < \{\vee\Lambda(T_i), T_i \in \pi(S)\}?1, 0\right)$

      **else**

        $\Lambda(S) \longleftarrow \Lambda(\pi(S))$ repeat $\Lambda$ when parent has single child

        $AC(S) = AC(\pi(S))$ Repeat flag

      **end**

    **end**

  **end**

**end**

The optimal value of $\lambda$ in case of Garrido has been obtained by a gradient search which starts with a upper and lower bounding value of $\lambda$, giving $\omega^h_\varphi(\pi), \omega^h_\partial(\pi)$ high data and constraint terms, and $\omega^l_\varphi(\pi), \omega^l_\partial(\pi)$ and the lower terms, such that the constraint is bounded between the two bounds $\omega^l_\partial(\pi), \omega^h_\partial(\pi)$. The new $\lambda$ iteratively recalculated by setting it to

$$\lambda' = \frac{\omega^l_\varphi(\pi) - \omega^h_\varphi(\pi)}{\omega^h_\partial(\pi) - \omega^l_\partial(\pi)} \tag{5.1}$$

And in each iteration one recalculates the optimal cut for each new $\lambda'$.

Algorithm 3 is not an efficient implementation, we present it for the sake of pedagogy to demonstrate how scale-increasingness works. One has linear complexity according to L.Guigues implementation, to extract all the $\lambda$-cuts. One should note here that the Guigues calculates the whole hierarchy of optimal cuts $\pi^*(\lambda)$, which is not the same as

**Algorithm 3:** Hierarchy of $\lambda$-cuts $H^*$.

**Data**: $H$, $\omega_\varphi(\mathcal{S}), \omega_\partial(\mathcal{S})$.
**Result**: Optimal Cut Hierarchy $H^*$
**begin**
  $i \longleftarrow 1$ indexing starts at 1
  $\Lambda(S) \longleftarrow \Lambda$-Func$(H, \omega_\varphi(\mathcal{S}), \omega_\partial(\mathcal{S}))$
  $N \longleftarrow$ # Parents in $H$ $\pi^*(\lambda) \longleftarrow$ OptimalCut$(H, \Lambda(1), \omega_\varphi(\mathcal{S}), \omega_\partial(\mathcal{S})$
  $\pi^*(\lambda)$ is Finest Optimal Cut
  **while** $|\pi^*(\lambda)| > 1$ **do**
    $i \longleftarrow i + 1$
    $\pi^*(\lambda) \longleftarrow$ OptimalCut$(H, \Lambda(i), \omega_\phi(\mathcal{S}), \omega_\partial(\mathcal{S}))$
    $H^*(i) \longleftarrow \pi^*(\lambda)$
  **end**
**end**

Salembier and Garrido, who calculate an optimal $\lambda^*$ given a constraint function value $\omega_\partial \leq C$.

### 5.2.2 Optimal Cut DP on BOP

In this section we first present the algorithm to create a Braid of partitions by composing a family of HOPs.

The algorithm 4 extracts an optimal cut given $q$-hierarchies and a $\lambda$. The Monitor hierarchy to compose the $q$-hierarchies is generated by algorithm 5, which uses the net opening operator. Though this might not be the only way to generate braids.

This demonstrates that given $q$-hierarchies one might encounter two possibilities:

1. Monitor Hierarchy with, single(full space) class, as unique class.
2. Non-trivial monitor hierarchy and thus Braid.

In case (1) one can now calculate the q-optimal cuts independently from the q-hierarchies, after which the optimal cut with the least energy is picked, in case of equal energies, one picks any of the equivalent optimal cuts. In case (2), the braid structure now ensures an partial optimal cut, that composes partial partitions from the q-hierarchies, and compares it with the monitoring hierarchy's parent, in the dynamic programming stage. The partial optimum is then either, the partial partition that is minimal in energy, or any of the equivalent partial partitions with equal energies (enforcing singularity here), or the parent. This would ensure a minimal $\lambda$-cut that could reach at least equal or better infimum w.r.t the original q-hierarchies, This betterment also impinges constraints upon the generation of the braid structure.

**Algorithm 4:** DP on BOP: Optimal-Cut-BOP($\{H_i, i \in [1,q]\}, \lambda, \omega_\varphi(\mathcal{S}), \omega_\partial(\mathcal{S})$)

**Data**: q-Hierarchies $\{H_1, H_2, ...H_q\}$, $\lambda$, $\omega_\phi(\mathcal{S})$, $\omega_\partial(\mathcal{S})$.
**Result**: Optimal cut $\pi^*(\lambda)$, $\omega_\varphi(\pi^*(\lambda))$, $\omega_\partial(\pi^*(\lambda))$
**begin**
  $H_{\text{monitor}} \longleftarrow$ generateBraid($\{H_1, H_2, ...H_q\}$) monitor hierarchy
  $\forall H \in \{H_1, H_2, ...H_q\}, N_i \longleftarrow |H_i|$
  $N_{\text{monitor}} \longleftarrow |H_{\text{monitor}}|$ # levels in $H_{\text{monitor}}$
  $TrivialBraidflag \longleftarrow |H_{\text{monitor}}(1)|$ # classes in leaves of $H_{\text{monitor}}$
  **if** *TrivialBraidflag* **then**
    \\Perform $q$-DPs independently, without the braid structure
    **for** $H \in \{H_i, i \in [1,q]\}$ **do**
      $[\pi_i^*(\lambda), \omega_\varphi(\pi_i^*(\lambda), \omega_\varphi(\pi_i^*(\lambda)] =$ Optimal-Cut$(H, \lambda, \omega_\varphi(\mathcal{S}), \omega_\partial(\mathcal{S}))$
      $\omega_i(\lambda) \longleftarrow \omega_\varphi(\pi_i^*(\lambda) + \lambda \cdot \omega_\varphi(\pi_i^*(\lambda))$
    **end**
    $i^* \longleftarrow \arg\min_i \omega_i(\lambda)$
    $\pi^*(\lambda) = \pi_{i^*}^*(\lambda)$
    $\omega_\varphi(\pi^*(\lambda)) = \omega_\varphi(\pi_{i^*}^*(\lambda))$
    $\omega_\partial(\pi^*(\lambda)) = \omega_\partial(\pi_{i^*}^*(\lambda))$
    **return**
  **else**
    \\Perform DP with monitor hierarchy
    $F(S) \longleftarrow 0, S \in H$ Set all flags to zero
    **for** $level \in [1, N_{monitor}]$ **do**
      **for** $S \in H_{monitor}(level)]$ **do**
        $\omega(S) \longleftarrow \omega_\varphi(S) + \lambda\omega_\partial(S)$
        **for** $\pi_i(S) \in H_i$ **do**
          $\omega_i(\pi(S)) \longleftarrow ComposeFunc(\omega_\varphi(\pi_i(S)), \omega_\partial(\pi_i(S)), \lambda)$
          **if** $(\omega(S) \leq \wedge_i \omega_i(\pi(S)))$ **then**
            $F(S) = 1$ Optimal Parent
          **else**
            $\pi^*(S) = \arg\min_i \omega_i(\pi(S))$
            $\forall T_i \in \pi^*(S), F(T_i) \longleftarrow 1$ Optimal Child
          **end**
        **end**
      **end**
    **end**
  **end**
  **for** $S \in \{\mathcal{S}\}$ **do**
    **if** $F(S)$ **then**
      $\pi^* \longleftarrow \pi^* \sqcup S$ add optimal classes to cut.
      $\omega_\varphi(\pi^*) = \omega_\varphi(\pi^*) + \omega_\varphi(S)$
      $\omega_\partial(\pi^*) = \omega_\partial(\pi^*) + \omega_\partial(S)$
    **end**
  **end**
**end**

The scale-function and hierarchy of $\lambda$-cuts algorithms do not change very much in their structure for the BOP otherwise.

**Algorithm 5:** Generate Monitor Composing q-hierarchies.

**Data**: $\{H_i\}, i \in [1, q]$.
**Result**: Monitor Hierarchy $H_{\text{monitor}}$
**begin**
  $\forall H \in \{H_i\}, i \in [1, n], N_i \longleftarrow |H_i|$
  $N_{\min} = \min_i N_i$
  \\Top-Down scan on q-hierarchies
  **for** $level = 1 : N_{\min}$ **do**
    $\forall H \in \{H_i\}, i \in [1, n], \pi_i \longleftarrow H(N_i - level)$
    \\Intersection set of all partitions
    $I = \bigcap_i \pi_i$
    \\Net opening on intersection set.
    $H_{\text{monitor}}(N_{\min} - level) \longleftarrow \gamma(I)$
  **end**
**end**

## 5.3 Intersection Graphs for Partition selection

The goal in this section is to identify that the Maximaly independent sets of intersection graphs corresponding to a family of partitions, enables one to extract partitions of the space, other than the ones in the family. Further on we re-introduce the family of "Partition Graphs" which explicity ensure this property for general coverings of the space.

First off, to being this section we follow the work in Brendel-Todorovic [22] which provides the formulation of image segmentation segmentation problem as a maximally weighted independent set(MWIS) problem on the intersection graph of segmentation *stack* generated by low level segmenters.

We note that the MWIS problem on the intersection graph corresponding to a HOP, consists in calculating a family of optimal cuts in the hierarchy. We also discuss a family of *partition graphs* modeling partitions of the input space given any general covering set, that has useful properties.

