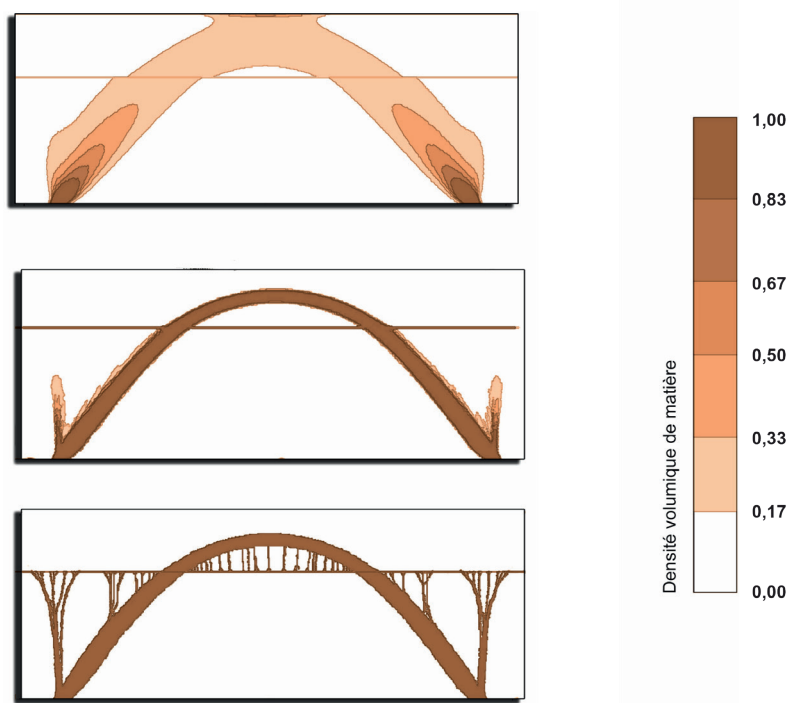


# Optimisation de formes en sciences de l'ingénieur

Méthodes et applications



sous la direction de Michaël PEIGNEY



Sous la direction de  
Michaël PEIGNEY

# Optimisation de formes en sciences de l'ingénieur

Méthodes et applications



Cet ouvrage a été préparé dans la cadre de l'opération de recherche MATEOPT (Matériaux et Énergie pour l'Optimisation des structures de génie civil), 2010-2014. Une partie des travaux présentés s'inscrit dans le projet européen QUIESST (Quieting the Environment for sustainable Surface Transport), 2008-2012.

Cet ouvrage a bénéficié de la relecture croisée de Pierre Charbonnier, Christophe Heinkelé, Michaël Peigney et Jean-Philippe Tarel. Le coordinateur remercie Pierre Argoul et Frédéric Bourquin pour leurs avis sur la pertinence et la qualité de cet ouvrage.

Les auteurs remercient également Corinne Brusque pour sa relecture attentive et ses remarques, ainsi que Daniel Bourbotte et Jorge Mariano pour leur travail sur la mise en forme et l'édition de cet ouvrage.

Ont participé à la rédaction de cet ouvrage :

Laurent Caraffa (IFSTTAR), Pierre Charbonnier (CEREMA), Jérôme Defrance (CSTB), Christophe Heinkelé (CEREMA), Thomas Leissing (CSTB), Mathias Paget (IFSTTAR), Michaël Peigney (IFSTTAR), Jean-Philippe Tarel (IFSTTAR)

*Comment citer cet ouvrage :*

*Peigney M. (dir.) Optimisation de formes en sciences de l'ingénieur. Marne-la-Vallée : Ifsttar, 2018. Ouvrages Scientifiques, OSI3 . 175 pages. ISBN 978-2-85782-744-3.*

*Comment citer une partie de cet ouvrage :*

*Auteur(s) de l'ouvrage. Titre du chapitre. in Peigney M. (dir.). Optimisation de formes en sciences de l'ingénieur. Marne-la-Vallée : Ifsttar, 2018. Ouvrages Scientifiques, OSI3 . Pagination première page–dernière page.*

**Institut français des sciences et technologies des transports, de l'aménagement et des réseaux - Ifsttar**

**14-20 boulevard Newton - Cité Descartes - Champs-sur-Marne - 77447 Marne-la-Vallée cedex 2**

[www.ifsttar.fr](http://www.ifsttar.fr)

**Les collections de l'Ifsttar - ouvrages scientifiques - Réf : OSI3**

**ISBN 978-2-85782-744-3- ISSN 2558-3018**

**Novembre 2018**



*Cet ouvrage est mis à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation commerciale - Pas de Modifications 4.0 International. Les termes de cette licence sont accessibles à l'adresse <http://creativecommons.org/licenses/by-nc-nd/4.0>*

# Sommaire

<b>Introduction</b> .....	7
<b>Chapitre 1.</b> Méthodes de dimensionnement robuste en conception de structures .....	11
1.1. Introduction .....	12
1.2. Problème de dimensionnement .....	13
1.3. L'optimisation topologique .....	16
1.4. Prise en compte des incertitudes .....	25
1.5. Étude d'un problème de dimensionnement robuste en milieu continu	27
1.6. Illustration numérique .....	38
1.7. Conclusion .....	42
1.8. Bibliographie .....	43
<b>Chapitre 2.</b> Optimisation d'un écran antibruit .....	45
2.1. Éléments de contexte .....	46
2.2. Outils numériques .....	46
2.3. Indicateurs de performance .....	52
2.4. Optimisation intrinsèque d'un écran antibruit .....	56
2.5. Optimisation globale de murs antibruit .....	60
2.6. Les outils obtenus .....	62
2.7. Vers une nouvelle méthode de conception environnementale de murs antibruit .....	66
2.8. Bibliographie .....	67
<b>Chapitre 3.</b> Microstructures dans les matériaux à changement de phase .	69
3.1. Introduction .....	69

3.2. Modélisation du problème .....	72
3.3. Lien avec l'optimisation de formes .....	76
3.4. Enveloppe convexe .....	78
3.5. Application à la transformation cubique-orthorombique .....	81
3.6. Microstructures laminées .....	84
3.7. Problèmes à 4 phases .....	89
3.8. Bornes inférieures non convexes sur $W$ .....	92
3.9. Transformations cubiques-monocliniques .....	95
3.10. Conclusion .....	100
3.11. Bibliographie .....	100
<b>Chapitre 4. Optimisation globale pour les contours actifs .....</b>	<b>103</b>
4.1. Introduction .....	103
4.2. Les contours actifs .....	105
4.3. Optimisation globale par recherche de chemin minimal .....	113
4.4. Optimisation globale pour les contours actifs région .....	120
4.5. Conclusion .....	125
4.6. Bibliographie .....	127
<b>Chapitre 5. Optimisation de fonctions pseudo-booléennes .....</b>	<b>137</b>
5.1. Introduction .....	138
5.2. Optimisation de fonctions pseudo-booléennes .....	140
5.3. Extension au cas multi-labels .....	155
5.4. Exemples d'utilisation .....	158
5.5. Conclusion .....	163
5.6. Bibliographie .....	164
<b>Conclusion .....</b>	<b>167</b>
<b>Liste des figures .....</b>	<b>169</b>
<b>Liste des tables .....</b>	<b>173</b>
<b>Fiche bibliographique .....</b>	<b>174</b>
<b>Publication data form .....</b>	<b>175</b>

# Présentation du coordinateur de l'ouvrage



Diplômé de l'École Polytechnique et docteur en mécanique, Michaël Peigney est Ingénieur en Chef des Ponts, Eaux et Forêts, et chercheur au laboratoire Navier (École des Ponts Paris Tech, IFSTTAR, CNRS), Université Paris-Est. Il enseigne à l'École Polytechnique, à l'École des Ponts Paris Tech ainsi qu'à l'ENSTA (École Nationale Supérieure des Techniques Avancées). Titulaire de l'habilitation à diriger des recherches, il mène une partie de ses travaux en collaboration avec des partenaires industriels. Ses thèmes de recherche concernent la mécanique des solides, notamment

le développement de méthodes numériques en mécanique non linéaire, les approches multi-échelles, les phénomènes de transformation de phase, et l'étude des matériaux intelligents.





# Introduction

L'exemple le plus souvent cité lorsqu'il s'agit de décrire ce qu'est l'optimisation de formes est issu de la mythologie gréco-romaine. Il se rapporte à la fondation de Carthage, aux environs de l'an 814 avant notre ère. Après le meurtre de son mari, la reine Didon s'enfuit en Tunisie, où elle demande une terre d'asile aux autochtones. Par dérision, ceux-ci lui attribuent « autant de terrain qu'elle pourrait en faire tenir dans une peau de bœuf ». Après réflexion, elle découpe la peau en fines lanières et les assemble pour former une cordelette. Le problème qui se pose à elle est alors d'entourer à l'aide de cette cordelette un territoire aussi vaste que possible.

En termes plus mathématiques, il s'agit, à périmètre donné, de déterminer la surface fermée d'aire maximale. La solution de ce problème, aussi appelé *inégalité iso-périmétrique*, est connue depuis l'antiquité : c'est le disque. Cet exemple académique met en évidence trois ingrédients qui caractérisent un problème d'optimisation de forme : un *modèle* décrivant la nature de la géométrie cherchée (courbe fermée dans l'exemple mentionné), une fonction *critère* à optimiser (ici, l'aire) et la présence d'éventuelles *contraintes* sur la géométrie, définissant un ensemble de solutions admissibles (périmètre fixé).

La bulle de savon est un autre exemple classique d'optimisation de formes. Dans ce cas, la forme de la bulle de savon est la solution d'un problème de surface minimale.

La forme de la bulle de savon est la solution d'un problème de surface minimale



(crédit photo : Pierre Charbonnier)

La notion de forme ou de structure optimale est naturellement très présente dans notre environnement. Dans le domaine du minéral, par exemple, elle permet d'expliquer la constitution de certains arrangements cristallins. En biologie, les lois de la physique constituent un déterminant important dans le développement

des organismes vivants (THOMPSON, 1992), complémentaires aux mécanismes de sélection naturelle.

De tout temps, l'être humain a cherché à optimiser la forme des objets qu'il fabrique afin de leur donner de meilleures propriétés mécaniques, aérodynamiques, acoustiques, ou simplement pour minimiser leur coût de fabrication. Ce type d'activité a connu un essor très important, notamment au XX<sup>e</sup> siècle, en lien avec les progrès de la technique et l'essor, plus récent, de l'informatique et du calcul scientifique intensif. Les besoins de l'industrie (automobile, aéronautique, électrotechnique, militaire, etc.) et du génie civil ont généré un grand nombre d'applications. En parallèle, la disponibilité de moyens de calcul de plus en plus performants a permis l'introduction de techniques de plus en plus efficaces. La montée en puissance des moyens de traitement numériques de données a également mis en évidence de nouvelles problématiques de recherche : les techniques d'optimisation de forme sont aujourd'hui très utilisées pour résoudre des problèmes inverses en reconstruction d'images numériques (contrôle non destructif, imagerie médicale), et plus généralement, dans les domaines de la vision artificielle et de la reconnaissance des formes.

L'optimisation de formes définit une discipline mathématique à part entière. En effet, les problèmes mathématiques qui se posent (existence d'une solution, conditions d'optimalité, propriétés géométriques ou qualitatives des solutions) sont souvent loin d'être triviaux. Ainsi, la première démonstration rigoureuse de l'inégalité iso-périmétrique dans le plan n'a été proposée qu'en 1841 par le mathématicien suisse Jakob Steiner. Et encore : elle s'est avérée incomplète car elle n'était pas accompagnée d'une preuve de l'existence de la solution, et n'a finalement été complétée qu'à la fin du XIX<sup>e</sup> siècle. De nos jours, certains problèmes liés à la recherche de surfaces minimales à volume fixé demeurent toujours non résolus (par exemple les surfaces de Willmore).

Un aspect primordial de l'optimisation de formes est celui du calcul, effectif ou approché, des solutions. Dans l'exemple du problème de la reine Didon, la solution peut être obtenue de manière théorique, mais ce n'est généralement pas le cas dans les applications traitées et l'on doit avoir recours au calcul numérique. L'*algorithme* constitue donc le quatrième ingrédient de l'optimisation de forme. Le plus souvent, il s'agit de faire évoluer une forme initiale de manière à l'améliorer, au sens du critère optimisé. Cela n'est pas toujours simple, du fait du caractère mal posé des problèmes rencontrés, et de la présence fréquente d'optima locaux. De ce point de vue, les problèmes d'optimisation peuvent se classer en trois catégories, par ordre de difficulté croissante. Dans le cas le plus simple, les formes, ou les structures, sont paramétrées à l'aide d'un petit nombre de variables. L'approche *paramétrique* a l'avantage de réduire l'espace de recherche, mais elle limite également l'espace des solutions admissibles. Un grand nombre d'algorithmes, déterministes ou stochastiques, peuvent être employés. Certains d'entre eux reproduisent des mécanismes naturels, physiques ou biologiques : on parle de méta-heuristiques (SIARRY, 2014). Dans une seconde approche, dite *géométrique*, on fait évoluer les

frontières discrétisées de l'objet, sans toutefois autoriser de modifications de topologie. Plus générique, elle est cependant limitée à l'obtention d'optimums locaux. De plus, son implantation numérique nécessite de remettre à jour périodiquement la représentation (le plus souvent, un maillage) utilisée pour la discrétisation et un raffinement du maillage ne conduit pas forcément à la convergence. Enfin, la troisième catégorie est celle de l'optimisation *topologique*. Permettant l'introduction de trous aussi bien que la fusion de composantes connexes, elle n'impose donc aucune restriction sur la forme optimale. Les techniques comme les lignes de niveaux (OSHER et SETHIAN, 1988), le gradient topologique (SOKOLOWSKI et ZOCHOWSKI, 1999) ou encore les méthodes d'homogénéisation (KOHN et al., 1986a; KOHN et al., 1986b) appartiennent à cette catégorie.

L'objectif de cet ouvrage n'est pas de développer les mathématiques de l'optimisation de forme, ni de décrire de manière approfondie l'ensemble des techniques d'optimisation associées. Pour cela, le lecteur pourra se référer à des ouvrages tels que (SOKOLOWSKI et ZOLÉSIO, 1992; DELFOUR et al., 2001; HENROT et al., 2005; ALLAIRE, 2007; NOVOTNY et al., 2012) pour les aspects mathématiques et (OSHER et FEDKIW, 2003; SIARRY, 2014; BENDSØE et al., 2003) en ce qui concerne les méthodes numériques.

Notre propos est ici de donner un aperçu de ces méthodes et de leurs applications pour les sciences de l'ingénieur, en mettant l'accent sur quelques contributions récentes de l'IFSTTAR, du CEREMA et du CSTB. Cet ouvrage se veut autosuffisant et accessible au lecteur non spécialiste de l'optimisation de formes.

Le chapitre 1 propose une présentation des méthodes d'homogénéisation et de leur application à la conception de structures mécaniques. L'auteur s'intéresse notamment à leur mise en œuvre en présence d'incertitudes sur les propriétés intrinsèques des matériaux.

Le chapitre 2 est également consacré à un problème d'optimisation de structure. Il s'agit ici d'écrans antibruit, dont les caractéristiques sont déterminées par une approche paramétrique mettant en œuvre une technique d'optimisation métaheuristique.

Le chapitre 3, nous permet de montrer comment l'étude du comportement des alliages à mémoire de forme est liée à la théorie de l'optimisation de formes, ce qui permet d'estimer les différentes microstructures susceptibles de se former lors des transformations de phase solide/solide sous l'effet de sollicitations thermo-mécaniques.

Le chapitre 4 s'intéresse à l'utilisation de modèles déformables pour la segmentation d'images numériques. En d'autres termes, il s'agit d'optimiser des formes mathématiques pour capturer des objets ou des structures d'intérêt dans des images. Les auteurs dressent un état de l'art des méthodes d'optimisation récentes permettant d'atteindre l'optimum global dans certaines situations.

Enfin, le chapitre 5 est consacré à la reconstruction 3D par stéréovision. L'objectif est ici d'estimer, à partir d'images numériques, la description tridimensionnelle d'objets ou de scènes, sous forme d'une carte de profondeur. Les données et les critères optimisés étant discrets, le formalisme employé est celui des graphes, conduisant à des algorithmes d'optimisation rapides et souvent globaux.

## Bibliographie

- Grégoire Allaire.** « Introduction à l'optimisation de formes ». In : *Conception optimale de structures*. Tome 58. Mathématiques & applications. Springer, 2007, pages 1-20. DOI : 10.1007/978-3-540-36856-4\_1.
- Martin Bendsøe et Ole Sigmund.** *Topology Optimization : Theory, Methods and Applications*. Springer, 2003.
- Michel C. Delfour et Jean-Paul Zolésio.** *Shape and geometries : analysis, differential calculus and optimization*. Advances in design and control. SIAM, 2001.
- Antoine Henrot et Michel Pierre.** *Variation et optimisation de formes : Une analyse géométrique*. Mathématiques et Applications. Springer, 2005.
- Robert V. Kohn et Gilbert Strang.** « Optimal design and relaxation of variational problems, I,II ». In : *Communications on Pure and Applied Mathematics* 39 (1986), pages 113-182.
- Robert V. Kohn et Gilbert Strang.** « Optimal design and relaxation of variational problems, III ». In : *Communications on Pure and Applied Mathematics* 39 (1986), pages 353-377.
- Antonio A. Novotny et Jan Sokoowski.** *Topological Derivatives in Shape Optimization*. Interaction of Mechanics and Mathematics. Springer, 2012.
- Stanley Osher et Ronald Fedkiw.** *Level Set Methods and Dynamic Implicit Surfaces*. New York : Springer-Verlag, 2003.
- Stanley Osher et J.A. Sethian.** « Fronts propagating with curvature-dependant speed : algorithms based on Hamilton-Jacobi formulations ». In : *Journal of Computational Physics* 79.1 (nov. 1988), pages 12-49.
- Patrick Siarry.** *Métaheuristiques : Recuits simulé, recherche avec tabous, recherche à voisinages variables, méthodes GRASP, algorithmes évolutionnaires, fourmis artificielles, essais particuliers et autres méthodes d'optimisation*. Algorithmes. Eyrolles, 2014.
- Jan Sokolowski et Antoni Zochowski.** « On the topological derivative in shape optimization ». In : *SIAM Journal on Control and Optimization* 37 (1999), pages 1251-1272.
- Jan Sokolowski et Jean-Paul Zolésio.** *Introduction to shape optimization : shape sensitivity analysis*. Tome 16. Springer Series in Computational Mathematics. Berlin, Heidelberg, New York : Springer Verlag, juil. 1992.
- D'Arcy Wentworth Thompson.** *On Growth and Form*. Cambridge paperbacks. Cambridge University Press, 1992.

# Chapitre 1

## Méthodes de dimensionnement robuste en conception de structures

Michaël PEIGNEY<sup>1</sup>

*Résumé – Nous nous intéressons dans ce chapitre à deux problèmes types en conception de structures mécaniques.*

*Le premier problème est celui de l'optimisation topologique : à quantité de matière donnée, il s'agit de déterminer la forme optimale de la structure de façon à minimiser une fonction objectif (par exemple la souplesse).*

*Le second problème, moins étudié, est celui du dimensionnement robuste : pour une géométrie donnée, il s'agit de vérifier qu'une fonction critère (par exemple un critère de résistance) reste satisfait en présence d'incertitudes sur les paramètres des matériaux (ou sur le chargement).*

*Ce chapitre vise à mettre en évidence le lien entre ces deux problèmes, de nature différente au premier abord. En particulier, dans les deux cas surgissent des difficultés liées à l'absence de maximum global et la multiplicité de maxima locaux. On montre comment les idées et méthodes développées dans le cadre de l'optimisation topologique peuvent être mises à profit pour étudier le problème de dimensionnement robuste.*

---

1. IFSTTAR

## 1.1. Introduction

La conception de structures mécaniques (ponts pour citer un exemple en génie civil) met en jeu différentes classes de paramètres : paramètres géométriques reliés à la forme de la structure, paramètres des matériaux, paramètres reliés au chargement appliqué. Dans un problème de dimensionnement, certains paramètres (ceux reliés au chargement par exemple) sont supposés fixés, et il s'agit de déterminer les paramètres restants (matériau et géométrie par exemple), de façon à atteindre un certain niveau de performance spécifié dans le cahier des charges (par exemple, la contrainte en chaque point de la structure doit rester inférieure à un certain seuil fixé par la réglementation). Dans certains cas, la performance de la solution obtenue est très sensible aux variations des paramètres : une faible variation de certains paramètres (géométrie, matériau ou chargement) entraîne une dégradation significative de la performance, qui peut alors ne plus atteindre le niveau requis. Cette situation pose problème dès lors que les paramètres sont mal maîtrisés. De telles incertitudes peuvent affecter aussi bien les paramètres de géométrie, de matériau ou de chargement. Par exemple, les tolérances dans les procédés de fabrication sont à l'origine d'incertitudes sur les dimensions exactes de la structure. Une méconnaissance partielle du comportement du matériau ou une fluctuation spatiale de nature aléatoire sont parfois à prendre en compte. Le chargement est également soumis à des fluctuations qui ne sont pas entièrement contrôlées. Cet état de fait est à l'origine du concept de dimensionnement robuste, consistant à déterminer une solution dont la performance est peu sensible aux fluctuations des paramètres.

Pour préciser cette définition, il est nécessaire de préciser la façon dont sont modélisées les fluctuations des dits paramètres. À cet égard, les approches probabilistes (ou stochastiques), consistant à décrire les fluctuations par des lois de probabilités, sont parmi les plus répandues (LEMAIRE, 2010). Dans un tel cadre, l'objectif de dimensionnement robuste consiste par exemple à garantir que la probabilité d'atteindre la performance ciblée est supérieure à un certain seuil prédéfini (proche de 1). D'autres approches sont envisageables. Par exemple, la théorie du calcul à la rupture - dont la pertinence pour le génie civil n'est plus à prouver - peut être interprétée comme une méthode de dimensionnement robuste (l'incertitude pris en compte affecte dans ce cas les propriétés matériaux) (SALENÇON, 1983). Une approche intermédiaire, dont la dénomination ne fait pas encore l'unanimité, consiste à décrire les incertitudes par le biais de bornes sur les paramètres (ELISHAKOFF et al., 2010). Cette approche (rencontrée dans la littérature sous différentes appellations comme par exemple, «performance garantie») a essentiellement été utilisée sur des structures à nombre fini de degrés de liberté. L'extension au milieu continu, peu étudiée, fait l'objet de ce chapitre. Un objectif est de mettre en évidence les liens avec l'optimisation topologique en conception de structures. Les méthodes et connaissances développées dans ce cadre peuvent alors être mises à profit pour aborder le problème de dimensionnement robuste en milieu continu.

Ce chapitre est organisé comme suit : après une formalisation du problème de dimensionnement, une présentation de l'optimisation topologique en conception

de structures est proposée. Les deux principales méthodes (méthodes d'homogénéisation et méthode SIMP) sont décrites, illustrées et comparées sur quelques exemples. La prise en compte des incertitudes est abordée quant à elle en 1.4. L'approche du dimensionnement robuste (au sens d'une performance garantie) est appliquée au problème de référence d'une poutre hétérogène en flexion. Cette étude, à la fois analytique et numérique, permet de faire le lien avec l'optimisation topologique.

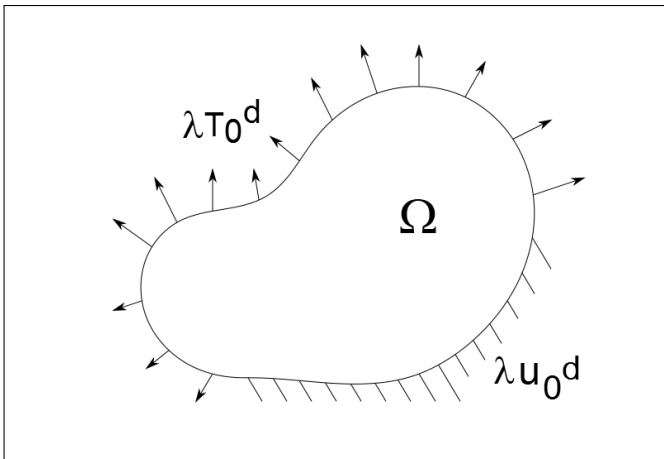
## 1.2. Problème de dimensionnement

### 1.2.1. Formulation générale

Considérons une structure occupant un domaine  $\Omega \subset \mathbb{R}^3$ , soumise à un chargement représenté par des forces surfaciques  $\lambda T_0^d$  appliquées sur une partie  $\Gamma_T \subset \partial\Omega$  de la surface, des déplacements  $\lambda u_0^d$  imposés sur  $\partial\Omega - \Gamma_U$ , et des forces volumiques  $\lambda f_0^d$  (figure 1.1). Les fonctions  $T_0^d$ ,  $u_0^d$ ,  $f_0^d$  sont données et  $\lambda \geq 0$  est un paramètre de chargement.

#### Figure 1.1.

Chargement mécanique pour le problème de dimensionnement



Pour simplifier la présentation, on suppose que le matériau est élastique linéaire, i.e. obéit localement à la loi de comportement

$$\sigma(x) = \mathbf{L}(x) : e(x) \quad (1.1)$$

où  $\sigma(x)$  est le tenseur des contraintes au point  $x$ ,  $e(x)$  est le tenseur des déformations et  $\mathbf{L}(x)$  est le tenseur d'élasticité. Ici et dans la suite, la notation « : » désigne le produit doublement contracté. En représentation matricielle, la relation (1.1) s'écrit  $\sigma_{ij} = \sum_{k,l=1}^3 L_{ijkl} e_{kl}$ . Sous l'hypothèse des petites perturbations, la

déformation  $e(\boldsymbol{x})$  est reliée au champ de déplacement  $\boldsymbol{u}(\boldsymbol{x})$  par

$$e(\boldsymbol{x}) = \frac{1}{2}(\nabla \boldsymbol{u} + \nabla^T \boldsymbol{u}). \quad (1.2)$$

Le champ de déplacement vérifie d'autre part la condition aux limites

$$\boldsymbol{u} = \lambda \boldsymbol{u}_0^d \text{ sur } \Gamma_u. \quad (1.3)$$

À l'équilibre, le champ de contraintes  $\boldsymbol{\sigma}$  vérifie

$$\operatorname{div} \boldsymbol{\sigma} + \lambda \boldsymbol{f}_0^d = 0 \text{ dans } \Omega, \quad \boldsymbol{\sigma} \cdot \boldsymbol{n} = \lambda \boldsymbol{T}_0^d \text{ sur } \Gamma_T. \quad (1.4)$$

Pour une répartition de matériau  $\boldsymbol{L}(\boldsymbol{x})$  donnée, le problème d'élasticité linéaire formé par les équations (1.1-1.4) admet une solution unique en contraintes et en déformations. Cette solution dépend linéairement du paramètre de chargement  $\lambda$ . On note  $\boldsymbol{\sigma}_0$  la solution en contraintes du problème (1.1-1.4) correspondant à  $\lambda = 1$ .

Dans un problème de dimensionnement, il est en général nécessaire de vérifier que le champ  $\boldsymbol{\sigma}$  obtenu respecte localement un critère lié à la résistance du matériau. Un tel critère est exprimé sous la forme  $g(\boldsymbol{\sigma}) \leq 0$  où  $g$  est une fonction scalaire. Un exemple de critère est exposé dans le paragraphe suivant.

### 1.2.2. Exemple d'un critère à la rupture

Pour les matériaux métalliques, un critère couramment utilisé est celui de Von Mises, donné par

$$g(\boldsymbol{\sigma}) = \sqrt{\frac{3}{2} \boldsymbol{\sigma}' : \boldsymbol{\sigma}' - \sigma_Y} \quad (1.5)$$

où  $\boldsymbol{\sigma}' = \boldsymbol{\sigma} - \frac{1}{3}(\operatorname{tr} \boldsymbol{\sigma})\boldsymbol{I}$  est le déviateur de  $\boldsymbol{\sigma}$  et  $\sigma_Y$  un paramètre caractéristique du matériau (contrainte maximale en traction). L'inégalité  $g(\boldsymbol{\sigma}) \leq 0$  doit être satisfaite en tout point. La condition exprimant la tenue de la structure peut donc se réécrire sous la forme

$$G(\Omega, \boldsymbol{L}) \leq 0 \quad (1.6)$$

où l'on a posé

$$G(\Omega, \boldsymbol{L}) = \sup_{\boldsymbol{x} \in \Omega} g(\boldsymbol{\sigma}). \quad (1.7)$$

Le scalaire  $G(\Omega, \boldsymbol{L})$  dépend de  $\Omega$  et  $\boldsymbol{L}$ , mais aussi du paramètre de chargement  $\lambda$ . Plus précisément, à  $(\Omega, \boldsymbol{L})$  donnés, la condition (1.6) est satisfaite pour tout  $\lambda \leq \lambda^{max}$  où  $\lambda^{max}$  (éventuellement nul) s'interprète comme le chargement maximum que peut supporter la structure. Il est possible d'expliciter le chargement  $\lambda^{max}$  si l'on suppose que  $\sigma_Y$  prend la même valeur en tout point : étant donné que la solution du problème (1.1-1.4) dépend linéairement de  $\lambda$ , on a  $g(\boldsymbol{\sigma}) = \lambda f(\boldsymbol{\sigma}_0) - \sigma_Y$  où

$$f(\boldsymbol{\sigma}) = \sqrt{\frac{3}{2} \boldsymbol{\sigma}' : \boldsymbol{\sigma}'} \quad (1.8)$$



et  $\sigma_0$  correspond aux champs de contrainte solution de (1.1-1.4) pour  $\lambda = 1$ . Par conséquent

$$G(\Omega, \mathbf{L}) = \lambda F(\Omega, \mathbf{L}) - \sigma_Y$$

où l'on a posé

$$F(\Omega, \mathbf{L}) = \sup_{\boldsymbol{\sigma} \in \Omega} f(\boldsymbol{\sigma}_0). \quad (1.9)$$

La condition (1.6) est donc vérifiée pour tout  $\lambda \leq \lambda^{max}$  avec

$$\lambda^{max} = \frac{\sigma_Y}{F(\Omega, \mathbf{L})}. \quad (1.10)$$

### 1.2.3. Optimisation en dimensionnement

Dans une démarche d'optimisation, le concepteur joue sur le matériau et/ou la géométrie de façon à minimiser une fonction objectif (par exemple le coût ou le volume de la structure), tout en vérifiant la condition (1.6). Le problème à résoudre prend ainsi la forme générique

$$\inf_{\substack{(\Omega, \mathbf{L}) \in \Sigma \\ G(\Omega, \mathbf{L}) \leq 0}} J(\Omega, \mathbf{L}) \quad (1.11)$$

où  $J$  est la fonction objectif et  $\Sigma$  désigne une classe de valeurs admissibles (choisie à l'avance) sur laquelle s'effectue l'optimisation.