### 5.3.1 Definitions

We briefly review the definitions of a maximal independent set on a graph, and the Maximally weighted independent set on a weighted graph, in this section.

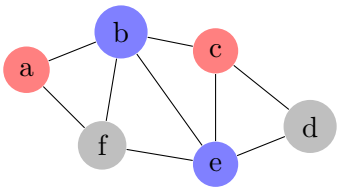



FIGURE 5.1: For the given graph, the maximal independent set of vertices are $\{\{a, c\}, \{b, e\}, \{d, f\}\}$. Addition of any other vertex would lead to an inclusion of an edge between the MIS vertex set. This characterizes its maximal nature.

Let the pair $G = (V, E)$ be an undirected finite graph where $V$ denotes the set of vertices and $E$ denotes the set of edges. If $G$ is connected and acyclic, then it is called a tree. A subset $I \subseteq V$ is said to be an independent set of $G$ if no two vertices are adjacent in $G$, or there is no pair of nodes in $I$ linked by an edge in $E$. Assume also that a positive value is associated to every node in the graph $\omega : V \to \mathbb{R}^+$. Now we define a maximal independent set (MIS).

**Definition 5.1.** *(Maximal Independent Set) Given an graph, an independent set consist of vertices in a graph, where no two vertices are adjacent. A maximal independent set is an independent set such that no more vertices can be added to the set without forcing the set to contain an edge.*

Figure 5.1 shows an example of a MIS. Now one can further define a weighted counterpart of the MIS: Maximally weighted independent set(MWIS). The weight of the independent set $I$ is defined to be the sum of the weights of vertices in the set: $\omega(I) = \sum_{S \in I} \omega(S)$. Now, maximally weighted independent set(MWIS) is an independent set with maximal weight.

In case of identical weights, the maximal property corresponds to the ordering relation, that there are no other set $I'$ that contains $I$. Finding a MWIS in a general graph is NP-hard. For particular cases of the graph structure, we can calculate the MWIS in polynomial time.

**Definition 5.2.** *An undirected graph $G = (V, E)$, where for any covering $\{S_1, S_2, ...S_n\}$ of the space $E$, on associates a node $v_i \in V$ of the graph for each set $S_i$, and connecting any two vertices $v_i, v_j$ when their corresponding sets have a non-empty intersection, i.e. $S_i \cap S_j \neq \emptyset$.*

One can distinguish different types of graphs by the nature of the intersection between two sets in the family covering the space:

- Intersection Graph: $S_i \cap S_j \neq \emptyset$

- Overlap Graph: $S_i \cap S_j \neq \emptyset$ and $S_i \nsubseteq S_j$ and $S_j \nsubseteq S_i$
- Containment Graph: $S_i \subseteq S_j$ and $S_j \subseteq S_i$
- Disjointedness graph: $S_i \cap S_j = \emptyset$

Following which various combinatorial problems are posed based on the characterization, such as intersection number, containment numbers, (minimum size of model in question) etc. Intersection graphs in general have been used in the area of computational geometry, with various types based on the object under study, e.g. intervals, polygons, lines or other geometric objects [75], as well as in abstract graph theory.

**MIS and hierarchies**:

There has been some notable work in using the maximal independent sets, on the one hand to create the hierarchies, while on the other hand to perform partition selection or in other words extract an optimal cut.

To create stochastic hierarchies of partitions with minimal height, Haxhimusa et al. [51, 53], create a graph pyramid from maximal independent vertex set of a base weighted pixel graph, while associating a uniform distribution of values to the vertex set. This is purely to generate a HOP, and not partition selection. The paper also studies the Maximal Independent Directed Edge Set(MIDES), to create the stochastic pyramid.

From experiments [22], it is notable that MWIS algorithm select meaningful segments from a provided set of low-level segmentations of the image. This of course depends on the features producing, based on the weights or energies assigned to each segment and to each neighboring segment pair the MWIS outputs a single, unique partition of the image. Brendel-Todorovic [22] applies MWIS onto the already calculated UCM hierarchy, segmentation performance is noted to go up by selecting a cut traversing several levels and not just a horizontal cut.

### 5.3.2 Partition Graphs

In this section we define a graph structure that would enumerate and extract the partitions in a HOP. This corresponds basically to the MIS of an Intersection graph corresponding to a HOP.

First defined in [73], where for given set or space $E$, the Maximum Independent Sets of a *Partition Graph* would correspond to partition of space $E$.

**Definition 5.3.** *(Partition Graph) Let $\mathcal{S} = \{S_1, S_2, ..., S_n\}$ be a family of distinct, non-empty subsets of some universal set $U$. Then intersection graph $G$ for the family $\mathcal{S}$, if*

1. *For any two vertices $x, y$ are adjacent iff $S_x \cap S_y \neq \emptyset$,*
2. *If the family covers $E$, i.e. $E = \bigcup_{S \in \mathcal{S}} S$,*
3. *If every maximal independent set $M$ on the vertices in the family $S_x : x \in M$ partitios of $E$,*

*Then $G$ is partition graph of $\mathcal{S}$ [73, 74].*

Thus for any MIS $M \subset V$ yields a partition,

$$E = \bigsqcup_{v \in M} S_v$$

Furthermore this results gives a bijection between the partitions from a HOP and the family of MIS of the partition graph corresponding to the classes from a HOP. Consider $\mathcal{M} := \{M\}$ to be the set of all MIS of a partition graph corresponding to a HOP $H$, then we can write:

$$\mathcal{M} \to \Pi(H, E) \tag{5.2}$$

One of the important properties from [73], we will be using in the case of hierarchies and later for braids, is that the subgraph of any partition graph will also be a partition graph for the subset of vertices considered in the subgraph. This property is simple to prove, though is vital to understand the nested structure that recursively is produced in each class in the hierarchy/braid. We will return to this point shortly. We also see later why the intersection graph for a braid does not form a partition graph, but needs some transformation so as to have the MIS-partition bijection as in equation 5.2. For further development on the properties of the graph one can find a good summary in [75].

### 5.3.3 Maximally weighted Independent Set on HOP Intersection Graph

Given a hierarchy of partitions(HOP) $H$ of space $E$, one can construct an intersection graph, where each class family $\mathcal{S} = S_i \in H$ is represented by node $v_i \in V$, and an edge between any two classes is added when they have non-zero intersection $S_i \cap S_j \neq \emptyset, \forall S_i, S_j \in H$. In case of hierarchies the intersection is necessarily implies an inclusion relation, though in case of braids this will not be the case.

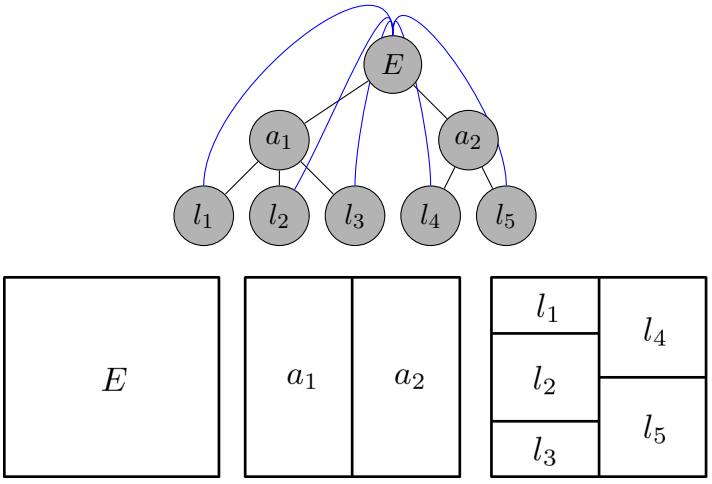



FIGURE 5.2: Typical Intersection Graph for a hierarchy of partitions $H$. The independent sets of such an intersection graph enumerates all the cuts in the hierarchy $H$: $\{\pi_1(E) = l_1, l_2, l_3, l_4, l_5, \pi_2(E) = a_1, l_4, l_5, \pi_3(E) = l_1, l_2, l_3, a_2, \pi_4(E) = a_1, a_2, \pi(E) = E\}$. This graph codes the intersection of classes in the hierarchal structure shown beside it.

The intersection graph corresponding corresponding to the classes of $H$, is in fact a containment graph, where an edge exists between any two vertices $V(S_i), V(S_j)$ is $S_i \subset S_j$ or $S_i \supset S_j$ [38]. This produces an edge between any child and all of its ancestors and vice versa. One can note that the whole space $E$ as a single class, if present, is linked to all the other nodes in the graph since its intersection with any class is not empty.

**Proposition 5.4.** *The intersection graph for a family of classes $\mathcal{S} = \{S|S \in H\}$ from any hierarchy of partitions, is a partition graph, and consequently the Maximum Independent Sets(MIS) of this Partition graph has a unique correspondence with the cuts in the hierarchy, $\forall x \in M$, $\exists \pi(E) \in \Pi_H(E)$.*

Note for brevity we reiterate here that, the maximal nature of a Independent set does not refer to the refinement or coarseness of the classes in the partition, and the impending refinement ordering of partitions. One can now add the following observation:

**MWIS for HOP**: Now, given a weight $\omega : V \to \mathbb{R}^+$ to each node, then the Maximally weighted Independent Set (MWIS) is the MIS with the largest weight, denoted by the family $\mathcal{M}^*$. This basically refers to the fact that there are more than one, MWIS that partition the space. When we have the singular energy $\omega$ 1.9, the exists only one MWIS, i.e. $\mathcal{M}^* = \{M^*\}$, which corresponds to the unique optimal cut.

An example of MWIS on the intersection graph formed for a given hierarchy of partitions is shown in figure 5.2.

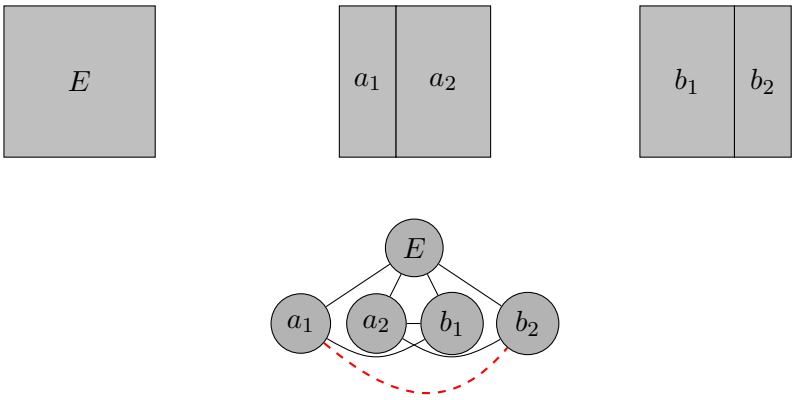



FIGURE 5.3: The MIS of the intersection graph corresponding to family of partitions from a Braid are shown in figure. They are $\{(a_1, a_2), (b_1, b_2), (a_1, b_2), (E)\}$. Node pair $(a_1, b_2)$ form a MIS but does not form a cover of the space, since if $A \sqcup Y \neq E$. Consequently the intersection graph is not a partition graph for the class family $\{a_1, a_2, b_1, b_2, E\}$. This is due to the missing edge shown in red, which when added transforms the graph in to a partition graph.

### 5.3.4 Intersection Graph for Braids

Using low-level segmenters like in [22], or super-pixel segmentations algorithms like SLIC [1] with increasing scale, one can create "stack" of partitions/segmentations $\Pi_{\text{stack}}$. and its intersection graph was calculated. For the purpose of robustness [22] to avoid thin or parasitic intersections, Brendel-Todorovic adds a constraint that the overlap between segments needs to be greater than a filtering threshold chosen, i.e., $S_i \cap S_j \geq t$. This also the reason why [22] needs morphological post-processing to remove soft-overlaps. One also needs to take into account that this threshold needs to be chosen smaller than size of classes in the stack, i.e. $t << |S_i|$.

**Intersection graphs on Braids**: Given a braid of partitions as in figure 5.3, one can find a situation where the soft overlaps are not respected anymore. The family of partitions from a braid consists of classes which may have multiple fathers with large overlaps, which may not produce partition graphs, and would require further operations to have a correspondence between the cuts of a braid and the Maximal independent sets of its intersection graph. To repair this situation when one does not have a covering, we basically reject this MIS by forcibly connecting disjoint classes, and recursing this process for the tuplets of classes from the disjoint class cones. This operation basically enforces the extraction of partitions and enforces them to be a covering of $E$.

We observe though that for a general stack of partitions, where small overlaps are possible between large classes, one would lose approximate good solutions by performing this recursive correction of the intersection graph, though this is still up for debate since the

class size are controlled by parameters in low-level segmentation algorithms. We leave this problem of repairing the intersection graphs "open" for further study.

Another point to note is the similary in the dynamic program substructure between the Breiman's BFOS algorithm [21] and that of the of extraction of the MIS of an intersection graph. Both have a recursive structure.

## 5.4 Max-Flow Min-Cut Analysis

In this section we briefly formulate the optimization problem over a graph, which is an undirected forest/tree, corresponding to first the HOP, Further we also show what could be the equivalent graph topology to model the optimal cut by min-cut in case of the Braid of Partitions.

### 5.4.1 Graph Structure for HOP

The definition of a flow through $G$ requires the data of a source and a sink. The particular shape of a pyramid leads us to take for source the family $A$ of all leaves, and for sink the whole space $E$. In flows, capacities are often allocated to the edges, and sometimes to the vertices. For the sake of comparison, in case of a hierarchy $H$, we will take the nodes. Now, in the graph case, one wants to maximize the flow, whereas above, both additive and sup-generated energies were the matter of minimizations. We must choose, and from now on we decide to maximize the hierarchical energies, i.e. to invert the ordering relations (e.g. in comparisons father/sons of the $h-$increasing case).

In a hierarchy, each leaf $a$ is connected to the root $E$ by a unique path $[a, .., E]$, strictly increasing, and different for each leave. For example, in Figure 5.4 we demonstrate a toy example with sample energies shown on a dendogram. Each node is given a capacity, which appears within it, as shown in figure. As long as two paths in this graph(tree) have no common node, the flows they carry are independent, and upper bounded by the lowest capacity along the portion where they are disjoint. When two such lines meet at some node, e.g. the two paths $[a_1, .., S]$ and $[a_1, .., S]$ which meet in $S$ in then one must adopt some law for composing them, which is exactly what the optimal cut algorithm performs in the dynamic program.

Consider for example the additive energies, which are the most similar to the flows over a directed graph. In this additive case, the capacities $\omega(T_i\}$ of the sons $\{T_i\}$ of $S$ are added, and compared to $\omega(S)$. The min of $\sum \omega(T_i\}$ and $\omega(S)$ gives the provisional capacity of the flow in $S$, and one pursues the climbing. At the end, the nodes of the

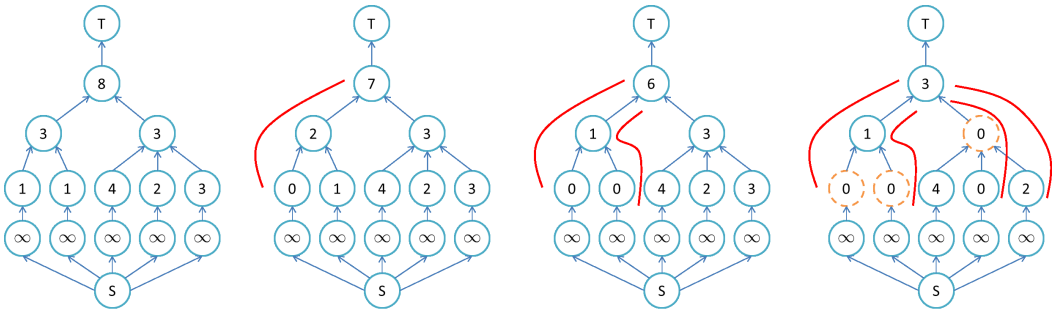



FIGURE 5.4: Flow on Hierarchy: The diagram shows the min-cut for a planar graph(a tree) representing the hierarchy. The source $S$ is connected to all the leaves by an infinite weight to force the flow through all leaves, while the sink is fixed at the root of the hierarchy. The iterations of the augmenting flow method are shown, where the minimum value on each path is subtracted from each node in the path, up till the point where we obtain a cut that separates $S$ and $T$. Each augmenting flow step saturates necessarily one edge of the tree. The max-flow optimization is equivalent to the climbing optimization.

optimal cut are those which are labelled 0, as depicted in Figure 5.4. Finally, we exactly obtain a min-cut in the graph-cut sense, but presented in another formalism, and we can state:

**Proposition 5.5.** *Given an additive energy $\omega$ the optimal cut of a hierarchy $H$ is exactly the min-cut for sources located at all leaves and a sink located at the root, over the weighted tree representing $H$.*

A final point to note is that the energies in case of HOP, are additive in case of calculating the max-flow. We will now extend this to a braid with a convenient graph decomposition.

### 5.4.2 Graph Structure for BOP

Let us consider a braid of partitions $B$ with a monitor $H$ included. For the sake of demonstration we will consider a braid created by composing two hierarchies $H_1, H_2$. Each class $S \in B$ is associated with a node, while the inclusion relation $S \subseteq a_i, a_i \in \pi(S)$ led to the edges, representing the parent-child relationships.

In figure 5.5 we demonstrate an elementary example of a braid composed of two hierarchies: $H_1 = \{\{E\}\{a_1 \sqcup a_2\}\}$ and $H_2 = \{\{E\}\{b_1 \sqcup b_2 \sqcup b_3\}\}$, where the monitoring class is the $\{E\}$. As seen in the demonstration from having a source node $S$ followed by $\infty$ nodes, provides us a graph that has a tree structure, over which the calculation for min-cut corresponds exactly to the min-cut on the HOP.