Nous pouvons distinguer différentes classes de problèmes selon la dimension de  $\Sigma$ . Si l'ensemble  $\Sigma$  est de dimension finie, on parle d'optimisation *paramétrique*. Si l'ensemble  $\Sigma$  est de dimension infinie, on parle d'optimisation *géométrique* ou *topologique*, selon que les changements de topologie (création de trous) sont pris en compte ou non.

### 1.2.4. Un exemple d'optimisation géométrique

Afin d'illustrer ces considérations par un exemple concret, étudions le dimensionnement d'une poutre en flexion pure. Le matériau constitutif est supposé homogène, élastique linéaire isotrope de module d'Young  $E$ . En considérant un repère orthonormal  $(\mathbf{u}_x, \mathbf{u}_y, \mathbf{u}_z)$  où  $\mathbf{u}_x$  est selon l'axe de la poutre, le champ de contraintes  $\boldsymbol{\sigma}$  est donné par  $\boldsymbol{\sigma} = -(M/I)y\mathbf{u}_x \otimes \mathbf{u}_x$  où  $M$  est le moment de flexion appliqué et  $I = \int_S y^2 dS$  est le moment d'inertie selon  $\mathbf{u}_y$ . Le critère de Von Mises donne alors  $f(\boldsymbol{\sigma}) = |My/I|$  et donc  $F(\Omega, E) = M(\sup_S y)/I$ .

Pour une géométrie donnée, la relation (1.10) donne le chargement maximal que la structure peut supporter. En l'occurrence la valeur maximale  $M^{max}$  du moment est donnée par  $M^{max} = I\sigma_Y/(\sup_S y)$ .

Le matériau constitutif étant donné, supposons qu'on cherche à minimiser la section, en se restreignant dans un premier temps à l'ensemble  $\Sigma$  formé par les sections rectangulaires  $([-h, h] \times [-H, H])$ . On a alors  $\sup_S y = h$  et  $I = 2Hh^3/3$ , donc  $M^{max} = (2/3)\sigma_Y Hh^2$ . Si la poutre est conçue pour supporter un moment

$M$  donné, il faut que les paramètres  $H$  et  $h$  soient tels que

$$M \geq \frac{2}{3} \sigma_Y H h^2.$$

À  $h$  donné, la surface de la section est minimale pour  $H = 3M/(2\sigma_Y h^2)$ .

On s'est restreint ici à des sections rectangulaires, paramétrées par 2 scalaires. Des formes plus complexes, paramétrées par  $n$  scalaires, peuvent être considérées de la même façon. Nous pouvons ainsi utiliser par exemple des splines fermées, c'est à dire des courbes  $\Gamma : [0, 1] \mapsto \mathbb{R}^2$  polynomiales par morceaux et telles que  $\Gamma(0) = \Gamma(1)$  (voir par exemple (CHRISTENSEN et al., 2009) pour plus de détails sur cette approche). Le problème (1.11) se traduit alors par un problème d'optimisation (sous contraintes) dans  $\mathbb{R}^n$ . Le cas extrême de cette démarche consiste à considérer toutes les courbes fermées possibles : il s'agit de l'optimisation géométrique (SOKOLOWSKI et al., 1992). L'ensemble  $\Sigma$  correspondant, de dimension infinie, est par exemple

$$\Sigma = \{\Gamma \in C^1([0, 1], \mathbb{R}^2) | \Gamma(0) = \Gamma(1)\}.$$

D'autres choix sont possibles en changeant la régularité (ici  $C^1$ ) imposée aux courbes  $\Gamma$ . Notons que les sections considérées sont nécessairement pleines (connexes). Nous pouvons éventuellement considérer une section perforée (non connexe) en rajoutant une deuxième (ou plus) courbe dans le paramétrage pour décrire la géométrie d'un (ou plus) trou intérieur.

Ce type d'approche ne permet pas la création ou la suppression de trous au cours de l'optimisation : leur nombre en est fixé à l'avance via le nombre de courbes utilisées dans la paramétrisation. Cette propriété reflète le fait que l'optimisation géométrique n'autorise pas le changement de topologie. Les méthodes d'optimisation dite topologique, présentées dans la suite, permettent de tels changements.

### 1.3. L'optimisation topologique

Le principe de l'optimisation topologique est de considérer l'ensemble de tous les domaines possibles (inclus dans un domaine de référence  $\Omega_0$ ), sans restriction de régularité spatiale ou de connexité. Pour un tenseur l'élasticité  $\mathbf{L}$  fixé, l'ensemble  $\Sigma$  est ainsi pris sous la forme

$$\Sigma = \{(\mathbf{L}, \Omega) | \Omega \subset \Omega_0\}. \quad (1.12)$$

Soit  $\chi$  la fonction caractéristique du domaine  $\Omega$ , définie par  $\chi(\mathbf{x}) = 1$  si  $\mathbf{x} \in \Omega$ , et  $\chi(\mathbf{x}) = 0$  sinon. L'ensemble  $\Sigma$  peut être réécrit sous la forme

$$\Sigma = \{(\mathbf{L}(\mathbf{x}) = \chi(\mathbf{x})\mathbf{L}, \Omega_0) | \chi : \Omega_0 \mapsto \{0, 1\}\}. \quad (1.13)$$

Dans l'écriture (1.12), le domaine  $\Omega$  est variable, et  $\mathbf{L}$  est fixe. À l'inverse, dans l'écriture (1.13), le domaine  $\Omega_0$  est fixe, et le champ de modules élastiques est variable et inhomogène. Le problème géométrique d'optimisation de forme est ainsi converti en un problème de répartition optimale de matière dans un domaine fixe. Il ne s'agit pas uniquement d'un jeu de réécriture formelle : l'interprétation du problème d'optimisation topologique comme un problème de distribution de matière s'avère fructueuse, permettant notamment d'utiliser la théorie de l'homogénéisation de matériaux et structures hétérogènes.

Ce type d'approche pour l'optimisation topologique est bien maîtrisée dans le cas où

$$J(\Omega, \mathbf{L}) = \int_{\Omega_0} \frac{1}{2} \boldsymbol{\sigma}(\mathbf{x}) : \mathbf{L}^{-1}(\mathbf{x}) : \boldsymbol{\sigma}(\mathbf{x}) d\mathbf{x} \quad (1.14)$$

et la condition de résistance  $G(\Omega, \mathbf{L}) \leq 0$  est remplacée par une contrainte de la forme

$$\int_{\Omega_0} \chi(\mathbf{x}) d\mathbf{x} = c|\Omega_0| \quad (1.15)$$

avec  $0 < c \leq 1$  donné. La contrainte (1.15) exprime le fait que le volume de la structure est fixé (et égal à  $c|\Omega_0|$ ). La fonction  $J$  définie par (1.14) est la complaisance (*compliance* en anglais) de la structure et en mesure la souplesse vis-à-vis du chargement considéré. Le problème (1.11) se réécrit alors

$$\min_{(\chi, \mathbf{L}) \in \Sigma} J(\Omega, \mathbf{L}) \quad (1.16)$$

$$\int_{\Omega_0} \chi = c|\Omega_0|$$

où

$$\Sigma = \{(\chi, \mathbf{L}(\mathbf{x})) | \mathbf{L}(\mathbf{x}) = \chi(\mathbf{x})\mathbf{L}; \chi : \Omega_0 \mapsto \{0, 1\}\}.$$

Ce problème consiste à chercher la structure de volume  $c|\Omega_0|$  qui soit la plus rigide possible pour le chargement considéré.

En général, le problème (1.16) n'a pas de solution : les suites minimisantes  $\chi_n$  présentent des oscillations à une longueur caractéristique  $l_n$  tendant vers 0 (quand  $n \rightarrow +\infty$ ) et ne convergent pas (au sens classique) vers une distribution optimale. Ceci correspond physiquement à la formation de microstructures à une échelle infiniment fine : ce phénomène est en tout point similaire à ce qui est présenté au chapitre 3 sur l'étude de microstructures optimales dans les matériaux à transformation de phase.

Ces considérations incitent à remplacer (1.16) par une formulation relaxée de la forme

$$\min_{(\chi, \mathbf{L}) \in \Sigma} J(\Omega, \mathbf{L}) \quad (1.17)$$

$$\int_{\Omega_0} \chi = c|\Omega_0|$$

où

$$\Sigma = \{(\rho, \mathbf{L}(\mathbf{x})) | \mathbf{L}(\mathbf{x}) \in \Lambda(\rho(\mathbf{x})); \chi : \Omega_0 \mapsto [0, 1]\}$$

et  $\Lambda$  désigne l'ensemble des tenseurs d'élasticité compatibles avec une densité  $\rho(x)$ . La variable  $\rho(x)$  est à valeurs dans l'intervalle  $[0, 1]$  et s'interprète comme une densité locale de matière. Toute la difficulté réside ainsi dans la construction de l'ensemble  $\Lambda$ . Deux principales méthodes ont été proposées à ce sujet : la méthode SIMP (*Solid Isotropic Material with Penalization*) et la méthode d'homogénéisation. Ces deux méthodes sont brièvement exposées dans la suite. On pourra se référer respectivement aux ouvrages d'Allaire (ALLAIRE, 2002), et de Bendsøe et Sigmund (BENDSØE et al., 2003) pour une présentation plus détaillée.

### 1.3.1. La méthode d'homogénéisation

#### 1.3.1.1. Formulation

L'idée principale de la méthode d'homogénéisation est de considérer l'ensemble  $\Lambda(\rho)$  formé par les tenseurs d'élasticité effective de tous les matériaux composites ayant une densité  $\rho$ . Par matériau composite on entend ici un matériau obtenu par micro-perforation du milieu initial de module  $\mathbf{L}$ . Définissons l'ensemble  $\Lambda(\rho)$  de façon plus précise : soit  $\tilde{\Omega}$  un domaine de référence, de volume unitaire. La microstructure d'un composite de densité  $\rho$  est définie par une fonction caractéristique  $\tilde{\chi} : \tilde{\Omega} \mapsto \{0, 1\}$  telle que  $\int_{\tilde{\Omega}} \tilde{\chi} = \rho$ . Pour toute déformation  $\bar{e}$  donnée, le problème d'élasticité linéaire défini par

$$\begin{cases} \mathbf{u} = \bar{e} \cdot \mathbf{x} & \text{sur } \partial\tilde{\Omega}, \\ \operatorname{div} \boldsymbol{\sigma} = 0 & \text{dans } \tilde{\Omega}, \\ \boldsymbol{\sigma}(\mathbf{x}) = \tilde{\chi}(\mathbf{x}) \mathbf{L} : \mathbf{e}(\mathbf{x}) & \text{dans } \tilde{\Omega}, \end{cases}$$

admet une solution unique, qui dépend linéairement de  $\bar{e}$ .

Il existe donc un opérateur  $\mathbf{L}(\tilde{\chi})$  tel que, pour tout  $\bar{e}$

$$\int_{\tilde{\Omega}} \boldsymbol{\sigma}(\mathbf{x}) d\mathbf{x} = \mathbf{L}(\tilde{\chi}) : \bar{e}.$$

Cet opérateur  $\mathbf{L}(\tilde{\chi})$  s'interprète comme le tenseur d'élasticité effectif du matériau composite associé à la microstructure  $\tilde{\chi}$ . L'ensemble  $\Lambda(\rho)$  est alors défini par

$$\Lambda(\rho) = \{ \mathbf{L}(\tilde{\chi}) \mid \tilde{\chi} : \tilde{\Omega} \mapsto \{0, 1\}; \int_{\tilde{\Omega}} \tilde{\chi} = \rho \}. \quad (1.18)$$

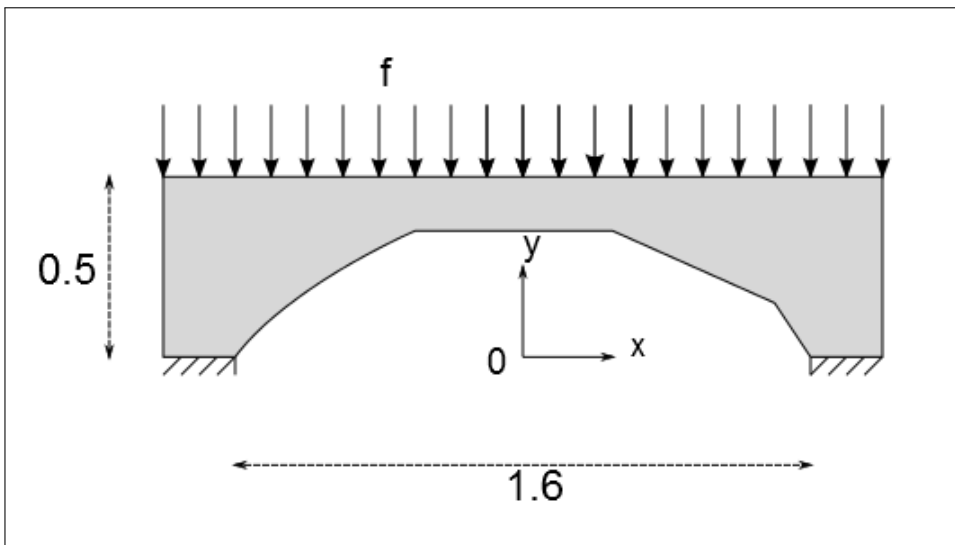
Avec ce choix de  $\Lambda(\rho)$ , on peut alors montrer que le problème (1.17) admet une solution. Une difficulté notable est que, dans le cas général, on ne dispose pas de l'expression explicite de l'ensemble défini par (1.18). Dans le cas de la minimisation de la complaisance, il est montré cependant qu'on peut, sans perte de généralité, remplacer (1.18) par le sous-ensemble explicite associé aux microstructures laminées (une présentation des microstructures laminées est donnée au chapitre 3).

### 1.3.1.2. Exemple

Considérons un problème bidimensionnel inspiré de la conception d'un pont (figure 1.2). On cherche à optimiser la forme d'un pont permettant de franchir une distance  $l_0 = 1.6$  et supportant un chargement représenté par une force linéique  $f = -\bar{e}_y$  appliquée en  $y = 0.5$ ,  $|x| \leq 1$  (toutes les forces et longueurs sont exprimées sous forme adimensionnelle). Le volume occupé par le pont est fixé à 0.25. On utilise la méthode d'homogénéisation pour chercher la forme de rigidité maximale vis-à-vis du chargement considéré. Pour ce faire, le domaine  $\Omega_0$  est choisi égal au rectangle  $-1 \leq x \leq 1$ ,  $0 \leq y \leq 1$ . Les conditions aux limites sont de type encastrement (déplacement nul) en  $0.8 \leq |x| \leq 1$ ,  $y = 0$ , et de type surface libre sur le reste de  $\partial\Omega_0$ .

**Figure 1.2.**

Recherche d'un pont optimal. Positionnement du problème



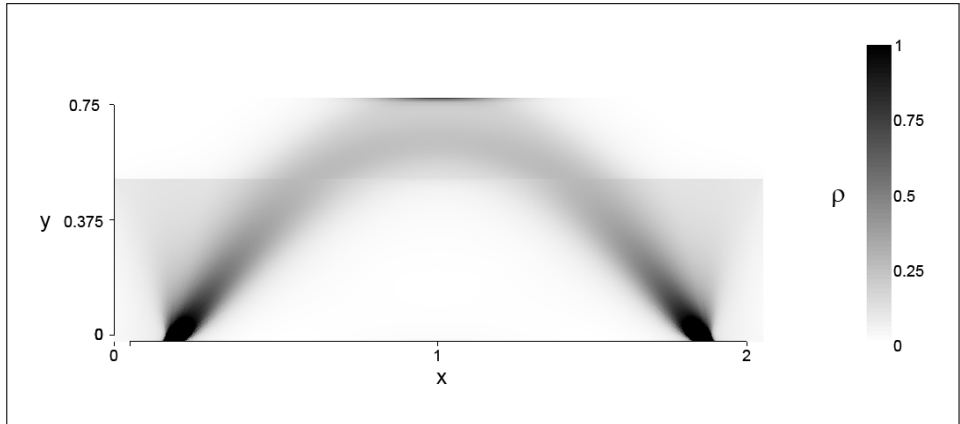
Le problème relaxé (1.17) est résolu à l'aide d'une discrétisation spatiale par éléments finis, couplé à un algorithme de minimisation alternée (voir par exemple (ALLAIRE et PANTZ, 2006) pour plus de détails).

La figure 1.3 montre la carte de densités  $\rho(x)$  solution du problème relaxé (1.17). Pour réaliser en pratique la distribution de matière représentée figure 1.3, il faut imaginer une distribution de matériau composite (laminés), dont la microstructure évolue de façon continue en fonction de la position.

Il est clair qu'une telle structure est difficile à réaliser en pratique. C'est pourquoi, en vue d'obtenir une forme facilement réalisable en pratique, un post-traitement est appliqué à la solution  $\rho(x)$  du problème relaxé, de façon à filtrer les densités intermédiaires et construire une distribution  $\bar{\rho}$  à valeurs dans  $\{0, 1\}$ .

**Figure 1.3.**

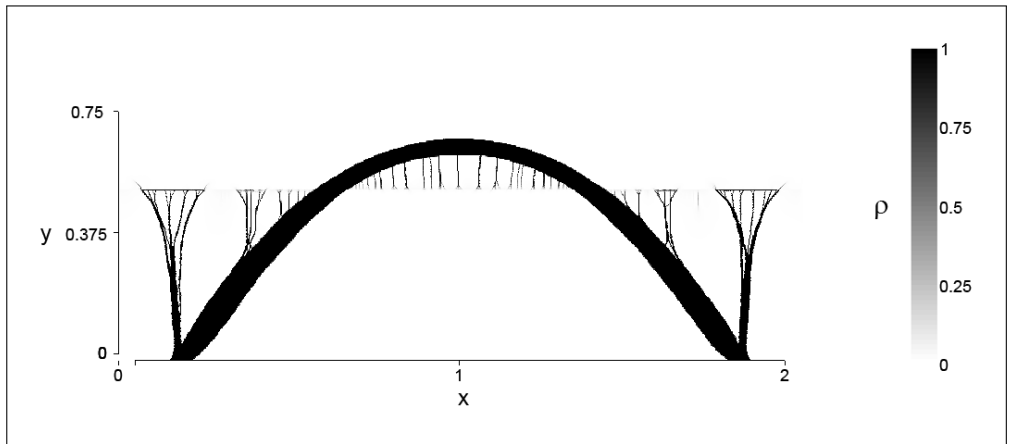
Recherche d'un pont optimal. Solution relaxée  $\rho$  obtenue par la méthode d'homogénéisation



La figure 1.4 montre la carte de densités  $\tilde{\rho}$  obtenue après filtrage. Ces résultats numériques ont été obtenus sous Freefem++ (PIRONNEAU et al., 2009), en utilisant la démarche exposée dans (ALLAIRE et PANTZ, 2006).

**Figure 1.4.**

Recherche d'un pont optimal. Solution  $\tilde{\rho}$  obtenue par homogénéisation, après filtrage



Il est important de souligner que ce filtrage dégrade nécessairement la solution : la complaisance associée à la distribution  $\tilde{\rho}$  est égale à 21,5, alors que la complaisance associée à la distribution  $\rho$  est égale à 18,4. De plus, la solution obtenue après filtrage est non unique et dépend notamment du maillage utilisé.

## 1.3.2. La méthode SIMP

### 1.3.2.1. Formulation

Dans la méthode SIMP, l'ensemble  $\Lambda(\rho)$  est choisi égal à

$$\Lambda(\rho) = \{\rho^p \mathbf{L}\}, \quad (1.19)$$

où  $p > 1$  est un exposant arbitraire.

La relation  $\mathbf{L}(\mathbf{x}) = \chi(\mathbf{x})\mathbf{L}$  du problème original est ainsi remplacée par

$$\mathbf{L}(\mathbf{x}) = \rho^p(\mathbf{x})\mathbf{L}, \quad (1.20)$$

où  $p$  est typiquement choisi assez grand pour pénaliser les densités intermédiaires (c'est-à-dire différentes de 0 et 1). Dans son esprit, cette approche est voisine de certaines techniques de pénalisation utilisées dans d'autres problèmes de mécanique non-différentiable (comme par exemple l'approche de Oden-Martins (ODEN et al., 1985) en mécanique du contact). Malgré son caractère différentiable, la résolution du problème (1.17) obtenue par la méthode SIMP présente certaines difficultés notables.

La principale est que, sans précaution supplémentaire, le problème (1.17) n'a pas de solution optimale : comme pour (1.16), les suites minimisantes présentent des oscillations à l'échelle d'une longueur caractéristique qui devient infiniment fine. Il est nécessaire d'introduire une échelle de longueur (choisie arbitrairement) dans le problème, ce qui exclue la formation de microstructures. Différentes techniques sont possibles pour introduire cette échelle de longueur. Une solution est d'imposer une condition de régularité aux fonctions  $\rho(\mathbf{x})$  qui apparaissent dans (1.17), de la forme

$$\int_{\Omega_0} \rho^2 + \|\nabla \rho\|^2 d\mathbf{x} \leq l, \quad (1.21)$$

où  $l$  est une constante donnée. Dans le cas tridimensionnel, on peut alors montrer l'existence de solution au problème (1.17) si  $p < 3$ .

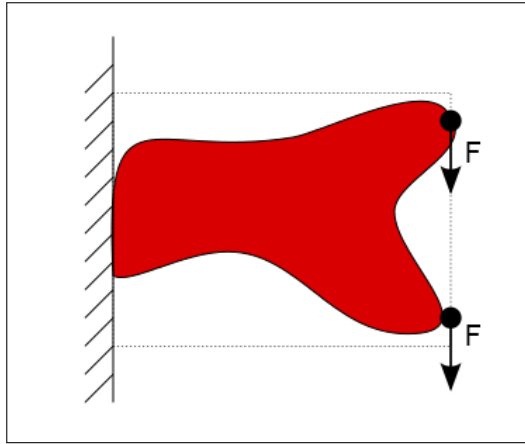
### 1.3.2.2. Exemple

Considérons l'exemple bidimensionnel d'une poutre console encastrée sur son extrémité gauche (figure 1.5).

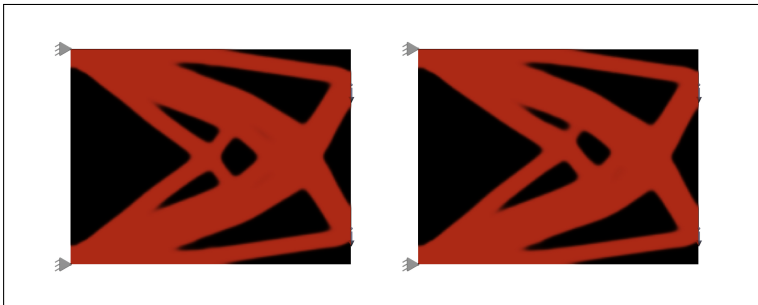
Le domaine  $\Omega_0$  est défini par  $0 \leq x \leq 1, 0 \leq y \leq 3/4$ . Le chargement est constitué de deux forces ponctuelles (égales et orientées selon  $\mathbf{u}_y$ ), appliquées en  $(1/8, 1)$  et  $(5/8, 1)$ . Pour ces conditions aux limites, on cherche la forme optimale  $\Omega \subset \Omega_0$  maximisant la rigidité, sous la contrainte  $|\Omega| = c|\Omega_0|$ . On choisit dans la suite  $c = 0.58$ .

La figure 1.6 montre les résultats obtenus par la méthode SIMP, après filtrage. Les deux champs  $\rho(\mathbf{x})$  représentés correspondent aux solutions obtenues partant de deux états initiaux différents. Ceci illustre la multiplicité des optima locaux au problème. Ces résultats numériques ont été obtenus à l'aide du logiciel (gratuit) Topopt (AAGE et al., 2013).

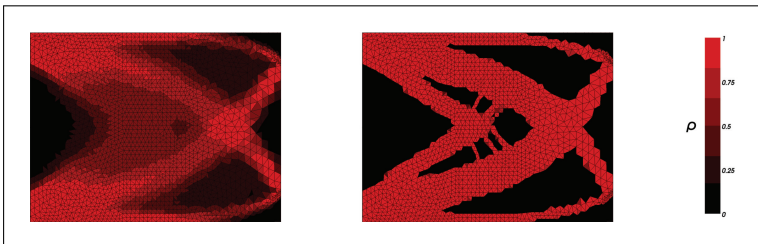
**Figure 1.5.**  
Optimisation d'une poutre console



**Figure 1.6.**  
Solutions obtenues par la méthode SIMP, pour deux états initiaux différents.  
Maillage à 4586 éléments



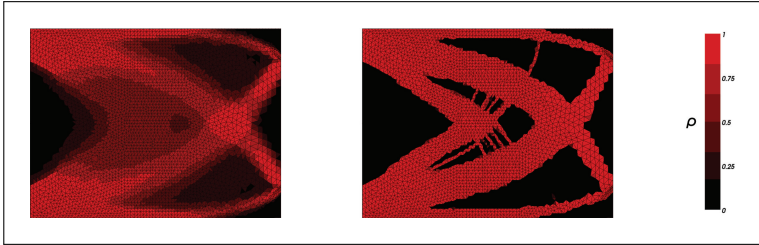
**Figure 1.7.**  
Solutions obtenues par la méthode d'homogénéisation, avant (gauche) et après (droite) filtrage. Maillage à 4586 éléments



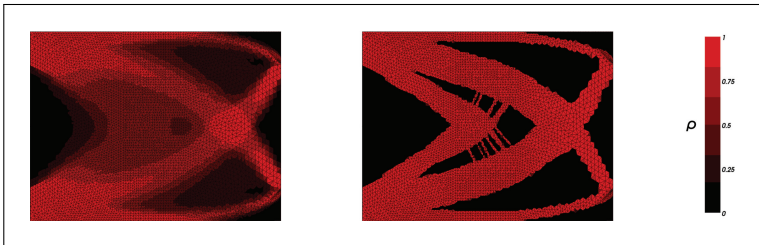


**Figure 1.8.**

Solutions obtenues par la méthode d'homogénéisation, avant (gauche) et après (droite) filtrage. Maillage à 8201 éléments

**Figure 1.9.**

Solutions obtenues par la méthode d'homogénéisation, avant (gauche) et après (droite) filtrage. Maillage à 12913 éléments



### 1.3.3. Synthèse sur les méthodes d'optimisation topologique

Les deux approches présentées reposent sur des philosophies différentes : alors que la méthode SIMP est essentiellement guidée par des arguments mathématiques de régularisation, la méthode d'homogénéisation repose sur une interprétation plus physique en termes de matériaux composites.

Dans certains cas, comme le problème de la poutre console, les deux méthodes donnent des résultats similaires. À ce sujet, on pourra comparer la figure 1.6 (méthode SIMP) avec la figure 1.7, qui montre les résultats obtenus avec la méthode d'homogénéisation (les maillages utilisés dans les deux calculs sont similaires).

On peut se demander s'il existe un lien entre les deux méthodes, et plus particulièrement si la méthode SIMP peut être interprétée dans le cadre de l'homogénéisation. La question est alors de savoir s'il existe un matériau composite de densité  $\rho$  dont le tenseur d'élasticité effectif est égal à  $\rho^p \mathbf{L}$ . La réponse est négative dans le cas général. Cependant, dans le cas où  $\mathbf{L}$  est isotrope (de module d'Young  $E$  et de coefficient de Poisson  $\nu_0$ ), il est montré (BENDSØE et al., 1999) qu'une telle microstructure existe dans le cas bidimensionnel si

$$p \geq \max\left\{\frac{2}{1-\nu_0}, \frac{4}{1+\nu_0}\right\} \quad (1.22)$$

et dans le cas tridimensionnel si

$$p \geq \max\left\{15 \frac{1 - \nu_0}{7 - 5\nu_0}, \frac{3}{2} \frac{1 - \nu_0}{1 - 2\nu_0}\right\}. \quad (1.23)$$

En termes de mise en oeuvre, la méthode SIMP paraît plus simple à première vue car l'ensemble  $\Lambda(\rho)$  est alors réduit à un singleton. Cette apparente simplicité doit être nuancée par le fait que, pour garantir l'existence d'une solution dans la méthode SIMP, il est nécessaire d'implémenter une condition supplémentaire de régularité (par exemple la condition (1.21)), ce qui n'est pas nécessaire dans la méthode d'homogénéisation. L'un dans l'autre, la mise en oeuvre des approches présente une complexité similaire.

Listons quelques remarques d'ensemble, valable pour les deux méthodes :

**Implémentation numérique :** Le problème (1.17) est généralement discrétisé en espace (par une méthode éléments finis) et résolu de façon itérative. Les problèmes obtenus sont de *grande taille* et relativement *coûteux en temps de calcul*; l'inconnue principale est un champ défini sur  $\Omega_0$  et à chaque itération, il faut recalculer la matrice de rigidité du système.

**Existence de solution optimale :** Le problème relaxé (1.17) fourni par la méthode d'homogénéisation admet toujours une solution. Ce n'est le cas pour la méthode SIMP que lorsque l'exposant  $p$  est suffisamment petit. Soulignons que, pour les deux méthodes, ces résultats d'existence sont étroitement liées au choix de  $J$  comme complaisance du système et ne s'étendent pas de façon systématique à d'autres choix de fonction  $J$  (ni à la prise en compte d'autres contraintes  $g_\alpha$ ).

**Multiplicité de minima locaux :** L'existence d'un minimum global au problème relaxé (1.17) n'exclue pas la présence de minima locaux. Tant pour la méthode SIMP que pour la méthode d'homogénéisation, on constate en pratique que les minima locaux sont nombreux. Ceci se manifeste par une forte dépendance de la solution obtenue par rapport à l'état initial (utilisé pour démarrer la minimisation). Il est en général difficile de garantir que la solution obtenue corresponde à un optimum global et non à un optimum seulement local.

**Dépendance au maillage :** Ce point est illustré sur les figures 1.7 à 1.9 qui, pour le problème de la poutre console, montrent les résultats obtenus par la méthode d'homogénéisation avec trois maillages de raffinement croissant. Sur chacune de ces figures, on représente à gauche le champ de densité  $\rho(x)$  solution du problème relaxé avant filtrage, et à droite le champ obtenu après filtrage. On observe que la solution du problème relaxé semble converger quand le maillage se raffine. Il n'en va pas de même pour les solutions filtrées; raffiner le maillage modifie sensiblement les détails de la structure obtenue, sans convergence apparente. Un même constat s'applique pour la méthode SIMP. Cette dépendance de la solution au maillage est liée à la multiplicité de minima locaux évoquée précédemment.