In the case of BOP, aside the binary decision between a parent class and child classes, one also needs a binary decision between the different child classes, in this case in figure 5.5

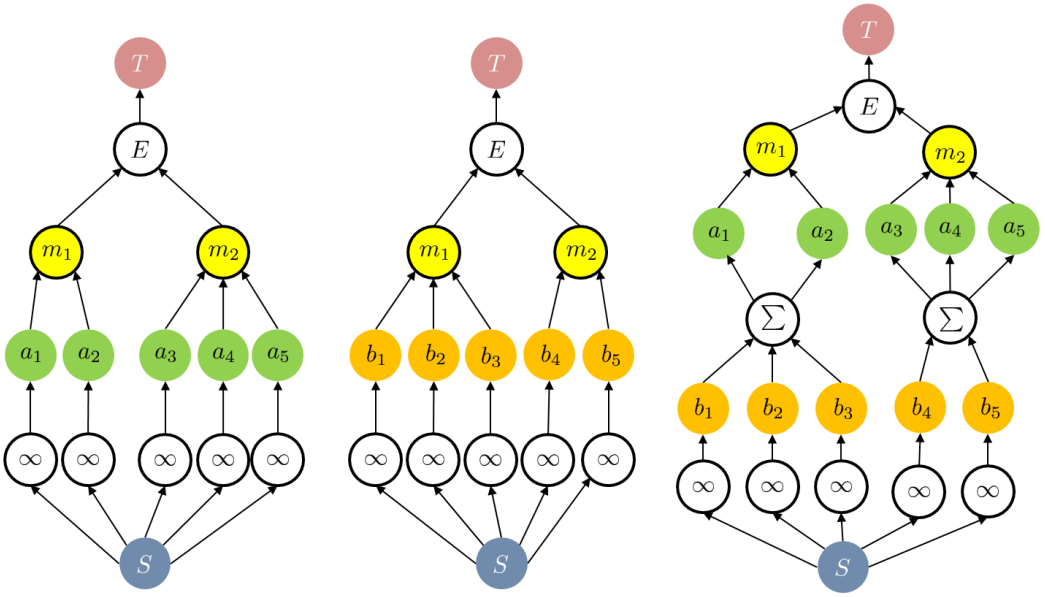



FIGURE 5.5: Series arrangement of of partial partitions in the formulation of max-flow on a braid of partitions.

between nodes $\{a_1, a_2\}$ and $\{b_1, b_2, b_3\}$. In case of extraction of cuts, this corresponds to a disjunctive(OR) choice, while in case of the flow this becomes and conjunction(AND) choice. This is demonstrated by placing these partial partitions in series configuration under the monitoring parent $\{E\}$. In case of a max-flow situation this corresponds to a saturation of either nodes $\{E\}$ or $\{a_1, a_2\}$ or $\{b_1, b_2, b_3\}$.

**Inference and Polytrees**: In conclusion it also important to mention the seminal work by Pearl J. [91], on probabilistic networks, and laid the foundation for modern day graphical models. He was also the first to have introduced a more general graph topology, namely the polytree, over which inference is tractable and can be calculated exactly. His algorithm of Belief Propogation(BP) operates again by local message passing algorithm, with different instances, namely the Forward-Backward algorithm, Viterbi Algorithm, Sum-product algorithm.

Furhtermore, a polytree is a directed acyclic graph, where there is a single path between any two given nodes of the graph, and where the first order markov-chain structure is still usable for a causal and tractable decomposition of a joint distribution [65]. A polyforest generalizes the polytree, as a family of disconnected polytrees. The point of interest now ahead of us is to develop relations between inference algorithms like belief propagation and the BFOS dynamic program. The local-global nature of such problems enable the use of simple message passing methods to obtain a global optimum. The future study is to perfom such studies on the braids.

## 5.5 Summary

This chapter has basically provides a compilation of the different algorithms and graph structures used in the domain of the hierarchical cuts problem.

**Chapter contribution summary**

- Algorithms to calculate a braid from a family of hierarchies, and extract the optimal cut for given energy.
- Review on super-pixel and review of MWIS formulation of image segmentation problem.
- Recalling the definition of "Partition Graphs", which enumerates the partitions from a HOP, transforming an intersection graph for BOP, into a Partition Graph.
- Max-flow min-cut formulation over a tree corresponding to a HOP and over a series of trees.

# Chapter 6

# Conclusion

## 6.1 Thesis Contributions

A brief overview of contributions are provided below.

- The thesis introduces the concept of an *energetic lattice* which is a pair $(H, \omega)$, of a hierarchy of partitions $H$, and energy $\omega$. This structure enables the use of axioms of a lattice, especially the in defining the existence and uniqueness of the minima.
- We introduce the Braid of Partitions, that is richer structure in between, the all possible sets of partitions and the hierarchy of partitions. This enables one to have a larger search space than the HOP, and possibly better minimums on the energetic lattice. Furthermore the braid allows for a multiscale overlapping segmentation hypothesis over the image domain, thus providing its possible use in optical flow and achieving better segmentations by structuring segmentation stacks from low level super-pixel segmenters. The structure of energetic lattice, which is valid for braids and uniquely for them, leads to various results of optimality on braids.
- The thesis generalizes the dynamic program used first in pruning decision trees by Breiman et al. [21], and later by Guigues et al. [13, 47, 100] for non-linear energies with the *h-increasingness* property.
- We provide the *scale-incresaingness* property energy $\omega_\lambda$ in the energetic lattice, and extend Guigues et al.'s [47] *causality* condition of monotonicity to non-linear energies.
- The thesis introduces for the problem of constrained optimization on partitions, the notion of *inf-modularity*. This condition plays a role similar to that of subadditivity, for set valued functions [12], while operating on partial partitions.

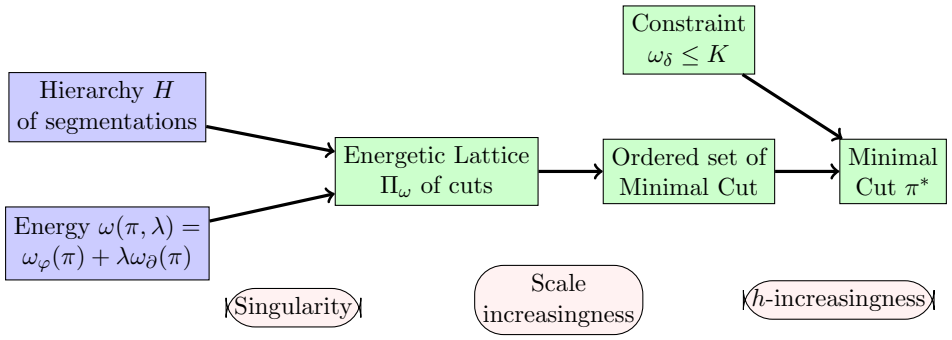



FIGURE 6.1: Chain of ideas for yielding a minimal cut in a hierarchy or a braid. (axioms in rounded boxes).

- We study the meaning of optimality in the Lagrangian Multiplier methods used by Guigues et al., Salembier et al., [47, 100], and demonstrate how it is a Lagrangian Relaxation, of the original hard constrained optimization problem on the HOP, and study its implications.
- We define a local Haussdorf based proximity measure between any two partitions. This is introduced to solve the problem of extracting a partition from a hierarchy $H$, that is closest to a ground truth partition $G$, where $G, h$ correspond to an input image. This local-global problem enables the application of the energetic lattice as well as the dynamic program with infimum composition. We also study the laws of composition for composition of ground truth partitions.
- In an apparently disjunct contribution, we show how to construct a hierarchy of partitions using a finite set of Jordan curves partitioning the image domain, and introducing a lattice structure over it. The method derives but delineates from the study on ground truth energies, where the lattice is moved from partial partitions to the family of jordan curves. With the lattice structure we introduce the *net opening* that extracts the smallest set of jordan curves, and the numerical opening. In the problem of extracting a proximal partition from a hierarchy w.r.t the ground truth, we calculate a new hierarchy, including the proximal partition, by ordering the set of Jordan curves, by the distance function of the ground truth.

To have a global picture of the progression of ideas with the energetic lattice and the lagrangian approach refer to Figure 6.1, and Figure 6.2.

## 6.2 Applications and Future perspectives

We shall briefly provide perspectives on how Braids can be used in the domain of hyperspectral image segmentation, when working on spatial-spectral methods [118] and

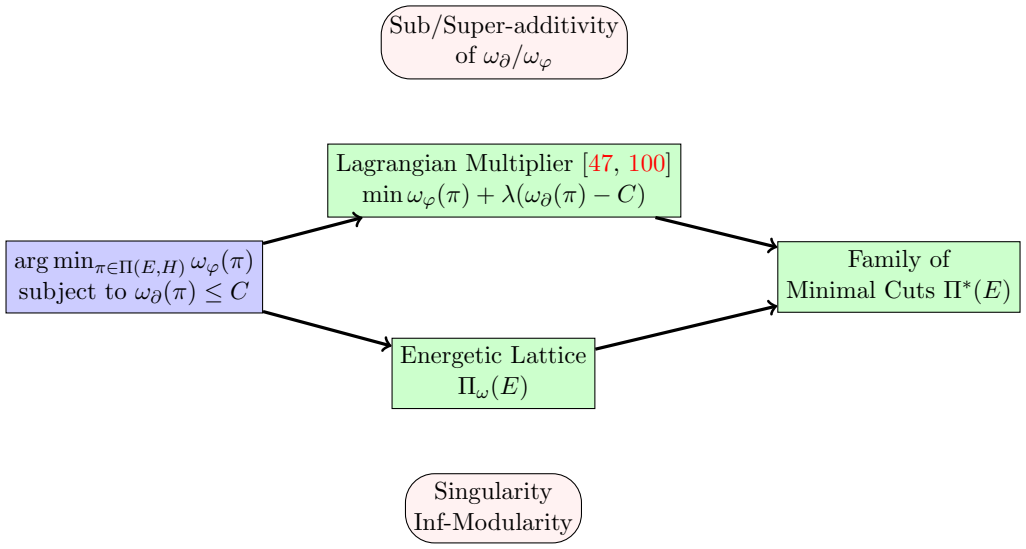



FIGURE 6.2: Two models to formulate the constrained optimization problem. The Lagrangian multipler's method minimizes the Lagrangian of the constrained optimization problem for each feasible $\lambda$, while the energetic lattice defines an order on the lattice structure of partitions. The infimum of the energetic lattice, in this case for the Lagrangian function $\omega(\pi, \lambda) = \omega_\varphi(\pi) + \lambda\omega_\partial(\pi)$, leads to the ordered set of optimal cuts. The conditions on the energy are presented in rounded boxes, so as to obtain a minimum on the HOP for both methods. One can see the counterpart conditions of the Lagrangian Multipliers method, are properties of functions needed to achieve a ordering of energies on the HOP partition strucure. Though its important to note that the energetic lattice can be formulated for the lagrangian linear case, while the axioms of $h$-increasingness and scale-incresingness, generalize subadditivity and superaddivity.