**Prise en compte de critère local sur  $\sigma$**  : Pour le dimensionnement de structures, il est souvent important de tenir compte d'un critère local sur  $\sigma$  (exprimant la condition de résistance du matériau), comme on l'a expliqué en début de ce chapitre. Un tel critère rend les difficultés (déjà substantielles) du problème d'optimisation topologique encore plus aiguës. Le critère de résistance étant une fonction locale de  $\sigma$ , il se traduit numériquement par un grand nombre de contraintes à respecter et rend donc le problème encore plus coûteux en temps de calcul. De plus, la présence d'une critère de résistance tend à multiplier les minima locaux et à bloquer les changements de topologie (i.e, en 2D, le nombre de trous) dans la minimisation. Afin d'illustrer ce phénomène, connu sous le nom de *phénomène de singularité* (KIRSCH, 1990), considérons que l'état optimal (s'il existe)  $\rho_0$  ne présente qu'un seul trou, et considérons une distribution  $\rho_1$  avec une topologie différente, comportant 2 trous. Soit  $\rho(\theta)$  une courbe reliant continûment  $\rho_0$  et  $\rho_1$  (avec  $\theta \in [0, 1]$ ,  $\rho(0) = \rho_0$ ,  $\rho(1) = \rho_1$ ). La distribution  $\rho(\theta)$  présente typiquement une zone dont l'épaisseur tend vers 0 quand  $\theta$  tend vers 0. Dans cette zone,  $\sigma$  a alors tendance à devenir très grand quand  $\theta$  tend vers 0, et le critère de résistance n'est alors plus satisfait. Ceci explique que, dans un schéma de minimisation, il est difficile de passer d'un état  $\rho_1$  à  $\rho_0$ ; le critère de résistance bloque le changement de topologie.

Chercher à surmonter ces difficultés fait l'objet d'une attention constante dans la communauté de l'optimisation topologique. Sans chercher à être exhaustif, citons par exemple l'approche introduite dans (DUYSINX et al., 1998) qui consiste à remplacer le critère local sur  $\sigma$  par un critère global. Bien qu'efficace en termes de coûts de calcul, une telle approche a l'inconvénient de ne pas être sensible aux concentrations locales de contraintes, souvent observées dans les calculs de structure (voir par exemple la figure 1.16 dans ce chapitre). Afin d'atténuer le phénomène de singularité, une stratégie proposée est de remplacer le critère  $g(\sigma)$  par une forme étendue  $g(\sigma, \rho)$  telle que  $g$  tend vers 0 quand  $\rho$  tend vers 0 (GENG DONG et al., 1997). Le choix de la fonction  $g(\sigma, \rho)$  est une question sensible, et le phénomène de singularité, bien qu'atténué, persiste dans certains cas.

On pourra notamment consulter les références (BRUGGI et al., 2012; HOLMBERG et al., 2013) pour quelques avancées relativement récentes sur ces questions. Malgré ces progrès, la prise en compte de critères locaux sur  $\sigma$  reste un défi majeur de l'optimisation topologique.

## 1.4. Prise en compte des incertitudes

Dans un problème de dimensionnement, les différentes données (comme les propriétés des matériaux et le chargement) sont souvent soumises à des incertitudes. La source de ces incertitudes est multiple : variabilité de nature aléatoire, manque d'information, paramètres de tolérance dans la fabrication, incertitudes de mesures, effets à long terme en sont quelques exemples. Une préoccupation plus récente est le changement climatique, qui peut jouer sur

les propriétés des matériaux et le chargement d'une manière seulement très partiellement connue.

Plusieurs approches existent pour intégrer de telles incertitudes dans le problème de dimensionnement. Ces approches se distinguent par la nature de l'information disponible sur les incertitudes. Dans le but principal de simplifier la présentation, nous nous concentrons dans la suite sur les incertitudes liées aux propriétés des matériaux. Plus précisément, la relation contrainte-déformation étant supposée linéaire, on considère une incertitude sur le tenseur d'élasticité  $\mathbf{L}(x)$ .

Les approches probabilistes abordent les incertitudes via la donnée de lois de probabilité sur  $\mathbf{L}$ , vu comme une variable aléatoire. On s'intéresse alors à calculer la probabilité de rupture, ou plus généralement à déterminer la loi de probabilité d'une quantité liée à la réponse de la structure (comme la complaisance, considérée précédemment). Afin de fixer les idées, considérons le cas où  $\mathbf{L}$  est isotrope et homogène, avec une incertitude sur le module d'Young  $E$  définie par une loi de probabilité  $p$ . La condition exprimant la résistance de la structure est

$$G(\Omega, E) = \sup_{\boldsymbol{\sigma} \in \Omega} g(\boldsymbol{\sigma}) \leq 0 \quad (1.24)$$

où  $g$  exprime le critère de résistance du matériau, et  $\boldsymbol{\sigma}$  est la solution du problème d'élasticité. Comme  $\boldsymbol{\sigma}$  dépend du champ  $\mathbf{L}$ , le membre de gauche dans (1.24) est, tout comme  $\mathbf{L}$ , une variable aléatoire. La probabilité de rupture est alors donnée par

$$\int_0^{+\infty} H(G(\Omega, E))p(E)dE$$

où  $H$  est la fonction de Heaviside. Il s'agit ensuite de vérifier que cette probabilité est inférieure à un seuil prédéfini par le concepteur.

Les approches probabilistes ne sont pas la seule façon de décrire les incertitudes, ni la plus adaptée dans tous les cas. Une autre approche est de considérer qu'on dispose uniquement de bornes (notées  $E_-$  et  $E_+$  sur  $E$ ). Il s'agit alors de vérifier que la condition (1.24) est satisfaite pour tout  $E \in [E_-, E_+]$ , c'est-à-dire que

$$\sup_{E_- \leq E(x) \leq E_+} G(\Omega, E) \leq 0. \quad (1.25)$$

De façon analogue, l'expression (1.10) du paramètre de chargement maximum devient

$$\lambda^{max} = \frac{\sigma_Y}{\sup_{E_- \leq E(x) \leq E_+} F(\Omega, E)}. \quad (1.26)$$

Tant que le paramètre de chargement  $\lambda$  reste inférieur à la valeur  $\lambda^{max}$  ainsi définie, il est garanti que la condition de résistance est vérifiée pour toutes les valeurs admissibles de  $E$ .

Les problèmes (1.25)-(1.26) présentent certaines similarités avec les problèmes d'optimisation topologique présentés précédemment. Ainsi, dans le cas général, la valeur de  $E$  dans (1.25) n'est pas nécessairement la même en tout point et la variable  $E$  dans (1.25) est donc une distribution de modules élastiques, comme dans 1.3. De même, comme dans 1.3, la fonction ( $F$  ou  $G$ ) à optimiser dépend de  $E$  via la solution d'un problème d'élasticité.

L'approche exprimée par (1.25-1.26) se rencontre dans la littérature sous différentes appellations : dimensionnement robuste (reliable design), worst-case scenario, anti-optimization (ELISHAKOFF et al., 2010). La majorité des travaux existants portent sur des systèmes à nombre fini de degrés de libertés, utilisant le calcul par intervalles (MOORE, 1966) pour résoudre le problème (1.25). L'extension au problème continu est peu étudiée.

La présentation des méthodes d'optimisation topologique pour la complaisance a mis en évidence les difficultés qui peuvent surgir lorsqu'on considère un problème d'optimisation portant sur une distribution de modules élastique. En particulier se posent les questions de l'existence d'une fonction  $E(x)$  atteignant le *supremum* dans (1.25) et du bien-fondé de problèmes discrétisés (existence de solutions, sensibilité au maillage, convergence).

Afin d'explorer ces questions, nous étudions dans la suite un problème-modèle de dimensionnement robuste pour le milieu continu, suffisamment simple pour être résolu de façon exacte. Les résultats obtenus sont complétés par une étude numérique, présentée en 1.6.

## 1.5. Étude d'un problème de dimensionnement robuste en milieu continu

Nous étudions la flexion composée d'une poutre en déformations planes (figure 1.10).

La poutre occupe le domaine rectangulaire  $0 \leq x \leq \eta$ ,  $-1 \leq y \leq 1$  (les longueurs sont exprimées sous forme adimensionnelle).

Une densité de force  $T(y)\mathbf{u}_y$  est appliquée sur l'extrémité  $x = 0$ .

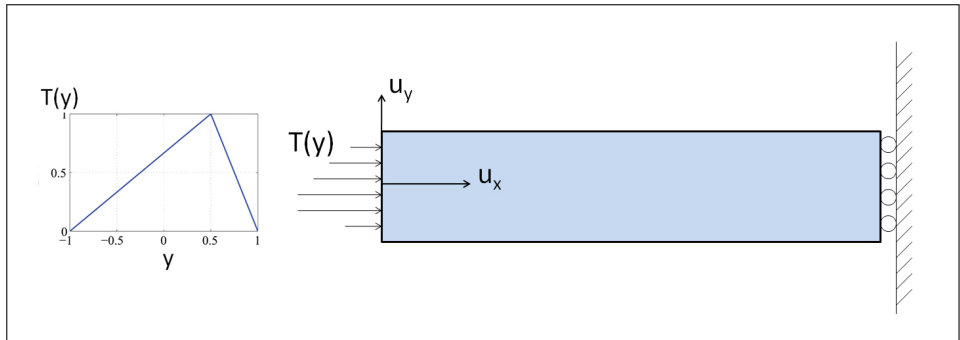
La fonction  $T(y)$  est de la forme  $T(y) = \lambda f_0(y)$  où  $\lambda$  est un paramètre de chargement et  $f_0$  la fonction linéaire par morceaux représentée figure 1.10.

L'extrémité  $x = \eta$  est en contact sans frottement avec le demi-espace  $x \geq \eta$ .

Le reste de la surface de la poutre est libre de contraintes, et il n'y a pas de forces volumiques.

**Figure 1.10.**

Dimensionnement robuste d'une poutre en flexion



Le matériau constitutif étant supposé linéaire isotrope, on considère que la valeur exacte de  $E$  est inconnue et l'on dispose seulement d'un encadrement  $E_- \leq E \leq E_+$ . De plus, la valeur de  $E$  n'est pas nécessairement la même en tout point. Pour permettre une résolution analytique du problème, on supposera que  $E$  dépend uniquement de  $y$ .

Dans ces conditions, la loi de comportement s'écrit localement

$$e(x, y) = \frac{1 + \nu}{E(y)} \boldsymbol{\sigma}(x, y) - \frac{\nu}{E(y)} (\text{tr } \boldsymbol{\sigma}) I, \quad (1.27)$$

où  $e(x, y)$  est relié au champ de déplacement  $\mathbf{u}(x, y)$  par les relations

$$e(x, y) = \begin{pmatrix} \frac{\partial u_x}{\partial x} & \frac{1}{2} \left( \frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \\ \frac{1}{2} \left( \frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) & \frac{\partial u_y}{\partial y} \end{pmatrix}. \quad (1.28)$$

À l'équilibre, le champ de contraintes  $\boldsymbol{\sigma}$  vérifie les relations

$$\text{div } \boldsymbol{\sigma} = 0 \text{ dans } \Omega, \quad [\boldsymbol{\sigma}] \cdot \mathbf{n} = 0 \text{ sur } \Sigma \quad (1.29)$$

où  $\Sigma$  désigne une surface de discontinuité (de normale  $\mathbf{n}$ ) pour  $\boldsymbol{\sigma}$ , et  $[\boldsymbol{\sigma}]$  désigne le saut de  $\boldsymbol{\sigma}$  au franchissement de  $\Sigma$ . Nous cherchons dans la suite à résoudre le problème de dimensionnement robuste (1.25) pour le critère de Von Mises (1.5). Plus précisément, on vise à déterminer le chargement maximal  $\lambda^{max}$  défini en (1.26).

### 1.5.1. Solution du problème d'équilibre au sens de Saint-Venant

Pour une distribution  $E(y)$  fixée, commençons par déterminer le champ de contraintes  $\boldsymbol{\sigma}_0$  solution du problème d'équilibre pour  $\lambda = 1$ . On cherche une solution sous la forme

$$\boldsymbol{\sigma}_0(x, y) = \sigma(y) \mathbf{u}_x \otimes \mathbf{u}_x$$

où  $\otimes$  désigne le produit tensoriel et  $\sigma(y)$  est une fonction à déterminer. Un tel champ est compatible avec les équations d'équilibre (1.29). En utilisant les

relations (1.27-1.28) :

$$\frac{\partial u_x}{\partial x} = f(y), \quad \frac{\partial u_y}{\partial y} = -\nu f(y), \quad \frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} = 0 \quad (1.30)$$

où l'on a posé  $f(y) = \sigma(y)/E(y)$ . Soit  $F$  est une primitive de  $f$ . En intégrant les deux premières équations dans (1.30) :

$$u_x(x, y) = xf(y) + g(y), \quad u_y(x, y) = -\nu F(y) + h(x) \quad (1.31)$$

où  $g$  et  $h$  sont deux fonctions indéterminées à ce stade. En substituant (1.31) dans la dernière équation de (1.30), on obtient la relation suivante liant  $g$  et  $h$  :

$$xf'(y) + g'(y) + h'(x) = 0. \quad (1.32)$$

En dérivant cette relation par rapport à  $x$ , on trouve  $f'(y) + h''(x) = 0$ , ce qui implique

$$-f'(y) = h''(x) = -\alpha$$

pour une certaine constante  $\alpha$ . On a donc

$$f(y) = \alpha y + \beta, \quad h(x) = -\alpha \frac{x^2}{2} + \gamma x + \delta$$

où  $\beta, \delta, \gamma$  sont des constantes. Remplacer ces expressions dans (1.32) donne l'expression de  $g$  :

$$g(y) = -\gamma y + \epsilon$$

où  $\epsilon$  est une constante. Les différentes constantes d'intégration sont déterminées par les conditions aux limites. En particulier la condition  $u_x = 0$  en  $x = \eta$  donne

$$\gamma = \eta\alpha, \quad \epsilon = -\eta\beta. \quad (1.33)$$

On arrive ainsi aux expressions suivantes pour le champ de contraintes et le champ de déplacement :

$$\begin{aligned} \sigma(y) &= E(y)(\alpha y + \beta), \\ u_x(x, y) &= x(\alpha y + \beta) - \eta\alpha y - \eta\beta, \\ u_y(x, y) &= -\nu\left(\frac{1}{2}\alpha y^2 + \beta y\right) - \alpha \frac{x^2}{2} + \eta\alpha x + \delta. \end{aligned} \quad (1.34)$$

Les deux constantes  $\alpha$  et  $\beta$  sont déterminées par la résultante  $N_0$  et le moment  $M_0$  des efforts appliqués en  $x = 0$ , définis par

$$N_0 = \int_{-1}^1 T_0(y) dy, \quad M_0 = \int_{-1}^1 y T_0(y) dy.$$

On a en effet le système

$$\begin{pmatrix} X_1 & X_0 \\ X_2 & X_1 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} N_0 \\ M_0 \end{pmatrix} \quad (1.35)$$

où

$$X_i = \int_{-1}^1 y^i E(y) dy. \quad (1.36)$$

Pour la fonction  $f_0$  représentée figure 1.10,  $N_0 = 1$ ,  $M_0 = 1/6$  et la résolution du système (1.35) donne

$$\alpha = \frac{6X_1 - X_0}{6(X_1^2 - X_0X_2)}, \quad \beta = \frac{X_1 - 6X_2}{6(X_1^2 - X_0X_2)} \quad (1.37)$$

Dans le cas général, le champ de contraintes  $\sigma$  obtenu ne vérifie pas la relation  $\sigma.n = 0$  en tout point de la surface  $x = 0$ . Le champ obtenu n'est donc pas solution du problème d'élasticité considéré. Cependant, par le principe de Saint-Venant, on sait que, sous réserve que le rapport  $\eta/h$  soit suffisamment grand, la solution obtenue est valable loin des surfaces  $x = 0$  et  $x = \eta$ .

Remarquons également que le champ de déplacement obtenu est différentiable (polynomial de degré 2) même si la distribution  $E(y)$  n'est pas continue. Le champ de déplacement ne dépend en effet de la distribution  $E(y)$  qu'à travers les constantes  $(\alpha, \beta)$ .

Notons que la solution en contraintes  $\sigma$  reste inchangée si le champ  $E(y)$  est multiplié par une constante. Nous pouvons donc supposer que  $E(y)$  est écrit sous forme adimensionnelle et soumis à la restriction  $1 \leq E(y) \leq r$  où  $r = \frac{E_+}{E_-}$  mesure le niveau d'incertitude sur le module d'Young.

À  $r$  donné, la distribution  $E(y)$  est cherchée dans l'ensemble

$$\mathcal{E}(r) = L^2([-1, 1], [1, r]).$$

Le problème de dimensionnement robuste (1.25) amène à résoudre

$$\sup_{E \in \mathcal{E}(r)} F(E) \quad (1.38)$$

où, compte-tenu de l'expression obtenue pour  $\sigma$ ,

$$F(E) = \sup_{-1 \leq y \leq 1} E(y) |\alpha y + \beta|. \quad (1.39)$$



La relation (1.26) donnant le chargement maximum  $\lambda^{max}$  devient

$$\lambda^{max}(r) = \frac{\sigma_Y}{\sup_{E \in \mathcal{E}(r)} F(E)}. \quad (1.40)$$

Commençons par étudier le cas déterministe, i.e.  $r = 1$ . Dans ce cas, l'ensemble  $\mathcal{E}(1)$  se réduit à la distribution  $E(y)$  égale à 1 en tout point (cette distribution est notée 1 dans la suite). On trouve alors  $X_0 = 2$ ,  $X_1 = 0$ ,  $X_2 = 2/3$  et  $\alpha = 1/4$ ,  $\beta = 1/2$ , donc  $F(1) = 3/4$  et

$$\lambda^{max}(1) = \frac{4}{3}\sigma_Y.$$

Dans (1.39), le *supremum* sur  $y$  est atteint en  $y = 1$  : ces points correspondant à la zone critique où le critère de rupture est atteint en premier.

Dans la suite, on cherche à résoudre de façon exacte le problème d'optimisation (1.38) pour  $r > 1$ . Dans ce cas, l'ensemble  $\mathcal{E}(r)$  est de dimension infinie, et la résolution exacte de (1.38) nécessite l'utilisation d'outils de l'analyse fonctionnelle. Cette étude est résumée dans les parties 1.5.2-1.5.3 et s'appuie sur des résultats classiques d'analyse (BONY, 2001 ; LARROUTOUROU et al., 1994). Il est possible en première lecture de passer directement à la partie 1.5.4, qui énonce le résultat principal et ses conséquences.

## 1.5.2. Quelques résultats préliminaires

### 1.5.2.1. Résultats d'analyse fonctionnelle

Pour tout  $E \in \mathcal{E}(r)$ , on note  $X(E) = (X_0(E), X_1(E), X_2(E))$  où  $X_i(E)$  est définie par l'intégrale (1.36). Soit  $\| \cdot \|$  la norme euclidienne usuelle dans  $\mathbb{R}^3$ . Nous pouvons facilement vérifier que la fonction  $X : \mathcal{E}(r) \mapsto \mathbb{R}^3$  est linéaire et continue.

Nous avons les 3 résultats suivants :

**Résultat 1** : Soit  $f : A \rightarrow \mathbb{R}$  une fonction continue, où  $A$  est une partie fermée bornée de  $\mathbb{R}^3$  contenant  $X(\mathcal{E}(r))$ . Alors

$$\sup_{E \in \mathcal{E}(r)} f(X(E)) \text{ est atteint.} \quad (1.41)$$

Le résultat suivant complète le résultat 1 dans le cas où  $f$  est différentiable :

**Résultat 2** : Soit  $\tilde{E} \in \mathcal{E}(r)$  atteignant  $\sup_{E \in \mathcal{E}(r)} f(X(E))$ . Si  $f$  est différentiable, alors  $\tilde{E}(y)$  est constant par morceaux, à valeurs dans  $\{1, r\}$  et vérifie les relations

$$\tilde{E}(y) = 1 \text{ si } P(y) < 0, \quad \tilde{E}(y) = r \text{ si } P(y) > 0 \quad (1.42)$$

où  $P$  est le polynôme du second degré défini par

$$P(y) = \sum_{i=0}^2 \frac{\partial f}{\partial X_i}(X(\tilde{E}))y^i. \quad (1.43)$$

Il sera fait fréquemment usage du résultat utile suivant :

**Résultat 3 :** Si  $f(X(1)) > 0$  alors il existe  $r_0 > 1$  tel que pour tout  $1 \leq r \leq r_0$  :

$$f(X(E)) > 0 \quad \forall E \in \mathcal{E}(r). \quad (1.44)$$

### 1.5.2.2. Démonstrations des résultats

Démontrons le résultat 1. Soit  $\{E_n\}$  une suite maximisante, c'est-à-dire telle que  $E_n \in \mathcal{E}(r)$  et  $f(X(E_n)) \rightarrow \sup_{E \in \mathcal{E}(r)} f(X(E))$  quand  $n \rightarrow \infty$ . L'ensemble  $\mathcal{E}(r)$  étant borné dans  $L^2([-1, 1], \mathbb{R})$ , il existe une suite extraite (encore notée  $\{E_n\}$ ) faiblement convergente dans  $L^2([-1, 1], \mathbb{R})$ , dont on note  $\tilde{E}$  la limite.

$$\int_{-1}^1 E_n(y)\phi(y)dy \rightarrow \int_{-1}^1 \tilde{E}(y)\phi(y)dy \quad \text{pour tout } \phi \in L^2([-1, 1], \mathbb{R}). \quad (1.45)$$

Comme  $\mathcal{E}(r)$  est convexe et fortement fermé dans  $L^2([-1, 1], \mathbb{R})$  (BONY, 2001),  $\mathcal{E}(r)$  est également faiblement fermé (LARROUTUROU et al., 1994). Par conséquent, on a  $\tilde{E} \in \mathcal{E}(r)$ .

La relation (1.45) appliquée avec  $\phi = y^i$  montre que

$$X_i(E_n) \rightarrow X_i(\tilde{E}).$$

Par continuité de  $f$ ,

$$f(X(E_n)) \rightarrow f(X(\tilde{E})).$$

Ceci montre que  $\tilde{E}$  atteint  $\sup_{E \in \mathcal{E}(r)} f(X(E))$ .

Démontrons maintenant le résultat 2. Soit  $E_0 \in \mathcal{E}(r)$  et  $E(t) = \tilde{E} + t(E_0 - \tilde{E})$ . Par convexité de  $\mathcal{E}(r)$ , on a  $E(t) \in \mathcal{E}(r)$  pour tout  $t \in [0, 1]$ . En posant  $g(t) = f(X(E(t)))$ , par définition de  $\tilde{E}$  :

$$g(t) \leq g(0) \quad \forall t \in [0, 1]$$

ce qui implique  $g'(0) \leq 0$ , i.e.

$$0 \geq \int_{-1}^1 P(y)(E_0 - \tilde{E})(y)dy. \quad (1.46)$$

Établissons maintenant les relations (1.42). Soit  $y_0$  tel que  $P(y_0) > 0$ . Par continuité de  $P$ , il existe un intervalle  $I$  dans  $[-1, 1]$  (non réduit à un singleton) tel

que  $P(y) > 0$  pour tout  $y \in I$ . Choisissons alors  $E_0$  sous la forme

$$E_0(y) = \tilde{E}(y) \text{ pour } y \notin I, E_0(y) = r \text{ pour } y \in I.$$

En remplaçant dans (1.46),

$$0 \geq \int_I P(y)(r - \tilde{E}(y)). \quad (1.47)$$

où  $P$  est le polynôme défini en (1.43). Comme  $P(y) > 0$  et  $r - \tilde{E}(y) \geq 0$  sur  $I$ , l'inégalité (1.47) implique nécessairement  $P(y)(r - \tilde{E}(y)) = 0$  sur  $I$ , i.e.  $\tilde{E}(y) = r$ . Le cas où  $P(y_0) < 0$  se traite de façon similaire.

Démontrons enfin le résultat 3. Posons  $\mathcal{F} = L^2([-1, 1], [0, 1])$ . On vérifie facilement que  $X(\mathcal{F})$  est un ensemble borné de  $\mathbb{R}^3$ , i.e. il existe  $M > 0$  tel que  $\|X(F)\| \leq M$  pour tout  $F \in \mathcal{F}$ . Tout  $E \in \mathcal{E}(r)$  s'écrit de façon unique sous la forme

$$E = (r - 1)F + 1$$

où  $F \in \mathcal{F}$ . Par linéarité de  $X$ , on a  $X(E) = (r - 1)X(F) + X(1)$  et donc

$$\|X(E) - X(1)\| \leq (r - 1)M.$$

Comme  $f$  est continue et  $f(X(1)) > 0$ ,  $f$  est strictement positive sur une boule de centre  $X(1)$  et de rayon  $\tau > 0$ .

Soit  $r_0 = 1 + \tau/M$ . Pour tout  $r \in [1, r_0]$  et  $E \in \mathcal{E}(r)$ , on a  $\|X(E) - X(1)\| \leq \tau$  et par suite  $f(E) > 0$ .

### 1.5.3. Étude de l'existence d'une solution optimale

D'après l'expression (1.39), il est clair que si la distribution  $E(y)$  est telle que

$$\alpha > 0, \beta > 0 \quad (1.48)$$

alors

$$F(E) \leq r(\alpha + \beta). \quad (1.49)$$

On a vu que la condition (1.48) est satisfaite en  $r = 1$ . Le résultat 3 (cf Eq.(1.44)), appliqués aux fonctions  $\alpha(E)$  et  $\beta(E)$ , montre alors qu'il existe  $r_0 > 1$  tel que (1.48) est satisfaite pour tout  $1 \leq r \leq r_0$  et  $E \in \mathcal{E}(r)$ . Pour de telles valeurs de  $r$ , la majoration (1.49) incite à maximiser la quantité  $f(E) = \alpha + \beta$  par rapport à  $E$  dans  $\mathcal{E}$ , en utilisant pour cela les résultats 1 et 2. En posant  $c_i(E) = (\partial f / \partial X_i)(E)$ ,

notons au préalable que

$$\begin{aligned} c_0(E) &= -\frac{X_1^2 - 7X_1X_2 + 6X_2^2}{6(X_1^2 - X_0X_2)^2}, \\ c_1(E) &= \frac{2X_0X_1 - 7X_0X_2 - 7X_1^2 + 12X_1X_2}{6(X_1^2 - X_0X_2)^2}, \\ c_2(E) &= -\frac{(X_0 - 6X_1)(X_0 - X_1)}{6(X_1^2 - X_0X_2)^2}. \end{aligned} \quad (1.50)$$

En particulier, pour  $r = 1$ ,  $c_0 = -1/4$ ,  $c_1 = -7/8$ ,  $c_2 = -3/8$ . Dans ce cas, on vérifie facilement que le polynôme  $P(y) = \sum_{i=0}^2 c_i y^i$  a les propriétés suivantes :

$$P \text{ a un unique zéro (noté } a) \text{ dans } [-1, 1] \text{ et décroît strictement sur } [-1, 1]. \quad (1.51)$$

En fait, la propriété (1.51) reste valable pour tout  $E \in \mathcal{E}(r)$  si  $r$  est suffisamment proche de 1. En effet, (1.51) est satisfaite si et seulement si les coefficients de  $P$  vérifient

$$c_2 < 0, \frac{-c_1}{2c_2} < -1, \Delta = c_1^2 - 4c_0c_2 > 0, -1 < \frac{-c_1 - \sqrt{\Delta}}{2c_2} < 1$$

ce qui peut se réécrire sous la forme

$$c_2 < 0, c_1 - 2c_2 < 0, 0 < c_2 - c_1 + c_0, 0 > c_2 + c_1 + c_0.$$

Ces conditions sont satisfaites pour  $E = 1$ . L'application répétée du résultat 3 aux fonctions  $c_2, c_1 - 2c_2, c_2 - c_1 + c_0, c_2 + c_1 + c_0$  montre qu'il existe  $r'_0 > 1$  tel que (1.51) est satisfaite pour tout  $1 \leq r \leq r'_0$  et  $E \in \mathcal{E}(r)$ . On pose  $R = \max\{r_0, r'_0\}$  et on s'intéresse dans la suite à  $r \leq R$ . Notons qu'une analyse plus détaillée montre qu'on peut prendre  $R = 4$ .

Soit  $r$  fixé dans  $[1, R]$  et  $E \in \mathcal{E}(r)$ . L'application des résultats 1 et 2 à la fonction  $f = \alpha + \beta$  montre qu'il existe  $\tilde{E}$  dans  $\mathcal{E}(r)$  tel que

$$\alpha + \beta \leq (\alpha + \beta)(\tilde{E}) \quad (1.52)$$

où  $\tilde{E}$  satisfait les conditions d'optimalité (1.42).

Or par (1.51),  $P(y) > 0$  pour  $y < a$  et  $P(y) < 0$  pour  $y > a$ . Par conséquent,

$$\begin{aligned} \tilde{E}(y) &= r \quad \text{pour } y < a, \\ \tilde{E}(y) &= 1 \quad \text{pour } y > a. \end{aligned} \quad (1.53)$$

La valeur de  $a$  dépend de  $r$  et est déterminée par la condition de cohérence  $P(a) = 0$ . Pour  $\tilde{E}$  de la forme (1.53), alors

$$\begin{aligned} X_0 &= 2r - (r-1)(1-a), \\ X_1 &= -\frac{1}{2}(a^2-1)(r-1), \\ X_2 &= \frac{1}{3}(2r - (r-1)(1-a^3)). \end{aligned}$$

En substituant dans (1.50), après calculs, la condition  $P(a) = 0$  se réécrit

$$2(a^2-1)^2(a(2a-7)-5)r - (a-5)(a+1)^4(2a-5)r^2 - (2a+3)(a-1)^5 = 0$$

qui se résout en

$$r = \frac{(a-1)^3}{(a-5)(a+1)^2}. \quad (1.54)$$

Cette équation définit  $a$  implicitement en fonction de  $r$ . La courbe  $r \mapsto a$  est représentée figure 1.11.

Les relations (1.53)-(1.54) déterminent complètement la distribution  $\tilde{E}$  maximisant  $f$  à  $r$  donné. La valeur correspondante de  $\alpha + \beta$  est égale, tous calculs faits, à  $(a-5)(6a-1)/12(a-1)^2$ . Comme  $\alpha$  et  $\beta$  sont positifs, pour tout  $E \in \mathcal{E}(r)$ ,

$$F(E) \leq r \frac{(a-5)(6a-1)}{12(a+1)^2}, \quad (1.55)$$

où  $a$  est donné par (1.54). Étant donné que l'inégalité (1.55) est valable pour tout  $E \in \mathcal{E}(r)$ , alors

$$\sup_{E \in \mathcal{E}(r)} F(E) \leq r \frac{(a-1)(6a-1)}{12(a+1)^2}. \quad (1.56)$$

Il y a en fait égalité dans (1.56). Pour tout  $b > 0$ , considérons en effet la distribution  $\tilde{E}(b)$  définie par

$$\begin{aligned} \tilde{E}(b)(y) &= \tilde{E}(y) && \text{pour } -1 \leq y \leq 1-b, \\ \tilde{E}(b)(y) &= r && \text{pour } 1-b < y \leq 1. \end{aligned} \quad (1.57)$$

Soit  $\{b_n\}$  une suite positive tendant vers 0, et  $E_n = \tilde{E}(b_n)$ . On vérifie facilement que  $X_i(E_n) \rightarrow X_i(\tilde{E})$  ( $i = 0, 1, 2$ ) quand  $n \rightarrow \infty$ , et par continuité que  $f(E_n) \rightarrow f(\tilde{E})$ . Comme  $E_n(1) = r$ , on a  $F(E_n) = \sup_y f(\sigma) = r(\alpha + \beta)$ . Par suite,  $\sup_y g(y)$  tend vers le membre de droite dans (1.55).