GIS(Geographic Information System) based spatial segmentation problems, for example in Kurtz et al. [66].

**Project with UMR ESPACE**: The methods studied in this thesis have already been applied to problems in (Geographic information systems) GIS, in particular to spatio-temporal and morphological analysis population the region of Provence-Alpes-Côte d'Azur (PACA) south of France. This work is in collaboration with Christine Voiron and the team at UMR ESPACE, Nice Sophia Antipolis. Problems purview include population density based hierarchical segmentation [14], multivariable hierarchical segmentations, autoroute proximity based segmentation and other problems involving hierarhical spatial structure and different types of constraints.

### 6.2.1 Multi-segmentations

Satellites and planes provide sequences of images in different wave bands. If there are many, like in airborne images, we suppose that they have been reduced to their first principal components. Each of these significant images is segmented individually. This results in the hierarchies $H^1 = \{\pi_1^1 ... \pi_n^1\}$ ... $H^k = \{\pi_1^k ... \pi_n^k\}$ which are distinct but

with a common base of leaves, and a same number of levels. How to synthesize these hierarchies, and to extract from the series of $H_k$ a unique (constrained or not) optimal cut ? Introduce the refinement supremum $H = \vee H_k$ of the $H_k$ , i.e. the hierarchy $H$ whose section $\pi$ at each level is the refinement supremum of the $\pi_k$ at the same level. The family $\{H_k\}$ of all sections of all $H_k$ turns out to be a braid of monitor $H$. At point $x$ the minimization processing will result in the class among the $S_k(x)$ which is locally the less energetic. We have already seen the work of Chanussot, Angulo, Santiago and Valero et al. [6, 46, 118] on how one enters the multidimensional case.

### 6.2.2 Thematic partitions

We now consider a partition of the Euclidean plane into thematic classes, like the GIS classification in the $P$ categories of water, forest, cultures, habitations, soil, etc... Each class is given one "color" taken among the $P$ categories. As $\mathbb{R}^2$ is topological, one can speak of frontiers and of adjacent classes. A matrix $M(p,q) = [m(p,q)]$ of similarities between adjacent classes is given. Two classes $S_p$ and $S_q$ merge if they are adjacent and if their similarity is higher or equal to $[m(p,q)]$. The similarity of $S_p$ and $S_q$ is an energy which can include physical parameters (nature of the soil, slope, etc.) and geometrical ones (length of the frontier, simplicity of the union $S_p \cup S_q$, distance to a highway, etc.). The merging of each $p \to q$ of the $P$ categories leads to $2\binom{P}{2}$ partitions. They form a braid $B$ because one can construct a monitor hierarchy $H$ by means of the Jordan net opening developed in Chapter 5. The maximal cut of this braid is a partition which optimizes the similarities, a processing which thus results in a clustering.

### 6.2.3 Thematic prediction

We continue which the previous situation, by adding the time dimension. We start from the partition of a region $Z$ in $P$ categories, and we want to estimate the partition of the same region at time $t'$. The classes are the same, but their categories may change between $t$ and $t'$. One can urbanize a class initially devoted to agriculture, or extend former soils to forest, etc. A transition matrix $A = [a(p,q)]$ has been estimated from other sources, and is given. This matrix is a priori different from the previous one $M$ . A class $S_p$ of category $p$ is changed into $S_q$ according to an energy associated with the categories of the neighbors of $S_p$ and with matrix $A$. It lead to $\binom{P}{2}$ partitions which form a braid $B$, as previously. The maximal cut of $B$ is then the best estimator w.r.t. the energetic choices and the matrix $A$.

Similar situations occur in binocular vision, or in motion estimation. In binocular vision, for example, an energy is associated with the lateral shifts, and the classes of the optimal cut provides zones of the space considered as located at the same depth.

### 6.2.4 Combination of earth images

When they focus on a same geographical zone $Z$, different sensors produce maps which indicate various informations, and with different accuracies. The same as for situation c) (thematic hierarchies) may be said here. The optimal cut of the braid provides here the optimal synthesis of the maps, w.r.t. the chosen criteria.

These problems have not been studied by braids yet they all can be approached by the braid theory, which of course does not excludes other approaches.

# Index

# List of Figures

# List of Tables

# Bibliography


[1] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S. Slic superpixels compared to state of the art superpixel methods. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(11):2274–2282, Nov 2012. ISSN 0162-8828. pages [166]

[2] Akcay, H. G. and Aksoy, S. Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE T. Geoscience and Remote Sensing*, 46(7):2097–2111, 2008. pages [21], [43], [44], [55], [74]

[3] Allène, C., Audibert, J.-Y., Couprie, M., Cousty, J., and Keriven, R. Some links between min-cuts, optimal spanning forests and watersheds. In Banon, G. J. F., Barrera, J., Braga-Neto, U. d. M., and Hirata, N. S. T., editors, *Proceedings, International Symposium on Mathematical Morphology, (ISMM)*, volume 1, pages 253–264. Instituto Nacional de Pesquisas Espaciais (INPE), October 2007. ISBN 978–85–17–00035–5. pages [42]

[4] Alvarez, L., Guichard, F., Lions, P.-L., and Morel, J.-M. Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis*, 123(3):199–257, 1993. pages [50], [76], [187]

[5] Andres, B., Kroeger, T., Briggman, K., Denk, W., Korogod, N., Knott, G., Koethe, U., and Hamprecht, F. Globally optimal closed-surface segmentation for connectomics. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *Computer Vision ECCV 2012*, volume 7574 of *Lecture Notes in Computer Science*, pages 778–791. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33711-6. pages [127]

[6] Angulo, J., Velasco-Forero, S., and Chanussot, J. Multiscale stochastic watershed for unsupervised hyperspectral image segmentation. In *Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009*, volume 3, pages III–93–III–96, July 2009. pages [174]

[7] Arbelaez, P. *Une approche metrique pour la segmentation d'images*. PhD thesis, CEREMADE, Universite Paris-Dauphine, France, 2005. pages [106]

[8] Arbelaez, P. Boundary extraction in natural images using ultrametric contour maps. In *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, CVPRW '06, pages 182–, Washington, DC, USA, 2006. IEEE Computer Society. ISBN 0-7695-2646-2. pages [124]

[9] Arbelaez, P. and Cohen, L. D. Constrained image segmentation from hierarchical boundaries. In *CVPR*. IEEE Computer Society, 2008. pages [112]

[10] Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33 (5):898–916, may 2011. ISSN 0162-8828. pages [9], [100], [104], [106], [114], [120], [147], [154]

[11] Arthur F. Veinott, J. *Lectures in Supply-Chain Optimization, Management Science and Engineering 361*. Department of Management Science and Engineering Stanford University, Stanford, California 94305, 1998. pages [2], [3]

[12] Bach, F. Learning with submodular functions: A convex optimization perspective. *Foundations and Trends in Machine Learning*, 6(2-3):145–373, 2013. pages [2], [52], [53], [171]

[13] Ballester, C., Caselles, V., Igual, L., and Garrido, L. Level lines selection with variational models for segmentation and encoding. *Journal of Mathematical Imaging and Vision*, 27(1):5–27, 2007. pages [2], [57], [59], [67], [76], [171]

[14] Baro, J., Côme, E., Aknin, P., and Bonin, O. Hierarchical and multiscale Mean Shift segmentation of population grid. In *22th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2013)*, Bruges, Belgique, Apr 2013. pages [173]

[15] Bell, E. T. Exponential numbers. *The American Mathematical Monthly*, 41(7): 411–419, 1934. ISSN 00029890. pages [2]

[16] Bertrand, G. On topological watersheds. *Journal of Mathematical Imaging and Vision*, 22(2-3):217–230, 2005. pages [127], [128], [129]

[17] Bertsekas, D. P. *Nonlinear Programming: 2nd Edition*. Athena Scientific, 2004. ISBN 1-886529-00-0. pages [75]

[18] Beucher, S. Watershed, hierarchical segmentation and waterfall algorithm. In Serra, J and Soille, P, editor, *Mathematical Morphology And Its Applications To*

*Image Processing*, volume 2 of *Computational Imaging And Vision*, pages 69–76, 1994. pages [129]

[19] Beucher, S. and Lantuéjoul, C. A topological approach to hierarchical segmentation using mean shift. In *International Workshop on image processing, real-time edge and motion detection/estimation*, 1979. pages [128]

[20] Boyd, S. and Vandenberghe, L. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004. ISBN 0521833787. pages [61], [65], [66], [72]

[21] Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. *Classification and Regression Trees*. Wadsworth, 1984. ISBN 0-534-98053-8. pages [3], [12], [13], [14], [15], [20], [42], [57], [60], [67], [69], [75], [76], [167], [171], [181]