Notons que la suite  $E_n$  converge dans  $L^2([-1, 1], \mathbb{R})$  vers  $\tilde{E}$ , mais  $F(E_n)$  ne converge pas vers  $F(\tilde{E})$ . En fait le *supremum* dans (1.56) n'est pas atteint : pour  $E = \tilde{E}$ , on a  $F(\tilde{E}) < X$  où  $X = \sup_{\mathcal{E}} F$ . Pour  $E \neq \tilde{E}$ , les conditions d'optimalité (1.53) ne sont pas satisfaites et par conséquent  $\alpha + \beta < (\alpha + \beta)(\tilde{E})$ , dont on déduit  $F(\tilde{E}) < X$ .

Nous arrivons ainsi au résultat principal suivant.

### 1.5.4. Résultat principal et conséquences

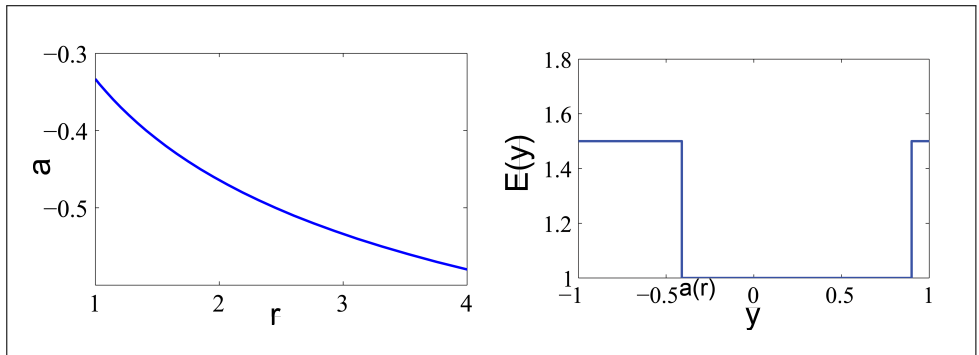
Pour tout  $1 \leq r < 4$ , on a

$$\sup_{E \in \mathcal{E}(r)} F(E) = r \frac{(a-1)(6a-1)}{12(a+1)^2}, \quad (1.58)$$

où  $a$  est défini implicitement en fonction de  $r$  par la formule (1.54) (voir figure 1.11 (gauche)). Le *supremum* dans (1.58) n'est pas atteint. Une suite maximisante est donnée par la suite  $\{E_n\}$  définie en (1.57) et représentée figure 1.11 (droite).

**Figure 1.11.**

Représentation de la fonction  $a(r)$  (gauche) et d'une suite maximisante (droite) pour un problème de dimensionnement robuste



On examine dans la suite quelques conséquences et interprétations de ce résultat.

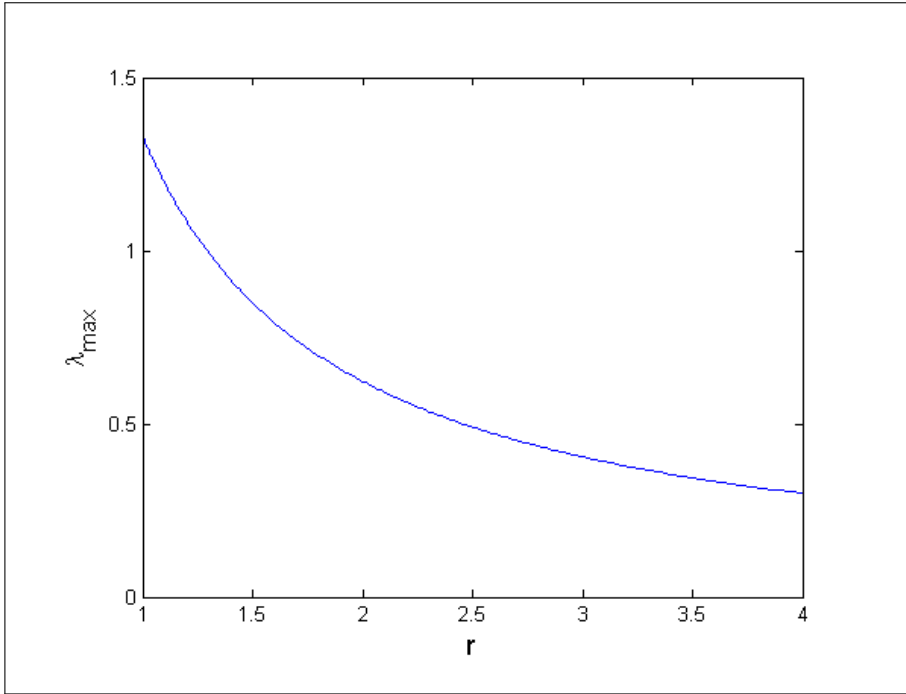
L'expression (1.58) donne la solution exacte du problème de dimensionnement robuste. En remplaçant (1.58) dans (1.26), en effet

$$\lambda^{max}(r) = \frac{\sigma_Y}{r} \frac{12(a+1)^2}{(a-1)(6a-1)} \quad (1.59)$$

qui s'interprète comme le chargement maximal au sens du dimensionnement robuste : tant que  $\lambda \leq \lambda^{max}(r)$ , le critère de résistance (1.5) est satisfait en tout point et pour toute distribution du module d'Young  $E(y)$  compatible avec un niveau d'incertitude donné  $r$ . La fonction  $r \mapsto \lambda^{max}(r)$  est représentée figure 1.12. Cette fonction est décroissante : plus le niveau d'incertitude est élevé, plus la prévision sur la tenue de la structure est pessimiste. Comme on l'observe figure 1.12, la sensibilité  $-d\lambda^{max}/dr$  est maximale en  $r = 1$ .

**Figure 1.12.**

Chargement maximal en fonction du niveau d'incertitude sur le module d'Young



Comme on l'a déjà mentionné, le *supremum* dans (1.58) n'est jamais atteint, même si l'incertitude est très faible. Afin d'interpréter ce phénomène, considérons une répartition arbitraire  $E(y)$  du module d'Young et modifions localement cette distribution autour d'un point  $y_0$ . Pour  $E^0 \in [1, r]$  et  $\delta y$  donnés, on considère la distribution  $E'$  définie par

$$E'(y) = E^0 \text{ si } y \in [y_0, y_0 + \delta y], \quad E'(y) = E(y) \text{ sinon.} \quad (1.60)$$

Lorsque  $\delta y$  est choisi très petit, les scalaires  $\alpha$  et  $\beta$  dans (1.34) ne sont pas perturbés (au premier ordre). En effet,  $\alpha$  et  $\beta$  sont déterminés par des intégrales sur la distribution  $E$  : une modification localisée, de la forme (1.60), a un effet négligeable. En conséquence,  $\sigma(y) = E(y)(\alpha y + \beta)$  n'est pas modifié (au premier ordre en  $\delta y$ ) hors de la zone de perturbation  $[y_0, y_0 + \delta y]$ . Par contre, dans la zone de perturbation, la contrainte  $\sigma(y_0)$  passe de la valeur  $E(y_0)(\alpha y_0 + \beta)$  à  $E^0(\alpha y_0 + \beta)$  : il y a une amplification locale de la contrainte, égale à  $E^0/E(y)$  et maximale pour  $E^0 = r$ .

Un tel phénomène d'amplification locale est précisément ce qui apparaît dans la suite maximisante  $\{E_n\}$  : la distribution  $E_n$  est une perturbation locale, de la forme (1.60), appliquée en  $y_0 = 1$  à la distribution  $\tilde{E}$  définie (1.57). Plus la taille

$b_n$  de la zone de perturbation est faible, plus la contrainte en  $y_0 = 1$  est amplifiée, tendant vers la limite  $(\alpha + \beta)r/\tilde{E}(1)$ .

Le fait que le *supremum* dans (1.58) n'est jamais atteint n'a pas uniquement une portée théorique : ce résultat a des répercussions directes (et néfastes) sur la résolution numérique du problème. Ce point est illustré plus en détails dans la suite.

## 1.6. Illustration numérique

### 1.6.1. Méthode de résolution

Nous pouvons montrer que le problème (1.26) peut se réécrire sous la forme

$$\lambda^{max}(r) = \sup_{G(\lambda) > 0} \lambda \quad (1.61)$$

où

$$G(\lambda) = \sup_{E \in \mathcal{E}(r)} \int_{-1}^1 \langle |\sigma(y)| - \lambda \rangle_+^2 dy \quad (1.62)$$

et  $\langle x \rangle_+$  désigne la partie positive de  $x$ , i.e.  $\langle x \rangle_+ = (x + |x|)/2$ . Du point de vue de la résolution numérique, l'ensemble  $\mathcal{E}(r)$  est remplacé par un sous-espace de dimension finie  $\mathcal{E}_h(r)$ , et l'on résout le problème

$$\lambda_h^{max}(r) = \sup_{G_h(\lambda) > 0} \lambda \quad (1.63)$$

où

$$G_h(\lambda) = \sup_{E \in \mathcal{E}_h(r)} \int_{-1}^1 \langle |\sigma(y)| - \lambda \rangle_+^2 dy. \quad (1.64)$$

Le problème d'optimisation obtenu (1.64) est ensuite résolu par un algorithme de gradient. Le problème (1.64) étant en dimension finie, le *supremum* dans (1.64) est atteint, ce qui n'est pas le cas pour (1.62).

Soit  $-1 = x_1 \leq x_2 \leq \dots \leq x_{n+1} = 1$  une subdivision de l'intervalle  $[-1, 1]$ .

On résout le problème (1.63) en considérant l'espace d'approximation  $\mathcal{E}_h(r)$  formé par les fonctions constantes sur chaque intervalle  $[x_i, x_{i+1}]$ . La table 1.1 montre les solutions obtenues à partir de trois états initiaux différents, représentés figure 1.13.

Ces résultats ont été obtenus avec  $n = 40$  points de discrétisation équi-répartis. Le paramètre  $r$  mesurant l'incertitude est fixé à 1.5. Comme on le constate sur la table 1.1, l'estimation de  $\lambda^{max}$  obtenue dépend de l'état initial considéré, et dans certains cas s'éloigne de la valeur exacte, donnée par (1.59) et égale à 0.856 pour  $r = 1.5$ .



Les distributions critiques obtenues sont représentées figure 1.14 et sont également sensibles à l'état initial. Ce comportement est la manifestation numérique du fait que le *supremum* dans (1.26) n'est pas atteint.

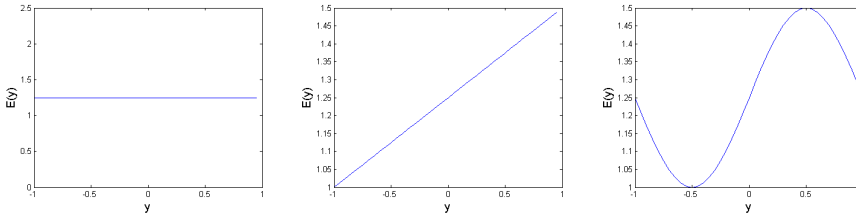
**Table 1.1.**

Résultats numériques ( $r = 1.5$ )

	état initial	$E_0$	$E_1$	$E_2$
problème (1.63)	$\lambda_h^{max}$	0.91	0.90	0.94
	nombre d'itérations	51	32	21
problème (1.67)	$\lambda_h^{max}$	0.86	0.86	0.86
	nombre d'itérations	31	22	16

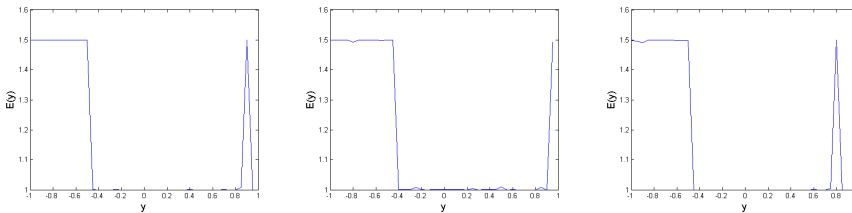
**Figure 1.13.**

États initiaux (notés  $E_0, E_1, E_2$ ) utilisés



**Figure 1.14.**

Distributions critiques obtenues pour chacun des 3 états initiaux  $E_0, E_1, E_2$



L'absence de solution optimale au problème continu en espace a été analysée en détails en 1.5.3 et interprétée en termes d'effets de microstructure. Cette interprétation suggère une reformulation du problème (1.61) sous la forme

$$\sup_{G^{hom}(\lambda) > 0} \lambda \tag{1.65}$$

où

$$G^{hom}(\lambda) = \sup_{E \in \mathcal{E}(r)} \int_{-1}^1 \langle |A(E(y))\sigma(y)| - \lambda \rangle_+^2 dy \tag{1.66}$$

et  $A(E(y)) = r/E(y)$ . L'idée est de voir la distribution  $E$  et la contrainte  $\sigma$  comme des grandeurs à l'échelle mésoscopique. Le terme  $A(E(y))$  représente le facteur d'amplification locale (au niveau microscopique) des contraintes, dont on a déterminé précédemment qu'il était égal à  $r/E(y)$ . Cette formulation est similaire à celle de la méthode d'homogénéisation : les champs considérés dans le problème d'optimisation sont des variables mésoscopiques, et l'effet de microstructure est pris en compte à travers le coefficient  $A(E)$ . Nous pouvons facilement établir que  $\lambda^{max}(r)$  est solution de (1.65) et que le *supremum* dans (1.66) définissant  $G^{hom}(\lambda^{max})$  est atteint (par la distribution  $\tilde{E}$  définie en (1.53)).

De façon analogue à (1.61), le problème (1.65) se discrétise en

$$\sup_{G_h^{hom}(\lambda) > 0} \lambda \tag{1.67}$$

où

$$G_h^{hom}(\lambda) = \sup_{E \in \mathcal{E}_h(r)} \int_{-1}^1 \langle |A(E(y))\sigma(y)| - a \rangle_+^2 dy. \tag{1.68}$$

On présente dans la table 1.1 les résultats obtenus par résolution numérique du problème, en conservant les mêmes états initiaux et les mêmes paramètres que ceux utilisés pour résoudre (1.63). Les valeurs obtenues pour  $\lambda_h^{max}$  sont systématiquement très proches de la valeur exacte (l'écart peut être réduit en choisissant une discrétisation plus fine). De plus la convergence est significativement plus rapide, comme on l'observe sur le nombre d'itérations. En outre, pour les 3 états initiaux, la distribution critique obtenue reste quasiment inchangée (aux erreurs numériques près) et proche de la distribution  $\tilde{E}$ .

Une autre voie pour régulariser le problème (1.61) consiste à imposer une condition de régularité sur les distributions  $E(y)$  considérées dans  $\mathcal{E}(r)$ . Comme on le voit figure 1.11, les suites maximisantes pour le problème (1.61) présentent des discontinuités. Dans certains cas, on peut considérer que de telles discontinuités ne sont pas autorisées physiquement. Par exemple, seules des variations continues peuvent être considérées comme admissibles. De telles restrictions *a priori* dépendent de la nature du problème, et plus précisément des informations disponibles sur la dépendance spatiale du module d'Young. L'ajout de telles conditions de régularité peut suffire (mais pas toujours) à garantir l'existence d'une distribution optimale en dimension infinie. Par exemple, on peut montrer que le *supremum* est atteint si l'on prend

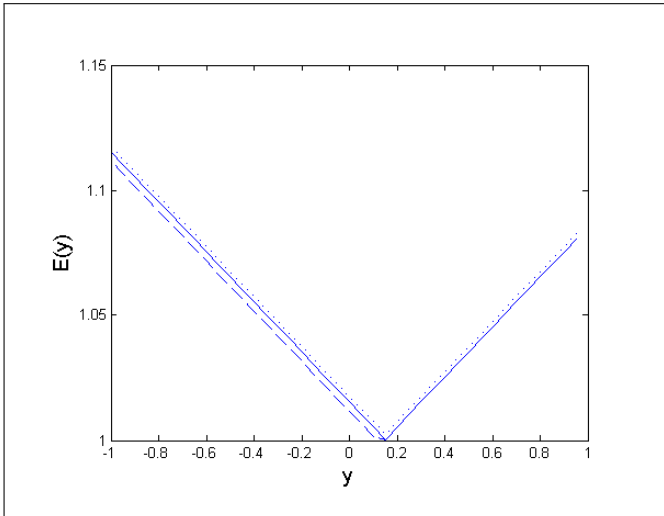
$$\mathcal{E}(r) = \{E : [-1, 1] \mapsto [1, r] \mid |E(y') - E(y)| \leq k|y' - y| \forall y, y' \in [-1, 1]\} \tag{1.69}$$

où  $k > 0$  est fixé. Par contre, si l'on fait le choix plus général  $\mathcal{E}(r) = C^0([-1, 1], [1, r])$  (ensemble des fonctions continues), alors le *supremum* n'est pas atteint. Notons que la définition (1.69) revient à introduire une échelle de longueur (représentée par le paramètre  $1/k$ ) dans le problème. De façon semblable à ce qui a été présenté précédemment sur la méthode SIMP, l'ajout

d'une échelle de longueur (réalisé de manière adaptée) supprime l'effet de microstructure et permet l'existence d'une solution optimale.

**Figure 1.15.**

Distributions critiques obtenues pour chacun des 3 états initiaux  $E_0$ ,  $E_1$ ,  $E_2$  (problème régularisé)



La figure 1.15 montre les distributions optimales obtenues en résolvant le problème (1.61) sur un ensemble  $\mathcal{E}(r)$  de la forme (1.69). On considère de nouveau les trois états initiaux représentés figure 1.14. Les solutions obtenues sont très proches, et correspondent toutes à une même valeur de  $\lambda^{max}$ , égale à 1,30.

Cette valeur est supérieure à celle obtenue pour le problème (1.65), ce qui est cohérent car on a restreint l'espace  $\mathcal{E}(r)$  sur lequel porte la maximisation. Soulignons que la contrainte de régularité (1.69) se traduit numériquement par un nombre de contraintes proportionnel au nombre de degrés de liberté de la discrétisation, et ralentit de fait considérablement l'optimisation.

### 1.6.2. Comparaison avec des simulations 2D

La figure 1.16 présente les champs de contraintes (valeur de la contrainte de Von Mises  $f(\sigma)$ ) obtenues par résolution du problème d'équilibre bidimensionnel (1.27-1.29) à l'aide d'une méthode par éléments finis. Les trois champs présentés correspondent à trois répartitions différentes du module d'Young dans la plaque :  $E(y) = E_0(y) = (1 + r)/2$  (gauche),  $E(y) = E_1(y)$  (milieu),  $E(y) = \tilde{E}_h(0.97)$  (droite). Les champs de contraintes sont tracés sur la configuration déformée (les déformations sont amplifiées pour une meilleure lisibilité).

À l'exception du voisinage immédiat des extrémités  $x = 0$  et  $x = \eta$ , le champ de contraintes obtenu est uniaxial et concorde avec la solution analytique. Dans cette zone (notée  $\Omega_1$ ), la contrainte  $f(\sigma)$  est maximale en  $y = 1$ , en accord avec

**Table 1.2.**

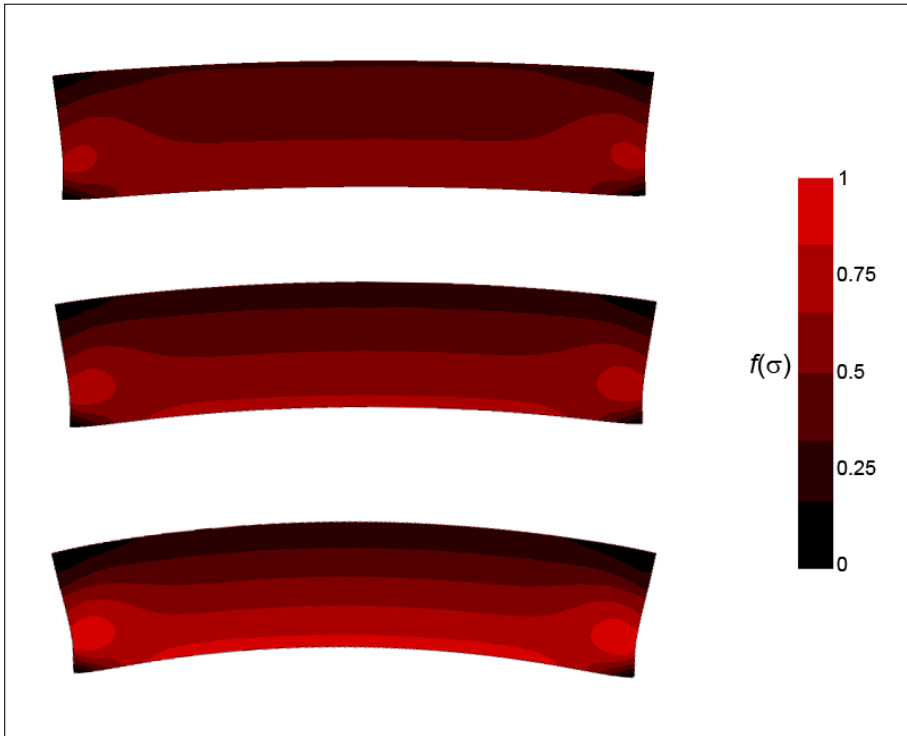
Valeur maximale de la contrainte de Von Mises

état initial	$E_0$	$E_1$	$\tilde{E}^h$
$\sup_{\Omega_1} f(\sigma)$	0.71	0.61	0.90
$\sup_{\Omega_2} f(\sigma)$	0.83	0.75	1.03

l'analyse présentée précédemment. La table 1.2 montre les valeurs  $\sup_{\Omega_1} f(\sigma)$  obtenues numériquement pour les trois distributions de module d'Young testées. Le cas le plus critique correspond à la distribution  $\tilde{E}_h$ , comme prévu par l'analyse théorique présentée précédemment. Pour cette configuration, la contrainte de Von Mises maximum est atteinte sur la surface inférieure de la poutre.

**Figure 1.16.**

Calcul bidimensionnel par éléments finis



## 1.7. Conclusion

Les méthodes d'optimisation de forme en conception de structures, dont on a donné un aperçu dans ce chapitre, ont acquis une maturité certaine et sont applicables sur certains problèmes industriels. Ces méthodes offrent des outils pour guider la conception, ou améliorer un design existant. Pour autant, l'objectif

ambitieux d'une conception optimale entièrement automatique reste difficile à atteindre du fait de la difficulté à atteindre un optimum global.

Concernant le dimensionnement robuste de structures en présence d'incertitudes – problème de nature différente au premier abord –, l'étude d'un problème-modèle a mis en évidence les difficultés qui surgissent lorsqu'on s'intéresse aux milieux continus.

En particulier, le fait que le *supremum* n'est pas atteint se traduit sur le plan numérique par la multiplicité d'optima locaux. Ce constat n'est pas complètement surprenant dans la mesure où, comme on l'a montré, le problème de dimensionnement robuste s'apparente à un problème d'optimisation de forme, pour lequel la multiplicité de minima locaux est une difficulté connue. S'inspirant des résultats de l'optimisation de forme, on a indiqué quelques pistes pour contourner les difficultés observées en dimensionnement robuste.

Une première méthode est d'effectuer un changement d'échelle, en introduisant un facteur d'amplification locale des contraintes dû à l'effet de microstructures. Cette approche est dans l'esprit de la méthode d'homogénéisation.

Une deuxième méthode est d'imposer une condition de régularité aux solutions recherchées, dans l'esprit de la méthode SIMP.

Ces deux approches ont prouvé leur efficacité sur un problème-modèle, mais leur extension à des situations plus générales, en 2 ou 3 dimensions, reste à étudier.

## Bibliographie

- Niels Aage, Morten Nobel-Jørgensen, Casper Schousboe Andreasen et Ole Sigmund.** « Interactive topology optimization on hand-held devices ». In : *Structural and Multidisciplinary Optimization* 47 (2013), pages 1-6.
- Grégoire Allaire.** *Shape optimization by the homogenization method*. Springer, 2002.
- Grégoire Allaire et O. Pantz.** « Structural optimization with FreeFem++ ». In : *Structural and Multidisciplinary Optimization* 32 (2006), pages 173-181.
- Martin Philip Bendsøe et Ole Sigmund.** « Material interpolation schemes in topology optimization ». In : *Archives of Applied Mechanics* 69 (1999), pages 635-654.
- Martin Philip Bendsøe et Ole Sigmund.** *Topology Optimization : Theory, Methods and Applications*. Springer, 2003.
- Jean-Michel Bony.** *Cours d'analyse*. Éditions de l'École Polytechnique, 2001.
- Matteo Bruggi et Pierre Duysinx.** « Topology optimization for minimum weight with compliance and stress constraints ». In : *Structural and Multidisciplinary Optimization* 46 (2012), pages 369-384.
- Peter W. Christensen et Anders Klarbring.** *An Introduction To Structural Optimization*. Springer, 2009.

- Pierre Duysinx et Martin Philip Bendsøe.** « Topology optimization of continuum structures with local stress constraints ». In : *International Journal for Numerical Methods in Engineering* 43 (1998), pages 1453-1478.
- Isaac Elishakoff et Makoto Ohsaki.** *Optimization and anti-optimization of structures under uncertainty*. Imperial College Press, 2010.
- Cheng Geng Dong et Xu Guo.** «  $\varepsilon$ -relaxed approach in structural topology optimization ». In : *Structural Optimization* 13 (1997), pages 258-266.
- Erik Holmberg, Bo Torstenfelt et Anders Klarbring.** « Stress constrained topology optimization ». In : *Structural and Multidisciplinary Optimization* (2013), pages 1-15.
- U. Kirsch.** « On singular topologies in optimum structural design ». In : *Structural Optimization* 2 (1990), pages 133-142.
- B. Larrouturou et P.L. Lions.** *Optimisation et analyse numérique*. Éditions de l'École Polytechnique, 1994.
- Maurice Lemaire.** *Structural reliability*. Tome 84. John Wiley & Sons, 2010.
- Ramon E. Moore.** *Interval Analysis*. Prince-Hall, Englewood Cliffs, NJ, 1966.
- J.T. Oden et J.A.C. Martins.** « Models and computational methods for dynamic friction phenomena ». In : *Computer Methods in Applied Mechanics and Engineering* 52 (1985), pages 527-634.
- Olivier Pironneau, Frédéric Hecht, Antoine Le Hyaric et Jacques Morice.** « Freefem++ ». In : See <http://www.freefem.org/ff+> (2009).
- Jean Salençon.** *Calcul à la rupture et analyse limite*. Paris : Presses ENPC, 1983, 366 p. : ill., 1983. ISBN : 2-85978-059-9.
- Jan Sokolowski et Jean-Paul Zolésio.** *Introduction to shape optimization*. Springer, 1992.

## Chapitre 2

# Optimisation d'un écran antibruit

Christophe HEINKELE<sup>1</sup>, Thomas LEISSING<sup>2</sup>, Jérôme DEFRANCE<sup>2</sup>  
**avec la contribution de :** Jean-Pierre CLAIRBOIS<sup>3</sup>, Francis GRAENNEC<sup>2</sup>,  
Philippe JEAN<sup>2</sup>

*Résumé – Dans le cadre d'un projet européen, la question de l'optimisation globale d'un écran antibruit a été abordée en croisant la performance acoustique, le coût et l'impact environnemental de l'ouvrage. Pour ce faire, des critères d'évaluation ont été mis en place pour chacun des trois domaines ce qui a nécessité de faire appel à une analyse en cycle de vie d'une série de matériaux et par conséquent à une base de données pour différentes familles d'écrans et pour différents matériaux d'un point de vue économique et environnemental.*

*Une première optimisation partielle a été réalisée afin d'optimiser chaque type d'écran sur le plan acoustique, en utilisant la méthode des éléments de frontières pour la prédiction de la performance et les algorithmes génétiques pour dégager un ouvrage optimal sur le plan acoustique. Les critères spécifiques ont été ensuite agrégés avec un système de notations afin de mettre en place des indicateurs globaux de performance. Ces indicateurs globaux permettent de pondérer les critères spécifiques entre eux avant de procéder à une optimisation globale. La méthode a été ensuite enrichie pour prendre en compte la topographie lors de l'insertion de l'ouvrage. Les résultats sont présentés sous la forme de radar-plot, ce qui permet de juger de la pertinence conjointe des critères retenus en amont.*

- 
1. CEREMA
  2. CSTB
  3. A-Tech

*Le principal résultat consiste en une méthode globale d'optimisation d'un écran antibruit en fonction de paramètres inhérents au décideur. La méthode présente l'avantage d'être à la fois souple et pragmatique.*

## 2.1. Éléments de contexte

Le travail présenté dans ce chapitre s'articule autour des résultats obtenus dans le projet européen de recherche QUIESST (CLAIRBOIS, 2012) consacré aux propriétés des écrans antibruit. L'objectif principal de QUIESST était la révision de la méthode Adrienne (CEN1793, 2003b; CEN1793, 2003c; CEN1793, 2003a) mais prévoyait également d'autres développements innovants afin d'améliorer la prise en compte et la mise en place des écrans antibruit. QUIESST s'est achevé en décembre 2012. Il y avait quatre lots techniques au sein de ce projet : le premier (WP2) s'est consacré à l'étude du champ lointain, le deuxième (WP3) a constitué une base de données en compilant les solutions disponibles sur le marché, le troisième (WP4) a analysé l'impact environnemental d'un écran antibruit et le quatrième (WP5) a proposé une méthode d'optimisation d'écrans antibruits. Le présent chapitre s'appuie sur les travaux de ce dernier lot piloté par le CSTB<sup>4</sup>.

Le travail a débuté par une bibliographie visant à recenser les modèles de propagation acoustique en extérieur ainsi que les techniques d'optimisation disponibles, afin de dresser un état de l'art des optimisations déjà réalisées dans le contexte des protections sonores du bruit extérieur. Nous renvoyons à (AUERBACH et al., 2010) pour plus de détails. Ce travail préliminaire a permis de cerner les outils déjà disponibles afin de faire un choix pour la poursuite du travail. La méthodologie a été mise en place progressivement dans (CHUDALLA, DEFRANCE et al., 2011), en se concentrant sur l'utilisation de la méthode des éléments de frontière et des algorithmes génétiques pour effectuer un classement des écrans antibruit par famille. Des séries d'optimisation restreintes à des cas purement acoustiques pour chaque famille d'écrans, présentées dans (CHUDALLA, CLAIRBOIS, DEFRANCE et LEISSING, 2012), ont permis d'apporter des réponses à une optimisation intrinsèque. La suite du travail (CHUDALLA, CLAIRBOIS, DEFRANCE, GRANNEC et al., 2012) s'est concentré sur la prise en compte des coûts environnementaux des matériaux et sur une approche globale sur des sites de configurations différentes (BOULEY et al., 2012). Tous ces rapports sont disponibles en ligne (<http://www.quiesst.eu>).