[22] Brendel, W. and Todorovic, S. Segmentation as maximum-weight independent set. In Lafferty, J. D., Williams, C. K. I., Shawe-Taylor, J., Zemel, R. S., and Culotta, A., editors, *NIPS*, pages 307–315. Curran Associates, Inc., 2010. pages [5], [155], [161], [163], [166]

[23] Calderero, F. and Marques, F. Region merging techniques using information theory statistical measures. *Trans. Img. Proc.*, 19(6):1567–1586, June 2010. ISSN 1057-7149. pages [156]

[24] Chan, T. F. and Vese, L. A. Active contours without edges. *Trans. Img. Proc.*, 10(2):266–277, feb 2001. ISSN 1057-7149. pages [97]

[25] Chiang, M. and Boyd, S. P. Geometric programming duals of channel capacity and rate distortion. *IEEE Transactions on Information Theory*, 50(2):245–258, 2004. pages [60]

[26] Chou, P., Lookabaugh, T., and Gray, R. Optimal pruning with applications to tree-structured source coding and modeling. *Information Theory, IEEE Transactions on*, 35(2):299–315, Mar 1989. ISSN 0018-9448. pages [57], [58], [67], [76]

[27] Ciomaga, A., Monasse, P., and Morel, J.-M. Image visualization and restoration by curvature motions. *Multiscale Modeling and Simulation*, 9(2):834–871, 2011. pages [149]

[28] Couprie, C., Grady, L., Najman, L., and Talbot, H. Power watersheds: A new image segmentation framework extending graph cuts, random walker and optimal spanning forest. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 731–738, Sept 2009. pages [42]

[29] Couprie, C., Grady, L., Najman, L., and Talbot, H. Power watershed: A unifying graph-based optimization framework. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(7):1384–1399, July 2011. ISSN 0162-8828. pages [42]

[30] Couprie, M. and Bertrand, G. Topological gray-scale watershed transformation. In *Optical Science, Engineering and Instrumentation'97*, pages 136–146. International Society for Optics and Photonics, 1997. pages [128]

[31] Cousty, J., Bertrand, G., Najman, L., and Couprie, M. Watershed cuts: Minimum spanning forests and the drop of water principle. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(8):1362–1374, Aug 2009. pages [42], [127], [128], [129]

[32] Cousty, J. and Najman, L. Incremental algorithm for hierarchical minimum spanning forests and saliency of watershed cuts. In Soille, P., Pesaresi, M., and Ouzounis, G. K., editors, *ISMM*, volume 6671 of *Lecture Notes in Computer Science*, pages 272–283. Springer, 2011. ISBN 978-3-642-21568-1. pages [98], [99], [100], [117], [120], [127], [129], [152], [154], [156], [181], [185]

[33] Cousty, J., Najman, L., and Serra, J. Raising in watershed lattices. In *ICIP*, pages 2196–2199. IEEE, 2008. pages [127], [129]

[34] Cover, T. M. and Chiang, M. Duality between channel capacity and rate distortion with two-sided state information. *IEEE Trans. Inf. Theor.*, 48(6): 1629–1638, Sep 2006. ISSN 0018-9448. pages [60]

[35] Delest, S., Boné, R., and Cardot, H. Hierarchical mesh segmentation using waterfall and dynamics. In *ISPA '07: Proceedings of the 5th International Symposium on Image and Signal Processing and Analysis*, pages 162–167, September 2007. pages [149]

[36] Diday, E. Spatial pyramidal clustering based on a tessellation. In Banks, D., McMorris, F., Arabie, P., and Gaul, W., editors, *Classification, Clustering, and Data Mining Applications*, Studies in Classification, Data Analysis, and Knowledge Organisation, pages 105–120. Springer Berlin Heidelberg, 2004. pages [24]

[37] Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2000. ISBN 0471056693. pages [12]

[38] Čenek, E. and Stewart, L. Maximum independent set and maximum clique algorithms for overlap graphs. *Discrete Appl. Math.*, 131(1):77–91, sep 2003. ISSN 0166-218X. pages [165]

[39] Everett, H. Generalized lagrange multiplier method for solving problems of optimum allocation of resources. *Operations Research*, 11(3):399–417, 1963. doi: 10.1287/opre.11.3.399. pages [17], [59], [61], [63], [67], [75], [76]

[40] Felzenszwalb, P. F. and Huttenlocher, D. P. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, Sep 2004. ISSN 0920-5691. pages [156]

[41] Felzenszwalb, P. and Schwartz, J. Hierarchical matching of deformable shapes. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, June 2007. doi: 10.1109/CVPR.2007.383018. pages [106]

[42] Garrido, L. *Hierarchical Region Based Processing of Images and Video Sequences: Application to Filtering, Segmentation and Information Retrieval.* PhD thesis, Universitat Politècnica de Catalunya (UPC), 2002. pages [3], [57], [59], [60], [91], [157]

[43] Gersho, A. and Gray, R. M. *Vector Quantization and Signal Compression.* Kluwer Academic Publishers, Norwell, MA, USA, 1991. ISBN 0-7923-9181-0. pages [3], [60]

[44] Gorelick, L., Schmidt, F. R., Boykov, Y., Delong, A., and Ward, A. D. Segmentation with non-linear regional constraints via line-search cuts. In Fitzgibbon, A. W., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *ECCV (1)*, volume 7572 of *Lecture Notes in Computer Science*, pages 583–597. Springer, 2012. ISBN 978-3-642-33717-8. pages [115]

[45] Grady, L. and Jolly, M.-P. Weights and topology: A study of the effects of graph construction on 3d image segmentation. In *Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention - Part I*, MICCAI '08, pages 153–161, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-85987-1. pages [136]

[46] Gueguen, L., Velasco-Forero, S., and Soille, P. Local mutual information for dissimilarity-based image segmentation. *Journal of mathematical imaging and vision*, 48(3):625–644, 2014. pages [174]

[47] Guigues, L. *Modéles multi-èchelles pour la segmentation d'images.* PhD thesis, Thése doctorale Universitè de Cergy-Pontoise, Dec 2003. pages [2], [3], [4], [16], [17], [18], [40], [42], [57], [58], [59], [67], [68], [76], [79], [91], [96], [171], [172], [173]

[48] Guigues, L., Men, H. L., and Cocquerez, J. P. The hierarchy of the cocoons of a graph and its application to image segmentation. *Pattern Recognition Letters*, 24 (8):1059–1066, 2003. pages [17], [156]

[49] Guigues, L., Cocquerez, J. P., and Men, H. L. Scale-sets image analysis. *International Journal of Computer Vision*, 68(3):289–317, 2006. pages [10], [40], [44], [49], [51], [55], [74], [100]

[50] Hanbury, A. How do superpixels affect image segmentation? In *Proceedings of the 13th Iberoamerican Congress on Pattern Recognition: Progress in Pattern Recognition, Image Analysis and Applications*, CIARP '08, pages 178–186, Berlin, Heidelberg, 2008. Springer-Verlag. ISBN 978-3-540-85919-2. pages [156]

[51] Haxhimusa, Y. *The structurally Optimal Dual Graph Pyramid and its Application in Image Partitioning*. PhD thesis, 2007. pages [163]

[52] Haxhimusa, Y. and Kropatsch, W. Segmentation graph hierarchies. In Fred, A., Caelli, T., Duin, R., Campilho, A., and de Ridder, D., editors, *Structural, Syntactic, and Statistical Pattern Recognition*, volume 3138 of *Lecture Notes in Computer Science*, pages 343–351. Springer Berlin Heidelberg, 2004. ISBN 978-3-540-22570-6. pages [127]

[53] Haxhimusa, Y., Glantz, R., and Kropatsch, W. G. Constructing stochastic pyramids by mides: Maximal independent directed edge set. In *Proceedings of the 4th IAPR International Conference on Graph Based Representations in Pattern Recognition*, GbRPR'03, pages 24–34, Berlin, Heidelberg, 2003. Springer-Verlag. ISBN 3-540-40452-X. pages [163]

[54] Haxhimusa, Y., Ion, A., and Kropatsch, W. G. Evaluating hierarchical graph-based segmentation. In *ICPR (2)*, pages 195–198. IEEE Computer Society, 2006. ISBN 0-7695-2521-0. pages [147]

[55] Huttenlocher, D. P., Klanderman, G. A., and Rucklidge, W. A. Comparing images using the hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(9):850–863, Sep 1993. ISSN 0162-8828. pages [120]

[56] Kappes, J. H., Speth, M., Andres, B., Reinelt, G., and Schnörr, C. Globally optimal image partitioning by multicuts. In Boykov, Y., Kahl, F., Lempitsky, V., and Schmidt, F., editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 6819 of *Lecture Notes in Computer Science*, pages 31–44. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-23093-6. pages [127]

[57] Kass, M., Witkin, A., and Terzopoulos, D. Snakes: Active contour models. *International Journal Of Computer Vision*, 1(4):321–331, 1988. pages [97]

[58] Kiran, B. R. and Serra, J. Fusions of ground truths and of hierarchies of segmentations. *Pattern Recognition Letters*, 47(0):63–71, 2014. ISSN 0167-8655. Advances in Mathematical Morphology. pages [i], [119], [149], [150]

[59] Kiran, B. R. and Serra, J. Global-local optimizations by hierarchical cuts and climbing energies. *Pattern Recognition*, 47(1):12 – 24, 2014. ISSN 0031-3203. pages [i], [7], [36], [38], [76], [95], [107], [108], [109], [120]

[60] Kiran, B. R. and Serra, J. Scale space operators on hierarchies of segmentations. In Kuijper, A., Bredies, K., Pock, T., and Bischof, H., editors, *SSVM*, volume 7893 of *Lecture Notes in Computer Science*, pages 331–342. Springer, 2013. ISBN 978-3-642-38266-6. pages [i], [119], [121], [126]

[61] Kiran, B. R. and Serra, J. Ground truth energies for hierarchies of segmentations. In Hendriks, C., Borgefors, G., and Strand, R., editors, *Mathematical Morphology and Its Applications to Signal and Image Processing*, volume 7883 of *Lecture Notes in Computer Science*, pages 123–134. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-38293-2. pages [i], [42], [45], [95], [120], [148], [181]

[62] Kiran, B. R., Serra, J., and Cousty, J. Climbing: A unified approach for global constraints on hierarchical segmentation. In Fusiello, A., Murino, V., and Cucchiara, R., editors, *ECCV 2012. Workshops and Demonstrations, HiPot: ECCV 2012 Workshop on Higher-Order Models and Global Constraints in Computer Vision*, volume 7585 of *Lecture Notes in Computer Science*, pages 324–334. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33884-7. pages [i], [7]

[63] Kiselman, C. O. Digital jordan curve theorems. In *Proceedings of the 9th International Conference on Discrete Geometry for Computer Imagery*, DGCI '00, pages 46–56, London, UK, UK, 2000. Springer-Verlag. ISBN 3-540-41396-0. pages [137], [183]

[64] Koepfler, G., Lopez, C., and Morel, J. M. A multiscale algorithm for image segmentation by variational method. *SIAM J. Numer. Anal.*, 31(1):282–299, feb 1994. ISSN 0036-1429. pages [123], [124], [125], [183]

[65] Koller, D. and Friedman, N. *Probabilistic Graphical Models: Principles and Techniques - Adaptive Computation and Machine Learning*. The MIT Press, 2009. ISBN 0262013193, 9780262013192. pages [169]

[66] Kurtz, C. *Approche collaborative segmentation-classification pour l'analyse descendante d'images multirésolutions*. PhD thesis, Université de Strasbourg (LSIIT UMR CNR/UdS 7005), September 2012. pages [173]

[67] Laganiére, R. A morphological operator for corner detection. *Pattern Recognition*, 31(11):1643–1652, 1998. pages [149]

[68] Leclerc, B. Description combinatoire des ultramétriques. *Mathématiques et Sciences Humaines*, 73:5–37, 1981. pages [23], [127]

[69] Malisiewicz, T. and Efros, A. A. Improving spatial support for objects via multiple segmentations. In *British Machine Vision Conference (BMVC)*, September 2007. pages [28]

[70] Martin, D. R. *An Empirical Approach to Grouping and Segmentation*. PhD thesis, EECS Department, University of California, Berkeley, Aug 2003. pages [1], [101], [104]

[71] Masci, J., Angulo, J., and Schmidhuber, J. A learning framework for morphological operators using counter–harmonic mean. In Hendriks, C., Borgefors, G., and Strand, R., editors, *Mathematical Morphology and Its Applications to Signal and Image Processing*, volume 7883 of *Lecture Notes in Computer Science*, pages 329–340. Springer Berlin Heidelberg, 2013. ISBN 978-3-642-38293-2. pages [46]

[72] Matheron, G. *Random Sets and Integral Geometry*. Wiley, 1975. pages [51]

[73] McAvaney, K., Robertson, J., and DeTemple, D. A characterization and hereditary properties for partition graphs. *Soochow Journal of Mathematics*, (2): 121–129, 1987. pages [163], [164]

[74] McAvaney, K., Robertson, J., and DeTemple, D. A characterization and hereditary properties for partition graphs. *Discrete Math.*, 113(1-3):131–142, 1993. ISSN 0012-365X. pages [164]

[75] McKee, T. A. and McMorris, F. R. *Topics in Intersection Graph Theory, SIAM Monographs on Discrete Mathematics and Applications 2*. Society for Industrial and Applied Mathematics(SIAM), 1999. ISBN 0-89871-430-3. pages [163], [164]

[76] Melin, E. Digital khalimsky manifolds. *Journal of Mathematical Imaging and Vision*, 33(3):267–280, 2009. pages [137], [183]

[77] Meyer, F. Topographic distance and watershed lines. *Signal Process.*, 38(1): 113–125, Jul 1994. ISSN 0165-1684. pages [128]

[78] Meyer, F. Adjunctions on the lattice of dendrograms. In Kropatsch, W. G., Artner, N. M., Haxhimusa, Y., and Jiang, X., editors, *GbRPR*, volume 7877 of *Lecture Notes in Computer Science*, pages 91–100. Springer, 2013. ISBN 978-3-642-38220-8. pages [9]

[79] Meyer, F. and Lerallut, R. Morphological operators for flooding, leveling and filtering images using graphs. In Escolano, F. and Vento, M., editors, *GbRPR*, volume 4538 of *Lecture Notes in Computer Science*, pages 158–167. Springer, 2007. ISBN 978-3-540-72902-0. pages [120], [129]

[80] Michael Patriksson, A. E., Niclas Andréasson. *Introduction to Continuous Optimization*. Studentlitteratur, Chalmers Publication Library, 2013. ISBN 9789144060774. pages [61], [62], [63]

[81] Monasse, P. and Guichard, F. Fast computation of a contrast-invariant image representation. *IEEE Trans. on Image Proc*, 9:860–872, 1998. pages [124]

[82] Morel, J. M. and Solimini, S. *Variational methods in image segmentation*. Birkhauser Boston Inc., Cambridge, MA, USA, 1995. ISBN 0-8176-3720-6. pages [95], [123], [124], [129]

[83] Mori, G., Ren, X., Efros, A. A., and Malik, J. Recovering human body configurations: Combining segmentation and recognition. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR'04, pages 326–333, Washington, DC, USA, 2004. IEEE Computer Society. pages [156]

[84] Movahedi, V. and Elder, J. Design and perceptual validation of performance measures for salient object segmentation. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 49–56, June 2010. doi: 10.1109/CVPRW.2010.5543739. pages [108], [120]

[85] Mumford, D. and Shah, J. Boundary detection by minimizing functionals. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1985. pages [95]

[86] Najman, L. On the equivalence between hierarchical segmentations and ultrametric watersheds. *Journal of Mathematical Imaging and Vision*, 40(3): 231–247, 2011. pages [129], [137], [156], [183]

[87] Najman, L. and Schmitt, M. Watershed of a continuous function. *Signal Process.*, 38(1):99–112, Jul 1994. ISSN 0165-1684. pages [128]

[88] Najman, L. and Schmitt, M. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(12): 1163–1173, 1996. pages [117], [119], [129], [132]

[89] Nguyen, H. T., Worring, M., and Van den Boomgaard, R. Watersnakes: energy-driven watershed segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(3):330–342, March 2003. pages [128]

[90] P., S. Connected operators based on tree pruning strategies. In Laurent, N. and Hugues, T., editors, *Mathematical morphology: from theory to applications*, chapter 7. ISTE-Wiley, June 2010. ISBN 9781848212152. pages [9], [10]

[91] Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988. ISBN 1558604790. pages [169]

[92] Peng, B. and Zhang, L. Evaluation of image segmentation quality by adaptive ground truth composition. In *Proceedings of the 12th European Conference on Computer Vision - Volume Part III*, ECCV 2012, pages 287–300, Berlin, Heidelberg, 2012. Springer-Verlag. ISBN 978-3-642-33711-6. pages [107], [113]

[93] Pont-Tuset, J. *Image Segmentation Evaluation and Its Application to Object Detection*. PhD thesis, Universitat Politècnica de Catalunya, BarcelonaTech, Barcelona, Feb 2014. pages [101], [102], [103]

[94] Pont-Tuset, J. and Marqués, F. Supervised assessment of segmentation hierarchies. In Fitzgibbon, A. W., Lazebnik, S., Perona, P., Sato, Y., and Schmid, C., editors, *ECCV (4)*, volume 7575 of *Lecture Notes in Computer Science*, pages 814–827. Springer, 2012. ISBN 978-3-642-33764-2. pages [101], [106], [107], [109], [120], [147]

[95] Ramchandran, K. and Vetterli, M. Best wavelet packet bases in a rate-distortion sense. *Image Processing, IEEE Transactions on*, 2(2):160–175, Apr 1993. ISSN 1057-7149. doi: 10.1109/83.217221. pages [58], [67], [76]

[96] Romstad, B. and Etzelmullera, B. Mean-curvature watersheds: A simple method for segmentation of a digital elevation model into terrain units. *Geomorphology*, 139-140(0):293–302, 2012. ISSN 0169-555X. pages [128]

[97] Ronse, C. Partial partitions, partial connections and connective segmentation. *J. Math. Imaging Vis.*, 32(2):97–125, Oct 2008. ISSN 0924-9907. pages [8]

[98] Ronse, C. Orders on partial partitions and maximal partitioning of sets. In Soille, P., Pesaresi, M., and Ouzounis, G., editors, *Mathematical Morphology and Its Applications to Image and Signal Processing*, volume 6671 of *Lecture Notes in Computer Science*, pages 49–60. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21568-1. pages [11]

[99] Ronse, C. Ordering partial partitions for image segmentation and filtering: Merging, creating and inflating blocks. *Journal of Mathematical Imaging and Vision*, 49(1):202–233, 2014. ISSN 0924-9907. pages [11]