## 2.2. Outils numériques

### 2.2.1. Le modèle numérique de propagation sonore

La méthode numérique utilisée pour prédire les niveaux sonores est la méthode des éléments de frontière (BEM en anglais, pour *Boundary Element Method*).

---

4. Centre Scientifique et Technique du Bâtiment



La BEM permet une résolution approchée de l'équation de Helmholtz 2.1 en imposant des conditions aux limites sur une structure plongée dans un fluide homogène sans viscosité et s'étendant à l'infini (figure 2.1). La pression  $p(t)$  est supposée harmonique et telle que  $p(t) = \text{Re}(pe^{-i\omega t})$ , il s'agit donc d'une méthode fréquentielle où  $p$  est l'amplitude complexe et  $\omega = 2\pi f$  la pulsation et  $f$  la fréquence. L'équation de Helmholtz s'écrit alors

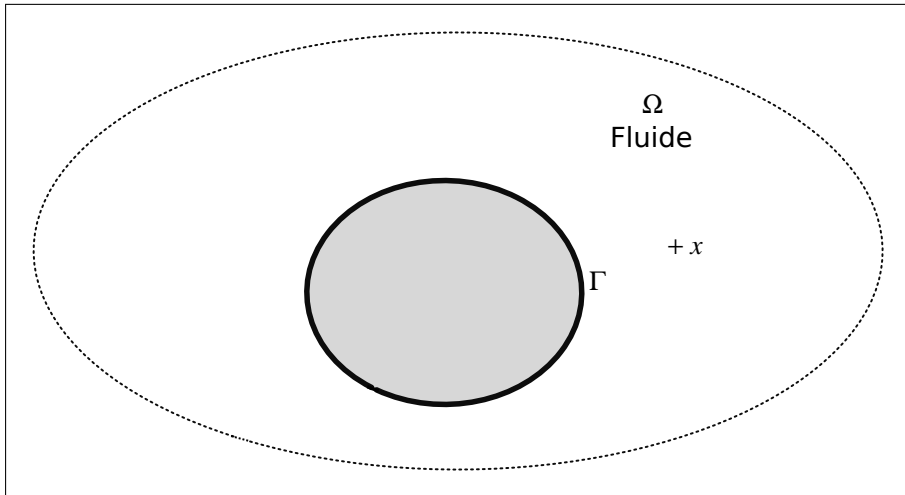
$$\Delta p + k^2 p = s \quad (2.1)$$

où  $k = \omega/c$  est le nombre d'onde,  $c$  la vitesse du son dans l'air et  $s$  le terme source. La résolution de ce problème permet de prédire la pression réfléchiée et diffusée par une structure dans un milieu homogène.

Les conditions aux limites sont définies sur  $\partial\Omega = \Gamma$  par des valeurs à imposer sur  $p$  et/ou ses dérivées partielles (figure 2.1).

### Figure 2.1.

Représentation symbolique du domaine  $\Omega$  ainsi que d'un objet plongé dans  $\Omega$  de contour  $\Gamma$



Si on considère le champ au point  $x$  résultant d'un point source placé en  $y$ , la solution générale du problème en dynamique est appelée fonction de Green  $G$  et vérifie 2.2 :

$$\Delta_{\mathbf{x}} G(\mathbf{x}, \mathbf{y}) + k^2 G(\mathbf{x}, \mathbf{y}) = -\delta(\mathbf{x} - \mathbf{y}), \quad (2.2)$$

où  $\delta$  désigne la distribution de Dirac. Cette solution générale possède des expressions analytiques dans de nombreux cas.

En multipliant 2.1 par  $G(\mathbf{x}, \mathbf{y})$  et 2.2 par  $p(\mathbf{y})$ , en soustrayant ces deux expressions puis en intégrant le tout sur  $\Omega$ , par rapport à un élément de surface  $dS(\mathbf{y})$ , :

$$\underbrace{\int_{\Omega} \delta(\mathbf{x} - \mathbf{y})p(\mathbf{y})dS_{\mathbf{y}}}_{1} = \underbrace{\int_{\Omega} p(\mathbf{y})\Delta G(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}}}_{2} - \underbrace{\int_{\Omega} G(\mathbf{x}, \mathbf{y})\Delta p(\mathbf{y})dS_{\mathbf{y}}}_{3} + \underbrace{\int_{\Omega} s(\mathbf{y})G(\mathbf{x}, \mathbf{y})dS_{\mathbf{y}}}_{4}.$$

Ainsi pour un point  $\mathbf{x}$  situé dans le domaine fluide  $\Omega$ , la définition de la distribution de Dirac montre que l'expression 1 est égale à  $p(\mathbf{x})$ . L'expression 4 est connue sur le domaine  $\Omega$  et est notée  $p_{\text{inc}}(\mathbf{x})$ . On applique alors la formule de Green aux intégrales de surfaces 2 et 3 pour obtenir la formule de Kirchoff 2.3 qui permet d'estimer le champ de pression en un point  $\mathbf{x}$  de  $\Omega$  uniquement à partir des valeurs de la pression  $p$  et de sa dérivée première (où  $\partial_{\mathbf{n}_y}$  désigne la dérivée par rapport à la normale) sur le contour  $\Gamma$  :

$$p(\mathbf{x}) = \int_{\Gamma} p(\mathbf{y}) \frac{\partial G}{\partial \mathbf{n}_y}(\mathbf{x}, \mathbf{y}) d\mathbf{y} - \int_{\Gamma} \frac{\partial p}{\partial \mathbf{n}_y}(\mathbf{y}) G(\mathbf{x}, \mathbf{y}) d\mathbf{y} + p_{\text{inc}}(\mathbf{x}). \quad (2.3)$$

Si on désigne par  $\tilde{p}$  une fonction test, la formulation variationnelle de 2.3 conduit à

$$\begin{aligned} \int_{\Gamma} \tilde{p}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} &= \int_{\Gamma} \int_{\Gamma} \tilde{p}(\mathbf{x}) p(\mathbf{y}) \frac{\partial G}{\partial \mathbf{n}_y}(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &\quad - \int_{\Gamma} \int_{\Gamma} \tilde{p}(\mathbf{x}) \frac{\partial p}{\partial \mathbf{n}_y}(\mathbf{y}) G(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &\quad + \int_{\Gamma} \tilde{p}(\mathbf{x}) p_{\text{inc}}(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (2.4)$$

L'étape suivante consiste à discrétiser le contour  $\Gamma = \bigcup_i \Gamma_i$  et d'appliquer la relation de Chasles dans 2.4 :

$$\begin{aligned} \sum_{i=1}^M \int_{\Gamma_i} \tilde{p}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} &= \sum_{i=1}^M \sum_{j=1}^M \int_{\Gamma_i} \int_{\Gamma_j} \tilde{p}(\mathbf{x}) p(\mathbf{y}) \frac{\partial G}{\partial \mathbf{n}_y}(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &\quad - \sum_{i=1}^M \sum_{j=1}^M \int_{\Gamma_i} \int_{\Gamma_j} \tilde{p}(\mathbf{x}) \frac{\partial p}{\partial \mathbf{n}_y}(\mathbf{y}) G(\mathbf{x}, \mathbf{y}) d\mathbf{y} d\mathbf{x} \\ &\quad + \sum_{i=1}^M \int_{\Gamma_i} \tilde{p}(\mathbf{x}) p_{\text{inc}}(\mathbf{x}) d\mathbf{x}. \end{aligned} \quad (2.5)$$

La discrétisation de  $\Gamma$  possède  $N$  nœuds et  $M$  éléments. On se donne une expression discrète de  $p$  et sa dérivée  $q$  sur des éléments de formes  $(N_i(\mathbf{x}))_{i \leq N}$  de la manière suivante :

$$p(\mathbf{x}) = \sum_{i=k}^N p_k N_k(\mathbf{x}) \quad q(\mathbf{x}) = \sum_{k=1}^N q_k N_k(\mathbf{x}). \quad (2.6)$$

On ré-injecte alors 2.6 dans 2.5, en prenant en compte les valeurs connues sur les  $\Gamma_i$ . On intègre ensuite les fonctions de formes ainsi que les fonctions  $G$  sur ces même  $\Gamma_i$ , et on regroupe les termes connus et les termes inconnus. On obtient un système linéaire en  $\mathbf{X} = (p_i, q_i)$  de la forme  $\mathbf{KX} = \mathbf{F}$  où la matrice  $\mathbf{K}$  est une matrice «pleine» ne possédant pas forcément les propriétés numériques permettant une inversion rapide.

L'application aux écrans antibruit de cette méthode intégrale a déjà donné lieu à plusieurs travaux (DUHAMEL, 1996 ; JEAN, 1998). La BEM permet de prendre en compte des formes d'écran complexes ainsi que des changements d'impédance : il faut pour cela s'assurer que la surface est correctement discrétisée. Le CSTB possède une grande expérience de cette méthode et en a développé une version optimisée (en tabulant certaines fonctions, comme la fonction de Hankel qui est une résolution de 2.2 dans certains cas), aboutissant ainsi à un outil de prédiction efficace, éprouvé et très rapide. Dans le cadre de ce travail, ce sont majoritairement des scènes 2D qui ont été traitées par la BEM. Ceci implique que la géométrie est invariante selon  $y$  et se résume au plan  $xz$ . Il en va de même pour la source, qui se retrouve être une ligne source cohérente.

Il est possible en outre de définir l'approche 2.5D décrite dans (DUHAMEL, 1996) en considérant une ligne source incohérente. Cette approche permet d'obtenir une solution 3D par une manipulation peu coûteuse numériquement. Reposant sur les propriétés de la transformation de Fourier, cette approche permet de limiter les calculs à des configurations 2D, ce qui constitue un gain en temps de calcul intéressant par rapport à un calcul 3D complet.

La principale faiblesse de la BEM tient dans le fait qu'elle n'est valable que dans un milieu homogène. Ainsi l'atmosphère doit rester homogène pour notre application et il n'est pas possible de prendre en compte les effets météorologiques. Numériquement, la BEM est de moins en moins rapide lorsque que l'on considère des fréquences élevées, car le maillage exigé doit être plus raffiné et cette augmentation du nombre de points alourdit les calculs.

La méthode Adrienne (cf 2.3.1) demandant de monter à 5000 Hz, et étant donné la condition d'échantillonnage de Shanon, il est nécessaire de monter à 11000 Hz pour passer du domaine fréquentiel au domaine temporel par transformation de Fourier inverse. Ces hautes fréquences constituent un frein chronophage aux méthodes d'optimisation décrites ci-après (2.2.4).

### 2.2.2. L'analyse en cycle de vie

L'analyse du cycle de vie (LCA pour Life Cycle Assessment en anglais) est une méthode normalisée (NF-P-01010, 2004 ; ES-PREN-15804, 2010 ; ISO-14044, 2006 ; ISO-14040, 2006) permettant d'évaluer l'impact environnemental d'un produit ou d'un service de façon quantifiée.

Elle fait apparaître la notion d'unité fonctionnelle, qui constitue une grandeur de référence dont la définition dans l'analyse pratiquée est définie par «Assurer la fonction d'un équipement de réduction du bruit pendant un an sur un mètre linéaire». Les matériaux considérés sont supposés être de très bonne qualité au sens où leur durabilité supposée est de 20 ans (correspondant au matériau le moins durable, à savoir le bois de construction).

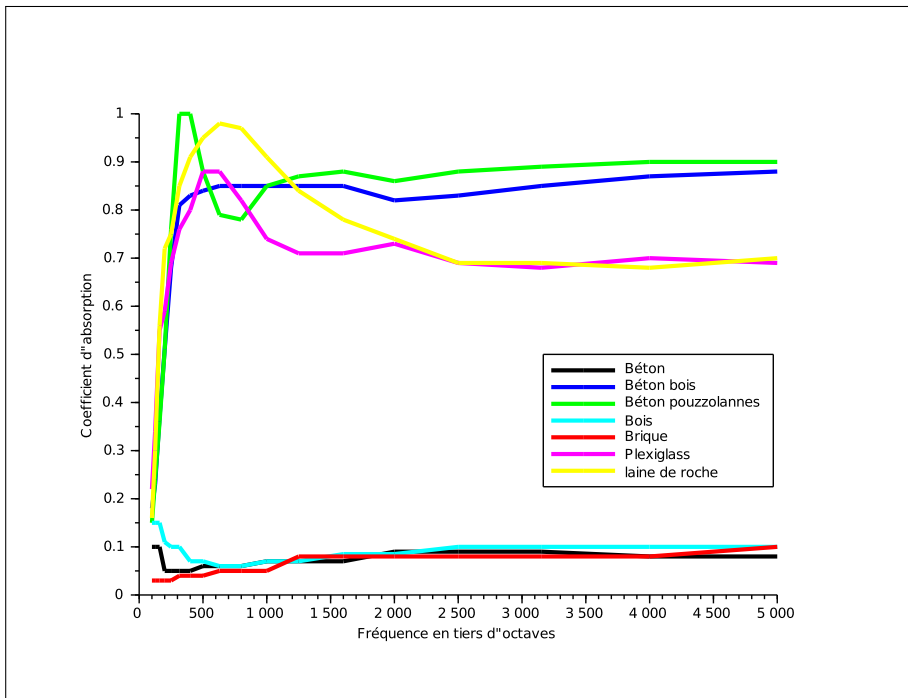
Les matériaux soumis à l'analyse sont ceux couramment utilisés dans la construction d'écrans antibruit. Ils sont au nombre de 9 : le béton armé, le béton de pouzzolane, le béton-bois, le bois, la brique, l'aluminium, le plexiglass, la laine de roche et l'acier perforé.

### 2.2.3. La description acoustique des matériaux

La description acoustique des matériaux est importante pour prendre en compte les propriétés absorbantes. En effet, plusieurs mécanismes de diffusion interviennent au niveau microscopique des matériaux, mais les modèles de comportement se situent généralement au niveau macroscopique. La BEM étant une méthode fréquentielle, il est naturel de se ramener à une impédance (i.e. un rapport pression/vitesse) qui sera prise en compte sur chaque surface discrétisée dans le modèle numérique. Ainsi les matériaux sont décrits acoustiquement sous la forme d'une impédance, elle-même déduite d'une mesure d'un coefficient d'absorption  $\alpha$  donné en tiers d'octave (figure 2.2).

**Figure 2.2.**

Courbes d'absorption des 9 matériaux retenus pour l'optimisation



Il ne s'agit pas d'une caractérisation, car aucun paramètre macroscopique n'a été utilisé au sein d'un modèle prédictif pour fournir les impédances. La méthode pour passer du coefficient  $\alpha$  à l'impédance est tirée de (MORSE et al., 1986). Elle suppose la norme de l'admittance petite devant 1, ce qui permet de supposer  $\alpha$  proportionnel à la partie réelle de l'admittance. Cette approche limite donc l'impédance à une grandeur réelle, ce qui ne reflète pas la réalité. Il est possible d'améliorer la description acoustique des matériaux en faisant appel à des modèles plus poussés, mais cela demanderait de caractériser tous les matériaux susceptibles d'intervenir dans l'optimisation.

#### 2.2.4. Les outils d'optimisation

Les techniques d'optimisations utilisés dans ce travail reposent sur des algorithmes évolutionnaires (*evolutionary computation* en anglais), c'est à dire qu'ils proposent successivement des solutions à un problème donné, dans l'optique de déterminer un résultat optimal. Ces algorithmes sont employés lorsque la taille de l'espace des possibles exclut un parcours exhaustif ou lorsque le domaine à explorer n'est pas convexe. Les différentes solutions proposées itérativement lors de la résolution sont issues d'une marche aléatoire, ce qui les classe dans la catégories des algorithmes stochastiques.

Deux outils principaux d'optimisation seront mis en œuvre dans ce travail, selon qu'on concentre sur un ou plusieurs objectifs (optimisation mono- ou multi-critères).

Pour une optimisation avec un unique objectif, on utilisera une stratégie d'évolution (ES en anglais pour *Evolution Strategy*), tandis que pour une optimisation à plusieurs objectifs, on s'orientera vers un algorithme génétique de tri non dominé (NSGA en anglais pour *Non-dominated Sorting Genetic Algorithm*).

La stratégie d'évolution employée est une stratégie binaire. Un individu (ici un écran caractérisé par ses dimensions géométriques et ses matériaux) est représenté par un vecteur de données et appartient à une certaine population. Cet individu est alors muté en modifiant le vecteur de données qui le caractérise. Cette «mutation» est opérée avec une loi uniforme de moyenne nulle et de variance fixée au départ (la plupart du temps par des contraintes physiques).

Nous obtenons alors un descendant dont nous comparons la performance avec son antécédent. Si elle est supérieure, le descendant remplace son antécédent dans la population qui servira à générer la population suivante. Le taux de mutation correspond à la variance de la loi normale de moyenne nulle et de variance fixée. Il est possible d'améliorer l'algorithme en procédant à un ajustement au cours des générations de la variance de la loi normale utilisée pour générer des descendants. Nous adaptions ainsi cette variance afin d'obtenir 1/5 de mutations performantes.

Les algorithmes génétiques de tri non dominé (SRINIVAS et al., 1994) sont basés sur un classement des individus sur plusieurs fronts. La première étape consiste en une génération aléatoire d'une population initiale d'individus parents. Cette

population est classée en plusieurs fronts de rangs différents. Chaque individu  $P$  est comparé à tous les autres individus et tous les individus non dominants sont rassemblés sur le front de rang 1, appelé front optimal de Pareto. Ce front est ensuite mis de côté pour classer les autres individus. Au sein de chaque front, un critère de sélection permet de générer des descendants selon un processus qu'il serait trop long à décrire ici et qu'on pourra consulter dans (SRINIVAS et al., 1994).

Cet algorithme favorise les solutions non dominées et il utilise une variété explicite des solutions, par contre il se révèle très gourmand en nombre d'évaluations de fonctions coût, ce qui dans notre cas d'optimisation rend son application impossible à mettre en pratique au vu des exigences de calculs dans le cadre général. C'est la raison pour laquelle on s'est orienté vers des optimisations partielles (voir la section 2.4), perdant sans doute sur le plan de la généralité du problème, mais gagnant certainement sur le plan de la mise en œuvre.

## 2.3. Indicateurs de performance

Préalablement à l'optimisation, il est nécessaire de se donner des critères quantifiables afin de sélectionner les candidats répondant au mieux aux dits critères. Les critères que nous allons mettre en avant seront déclinés au travers de fonctions coût et sont issus des outils exposés ci-dessus (cf section 2.2). Pour les critères acoustiques, il est possible de s'appuyer sur des grandeurs physiques par nature aisément quantifiable. Ces critères sont présentés dans 2.3.1. Les critères économiques sont décrits dans 2.3.2 et restent les plus faciles à appréhender et à quantifier. Par contre les critères environnementaux exigent plus de travail, que ce soit en terme de sélection des critères et de collecte de données. En effet la quantification d'un critère environnemental, après analyse en cycle de vie, repose sur une étude amont pour chaque composant de l'ouvrage, et ce travail de recensement est difficilement automatisable et demande une expertise bien spécifique. Les résultats sur les critères retenus sont donnés dans 2.3.3.

### 2.3.1. Les critères acoustiques

Les grandeurs acoustiques considérées ici sont des niveaux calculés dans les 18 bandes de tiers d'octave comprises entre 100 Hz et 5 kHz, selon la norme CEN1793-5 :2003.  $f_i$  désigne la fréquence centrale de la bande de tiers d'octave  $i$ ,  $\Delta f_i$  en désigne la largeur.  $\mathcal{F}$  désigne la transformée de Fourier. La méthode de calcul reproduit ici la méthode Adrienne (CEN1793, 2003c), qui permet de mesurer in situ l'absorption et la transmission sonore de tout type de dispositif antibruit. Les simulations numériques produisant le champ de pression aux abords du dispositif, on applique des fenêtrages aux signaux pour séparer l'onde directe de l'onde réfléchie sur l'écran, ainsi que les autres réflexions parasites. Ainsi ces fenêtrages apparaissent dans chaque expression sous la forme d'une fonction temporelle  $w_*(t)$  que l'on multiplie au signal simulé.

Considérons à présent chaque indice acoustique dans une bande de fréquence donnée.

### 2.3.1.1. Indice de réflexion dans la bande $i$

$RI_i$  désigne l'indice de réflexion et est donné par l'équation 2.7. Le signal réfléchi est évalué avec la BEM. La norme CEN1793-5 :2003 prévoit 9 positions angulaires de microphone entre  $30^\circ$  et  $150^\circ$ , en considérant l'incidence normale à  $90^\circ$ .  $n_j$  désigne donc le nombre de positions angulaires considérées pour obtenir une moyenne sur l'ensemble de ces positions.

Nous avons :

$$RI_i = \frac{\frac{1}{n_j} \sum_{k=1}^{n_j} \int_{\Delta f_i} |\mathcal{F} [\sqrt{t} h_{r,k}(t) w_r(t)] (f)|^2 df}{\int_{\Delta f_i} |\mathcal{F} [\sqrt{t} h_i(t) w_i(t)] (f)|^2 df} \quad (2.7)$$

où  $h_{r,k}$  désigne le signal réfléchi pour la position angulaire  $k$ ,  $w_r$  la fenêtre de l'onde réfléchie,  $h_i$  l'onde incidente et  $w_i$  la fenêtre du signal incident.

Il faut noter que nous avons employé l'expression faisant intervenir  $\sqrt{t}$  dans les intégrales au numérateur et dénominateur, car les simulations sont réalisées en 2D avec la BEM, ce qui nécessite ce facteur correctif.

### 2.3.1.2. Indice d'absorption dans la bande $i$

$SL_i$  désigne l'indice de transmission et est donné par l'équation 2.8. Le signal transmis est évalué avec la TMM. La nouvelle méthode prévoit une grille de 3x3 microphones placée du coté opposé à l'émission.  $n_j$  désigne donc l'ensemble des microphones sur la grille et permet d'obtenir une moyenne sur l'ensemble de ces positions.

Nous avons :

$$SL_i = \frac{\frac{1}{n_j} \sum_{k=1}^{n_j} \int_{\Delta f_i} |h_{t,k}(t) w_t(t)|^2 df}{\int_{\Delta f_i} |h_i(t) w_i(t)|^2 df} \quad (2.8)$$

où  $h_{t,k}$  désigne le signal transmis pour la position angulaire  $k$ ,  $w_t$  la fenêtre de l'onde transmise,  $h_i$  l'onde incidente et  $w_i$  la fenêtre du signal incident.

### 2.3.1.3. Indice de diffraction dans la bande $i$

$DL_i$  désigne l'indice de diffraction et est donné par l'équation 2.9. Le signal diffracté est évalué avec la BEM. La norme (CEN1793, 2003c) préconise 4 positions de microphones de part et d'autre de l'écran.  $n_j$  désigne donc ces quatre positions de microphone.

Nous avons :

$$DI_i = \frac{\frac{1}{n_j} \sum_{k=1}^{n_j} \left(\frac{d_k}{d_i}\right)^2 \int_{\Delta f_i} |\mathcal{F}[h_{d,k}(t) w_{d,k}(t)](f)|^2 df}{\int_{\Delta f_i} |\mathcal{F}[h_i(t) w_i(t)](f)|^2 df} \quad (2.9)$$

où  $h_{d,k}$  désigne le signal au microphone  $k$ , et  $w_{d,k}$  la fenêtre spécifique à chaque microphone,  $h_i$  l'onde incidente et  $w_i$  la fenêtre du signal incident.  $d_i$  et  $d_k$  sont des facteurs de correction géométrique.

En pratique, l'indice pertinent pour la diffraction est obtenu en évaluant le gain par insertion en considérant la différence de l'indice  $DI_i^m$  en présence de l'écran avec l'indice  $DI_i^0$  sans l'écran. Nous noterons ainsi  $\Delta DI_i = DI_i^m - DI_i^0$ .

#### 2.3.1.4. Indicateurs globaux

Il reste à sommer chaque contribution fréquentielle pour chaque indice sur l'ensemble du spectre en tiers d'octave en introduisant les pondérations ( $L_i$ ) dans chaque bande de fréquence.

En posant  $K = \sum_{i=1}^{18} 10^{L_i/10}$  :

$$DL_{RI} = -10 \log \frac{1}{K} \sum_{i=1}^{18} RI_i 10^{L_i/10} \quad (2.10)$$

$$DL_{SI} = -10 \log \frac{1}{K} \sum_{i=1}^{18} SL_i 10^{L_i/10} \quad (2.11)$$

$$DL_{\Delta DI} = -10 \log \frac{1}{K} \sum_{i=1}^{18} (DI_i^m - DI_i^0) 10^{L_i/10}. \quad (2.12)$$

### 2.3.2. Les critères économiques

L'aspect économique est un critère essentiel dans l'élaboration d'un écran antibruit, c'est pourquoi des critères supplémentaires sont nécessaires pour intégrer le coût financier d'un ouvrage. Trois paramètres ont été introduits, à savoir le coût de construction, le coût de maintenance ainsi que le coût de démolition.

La table 2.1 donne le coût de construction et de démolition (matériaux, main d'œuvre, transport) d'un écran antibruit pour différentes hauteurs, en partant de l'hypothèse que ces coûts sont indépendants des matériaux utilisés (voir par exemple la fiche technique CERIB, 2017 pour des informations détaillées sur les coûts liés au béton plein).



Dans la table 2.1, la valeur obtenue est ramenée au coût de la solution de référence de l'écran de 4 m de haut et de 10 cm d'épaisseur en béton plein. La même normalisation est adoptée pour les coûts de maintenance et de démolition.

**Table 2.1.**

Exemple de coût de construction et de démolition en fonction de la hauteur de l'écran

Hauteur de l'écran	Coût de construction en €/ m	Coût de démolition en €/ m
1	825	278
2	1449	312
3	2066	345
4	2678	379
10	6441	579

La table 2.2 donne quelques exemples de coûts d'entretien des ouvrages en fonction des matériaux utilisés. Le choix des matériaux intervient donc ici pour les critères économiques.

**Table 2.2.**

Exemple de coût de maintenance des matériaux utilisés

Matériau	Coût de maintenance en €/ m <sup>2</sup> / an
Béton et brique	2.93
Bois	2.44
Plaque de métal	1.46
Thermoplastique transparent	1.48

Il est à noter que le coût de réutilisation des matériaux n'a pas été pris en compte dans le cadre de ce projet, ce qui aurait pour conséquence de faire baisser éventuellement les coûts de construction initiaux en cas de recyclage ainsi que les coûts de démolition une fois l'ouvrage «recyclé».

Tous ces coûts sont fixés avant l'optimisation, et donc ne prennent pas en compte les variations économiques qui peuvent intervenir durant la vie de l'ouvrage ni les réparations éventuelles en cas de détérioration accidentelle.

### 2.3.3. Les critères environnementaux

La sélection des critères environnementaux s'est appuyée sur une analyse en cycle de vie des matériaux considérés en veillant à l'indépendance et à la pertinence environnementale des critères retenus. Les critères retenus sont au nombre de 4 : l'énergie nécessaire à la fabrication du matériau, la quantité de CO<sub>2</sub> dégagée, la perte de matériau lors de la mise en œuvre et la consommation en eau. Le résultat de ce travail d'analyse du cycle de vie sur les 9 matériaux cités au 2.2.2 conjointement avec ces quatre critères nous amène à la table 2.3,

qui présente les résultats sur la base d'une production et du transport de 1 tonne de matériau sur 100 km.

**Table 2.3.**

Critères environnementaux liés aux matériaux considérés

Indicateurs environnementaux				
Matériau	Énergie Mj	CO <sub>2</sub> dégagée kg CO <sub>2</sub>	perte à la mise en œuvre en kg	Consommation en eau en litre
Béton armé	1,14E+3	1,43E+02	2,42E+01	1,82E+03
Béton bois	7,87E+03	3,67E+02	4,03E+01	1,69E+03
Béton pouzzolannes	5,46E+3	6,18E+01	5,35E+00	2,02E+02
Brique	3,02E+3	2,51E+02	7,17E+00	4,89E+02
Plexiglass	1,45E+5	8,40E+03	1,07E+02	2,03E+04
Laine de roche	1,92E+4	1,05E+03	3,39E+02	1,00E+04
Bois	2,22E+4	1,49E+02	2,47E+01	1,27E+03
Aluminium	1,46E+5	9,28E+03	2,56E+03	4,39E+04
Acier	3,40E+4	2,29E+03	2,21E+03	2,50E+04

Pour chaque écran étudié, on établit le rapport des unités fonctionnelles de l'écran et de la solution de référence (écran droit en béton de 10 cm d'épaisseur et de 4 m de hauteur).

## 2.4. Optimisation intrinsèque d'un écran antibruit

La démarche débute par une optimisation partielle, c'est à dire que l'écran antibruit est isolé de son contexte extérieur et que seules les performances de l'écran sont considérées. Cette première optimisation est désignée par le terme d'*optimisation intrinsèque* et permet de hiérarchiser les candidats sur le seul plan des performances acoustiques.

### 2.4.1. Définition

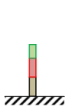
On désigne par optimisation intrinsèque la démarche qui vise à sélectionner le meilleur candidat d'écran possible sans tenir compte de son environnement au sens général (type de source, topographie, environnement urbain ou rural, etc). Cette optimisation est la plus courante lorsque l'on se concentre uniquement sur des critères acoustiques. Cette première optimisation a été menée en faisant varier formes et matériaux pour chaque famille d'écrans antibruit recensée. La technique utilisée pour chaque écran est la BEM (et la TMM pour une des sortes de famille) conjointement à une stratégie d'évolution. Les matériaux possibles sont ceux de la liste donnée ci-dessus (cf 2.2.2).

Pour chaque type d'écran, un procédé de construction géométrique permet de générer un écran candidat en fonction d'un ensemble de paramètres propre à chaque famille. Nous décrivons 4 familles particulières dans la table 2.4.

Ces 4 familles sont celles retenues pour l'optimisation globale (cf 2.5). Les intervalles avec une virgule désignent des intervalles continus de réels, tandis que les doubles points décrivent des ensembles d'entiers. Les puissances sur les intervalles indiquent le nombre de paramètres variant dans cet intervalle. Les couleurs sur les écrans indiquent l'emplacement des matériaux. Cette description condensée des familles d'écrans ne doit pas masquer l'ensemble des paramètres sur lesquels il est possible d'influer pour obtenir une solution optimale.

**Table 2.4.**

Description de 4 familles d'écran donnant la géométrie de l'écran ainsi que les positionnements possibles des matériaux sur les surfaces



Paramètre	unité	intervalle
largeur de l'écran	mètre	[0.1, 0.5]
longueur de l'écran	mètre	[1, 5]
inclinaison	degrés	[-5, +5]
matériaux	n°	[1..9] <sup>[2..5]</sup>
hauteurs des panneaux	mètre	[1, 5] <sup>[2..5]</sup>



Paramètre	unité	intervalle
largeur de l'écran	mètre	[0.1, 0.5]
longueur de l'écran	mètre	[1, 5]
inclinaison	degrés	[-5, +5]
matériaux	n°	[1..9] <sup>3</sup>
taille des aspérités	mètre	[0.08, 0.17]
angles des aspérités	degrés	[10, 80] <sup>2</sup>



Paramètre	unité	intervalle
largeur de l'écran	mètre	[0.1, 0.5]
longueur de l'écran	mètre	[1, 5]
inclinaison	degrés	[-5, +5]
matériaux	n°	[1..9]
taille des courbes	mètre	[0.08, 0.17]
angle des courbes	degrés	[10, 80] <sup>2</sup>



Paramètre	unité	intervalle
largeur de l'écran	mètre	[0.1, 0.5]
longueur de l'écran	mètre	[1, 5]
inclinaison	degrés	[-5, +5]
matériaux	n°	[1..9] <sup>2</sup>
Points de Bézier 1	mètre	[-1.5, -0.1] <sup>2</sup>
Points de Bézier 2	mètre	[0.1, 1.5] <sup>2</sup>

## 2.4.2. Performances attendues

Les performances intrinsèques d'un écran antibruit sont évaluées selon 3 critères qui consistent à ramener les gains relatifs du mur candidat par rapport aux critères d'un écran de référence de 10 cm d'épaisseur et de 4 m de haut en béton plein. On détermine ainsi des seuils de performance à atteindre comme le résume la table 2.5.

**Table 2.5.**

Seuils de performance à atteindre en termes de gain relatif pour les trois critères acoustiques

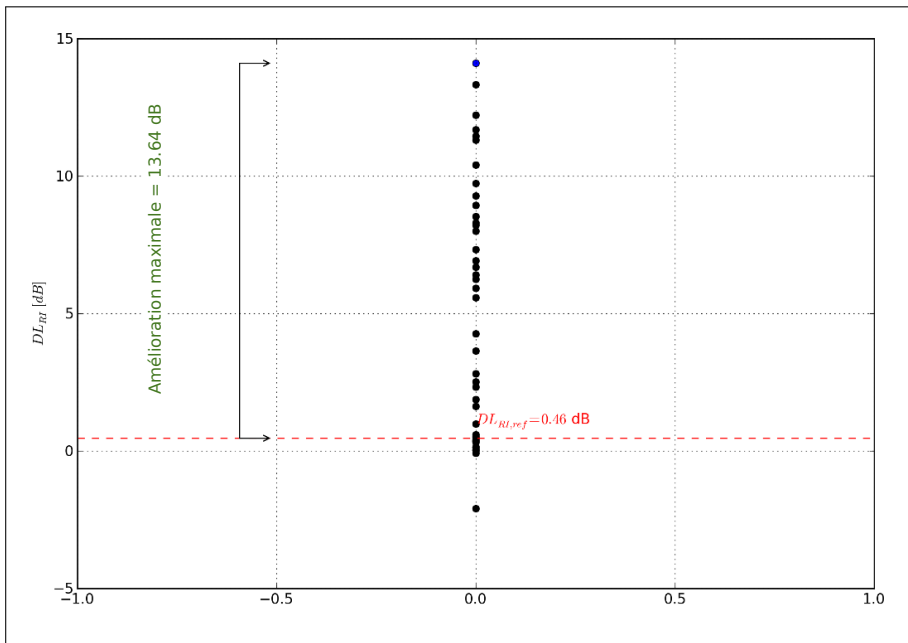
	Gain relatif	Seuil de performance
En réflexion	$DL_{RI} - DL_{RI}^{ref}$	12 dB(A)
En absorption	$DL_{SL} - DL_{SL}^{ref}$	60 dB(A)
En diffraction	$DL_{DI} - DL_{DI}^{ref}$	12 dB(A)

Ces seuils de performances ont été fixés dans l'esprit d'améliorer grandement la solution de référence, montrant ainsi que la modification des formes et des matériaux des écrans constituent des leviers efficaces d'optimisation.

### 2.4.3. Résultats d'optimisation intrinsèque

La figure 2.3 représente les gains obtenus pour chacun des écrans générés afin de sélectionner le meilleur candidat. Il n'y a qu'un seul critère pris en compte ici pour l'optimisation, à savoir le gain en réflexion. Nous pouvons obtenir ici un gain allant jusqu'à 13,64 dB par rapport à la solution de référence.

**Figure 2.3.**  
Exemple d'optimisation avec un seul critère



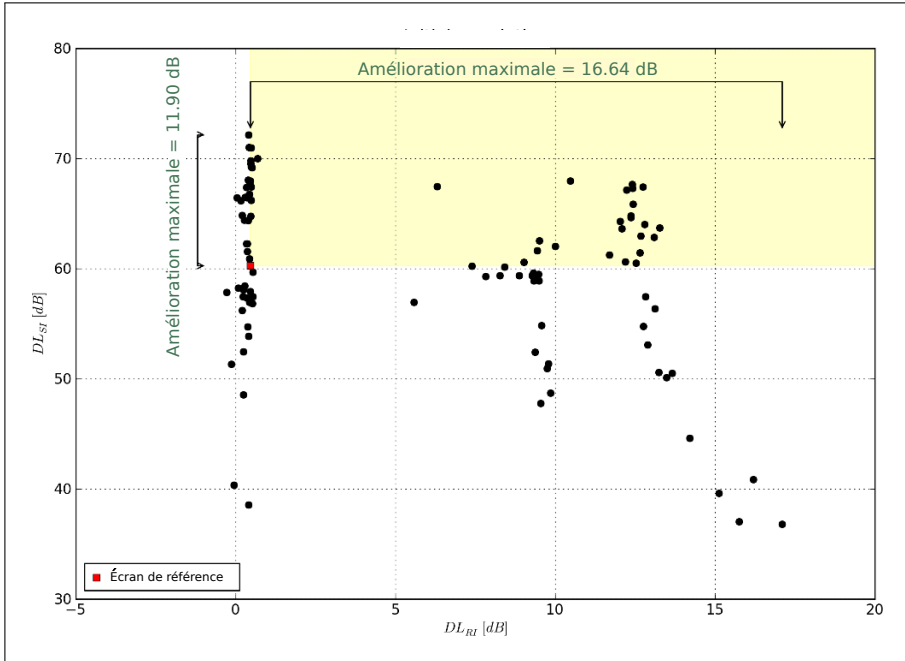
La figure 2.4 quant à elle s'intéresse à une optimisation à deux critères de façon conjointe, ce qui permet de dégager une zone d'amélioration (en jaune). Cette optimisation montre qu'il est possible de dégager de meilleurs candidats que le mur de référence.

Par contre, il est nécessaire d'établir une priorité entre les critères si l'on vise à ne retenir qu'un seul candidat. Le critère final dans le cas d'une optimisation multi-critères retenu par la suite sera la somme maximale des gains sur les deux critères acoustiques.

Avec les méthodes prédictives choisies, notamment la BEM, il n'est malheureusement pas possible d'évaluer tous les critères acoustiques en 2D décrits en 2.3.1 pour toute les familles d'écrans retenues. La table 2.6 indique les possibilités de calcul d'indicateur pour chaque famille avec les méthodes numériques choisies. Une croix signifie que le critère ne peut pas être évalué,

0 indique que le critère est immédiatement atteint lors de la première génération des candidats au cours de l'optimisation.

**Figure 2.4.**  
Exemple d'optimisation avec 2 critères



**Table 2.6.**  
Nombre de populations de candidats générées par la méthode d'optimisation, par critère acoustique et par type d'écran (l'impossibilité de calcul du critère est indiquée par le symbole X)









$\Delta DL_{Sj}$	0	X	0	X	X	X	X	X	X
$\Delta DL_{RI}$	1	3	2	4	2	4	4	2	2
$DL_{\Delta D}$	X	X	X	X	X	X	X	1	2

Pour une hauteur d'écran donnée de 4 m, chaque famille a fait l'objet d'une optimisation intrinsèque sur chaque critère acoustique. Ces optimisations successives permettent d'évaluer les gains obtenus en séparant les critères. La table 2.7 donnent les résultats obtenus et force est de constater que les résultats

de l'écran de référence peuvent être améliorés de manière significative sur les indicateurs définis.

**Table 2.7.**

Gains en dB(A) apportés par l'optimisation pour chaque critère acoustique et chaque famille pour une hauteur de 4 m

	Hauteur du mur									
$\Delta DL_{SI}$	4 m	11.9	X	40.3	X	X	X	X	X	X
$\Delta DL_{RI}$	4 m	16.6	15.6	15.5	17.3	15.7	17.1	17.6	15.9	14.1
$DL_{\Delta DI}$	4 m	X	X	X	X	X	X	X	12.7	14.3

Ainsi ces résultats d'optimisation intrinsèques permettent d'une part, de constater la pertinence de la recherche d'une solution optimale et d'autre part, de réduire les intervalles de recherche lors des optimisations globales.

## 2.5. Optimisation globale de murs antibruit

### 2.5.1. Principe de la démarche

L'optimisation globale ou extrinsèque consiste à tenir compte de l'environnement extérieur et à mettre en jeu des paramètres n'ayant pas trait à l'acoustique.

La méthode reste dans l'esprit d'une stratégie d'évolution, c'est à dire qu'un ensemble de candidats est généré parmi lesquels sont sélectionnés progressivement par mutation des candidats solutions du problème posé. Les conditions extérieurs à l'ouvrage, à savoir la source de bruit, l'environnement urbain ou rural, ainsi que la topographie sont des facteurs à prendre en compte en amont avant de procéder à la recherche d'une solution optimale. L'environnement au sens large de l'ouvrage est ainsi pris en compte directement dans la méthode globale.

Une telle optimisation globale met en jeu un grand nombre de critères qu'il est difficile de relier entre eux. Pour cette raison, on procède dans un premier temps à une agrégation des information sous la forme d'un indicateur spécifique à chaque composante de l'optimisation, à savoir les aspects acoustiques, les aspects environnementaux et les aspects économiques. La mise en place de ces indicateurs agrégés est décrite au 2.5.2.

### 2.5.2. Les indicateurs globaux ACOU, ENV et COST

Afin de d'aider à la prise de décision, il est essentiel d'agréger l'ensemble des informations au travers d'indicateurs en nombre réduits. Pour ce faire, un

système de notation a été mis en place pour chacun des 3 domaines (acoustique, environnemental et économique). Trois indicateurs ont ainsi été définis, chacun étant défini par rapport à l'écran de référence de 4 m de 10 cm d'épaisseur en béton plein.

Le premier est l'indicateur pour les aspects acoustiques  $\Delta L$ , défini par la moyenne des gains en dB(A) du candidat évalué sur les critères introduits au 2.3.1 par rapport à l'écran de référence (cet indicateur possède une unité).

Le deuxième indicateur concerne les aspects environnementaux  $X_{env}$  et le troisième concerne les aspects économiques  $X_{cost}$ . Ils sont définis comme la moyenne des critères définis au 2.3.2 et au 2.3.3, quotientées par la moyenne des critères obtenus avec l'écran de référence. Chacune de ces grandeurs, avec ou sans dimension, est transformée en une note allant de 0 à 4. Les tables 2.8, 2.9 et 2.10 donnent les systèmes de notation utilisés pour déterminer les indicateurs ACOU, ENV et COST.

**Table 2.8.**

Système de notation pour l'indicateur ACOU

Gain $\Delta L$ du niveau (en dB(A))	$\Delta L < 1$	$1 < \Delta L < 2$	$2 < \Delta L < 4$	$4 < \Delta L < 6$	$\Delta L > 6$
Note	0	1	2	3	4

**Table 2.9.**

Système de notation pour l'indicateur ENV

$x = X_{env}$	$x < 25\%$	$25\% < x < 50\%$	$50\% < x < 75\%$	$75\% < x < 100\%$	$100\% < x$
Note	4	3	2	1	0

**Table 2.10.**

Système de notation pour l'indicateur COST





$x = X_{cost}$	$x < 60\%$	$60\% < x < 70\%$	$70\% < x < 80\%$	$80\% < x < 90\%$	$90\% < x$
Note	4	3	2	1	0

### 2.5.3. Choix du mur et de son environnement extérieur

La table 2.11 présente les choix possibles pour l'environnement extérieur. La définition ces choix permet d'adapter au mieux une optimisation d'un mur antibruit dans son environnement et son contexte.

**Table 2.11.**

Possibilités considérées pour l'optimisation extrinsèque

Famille de murs				
Topographie	déblai	plat	remblai	
Environnement	urbain	rural		
Source	ferroviaire	route		

Le choix de la famille conditionne les paramètres géométriques et les matériaux utilisés lors de la génération des candidats. La topographie, l'environnement urbain ou rural ainsi que les choix de la source modifient l'application de la BEM, surtout lors de la discrétisation des surfaces avec leur impédance propre.

### 2.5.4. Récapitulatif de la méthodologie d'optimisation globale

La figure 2.5 récapitule sur un schéma la démarche d'une optimisation globale. Il s'agit de faire évoluer une population de murs d'une même famille, afin de dégager un individu optimal pour les critères que l'on se donne. Ces critères sont matérialisés par les 3 indicateurs ACOU, ENV et COST, se référant respectivement à l'acoustique, l'impact sur l'environnement et le coût du mur antibruit.

Ces indicateurs concentrent plusieurs informations et sont chacun calculés avec des méthodes différentes. Les aspects acoustiques sont issus de la simulation numérique par BEM et des données sur l'impédance des matériaux mis en oeuvre. L'impact environnemental et le coût sont évalués en fonction des bases de données créées pour ce travail d'optimisation. La figure suivante représente la méthode globale d'optimisation, avec un exemple de résultat possible.

## 2.6. Les outils obtenus

### 2.6.1. Pondération des indicateurs

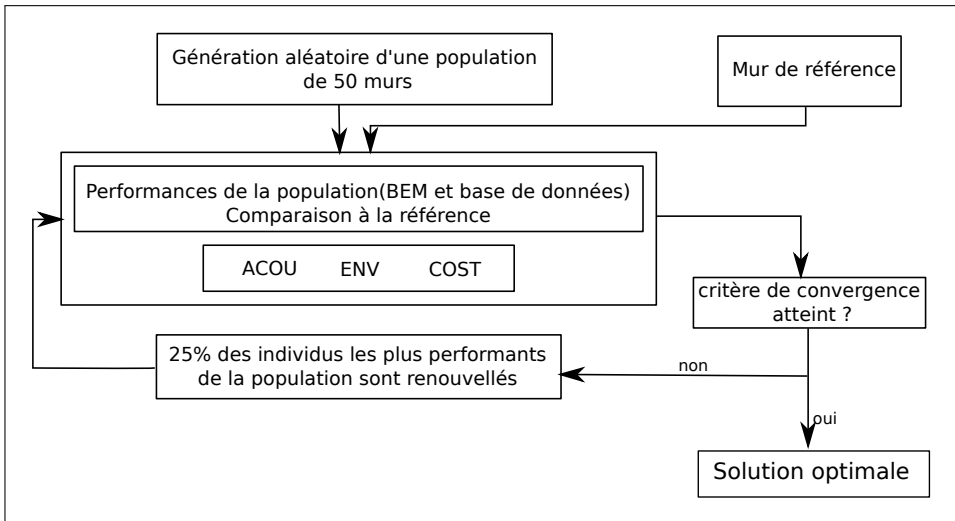
L'utilisateur de la présente méthode peut également être amené à modifier l'importance relative des indicateurs ACOU, ENV, ou COST. En effet, il est possible d'attribuer une pondération à chacun de ces indicateurs selon 3 pourcentages. 0% signifie que l'aspect doit être occulté et que l'utilisateur ne désire pas en tenir compte. 50% signifie que l'importance de l'indicateur est moyenne et 100% indique que ce paramètre est de la plus haute importance.

Cette approche permet ainsi au décideur de positionner le barycentre de son projet entre ces trois aspects. Ainsi, le choix de privilégier un aspect par rapport à l'autre est quantifié et permet d'orienter la décision, surtout lors de la phase très



**Figure 2.5.**

Schéma récapitulatif d'une optimisation globale, une fois les paramètres d'environnement fixés



importante d'avant projet où il s'agit d'évaluer conjointement l'efficacité, le coût et retombées environnementales de l'ouvrage.

### 2.6.2. Constitution d'une base de données sur les écrans antibruit

Il y a quatre choix à effectuer avant une optimisation globale, à savoir le type de source, l'environnement rural ou urbain, la topographie et la famille d'écrans étudiée. De plus, pour l'indicateur ACOU, il est possible de choisir entre 3 configurations sur la position des récepteurs (tous du même coté que la source, tous du coté opposé à la source, ou de part et d'autre de la source). Enfin, la pondérations des indicateurs ACOU, ENV et COST à 0%, 50% ou 100% rajoute encore 3 possibilités supplémentaires.

Il est alors facile d'entrevoir le dénombrement exhaustif de l'ensemble des situations possibles. Ainsi les choix disponibles mènent à  $2 \times 2 \times 3 \times 4 \times 3 \times 3 = 432$  cas qui, dans le cadre du projet européen QUIESST, ont tous fait l'objet de l'optimisation globale décrite en 2.5.4. Il est sans doute inutile de préciser que cette étape a été une phase très chronophage au sein du projet.

Les informations ainsi recueillies ont été ainsi répertoriées dans une base de données qui constitue un des produits essentiels du travail réalisé par le projet. Cette base de données est ainsi un outil permettant de cibler une solution d'écran à mettre en place en intégrant les données environnementales extérieures et en précisant les priorités éventuelles des aspects acoustiques ou environnementaux ou économiques dans le choix de la solution finale. Cette base de données, qui porte sur les écrans, devient un véritable outil d'aide à la décision et peut être évidemment affinée et complétée dans le futur.

### 2.6.3. Présentation des résultats d'une optimisation

Les indicateurs ACOU, ENV et COST ayant des significations très différentes, il a fallu trouver un système de représentation adaptée. Ainsi les performances des 3 indicateurs ont été placées sur un cercle composé lui même de 5 cercles concentriques correspondant aux notes des indicateurs (tables 2.8, 2.9 et 2.10) : on trace alors un triangle ayant pour sommets les notes obtenues. Appelée *radar-plot* en anglais, cette représentation permet d'évaluer graphiquement la performance d'un ouvrage vis-à-vis des 3 indicateurs.

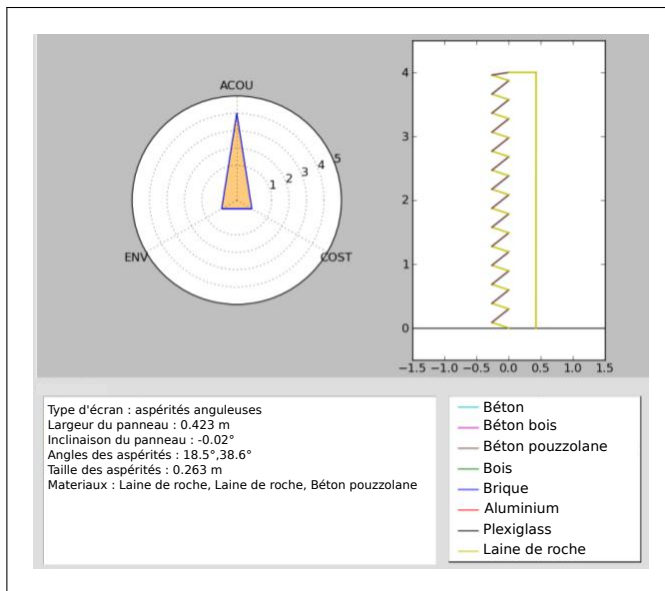
Cette représentation facilite la prise de décision et permet éventuellement de sélectionner un candidat dans la base de données. Des exemples de cette représentation sont donnés en 2.6.4.

### 2.6.4. Exemples de résultats d'optimisation globale

Nous allons présenter ici trois résultats d'optimisation globale en modifiant uniquement les importances attribuées aux critères ACOU, ENV et COST. Ces trois résultats ont en commun la famille de murs choisie (aspérités anguleuses), la topographie, ainsi que le type de source et les positions de ces dernières et des récepteurs (de part et d'autre de l'écran).

**Figure 2.6.**

Résultat de l'optimisation globale en pondérant l'indicateur ACOU à 100%

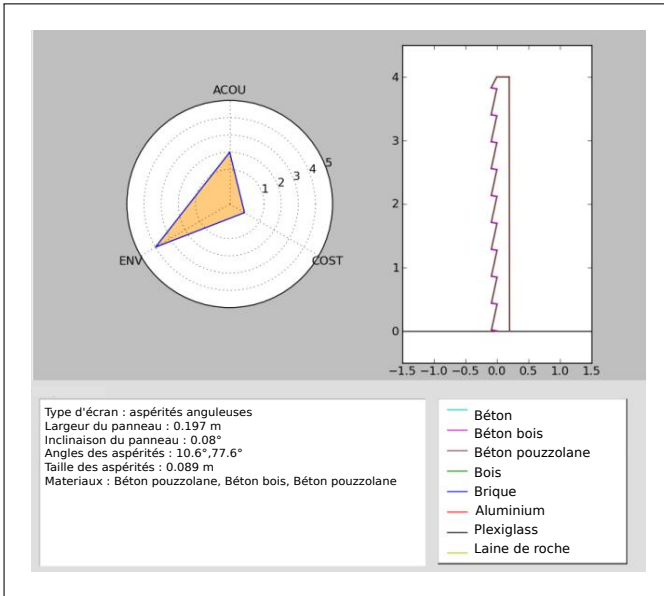


Il est facile d'identifier sur les figures 2.6, 2.7 et 2.8 les résultats obtenus en fonction des pondérations attribuées. Il est intéressant de noter que les matériaux et les formes des écrans obtenus diffèrent fortement, preuve que le choix en amont de l'optimisation est déterminant dans les solutions obtenues et que

la méthode d'optimisation globale est de ce point de vue discriminante et par conséquent efficace.

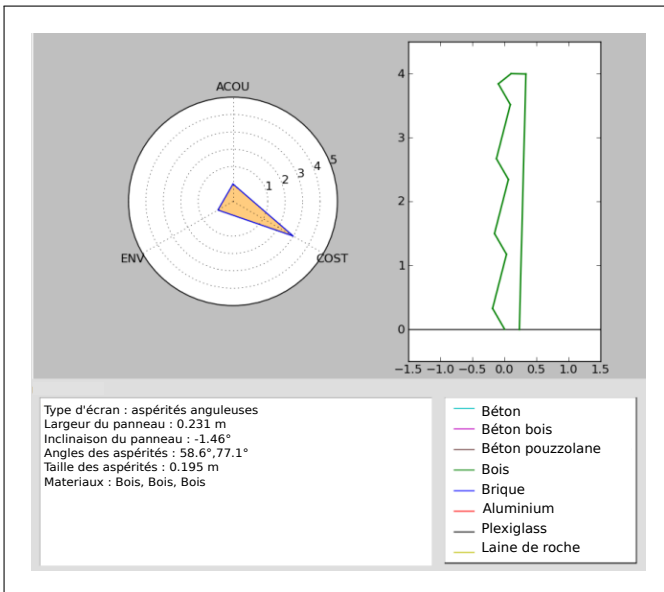
**Figure 2.7.**

Résultat de l'optimisation globale en pondérant l'indicateur ENV à 100%



**Figure 2.8.**

Résultat de l'optimisation globale en pondérant l'indicateur COST à 100%



## 2.7. Vers une nouvelle méthode de conception environnementale de murs antibruit

La méthode décrite dans ce chapitre constitue une optimisation de formes au sens large, intégrant des paramètres géométriques et une variété de matériaux différents et définissant des critères spécifiques à un problème posé en amont de la question de l'optimisation. Le cadre d'application de cette méthode s'en trouve élargi car il intègre non seulement des performances physiques (d'un point de vue acoustique) et des performances économiques, mais également des performances environnementales.

Le travail d'agrégation des indicateurs (éventuellement pondérés) dans un référentiel commun sous forme de note de performance est au cœur de l'originalité de ce travail. Il est possible de parler de conception éco-optimale d'écran antibruit, dans le sens où l'optimalité est acquise selon une contrainte économique, environnementale et acoustique.

Cette méthode a donnée naissance à l'utilisation de plusieurs outils d'aide à la décision comme la base de données issue de la combinatoire des paramètres au choix de l'utilisateur, de la possibilité de pondérer le problème vis-à-vis des indicateurs ACOU, ENV et COST ainsi que la représentation sous forme de *radar-plot* qui permet de faire rapidement un choix sur la solution obtenue. L'outil global d'optimisation trouve ainsi une résonance sur le plan pratique car il permet de combiner facilement des critères et des facteurs de décision entre plusieurs domaines dans le but d'aider les décideurs à faire un choix et à pouvoir le motiver.

Il faut remarquer que cette méthode reste encore figée dans le temps. Le prix des matériaux a été fixé et l'inflation n'est pas prise en compte (par exemple). Il peut exister également des matériaux disponibles à proximité de l'ouvrage en projet et qui peuvent satisfaire des critères uniquement de manière locale. Par ailleurs, la technique peut évoluer et aboutir à de nouveaux types d'écran utilisant de nouveaux matériaux plus performants selon les critères retenus. Cependant, la méthode proposée garde un caractère relativement universel et ne dépend pas de la technique employée : il est possible d'insérer de nouveaux matériaux, de refaire les calculs avec une nouvelle famille de murs ou encore d'optimiser un écran avec une autre source et même avec un autre outil prédictif de performances acoustiques. De nombreuses extensions techniques restent possibles au niveau des besoins d'un bureau d'études par exemple, sans remettre en cause la démarche d'optimisation globale.

La méthode de conception éco-optimale d'écran antibruit présentée ici garde une grande souplesse d'adaptabilité aux problèmes qui peuvent se poser dans la pratique. Les enjeux actuels d'économie d'énergie pousseront sans nul doute dans l'avenir à envisager toute conception comme éco-optimale et le présent chapitre montre clairement une des voies possibles en ce sens.

## Bibliographie

- Markus Auerbach, Marine Baulac, Jérôme Defrance, Philippe Jean, Guillaume Dutilleux, Christophe Heinkelé, Erik Salomons et Greg Watts.** *State of the art of existing optimisation methods and noise prediction approaches.* Rapport technique FP7-SST-2008-RTD-1. WP5 - Holistic optimisation et global noise impact, avr. 2010, page 100.
- Simon Bouley, Michael Chudalla, Jean-Pierre Clairbois, Jérôme Defrance, Frits Van Der Eerden, Thomas Leissing et Cristo Padmos.** *Global impact on noise abatement.* Rapport technique FP7-SST-2008-RTD-1. WP5 - Holistic optimisation et global noise impact, oct. 2012, page 67.
- CEN1793.** *Road traffic noise reducing devices. Test method for determining the acoustic performance Part 6 : Intrinsic characteristics - In situ values of airborne sound insulation under direct sound field conditions.* Rapport technique 1793-5 :2003. Avenue Marnix 17, B-1000 Brussels, Belgium : European Committee for Standardization, 2003.
- CEN1793.** *Road traffic noise reducing devices. Test method for determining the acoustic performance. Part 4 : Intrinsic characteristics In situ values of sound diffraction.* Rapport technique 1793-4 :2003. Avenue Marnix 17, B-1000 Brussels, Belgium : European Committee for Standardization, 2003.
- CEN1793.** *Road traffic noise reducing devices. Test method for determining the acoustic performance. Part 5 : Intrinsic characteristics - In situ values of sound reflection and airborne sound insulation.* Rapport technique 1793-5 :2003. Avenue Marnix 17, B-1000 Brussels, Belgium : European Committee for Standardization, 2003.
- CERIB.** *Panneau architectural plein en béton : fiche de déclaration environnementale et sanitaire, conforme à la norme NF-EN-15804+A1 et son complément national NF-EN-15804/CN.* Rapport technique. Centre d'Études et de Recherches de l'Industrie du Béton, 2017.
- Michael Chudalla, Jean-Pierre Clairbois, Jérôme Defrance, Francis Grannec, Thomas Leissing, Dorien Lutgendorf et Cristo Padmos.** *Application to extrinsic acoustic optimizations and holistic optimizations.* Rapport technique FP7-SST-2008-RTD-1. WP5 - Holistic optimisation et global noise impact, août 2012, page 60.
- Michael Chudalla, Jean-Pierre Clairbois, Jérôme Defrance et Thomas Leissing.** *Application to intrinsic acoustic optimizations.* Rapport technique FP7-SST-2008-RTD-1. WP5 - Holistic optimisation et global noise impact, juin 2012, page 29.
- Michael Chudalla, Jérôme Defrance, Guillaume Dutilleux, Julien Hans, Christophe Heinkelé, Thomas Leissing, Cristo Padmos et Nicoletta Schiopu.** *Setting a methodology for a holistic optimisation of noise reducing devices.* Rapport technique FP7-SST-2008-RTD-1. WP5 - Holistic optimisation et global noise impact, mai 2011, page 30.

- Jean-Pierre Clairbois.** *Quiests outcomes*. Déc. 2012. URL : [https://cordis.europa.eu/project/rcn/93636\\_fr.html](https://cordis.europa.eu/project/rcn/93636_fr.html).
- Denis Duhamel.** « Efficient calculation of the three-dimensional sound pressure field around a noise barrier ». In : *Journal of Sound and Vibration* 197.5 (1996), pages 547-571. ISSN : 0022-460X. DOI : 10.1006/jsvi.1996.0548.
- ISO-14040.** *Environmental management - Life cycle assessment - Principles and framework*. Rapport technique 14040. Geneva, Switzerland : International Organization for Standardization, 2006.
- ISO-14044.** *Environmental management - Life cycle assessment - Requirements and guidelines*. Rapport technique 14044. Geneva, Switzerland : International Organization for Standardization, 2006.
- Philippe Jean.** « A variational approach for the study of outdoor sound propagation and application to railway noise ». In : *Journal of Sound and Vibration* 212.2 (1998), pages 275-294. DOI : 10.1006/jsvi.1997.1407.
- Philip M. Morse et K. Uno Ingard.** *Theoretical Acoustics*. Princeton University Press, 1986. ISBN : 0-691-08425-4.
- NF-P-01010.** *Qualité environnementale des produits de construction - Déclaration environnementale et sanitaire des produits de construction (Environmental quality of construction products - Environmental and health declaration of construction products)*. Rapport technique. AFNOR, 2004, 48p.
- ES-prEN-15804.** *Sustainability of construction works - Environmental product declarations - core rules for the product category of construction products*. Rapport technique prEN 15804. CEN, 2010, 42p.
- N. Srinivas et Kalyanmoy Deb.** « Multiobjective Optimization Using Nondominated Sorting in Genetic Algorithms ». In : *Evol. Comput.* 2.3 (sept. 1994), pages 221-248. ISSN : 1063-6560. DOI : 10.1162/evco.1994.2.3.221.