[100] Salembier, P. and Garrido, L. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval. *IEEE Trans. on Image Processing*, 9(4):561–576, 2000. ISSN 1057-7149. pages [2], [3], [4], [10], [17], [19], [42], [55], [57], [59], [67], [74], [76], [79], [100], [171], [172], [173]

[101] Serra, J. *Image Analysis and Mathematical Morphology*. Academic Press, Inc., Orlando, FL, USA, 1982. ISBN 0126372403. pages [3], [122], [127], [135]

[102] Serra, J. Hausdorff distance and interpolations. In *H.Heijmans and J. Roerdink Eds.*, volume Mathematical Morphology and its Applications to Image and Signal Processing. Kluwer, 1998. pages [105]

[103] Serra, J. and Kiran, B. R. Optima on hierarchies of partitions. In Hendriks, C. L. L., Borgefors, G., and Strand, R., editors, *ISMM*, volume 7883 of *Lecture Notes in Computer Science*, pages 147–158. Springer, 2013. ISBN 978-3-642-38293-2, 978-3-642-38294-9. pages [i], [7], [95]

[104] Serra, J., Kiran, B. R., and Cousty, J. Hierarchies and climbing energies. In Alvarez, L., Mejail, M., Gomez, L., and Jacobo, J., editors, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, volume 7441 of *Lecture Notes in Computer Science*, pages 821–828. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-33274-6. pages [i], [7]

[105] Shoham, Y. and Gersho, A. Efficient bit allocation for an arbitrary set of quantizers [speech coding]. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 36(9):1445–1453, Sep 1988. ISSN 0096-3518. doi: 10.1109/29.90373. pages [58], [59], [60], [63], [67], [76]

[106] Sinop, A. and Grady, L. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, Oct 2007. pages [42]

[107] Soille, P. Constrained connectivity for hierarchical image partitioning and simplification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1132–1145, 2008. ISSN 0162-8828. pages [42], [43], [53], [74]

[108] Soille, P. and Grazzini, J. Constrained connectivity and transition regions. In Wilkinson, M. H. and Roerdink, J., editors, *Mathematical Morphology and Its Application to Signal and Image Processing*, volume 5720 of *Lecture Notes in Computer Science*, pages 59–69. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-03612-5. pages [21], [43], [53], [55]

[109] Soille, P. and Najman, L. On morphological hierarchical representations for image processing and spatial data clustering. In Köthe, U., Montanvert, A., and Soille, P., editors, *WADGMM*, volume 7346 of *Lecture Notes in Computer Science*, pages 43–67. Springer, 2010. ISBN 978-3-642-32312-6. pages [156]

[110] Stahl, J. S. and Wang, S. Globally optimal grouping for symmetric closed boundaries by combining boundary and region information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(3):395–411, 2008. ISSN 0162-8828. pages [141], [149]

[111] Szeliski, R. *Computer Vision: Algorithms and Applications*. Springer-Verlag New York, Inc., New York, NY, USA, 1st edition, 2010. ISBN 1848829345, 9781848829343. pages [156]

[112] Topkis, D. M. *Supermodularity and Complementarity. Frontiers of Economic Research*. Princeton University Press, Princeton, NJ, 1998. ISBN 9781400822539. pages [2]

[113] Trevor Hastie, J. F., Robert Tibshirani. *The Elements of Statistical Learning (2nd edition): Data Mining, Inference, and Prediction, Second Edition*. Springer Series in Statistics, 2009. ISBN 978-0387848570, 0387848576. pages [12], [14]

[114] Unnikrishnan, R., Pantofaru, C., and Hebert, M. A measure for objective evaluation of image segmentation algorithms. In *Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 34–34, June 2005. pages [29], [180]

[115] Unnikrishnan, R., Pantofaru, C., and Hebert, M. Toward objective evaluation of image segmentation algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):929–944, June 2007. pages [28], [29], [102], [103], [144], [180]

[116] Vachier, C. and Meyer, F. The viscous watershed transform. *Journal of Mathematical Imaging and Vision*, 22(2-3):251–267, 2005. pages [117], [129]

[117] Vachier C., M. F. Extinction value: a new measurement of persistence. In *IEEE Workshop on Nonlinear Signal and Image Processing.*, page 254–257, 1995. pages [129]

[118] Valero, S. *Hyperspectral image representation and Processing with Binary Partition Trees*. phd, Universitat Politècnica de Catalunya (UPC), 2011. pages [18], [42], [45], [47], [172], [174]

[119] Valero, S., Salembier, P., and Chanussot, J. Hyperspectral image representation and processing with binary partition trees. *IEEE Transactions on Image Processing*, 22:1430 – 1443, April 2013 2013. pages [18]

[120] Vanhamel, I., Pratikakis, I., and Sahli, H. Multiscale gradient watersheds of color images. *Image Processing, IEEE Transactions on*, 12(6):617–626, June 2003. ISSN 1057-7149. pages [129]

[121] Veganzones, M., Tochon, G., Dalla-Mura, M., Plaza, A., and Chanussot, J. Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation. *Image Processing, IEEE Transactions on*, 23(8): 3574–3589, Aug 2014. ISSN 1057-7149. pages [42], [47]

[122] Vincent, L. and Soille, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 13 (6):583–598, June 1991. ISSN 0162-8828. pages [128]

[123] Xu, Y. *Tree-based shape spaces: Definition and applications in image processing and computer vision*. PhD thesis, Université Paris-Est, Marne-la-Vallée, France, Dec 2013. pages [18]

[124] Xu, Y., Géraud, T., and Najman, L. Context-based energy estimator: Application to object segmentation on the tree of shapes. In *Proceedings of the 19th International Conference on Image Processing (ICIP)*, Orlando, Florida, USA, Oct 2012. IEEE. pages [97]

# Abstract

Hierarchical segmentation has been a model which both identifies with the cosntruct of extracting a tree structured model of the image, while also interpretting it as an optimization problem of the optimal scale selection. Hierarchical processing is an emerging field of problems in computer vision and hyperspectral image processing community, on account of its ability to structure high-dimensional data.

Chapter 1 discusses two important concepts of *Braids* and *Energetic lattices*. Braids of partitions is a richer hierarchical partition model that provides mutliple locally non-nested partitioning, while being globally a hierarchical partitioning of the space. The problem of optimization on hierarchies and further braids are non-tractable due the combinatorial nature of the problem. We provide conditions, of *h-increasingness*, *scale-increasingness* on the energy defined on partitions, to extract unique and monotonically ordered minimal partitions. Furthermore these conditions are found to be coherent with the Braid structure to perform constrained optimization on hierarchies, and more generally Braids. Chapter 2 demonstrates the Energetic lattice, and how it generalizes the Lagrangian formulation of the constrained optimization problem on hierarchies.

Finally in Chapter 3 we apply the method of optimization using energetic lattices to the problem of extraction of segmentations from a hierarchy, that are proximal to a ground truth set. Chapter 4 we show how one moves from the energetic lattice on hierarchies and braids, to a numerical lattice of Jordan Curves which define a continous model of hierarhical segmentation. This model enables also to compose different functions and hierarchies.

**Keywords**: Hierarchical segmentation, Lagrangian Multipliers, Lattice optimiztaion, Mathemtical Morphology.

# Résumé

La segmentation hiérarchique est une méthode pour produire des partitions qui représentent une même image de manière de moins en moins fine. En même temps, elle sert d'entrée à la recherche d'une partition optimale, qui combine des extraits des diverses partitions en divers endroits. Le traitement hiérarchique des images est un domaine émergent en vision par ordinateur, et en particulier dans la communauté qui étudie les images hyperspectrales et les SIG, du fait de son capacité à structurer des données hyper-dimensionnelles.

Le chapitre 1 porte sur les deux concepts fondamentaux de tresse et de treillis energétique. La tresse est une notion plus riche que celle de hierarchie de partitions, en ce qu'elle incorpore, en plus, des partitions qui ne sont pas emboîtées les unes dans les autres, tout en s'appuyant glolalement sur une hiérarchie. Le treillis énergétique est une structure mixte qui regroupe une tresse avec une énergie, et permet d'y défninr des éléments maximaux et minimaux. Lorsqu'on se donne une énergie, trouver la partition formée de classes de la tresse (ou de la hiérarchie) qui minimise cette énergie est un problème insoluble, de par sa complexité combinatoriale. Nous donnons les deux conditions de h-croissance et de croissance d'échelle, qui garantissent l'existence, l'unicité et la monotonie des solutions, et conduisent à un algorithme qui les détermine en deux passes de lecture des données.

Le chapitre 2 reste dans le cadre précédent, mais étudie plus spécifiquement l'optimisation sous contrainte. Il débouche sur trois généralisations du modèle Lagrangien. Le chapitre 3 applique l'optimisation par treillis énergétique au cas de figure où l'energie est introduite par une "vérité terrrain", c'est à dire par un jeu de dessins mauel, que les partitions optimales doivent serrer au plus près.

Enfin, le chapitre 4 passe des treills énergétiques à ceux des courbes de Jordan dans le plan euclidien, qui définissent un modèle continu de segmentations hierarchiques. Il permet entre autres de composer les hiérarchies avec diverses fonctions numériques.

**Mot Clé**: Segmentation hiérarchique, Multiplicateurs de Lagrange, Optimisation dans les treillis, Morphologie mathématique.