## Chapitre 3

# Microstructures à énergie minimale dans les matériaux à changement de phase

Michaël PEIGNEY<sup>1</sup>

*Résumé – Dans certains matériaux comme les alliages à mémoire de forme, on observe une transformation de phase solide/solide entre différentes structures cristallographiques. Cette transformation de phase est pilotée par les sollicitations thermomécaniques appliquées et donne lieu à la formation spontanée de microstructures.*

*Nous présentons dans ce chapitre un cadre théorique pour l'étude de ce phénomène. L'idée maîtresse est que les microstructures à l'équilibre minimisent l'énergie élastique totale. Nous sommes ainsi amenés à résoudre un problème d'optimisation de formes portant sur la répartition spatiale des différentes phases. Bien que la solution exacte soit en général hors de portée, on présente différentes bornes qui permettent d'estimer les microstructures susceptibles de se développer.*

### 3.1. Introduction

Certains alliages métalliques (NiTi ou CuAlNi par exemple) ont la spécificité de présenter un effet mémoire de forme : un échantillon déformé à basse température retrouve sa forme initiale une fois chauffé. Ce comportement original ouvre notamment la voie à des utilisations de ces matériaux comme actionneurs.

---

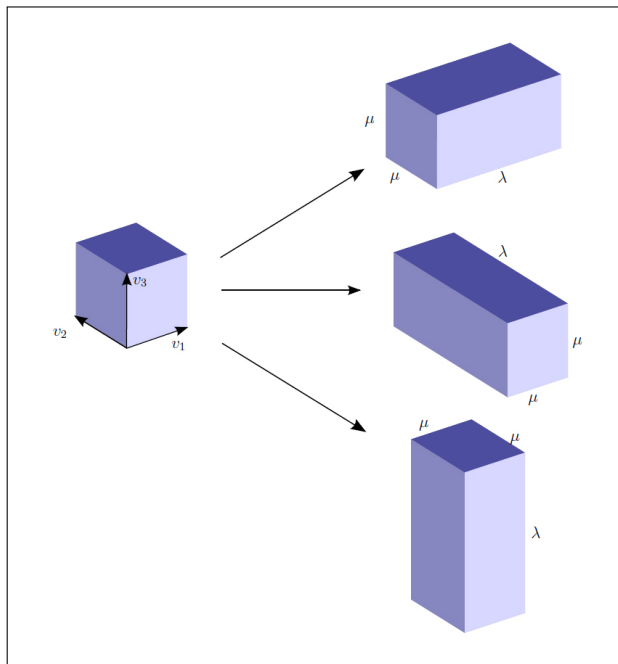
1. IFSTTAR

Un exemple en génie civil est la conception de façades actives pour la protection solaire (HANNEQUART et al., 2017).

L'effet mémoire de forme est la manifestation macroscopique d'un phénomène de changement de phase solide/solide entre deux structures cristallographiques différentes, nommées austénite (stable à haute température) et martensite (stable à basse température). En termes de structure cristallographique, l'austénite a un degré de symétrie supérieur à la martensite. On est ainsi amené à distinguer plusieurs variantes de martensite, se distinguant par l'orientation de la structure martensitique par rapport à la structure austénitique. À chacune de ces variantes correspond une déformation de transformation, décrivant la déformation entre la structure austénite et la structure martensitique.

La figure 3.1 illustre ces considérations dans le cas d'une transformation cubique - tétragonale, c'est-à-dire dans le cas où l'austénite a une structure cubique (d'axes  $v_1, v_2, v_3$ ) tandis que la martensite a une structure tétragonale obtenue par dilatation d'un rapport  $\lambda$  le long d'un des axes  $v_i$ , et d'un rapport  $\mu$  sur  $v_i^\perp$ .

**Figure 3.1.**  
Transformation cubique-tétragonale



Dans ce cas, il y a trois variantes de martensite, se distinguant par l'axe sur lequel est appliqué la dilatation de rapport  $\lambda$ . Les déformations de transformation



correspondantes s'écrivent, en représentation matricielle dans la base  $(v_1, v_2, v_3)$ ,

$$\begin{pmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \mu \end{pmatrix}, \begin{pmatrix} \mu & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{pmatrix}, \begin{pmatrix} \mu & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \lambda \end{pmatrix}.$$

Alors que les trois directions  $(v_1, v_2, v_3)$  sont indiscernables dans la structure cubique de l'austénite, seules deux d'entre elles demeurent indiscernables dans la structure tétragonale de la martensite.

La transformation cubique-tétragonale s'applique par exemple aux alliages MnCu et MnTi. De façon plus générale, les structures cristallographiques de l'austénite et de la martensite dépendent du matériau considéré.

Il en va donc de même pour le nombre de variantes de martensite et les déformations de transformation associées. Outre la transformation cubique-tétragonale, d'autres exemples courants sont les transformations cubique-orthorombique (CuAlNi) et cubique-monoclinique (TiNi), correspondant respectivement à 6 et 12 variantes de martensite.

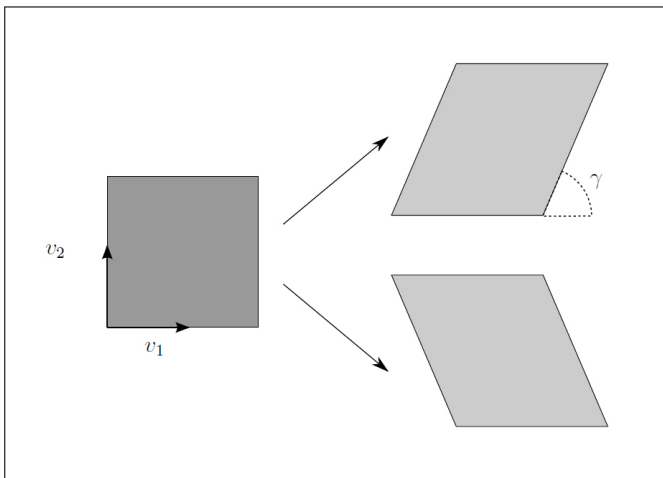
Afin d'expliquer le lien entre transformation de phase et effet mémoire de forme, considérons l'exemple simple d'un matériau avec deux variantes martensitiques. Les déformations de transformation sont de la forme

$$\begin{pmatrix} 1 & \pm\gamma & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

dans la base  $(v_1, v_2, v_3)$  (figure 3.2).

**Figure 3.2.**

Transformation tétragonale-orthorombique



Ce cas de figure correspond à la transformation orthorombique-monoclinique, observée par exemple dans l'alliage  $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$  (ANDERSEN et al., 1990).

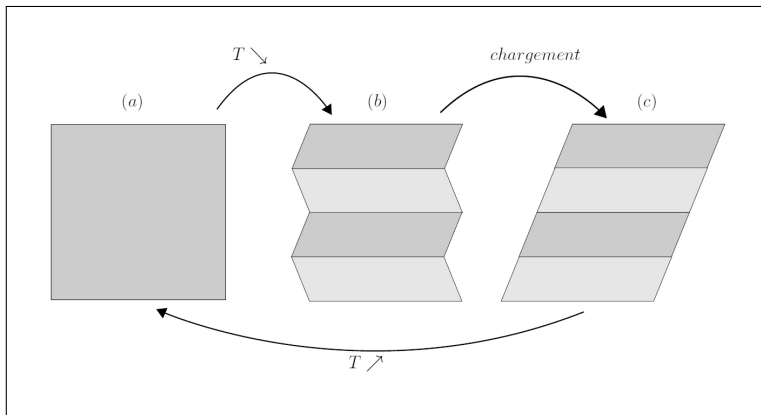
En refroidissant un échantillon libre de contraintes, le matériau passe d'un état austénitique (stable à haute température) à un état martensitique (stable à basse température), dans lequel les différentes variantes de martensite se développent et s'arrangent spatialement pour produire un état de contrainte nulle et à déformation macroscopique négligeable (état (b) sur la figure 3.3).

Déformer mécaniquement l'échantillon entraîne une réorientation des variantes, c'est-à-dire un changement de phase martensite/martensite (état (c)). Ce changement de phase s'effectue au profit des variantes les mieux orientées par rapport au chargement appliqué.

Après décharge, il subsiste une déformation résiduelle produite par l'effet collaboratif des déformations microscopiques dans chaque variante. Chauffer l'échantillon provoque un changement de phase de la martensite vers l'austénite, restaurant la géométrie initiale avant sollicitation (état (a)).

### Figure 3.3.

Interprétation de l'effet mémoire de forme en termes de changement de phase ( $T$  désigne la température)



Il est clair que si la déformation imposée est trop importante, des déformations plastiques peuvent apparaître et l'effet mémoire de forme est potentiellement perdu : la déformation imposée n'est pas *recouvrable*. L'ensemble des déformations recouvrables joue un rôle essentiel dans l'effet mémoire de forme (BHATTACHARYA et KOHN, 1997). Sa prédiction théorique est l'objet principal de ce chapitre. On utilisera pour cela le cadre de la théorie de l'élasticité non linéaire, en petites perturbations.

## 3.2. Modélisation du problème

La configuration austénitique est prise comme référence. Soient  $\mathbf{u}$  le déplacement et  $\mathbf{e} = (\nabla \mathbf{u} + \nabla^T \mathbf{u})/2$  le tenseur des déformations linéarisées. Les

déformations de transformation (linéarisées) sont notées  $e_1, \dots, e_n$  où  $n$  est le nombre de variantes. Par exemple, pour la transformation cubique-tétraogonale représentée figure 3.1, les trois déformations  $e_i$  sont données par

$$e_1 = \begin{pmatrix} \lambda' & 0 & 0 \\ 0 & \mu' & 0 \\ 0 & 0 & \mu' \end{pmatrix}, e_2 = \begin{pmatrix} \mu' & 0 & 0 \\ 0 & \lambda' & 0 \\ 0 & 0 & \mu' \end{pmatrix}, e_3 = \begin{pmatrix} \mu' & 0 & 0 \\ 0 & \mu' & 0 \\ 0 & 0 & \lambda' \end{pmatrix}.$$

avec  $\lambda' = \lambda - 1$  et  $\mu' = \mu - 1$ . De façon générale, les  $n$  déformations  $e_i$  sont reliées par symétrie : pour tout  $(i, j)$  il existe une rotation  $R_{ij}$  telle  $e_j = R_{ij}^T e_i R_{ij}$ .

On considère que chaque phase a un comportement élastique, régi par une fonction énergie  $w_i$  ( $i = 0$  correspond à l'austénite,  $i = 1, \dots, n$  correspond aux variantes de martensite). Pour simplifier la présentation, on supposera dans la suite que les fonctions  $w_i$  ( $i = 0, \dots, n$ ) ont la forme

$$w_i(e) = \frac{1}{2}(e - e_i) : L : (e - e_i) + m_i \quad (3.1)$$

où  $L$  est symétrique défini positif et  $e_i$  désigne la déformation de transformation pour la variante  $i$  (en adoptant la condition  $e_0 = 0$ ). L'expression (3.1) traduit un comportement élastique linéaire pour chaque phase : la contrainte  $\sigma$  est donnée par la relation

$$\sigma = w'_i(e) = L : (e - e_i).$$

Le coefficient  $m_i$  dans (3.1) est l'énergie minimale de la phase  $i$  et prend la même valeur pour toutes les variantes de martensite, c'est-à-dire pour  $i = 1, \dots, n$ . Ce coefficient dépend de la température  $T$ . Plus précisément, il existe une température critique  $T^{crit}$  tel que, pour tout  $T < T^{crit}$ ,

$$m_1 = \dots = m_n < m_0. \quad (3.2)$$

Pour  $T > T^{crit}$ ,

$$m_1 = \dots = m_n > m_0. \quad (3.3)$$

La température critique  $T^{crit}$  est la température d'équilibre entre les phases (également appelée température de transition), c'est-à-dire la température à laquelle austénite et martensite ont la même énergie minimale. Les relations (3.2-3.3) traduisent le fait que l'austénite est stable pour les températures supérieures à  $T^{crit}$ , tandis que la martensite est stable pour les températures inférieures à  $T^{crit}$ .

Dans la théorie de l'élasticité appliquée au changement de phase, on pose comme principe de base qu'à l'équilibre le système minimise son énergie à toutes les échelles, aussi bien microscopique (c'est-à-dire localement) que macroscopique (c'est-à-dire au niveau de l'échantillon). Au niveau microscopique, on considère ainsi que la fonction énergie  $w$  du matériau est donnée par

$$w(e) = \min_{0 \leq i \leq n} w_i(e). \quad (3.4)$$

Cette relation exprime le fait que le matériau, soumis à une déformation locale imposée  $e$ , "choisit" la phase d'énergie minimale. Afin d'écrire ce principe de minimisation au niveau macroscopique, considérons un échantillon occupant un domaine  $\Omega$  et soumis à des déplacements  $\bar{e}.x$  sur la frontière  $\partial\Omega$  (où  $\bar{e}$  est donné et s'interprète comme une déformation moyenne). L'énergie totale du système est donnée par

$$W = \int_{\Omega} w(e)d\omega. \tag{3.5}$$

À l'équilibre, le système minimise (3.5) par rapport à l'ensemble des champs de déformations  $e(x)$  compatibles avec les conditions aux limites. Cet ensemble est défini par

$$A(\bar{e}) = \{e|\exists u(x) \text{ tel que } e = (\nabla u + \nabla^T u)/2 \text{ dans } \Omega; u(x) = \bar{e}.x \text{ sur } \partial\Omega\}. \tag{3.6}$$

Nous sommes ainsi amenés à étudier le problème

$$W(\bar{e}) = \inf_{e \in A(\bar{e})} \frac{1}{|\Omega|} \int_{\Omega} w(e)d\omega. \tag{3.7}$$

Une difficulté vient du fait que l'*infimum* dans (3.7) n'est en général pas atteint. Afin d'illustrer cette propriété, considérons l'exemple du matériau à deux variantes représenté figure 3.2 : l'énergie  $w$  est de la forme (3.1)-(3.4) avec

$$e_1 = \begin{pmatrix} 0 & \frac{\gamma}{2} & 0 \\ \frac{\gamma}{2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 & -\frac{\gamma}{2} & 0 \\ -\frac{\gamma}{2} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.8}$$

et

$$m_1 = m_2 = 0, m_0 > 0. \tag{3.9}$$

Nous montrerons plus loin que  $W(0) = 0$  (voir 3.6.2). Vérifions ici que l'*infimum* dans (3.7) n'est pas atteint pour  $\bar{e} = 0$ .

S'il existe  $u \in A(0)$  tel que  $\int_{\Omega} w(e) = 0$  alors, comme  $w$  est positive, Nous avons nécessairement  $w(e) = 0$  en tout point, c'est-à-dire  $e(x) \in \{e_1, e_2\}$  pour tout  $x$  dans  $\Omega$ . Nous avons donc  $u_{1,1} = u_{2,2} = 0$  dans tout  $\Omega$ , ce qui implique que  $u_1$  (resp.  $u_2$ ) est fonction uniquement de  $x_2$  (resp.  $x_1$ ). La condition  $u = 0$  sur  $\partial\Omega$  implique alors  $u = 0$  dans tout  $\Omega$ , en contradiction avec la condition  $e \in \{e_1, e_2\}$ . Ceci montre qu'il n'existe pas de champ  $e \in A(0)$  tel que  $W(0) = (\int_{\Omega} w(e))/|\Omega|$ .

En termes mathématiques, le fait que l'*infimum* dans (3.7) n'est pas atteint signifie que les suites minimisantes ne convergent pas. Ceci correspond physiquement à la formation de microstructures infiniment fines. Ce phénomène, étroitement lié au fait que l'énergie  $w$  définie par (3.4) n'est pas convexe, rend la détermination de  $W(\bar{e})$  particulièrement délicate. De fait l'expression exacte de  $W$  n'est connue que dans quelques cas particuliers (KOHN, 1991 ; SMYSHLYAEV et al., 1998) qui

seront détaillés plus tard dans ce chapitre. La résolution de (3.7) dans le cas général reste un problème encore largement ouvert.

La fonction  $W$  définie par (3.7) est nommée *relaxation* ou *quasi-convexification* (ou encore enveloppe quasiconvexe) de l'énergie microscopique  $w$ . Cette dernière dénomination est justifiée par le fait que  $W$  est quasi-convexe (DACOROGNA, 2008), c'est-à-dire qu'elle vérifie

$$W(\bar{e}) \leq \frac{1}{|\Omega|} \int_{\Omega} W(e) d\omega \text{ pour tout } e \in A(\bar{e}). \quad (3.10)$$

En termes physiques, la fonction  $W$  peut être interprétée comme l'énergie du matériau au niveau macroscopique (DACOROGNA, 1982). Une propriété remarquable est que  $W$  est indépendant du domaine  $\Omega$  considéré dans la définition (3.7).

Examinons quelques relations simples entre  $w$  et  $W$ . Pour tout  $\bar{e}$ , le champ de déformations uniforme égal à  $\bar{e}$  en tout point est dans  $A(\bar{e})$ , donc la définition (3.7) donne

$$W(\bar{e}) \leq w(\bar{e}) \quad (3.11)$$

dont nous déduisons en particulier que  $\min W \leq \min w$ .

Par ailleurs, la définition (3.7) montre que  $\min w \leq W(\bar{e})$  pour tout  $\bar{e}$ , ce qui implique  $\min w \leq \min W$ . On en conclut que

$$\min w = \min W. \quad (3.12)$$

Posons

$$\mathcal{K} = \{e | w(e) = \min w\}$$

et

$$Q\mathcal{K} = \{\bar{e} | W(\bar{e}) = \min W\}.$$

L'ensemble  $\mathcal{K}$  est l'ensemble des déformations minimisant l'énergie microscopique  $w$ , alors que  $Q\mathcal{K}$  est l'ensemble des déformations minimisant l'énergie macroscopique.

De la relation (3.12) on déduit

$$\mathcal{K} \subset Q\mathcal{K}. \quad (3.13)$$

L'ensemble  $\mathcal{K}$  est simple à expliciter : d'après (3.1) et (3.4),  $\min w = \min_i m_i$ , si bien que l'ensemble  $\mathcal{K}$  est constitué des déformations de transformation  $e_i$  telles que  $m_i = \min_j m_j$ . Pour les températures supérieures à  $T^{crit}$ , on a

$$\mathcal{K} = \{e_0\}$$

et l'on peut alors montrer que

$$Q\mathcal{K} = \{e_0\}.$$

Pour les températures inférieures à  $T^{crit}$ , on a

$$\mathcal{K} = \{e_1, \dots, e_n\}$$

où  $e_i$  est la déformation de transformation pour la variante  $i$  de martensite. L'expression de l'ensemble  $Q\mathcal{K}$  est, dans ce cas, beaucoup plus complexe à obtenir. L'inclusion (3.13) est stricte, et l'on verra même que, pour la plupart des matériaux,  $Q\mathcal{K}$  est "beaucoup plus grand" que  $\mathcal{K}$ . Pour  $T < T^{crit}$ , les déformations dans  $Q\mathcal{K}$  s'identifient aux déformations recouvrables mentionnées en introduction. La prédiction théorique de l'ensemble  $Q\mathcal{K}$  est un objectif des travaux présentés dans ce chapitre.

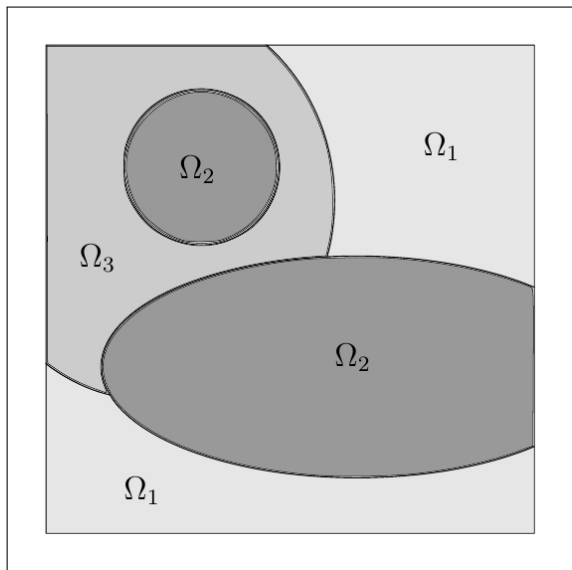
### 3.3. Lien avec l'optimisation de formes

Déterminer  $W$  revient à résoudre un problème d'optimisation de formes sur la répartition géométrique des différentes phases. L'objet de ce paragraphe est de mettre ce lien en évidence. La répartition des différentes phases peut être représentée par les fonctions caractéristiques  $\chi_i$  des domaines  $\Omega_i$  occupés par chaque phase  $i$ .

La fonction  $\chi_i$  est définie par  $\chi_i(\mathbf{x}) = 1$  si  $\mathbf{x} \in \Omega_i$  et  $\chi_i(\mathbf{x}) = 0$  sinon. Les domaines  $\Omega_i$  sont deux à deux disjoints et forment un recouvrement de  $\Omega$  (figure 3.4).

**Figure 3.4.**

Répartition géométrique des phases dans le domaine  $\Omega$  (ici carré).



Par conséquent,  $\chi = (\chi_1, \dots, \chi_n)$  appartient à l'ensemble  $C$  défini par

$$C = \{(\chi_1, \dots, \chi_n) | \chi_i(\mathbf{x}) \in \{0, 1\}; \sum_{i=0}^n \chi_i(\mathbf{x}) = 1 \text{ dans } \Omega\}.$$

Nous utiliserons dans la suite le terme de *microstructure* pour désigner une répartition géométrique des phases.

Pour tout  $e$  et pour tout  $\chi$  dans  $C$ ,  $w(e) \leq \sum_i \chi_i(\mathbf{x}) w_i(e)$ . Donc

$$W(\bar{e}) \leq \inf_{\chi \in C} \inf_{e \in A(\bar{e})} \frac{1}{|\Omega|} \int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x}) w_i(e) d\omega. \quad (3.14)$$

L'inégalité inverse est également valable.

Considérons en effet un champ  $\tilde{e}$  donné dans  $A(\bar{e})$ . Pour tout  $\mathbf{x}$ , on peut choisir  $i(\mathbf{x})$  tel que  $w(\tilde{e}(\mathbf{x})) = w_{i(\mathbf{x})}(\tilde{e}(\mathbf{x}))$ . Les domaines  $\Omega_i = \{\mathbf{x} \in \Omega | i(\mathbf{x}) = i\}$  sont deux à deux disjoints. Définissons  $\tilde{\chi}_i$  la fonction caractéristique de  $\Omega_i$ .

La fonction  $\tilde{\chi} = (\tilde{\chi}_1, \dots, \tilde{\chi}_n)$  appartient à l'ensemble  $C$  et l'on a  $w(\tilde{e}) = \sum_i \tilde{\chi}_i(\mathbf{x}) w_i(\tilde{e})$ . Donc

$$\int_{\Omega} w(\tilde{e}) d\omega = \int_{\Omega} \sum_{i=0}^n \tilde{\chi}_i(\mathbf{x}) w_i(\tilde{e}) d\omega \geq \inf_{\chi \in C} \int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x}) w_i(\tilde{e}) d\omega$$

Cette inégalité vaut pour tout  $\tilde{e} \in A(\bar{e})$ . En prenant l'*infimum* sur  $\tilde{e} \in A(\bar{e})$  :

$$W(\bar{e}) \geq \inf_{\chi \in C} \inf_{e \in A(\bar{e})} \frac{1}{|\Omega|} \int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x}) w_i(e) d\omega. \quad (3.15)$$

La comparaison avec (3.14) montre qu'il y a égalité dans (3.15). Ainsi,

$$W(\bar{e}) = \inf_{\chi \in C} W(\bar{e}; \chi) \quad (3.16)$$

où l'on a posé, pour tout  $\chi$  dans  $C$ ,

$$W(\bar{e}; \chi) = \inf_{e \in A(\bar{e})} \frac{1}{|\Omega|} \int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x}) w_i(e) d\omega. \quad (3.17)$$

Interprétons ces relations : pour une microstructure  $\chi$  donnée, déterminer  $W(\bar{e}; \chi)$  revient à résoudre le problème d'élasticité linéaire

$$\begin{aligned} e &\in A(\bar{e}), \\ \operatorname{div} \sigma &= 0 \text{ dans } \Omega, \\ \sigma(\mathbf{x}) &= w'_i(e(\mathbf{x})) = \mathbf{L}_i : (e(\mathbf{x}) - e_i) \text{ si } \mathbf{x} \in \Omega_i. \end{aligned} \quad (3.18)$$

La quantité  $W(\bar{e}, \chi)$  est l'énergie élastique pour la microstructure  $\chi$  et la déformation macroscopique imposée  $\bar{e}$ . La recherche de  $W(\bar{e})$  revient à déterminer la microstructure  $\chi$  d'énergie élastique minimale pour la déformation  $\bar{e}$  donnée. Il s'agit en ce sens d'un problème d'optimisation de formes, portant sur la microstructure  $\chi$  : nous cherchons à répartir les différentes phases dans le domaine  $\Omega$  de façon à obtenir une énergie de déformation élastique minimale.

Une telle répartition dépend de la déformation moyenne  $\bar{e}$  imposée. Par exemple, si  $\bar{e}$  est pris égal à une des déformations de transformation (par exemple  $e_1$ ), alors une répartition uniforme de la phase 1 est optimale (et donne une énergie nulle).

La fonction  $W$  joue un rôle clé dans l'étude des matériaux à changement de phase car elle contient toute l'information sur le comportement macroscopique, en particulier la relation contrainte-déformation et son évolution avec la température. De plus, pour une déformation  $\bar{e}$  donnée, les suites minimisantes dans (3.16) renseignent sur les microstructures qui se développent dans le matériau.

Cependant, le problème d'optimisation de formes (3.16) qui caractérise  $W$  ne peut en général être résolu de façon exacte. On s'attachera dans la suite à obtenir un encadrement de  $W$ , en étudiant des minorations et des majorations de  $W$ .

Une majoration explicite et très utile est donnée par l'enveloppe convexe de  $w$ , introduite dans la section suivante. Cette borne joue en pratique un rôle important, et fournit dans certains cas la valeur exacte de l'ensemble  $Q\mathcal{K}$ , comme on l'illustre dans la section 3.5.

Des minorations de  $W$  sont obtenues en considérant une classe de microstructures particulières, à savoir les microstructures laminées. Cette approche fait l'objet de la section 3.6.

### 3.4. Enveloppe convexe

Il a été établi plus haut que  $W(\bar{e}) = \inf_{\chi} W(\bar{e}; \chi)$  où  $W(\bar{e}; \chi)$  s'interprète comme l'énergie potentielle en déplacements pour le problème d'élasticité linéaire (3.18). Or les théorèmes de minimum bien connus en élasticité linéaire permettent d'obtenir une borne inférieure sur l'énergie potentielle en déplacement (SALENCON, 2007). Considérons en effet l'énergie potentielle en contraintes  $W^*$  pour le problème d'élasticité (3.18), définie par

$$|\Omega|W^* = \inf_{\sigma \mid \text{div } \sigma = 0} \left\{ \int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x}) w_i^*(\sigma) d\omega - \int_{\partial\Omega} (\sigma \cdot \mathbf{n}) \cdot (\bar{e} \cdot \mathbf{x}) d\Gamma \right\} \quad (3.19)$$

où  $w_i^* = \sup_{\mathbf{e}} \sigma : \mathbf{e} - w_i(\mathbf{e})$  est l'énergie complémentaire (transformée de Legendre-Fenchel) associée à  $w_i$ . Dans le cas présent :

$$w_i^*(\sigma) = \frac{1}{2} \sigma : \mathbf{L}^{-1} : \sigma + \sigma : \mathbf{e}_i - m_i. \quad (3.20)$$



Par ailleurs, notons que (3.19) peut se réécrire

$$|\Omega|W^* = \inf_{\sigma | \operatorname{div} \sigma = 0} \int_{\Omega} \sum_i \chi_i(\mathbf{x}) w_i^*(\sigma) d\omega - |\Omega| \bar{\mathbf{e}} : \bar{\sigma}$$

où

$$\bar{\sigma} = \frac{1}{|\Omega|} \int_{\Omega} \sigma d\omega$$

désigne la valeur moyenne de  $\sigma$ . La formule de Clapeyron (SALENCON, 2007) donne

$$W(\bar{\mathbf{e}}; \chi) = -W^*$$

et donc

$$-|\Omega|W(\bar{\mathbf{e}}; \chi) = \inf_{\sigma | \operatorname{div} \sigma = 0} \int_{\Omega} \sum_i \chi_i(\mathbf{x}) w_i^*(\sigma) d\omega - |\Omega| \bar{\mathbf{e}} : \bar{\sigma}. \quad (3.21)$$

En se restreignant à des champs  $\sigma$  uniformes dans (3.21), on obtient l'inégalité

$$-W(\bar{\mathbf{e}}; \chi) \leq \inf_{\bar{\sigma}} \sum_i \theta_i w_i^*(\bar{\sigma}) - \bar{\mathbf{e}} : \bar{\sigma}$$

où

$$\theta_i = \frac{1}{|\Omega|} \int_{\Omega} \chi_i(\mathbf{x}) d\omega$$

s'interprète comme la fraction volumique de phase  $i$  dans la microstructure  $\chi$  considérée. Observons que  $\theta = (\theta_0, \dots, \theta_n)$  appartient à l'ensemble  $\mathcal{T}$  défini par

$$\mathcal{T} = \{(\theta_0, \dots, \theta_n) | \theta_i \geq 0; \sum_{i=0}^n \theta_i = 1\}. \quad (3.22)$$

En remplaçant  $w_i^*$  par son expression (3.20) :

$$-W(\bar{\mathbf{e}}; \chi) \leq \inf_{\bar{\sigma}} \frac{1}{2} \bar{\sigma} : \mathbf{L}^{-1} : \bar{\sigma} - \bar{\sigma} : (\bar{\mathbf{e}} - \sum_{i=0}^n \theta_i \mathbf{e}_i) - \sum_{i=0}^n \theta_i m_i.$$

Le terme de droite est une fonction quadratique convexe en  $\bar{\sigma}$ .

En calculant l'*infimum* par rapport à  $\bar{\sigma}$  :

$$CW(\bar{\mathbf{e}}, \theta) \leq W(\bar{\mathbf{e}}; \chi)$$

avec

$$CW(\bar{\mathbf{e}}, \theta) = \frac{1}{2} (\bar{\mathbf{e}} - \sum_{i=0}^n \theta_i \mathbf{e}_i) : \mathbf{L}^{-1} : (\bar{\mathbf{e}} - \sum_{i=0}^n \theta_i \mathbf{e}_i) + \sum_{i=0}^n \theta_i m_i. \quad (3.23)$$

Il s'ensuit que

$$CW(\bar{e}) \leq W(\bar{e}) \tag{3.24}$$

avec

$$CW(\bar{e}) = \inf_{\theta \in \mathcal{T}} CW(\bar{e}, \theta). \tag{3.25}$$

Nous pouvons facilement vérifier que la fonction  $CW$  ainsi définie est *convexe*. Par ailleurs les relations (3.11) et (3.24) montrent que  $CW$  est une *borne inférieure* sur  $w$ . En fait  $CW$  est la plus grande fonction combinant ses deux propriétés, c'est-à-dire la meilleure borne inférieure convexe sur  $w$ . En effet, si  $v$  est une fonction convexe telle que  $v \leq w$ , alors  $w_i^* \leq v^*$  pour tout  $i$  et donc

$$\frac{1}{|\Omega|} \int_{\Omega} \sum_i \chi_i(\mathbf{x}) w_i^*(\bar{\sigma}) d\omega + \bar{e} : \bar{\sigma} \leq v^*(\bar{\sigma}) + \bar{e} : \bar{\sigma}$$

pour tout  $\bar{e}$  et tout  $\bar{\sigma}$ . En particulier, en prenant  $\bar{\sigma} = (\partial v / \partial e)(\bar{e})$  :

$$\frac{1}{|\Omega|} \int_{\Omega} \sum_i \chi_i(\mathbf{x}) w_i^*(\bar{\sigma}) d\omega + \bar{e} : \bar{\sigma} \leq -v(\bar{e}).$$

Or le raisonnement précédent a montré que le terme de gauche est minoré par  $-CW(\bar{e})$ . Nous arrivons donc à  $v \leq CW$ . Ceci prouve que la fonction  $CW$  est la plus grande fonction convexe majorée par  $w$ , c'est-à-dire *l'enveloppe convexe* de  $w$ .

Cette borne  $CW$  sur l'énergie  $W$  permet d'obtenir une borne sur l'ensemble des déformations à énergie macroscopique minimale. Considérons en effet  $\bar{e}$  une déformation minimisant  $W$  à  $T < T^{crit}$  (on a alors  $\min w = m_1 = m_n < m_0$ ). D'après (3.23), nécessairement  $CW(\bar{e}) = m$ , i.e il existe  $\theta \in \mathcal{T}$  tel que  $\theta_0 = 0$  et

$$\bar{e} = \sum_{i=1}^n \theta_i e_i$$

Ainsi

$$Q\mathcal{K} \subset C\mathcal{K} \tag{3.26}$$

où

$$C\mathcal{K} = \left\{ \sum_{i=1}^n \theta_i e_i \mid \theta \in \mathcal{T}; \theta_0 = 0 \right\} \tag{3.27}$$

s'interprète comme l'enveloppe convexe de  $\mathcal{K}$ , c'est-à-dire le plus petit ensemble convexe contenant  $\mathcal{K}$ . La relation (3.26) montre que l'ensemble  $C\mathcal{K}$  est une borne supérieure (au sens de l'inclusion d'ensembles) sur  $Q\mathcal{K}$ . Attardons-nous sur la dimension de  $C\mathcal{K}$ . La définition (3.27) montre que

$$C\mathcal{K} \subset \text{vect } \mathcal{K}$$

**Table 3.1.**  
Transformation cubique-orthorombique

$e_1$	$e_2$	$e_3$
$\begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha & \delta \\ 0 & \delta & \alpha \end{pmatrix}$	$\begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha & -\delta \\ 0 & -\delta & \alpha \end{pmatrix}$	$\begin{pmatrix} \alpha & 0 & \delta \\ 0 & \beta & 0 \\ \delta & 0 & \alpha \end{pmatrix}$
$e_4$	$e_5$	$e_6$
$\begin{pmatrix} \alpha & 0 & -\delta \\ 0 & \beta & 0 \\ -\delta & 0 & \alpha \end{pmatrix}$	$\begin{pmatrix} \alpha & \delta & 0 \\ \delta & \alpha & 0 \\ 0 & 0 & \beta \end{pmatrix}$	$\begin{pmatrix} \alpha & -\delta & 0 \\ -\delta & \alpha & 0 \\ -0 & 0 & \beta \end{pmatrix}$

où

$$\text{vect } \mathcal{K} = \left\{ \sum_{i=1}^n x_i e_i \mid x_i \in \mathbb{R} \right\}$$

est l'espace vectoriel engendré par  $\mathcal{K}$ . Nous avons  $\dim \text{vect } \mathcal{K} \leq n$ , avec égalité si les déformations  $e_1, \dots, e_n$  sont linéairement indépendantes.

Par ailleurs, comme on l'a noté en 3.2, les déformations de transformation  $e_1, \dots, e_n$  sont reliées par symétrie : pour tout  $(i, j)$ , il existe une rotation  $R_{ij}$  telle que  $e_j = R_{ij}^T e_i R_{ij}$ . Cette relation implique que  $\text{tr } e_i = \text{tr } e_j$  : les déformations de transformation ont même trace.

En notant  $t$  cette valeur commune,

$$\text{vect } \mathcal{K} \subset \{ \bar{e} \in \mathbb{R}_s^{3 \times 3} \mid \text{tr } \bar{e} = t \}$$

ce qui montre que  $\dim \text{vect } \mathcal{K} \leq 5$ . Par suite on arrive à

$$\dim Q\mathcal{K} \leq \dim C\mathcal{K} \leq \min(n, 5). \tag{3.28}$$

### 3.5. Application à la transformation cubique-orthorombique

Une application importante des résultats obtenus dans la section précédente concerne la transformation cubique-orthorombique, observée notamment dans l'alliage CuAlNi. Il y a dans ce cas 6 déformations de transformation, listées dans la table 3.1. Les représentations matricielles données dans cette table sont relatives à la base orthonormale  $(v_1, v_2, v_3)$  du réseau cubique de l'austénite.

L'expression de  $C\mathcal{K}$  peut être calculée de façon explicite en utilisant la définition (3.27) (BHATTACHARYA et KOHN, 1997). On obtient que les déformations  $\bar{e}$  dans

$CK$  sont caractérisées par

$$\begin{aligned} \text{tr } \bar{e} &= 2\alpha + \beta; \\ \min(\alpha, \beta) &\leq \bar{e}_{ii} \leq \max(\alpha, \beta) \text{ pour } i = 1, 2, 3; \\ |\bar{e}_{jk}| &\leq \frac{\bar{e}_{ii} - \alpha}{\beta - \alpha} \delta \text{ pour tout } \{i, j, k\} \text{ permutation of } \{1, 2, 3\}. \end{aligned} \quad (3.29)$$

Or  $\dim CK \leq 5$  en vertu de la relation (3.28). Or on peut observer à partir de l'expression (3.29) que  $CK$  contient une boule centrée sur  $(2\alpha + \beta)\mathbf{I}/3$  et de rayon  $\min(\delta, 2|\alpha - \beta|)/3$ . Par conséquent,

$$\dim CK = 5.$$

Nous pouvons facilement vérifier que les déformations de transformation  $e_i$  dans (3.1) sont compatibles deux à deux. Nous pouvons alors montrer (BHATTACHARYA, 1993) que  $QK = CK$  : l'enveloppe convexe de  $K$  donne la valeur exacte de l'ensemble des déformations à énergie minimale.

Afin d'illustrer ces résultats, considérons des déformations  $\bar{e}(\omega, \tau)$  de la forme

$$\bar{e}(\omega, \tau) = \frac{1}{3}(2\alpha + \beta)\mathbf{I} + \tau(\mathbf{u}(\omega) \otimes \mathbf{v}(\omega) + \mathbf{v}(\omega) \otimes \mathbf{u}(\omega)), \quad (3.30)$$

avec

$$\mathbf{u}(\omega) = \cos(\omega)\mathbf{v}_1 + \sin(\omega)\mathbf{v}_2, \quad \mathbf{v}(\omega) = -\sin(\omega)\mathbf{u}_1 + \cos(\omega)\mathbf{v}_2.$$

La représentation matricielle de  $\bar{e}(\omega, \tau)$  dans la base  $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$  est

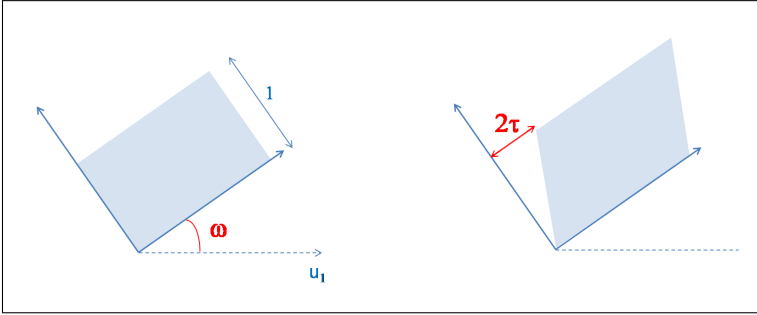
$$\bar{e}(\omega, \tau) = \frac{2\alpha + \beta}{3}\mathbf{I} + \tau \begin{pmatrix} -\sin 2\omega & \cos 2\omega & 0 \\ \cos 2\omega & \sin 2\omega & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.31)$$

La déformation  $\bar{e}(\omega, \tau)$  est obtenue en refroidissant un échantillon libre de contrainte jusqu'à une température inférieure à  $T^{crit}$  puis en appliquant un glissement simple d'amplitude  $2\tau$  entre les directions  $\mathbf{u}(\omega)$  et  $\mathbf{v}(\omega)$  (figure 3.5).

Refroidir un échantillon libre de contraintes produit en effet un état auto-accommodé (évoqué en introduction), dans lequel l'austénite est absente et toutes les variantes martensitiques apparaissent en même fraction volumique. Il en résulte une déformation macroscopique égale à  $(2\alpha + \beta)/3\mathbf{I}$ . On notera que la déformation dans l'état auto-accommodé est alors négligeable si  $(2\alpha + \beta)/3$  est très petit, ce qui est le cas pour les matériaux courants. Par exemple, pour l'alliage CuAlNi,  $\alpha = 0.0425$ ,  $\beta = -0.0822$ ,  $\delta = 0.0194$  (BHATTACHARYA, 1991), ce qui correspond à une déformation  $(2\alpha + \beta)/3$  inférieure à 0.01%.

**Figure 3.5.**

Exemple de déformations (à gauche la configuration avant déformation, à droite la configuration après déformations)



On cherche les valeurs  $(\omega, \tau)$  pour lesquelles la déformation  $\bar{e}(\omega, \tau)$  est recouvrable, c'est-à-dire pour lesquelles  $\bar{e}(\omega, \tau) \in Q\mathcal{K}$ . L'expression (3.29) permet directement d'avoir le résultat : on obtient que  $\bar{e}(\omega, \tau)$  est recouvrable tant que  $|\tau| \leq C\tau(\omega)$  où  $C\tau(\omega)$  est défini par

$$C\tau(\omega) = \min\left(\frac{A}{|\sin 2\omega|}, \frac{\delta}{3|\cos 2\omega|}\right) \quad (3.32)$$

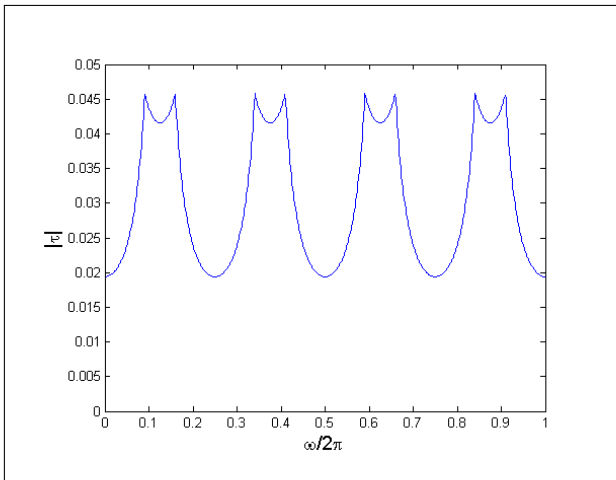
avec

$$A = \min\left(\frac{2\alpha + \beta}{3} - \min(\alpha, \beta), -\frac{2\alpha + \beta}{3} + \max(\alpha, \beta)\right).$$

La courbe  $C\tau$  est représentée figure 3.6 pour l'alliage CuAlNi. Les valeurs correspondantes des paramètres  $\alpha, \beta, \delta$  sont  $\alpha = 0.0425, \beta = -0.0822, \delta = 0.0194$  (BHATTACHARYA, 1991).

**Figure 3.6.**

Valeurs de  $(\omega, \tau)$  telles que  $\bar{e}(\omega, \tau) \in Q\mathcal{K}$  (CuAlNi)



La valeur  $C\tau(\omega)$  dépend de  $\omega$  car le matériau n'est pas isotrope. Nous pouvons cependant remarquer qu'il existe une valeur  $\tau_0$  (proche de 0.02 pour l'exemple étudié) telle que  $\bar{e}(\omega, \tau) \in Q\mathcal{K}$  pour tout  $|\tau| \leq \tau_0$ .

En conclusion, la transformation cubique-orthorombique offre un exemple de situation où la borne convexe  $CW$  donne une bonne estimation de l'énergie  $W$ . Ceci explique que, pour des matériaux obéissant à cette transformation, la fonction énergie  $CW$  a été utilisée dans plusieurs travaux pour étudier la réponse de structures à l'échelle macroscopique (GOVINDJEE et MIEHE, 2001 ; Michaël PEIGNEY et al., 2011 ; Michaël PEIGNEY et al., 2013).

Pour d'autres matériaux (et en particulier pour les alliages TiNi, qui sont les plus couramment utilisés), les déformations de transformation ne sont pas deux à deux compatibles. La fonction  $W$  diffère alors de  $CW$  de façon significative, et il devient nécessaire d'utiliser d'autres bornes sur  $CW$ .

### 3.6. Microstructures laminées

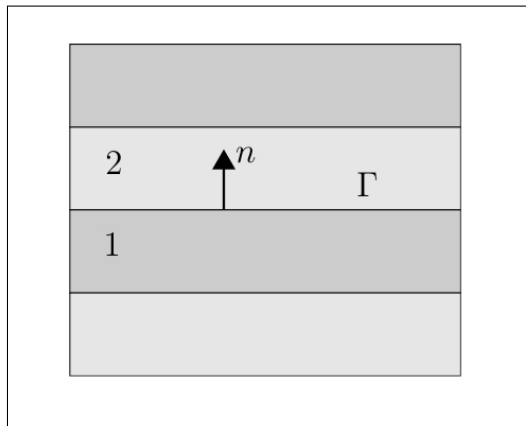
Parmi l'ensemble des microstructures dans  $C$ , les microstructures de type laminé jouent un rôle essentiel dans l'étude de  $W$ . Un laminé consiste en un empilement de couches parallèles, où chaque couche est homogène (c'est-à-dire occupée par une seule phase).

Le vecteur normal  $n$  aux interfaces entre couches successives est appelé direction de lamination (figure 3.7). Les microstructures laminées sont caractérisées par des fonctions  $\chi$  ne dépendant de  $x$  qu'à travers la composante scalaire  $x \cdot n$ , ce qu'on notera sous la forme

$$\chi(x) = \chi'(x \cdot n) \tag{3.33}$$

où  $\chi'$  est une fonction d'une variable réelle.

**Figure 3.7.**  
Structure laminée



Par extension, un champ de déformation  $e(x)$  est dit de type laminé s'il vérifie la même forme de dépendance spatiale que (3.33), à savoir

$$e(x) = e'(x.n) \quad (3.34)$$

où  $e'$  est une fonction d'une variable réelle. Un tel champ  $e(x)$  est constitué d'un empilement de couches parallèles où  $e(x)$  est constant.

### 3.6.1. Condition d'Hadamard

Soit  $e(x)$  un champ de déformation de type laminé. Pour que  $e \in A(\bar{e})$ , il est nécessaire que  $e(x)$  dérive d'un champ de déplacement, c'est-à-dire qu'il existe un déplacement  $u(x)$  tel que  $e = (\nabla u + \nabla^T u)/2$  dans  $\Omega$ . Considérons deux couches successives  $\Omega_1$  et  $\Omega_2$ , où  $e(x)$  prend respectivement les valeurs  $e_1$  et  $e_2$ . Soit  $\Gamma$  l'interface entre  $\Omega_1$  et  $\Omega_2$ . La normale  $n$  à  $\Gamma$  est orientée de  $\Omega_1$  vers  $\Omega_2$ . Dans  $\Omega_i$  ( $i = 1, 2$ ), nécessairement

$$u(x) = e^i . x + \omega^i \wedge x + b^i \quad (3.35)$$

où  $(\omega^i, b^i)$  sont constants. Le terme  $\omega^i \wedge x + b^i$  s'interprète comme un déplacement de corps rigide.

Par ailleurs, en un point  $x$  de  $\Gamma$ , la continuité du déplacement implique qu'il existe un vecteur  $a$  tel que

$$e_2 - e_1 = a \otimes n + n \otimes a \quad (3.36)$$

où  $n$  est la normale à  $\Gamma$  en  $x$ .

La condition (3.36) est appelée condition d'Hadamard. En utilisant cette condition, on obtient que les vecteurs  $(\omega^2, b^2)$  et  $(\omega^1, b^1)$  dans (3.35) sont liés par les relations

$$\begin{aligned} \omega_2 - \omega_1 &= n \wedge a \\ b_2 - b_1 &= -2(x.n)a. \end{aligned} \quad (3.37)$$

Deux déformations données  $(e_1, e_2)$  sont dites géométriquement compatibles si elles vérifient (3.36) pour un certain couple de vecteurs  $(a, n)$ . Dans ce cas, il est possible de construire des champs de déformation intégrable et à valeurs dans  $\{e_1, e_2\}$ . Ces champs de déformations sont des laminés dont la direction de lamination est solution de (3.36). Les expressions (3.35) et (3.37) déterminent alors l'expression du champ de déplacement.

### 3.6.2. Convexité de rang 1

La construction détaillée précédemment joue un rôle essentiel pour établir l'inégalité

$$W(\theta \bar{e}_1 + (1 - \theta) \bar{e}_2) \leq \theta W(\bar{e}_1) + (1 - \theta) W(\bar{e}_2), \quad (3.38)$$

valable pour toutes déformations compatibles  $(\bar{e}_1, \bar{e}_2)$  et pour tout  $\theta \in [0, 1]$ . En fait  $W$  vérifie une propriété plus forte, à savoir

$$W(\theta \bar{e}_1 + (1 - \theta) \bar{e}_2) \leq \theta W(\bar{e}_1) + (1 - \theta) W(\bar{e}_2), \quad (3.39)$$

pour toutes déformations compatibles  $(\bar{e}_1, \bar{e}_2)$  et pour tout  $\theta \in [0, 1]$ . Étant donné que  $W \leq w$  (cf Eq. 3.11), la propriété (3.38) est une conséquence directe de (3.39).

L'inégalité (3.39) n'est pas sans rappeler la propriété caractéristique de la convexité : une fonction convexe vérifie (3.39) pour tout  $(\bar{e}_1, \bar{e}_2)$  et  $\theta \in [0, 1]$ . La fonction  $W$  vérifie une condition plus faible que la convexité, dans la mesure où l'inégalité (3.39) n'est satisfaite que si  $\bar{e}_1$  et  $\bar{e}_2$  sont compatibles. Une telle fonction est dite *convexe de rang 1*.

Pour illustrer cette propriété, reprenons l'exemple du matériau à deux variantes introduit (3.8)-(3.9). Alors  $w(0) = \min(m_0, 1/2e_1 : L : e_1) > 0$ . L'inégalité (3.39) utilisée avec  $\bar{e}_i = e_i$  et  $\theta = 1/2$  montre que  $W(0) = 0$ . Donc  $W(0) < w(0)$ , ce qui illustre le fait que  $W$  est strictement inférieure à  $w$ .

Afin de justifier l'origine de (3.38), considérons un champ de déformations  $e(x)$  de type laminé, alternant des couches d'épaisseur  $\theta h$  où  $e(x) = e_1$  avec des couches d'épaisseur  $(1 - \theta)h$  où  $e(x) = e_2$ . Il est tentant de chercher à démontrer (3.38) en utilisant  $e$  dans la définition de  $W$  : ce champ est compatible et donne  $\int_{\Omega} w(e) = \theta w(e_1) + (1 - \theta)w(e_2)$ .

Comme le champ ne vérifie pas la condition limite sur le bord, une «zone de transition»  $\Gamma$  est introduite au voisinage et le déplacement  $y$  est modifié de façon à respecter la condition aux limites, tout en s'assurant que la déformation reste bornée quand la taille de  $\Gamma$  tend vers 0. Plus de détails sur cette construction dans (BHATTACHARYA, 2003). Les propriétés (3.38)-(3.39) seront utilisées dans la suite pour obtenir des bornes sur  $W$  et sur les déformations à énergie minimale.

### 3.6.3. Lamination séquentielle

La relation (3.38) permet de construire une borne supérieure sur  $W$ . En effet, pour une déformation  $\bar{e}$  donnée, on a l'inégalité  $W(\bar{e}) \leq \theta w(\bar{e}_1) + (1 - \theta)w(\bar{e}_2)$  pour tout  $(\bar{e}_1, \bar{e}_2, \theta)$  tel que

$$(\bar{e}_1, \bar{e}_2) \text{ compatibles, } \theta \in [0, 1], \bar{e} = \theta \bar{e}_1 + (1 - \theta) \bar{e}_2. \tag{3.40}$$

Soit  $\Pi(\bar{e})$  l'ensemble des valeurs vérifiant (3.40). Alors

$$W(\bar{e}) \leq R_1 W(\bar{e}) \tag{3.41}$$

avec

$$R_1 W(\bar{e}) = \inf_{(\bar{e}_1, \bar{e}_2, \theta) \in \Pi(\bar{e})} \theta w(\bar{e}_1) + (1 - \theta)w(\bar{e}_2).$$

La fonction  $R_1 W$  ainsi définie est donc une borne supérieure sur l'énergie effective  $W$ . Pour tout  $(\bar{e}_1, \bar{e}_2, \theta) \in \Pi(\bar{e})$ , on sait que l'énergie  $\theta w(\bar{e}_1) + (1 - \theta)w(\bar{e}_2)$  est atteinte par une microstructure laminée infiniment fine (selon la construction évoquée en 3.6.2). La quantité  $R_1 W(\bar{e})$  peut donc être interprétée comme l'énergie minimale atteignable par l'ensemble des microstructures laminées.



En combinant (3.39) et (3.41), on observe que

$$W(\bar{e}) \leq \theta R_1 W(\bar{e}_1) + (1 - \theta) R_1 W(\bar{e}_2)$$

pour tout  $(\bar{e}_1, \bar{e}_2, \theta) \in \Pi(\bar{e})$ . Par suite,

$$W(\bar{e}) \leq R_2 W(\bar{e}), \tag{3.42}$$

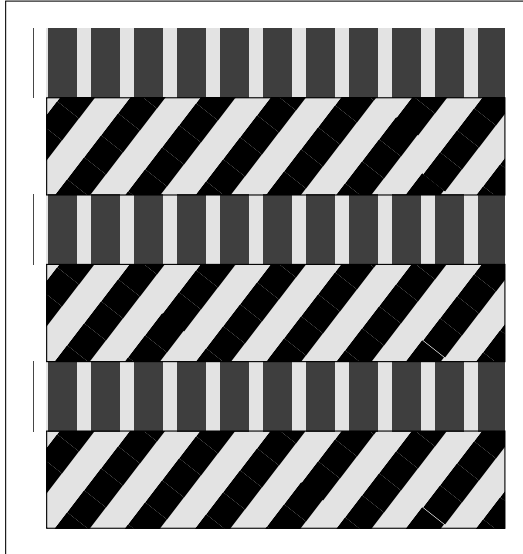
avec

$$R_2 W(\bar{e}) = \inf_{(\bar{e}_1, \bar{e}_2, \theta) \in \Pi(\bar{e})} \theta R_1 W(\bar{e}_1) + (1 - \theta) R_1 W(\bar{e}_2).$$

Comme il a été vu précédemment, l'énergie  $\theta R_1 W(\bar{e}_1) + (1 - \theta) R_1 W(\bar{e}_2)$  est atteinte par un champ de déformation laminé, à valeurs dans  $\{\bar{e}_1, \bar{e}_2\}$  et dont le pas de lamination tend vers 0. Pour chaque déformation  $\bar{e}_i$ , l'énergie  $R_1 W(\bar{e}_i)$  est elle-même atteinte par une microstructure laminée. L'énergie  $\theta R_1 W(\bar{e}_1) + (1 - \theta) R_1 W(\bar{e}_2)$  est ainsi réalisée par un laminé de rang 2 (figure 3.8).

**Figure 3.8.**

Structure laminée de rang 2



Cet argument peut être poursuivi de façon séquentielle : pour tout  $r \geq 1$ , on a

$$W(\bar{e}) \leq R_{r+1} W(\bar{e})$$

avec

$$R_{r+1} W(\bar{e}) = \inf_{(\bar{e}_1, \bar{e}_2, \theta) \in \Pi(\bar{e})} \theta R_r W(\bar{e}_1) + (1 - \theta) R_r W(\bar{e}_2). \tag{3.43}$$

Cette relation est également valable pour  $r = 0$  si l'on adopte la convention  $R_0 W = w$ . La quantité  $R_r W(\bar{e})$  est l'énergie minimale atteignable par les microstructures laminées de rang  $r$ . Comme conséquence immédiate de la

définition (3.43), on observe que  $R_{r+1}W \leq R_rW$  : chaque étape de lamination supplémentaire améliore potentiellement la borne obtenue sur  $W$ . On observera cependant que le coût de calcul augmente (en loi puissance) avec  $r$ .

En pratique, pour les problèmes associés aux alliages les plus courants, il est difficile de dépasser  $r = 2$  (GOVINDJEE, HACKL et al., 2007). Par ailleurs, en général  $W < \inf_r R_rW$ , ce qui exprime le fait que les microstructures optimales ne se limitent pas à la classe des laminés.

L'exploitation de ces bornes est plus facile si l'on se restreint à étudier les déformations d'énergie minimale. Plaçons-nous à  $T < T^{crit}$  et notons  $m$  la valeur minimale de  $w$ , atteinte pour  $e \in \mathcal{K}$ . Supposons que deux déformations  $e_i$  et  $e_j$  dans  $\mathcal{K}$  soient compatibles. Pour tout  $\theta \in [0, 1]$ , par (3.41)

$$m \leq W(\theta e_i + (1 - \theta)e_j) \leq \theta w(e_i) + (1 - \theta)w(e_j) = m$$

donc  $\theta e_i + (1 - \theta)e_j \in Q\mathcal{K}$ . Il s'ensuit que  $R_1\mathcal{K} \subset Q\mathcal{K}$  où

$$R_1\mathcal{K} = \bigcup_{e_i \in \mathcal{K} \text{ et } e_j \in \mathcal{K} \text{ compatibles}} [e_i, e_j].$$

Cet ensemble est une borne inférieure (au sens de l'inclusion d'ensembles) sur  $Q\mathcal{K}$ . C'est l'ensemble des déformations réalisables par des laminés de rang 1. Il correspond également à l'ensemble des déformations minimisant la fonction  $R_1W$ .

Comme précédemment, cette construction peut être poursuivie de façon séquentielle :

$$R_r\mathcal{K} \subset Q\mathcal{K}$$

où  $R_r\mathcal{K}$  est l'ensemble des déformations réalisables par des laminés de rang  $r$ . Cet ensemble correspond à l'ensemble des déformations minimisant la fonction  $R_rW$  définie plus haut. De plus, pour tout  $r \geq 1$

$$R_{n+1}\mathcal{K} = \bigcup_{e_i \in R_n\mathcal{K} \text{ et } e_j \in R_n\mathcal{K} \text{ compatibles}} [e_i, e_j].$$

Cette relation reste valable pour  $r = 0$  en adoptant la convention  $R_0\mathcal{K} = \mathcal{K}$ .

De façon générale, on notera que  $R_0\mathcal{K}$  est un ensemble discret, c'est-à-dire un ensemble de dimension 0. L'ensemble  $R_1\mathcal{K}$  est formé par un ensemble de segments, c'est-à-dire un ensemble de dimension 1. De la même façon,  $R_2\mathcal{K}$  est une surface (dans l'espace des déformations), c'est-à-dire une variété de dimension 2. De façon itérative

$$\dim R_r\mathcal{K} \leq r. \tag{3.44}$$

### 3.7. Problèmes à 4 phases

Afin d'illustrer les bornes par lamination  $R_r\mathcal{K}$ , considérons un problème à 4 phases  $\mathcal{K} = \{e_1, e_2, e_3, e_4\}$ . On choisit

$$\begin{aligned} e_1 &= \begin{pmatrix} \alpha & \delta & \epsilon \\ \delta & \alpha & \epsilon \\ \epsilon & \epsilon & \beta \end{pmatrix}, & e_2 &= \begin{pmatrix} \alpha & -\delta & -\epsilon \\ -\delta & \alpha & \epsilon \\ -\epsilon & \epsilon & \beta \end{pmatrix}, \\ e_3 &= \begin{pmatrix} \alpha & -\epsilon & \delta \\ -\epsilon & \beta & -\epsilon \\ \delta & -\epsilon & \alpha \end{pmatrix}, & e_4 &= \begin{pmatrix} \alpha & \epsilon & -\delta \\ \epsilon & \beta & -\epsilon \\ -\delta & -\epsilon & \alpha \end{pmatrix}. \end{aligned} \quad (3.45)$$

On a vu que  $R_r\mathcal{K} \subset Q\mathcal{K} \subset C\mathcal{K}$ . Toute déformation  $\bar{e}$  dans  $C\mathcal{K}$  admet une représentation de la forme

$$\bar{e} = \sum_{i=1}^3 \theta_i e_i + \left(1 - \sum_{i=1}^3 \theta_i\right) e_4 \quad (3.46)$$

où  $(\theta_1, \theta_2, \theta_3)$  appartient au tétraèdre  $\mathcal{T}'_3$  défini par

$$\mathcal{T}'_3 = \left\{ (\theta_1, \theta_2, \theta_3) \mid \theta_i \geq 0; \sum_{i=1}^3 \theta_i \leq 1 \right\}.$$

Sauf cas particulier, les tenseurs  $e_1, e_2, e_3$  et  $e_4$  dans (3.46) sont linéairement indépendants. La décomposition (3.46) d'un tenseur  $\bar{e}$  est donc unique. En d'autres termes, la fonction  $\mathcal{F}$  définie par

$$\begin{aligned} \mathcal{F}: \mathcal{T}'_3 &\longrightarrow C\mathcal{K} \\ (\theta_1, \theta_2, \theta_3) &\longrightarrow \sum_{i=1}^3 \theta_i e_i + \left(1 - \sum_{i=1}^3 \theta_i\right) e_4 \end{aligned} \quad (3.47)$$

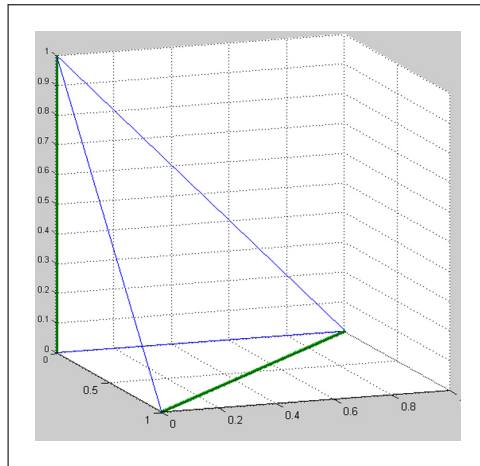
définit une bijection entre  $\mathcal{T}'_3$  et  $C\mathcal{K}$ . En pratique, cette fonction  $\mathcal{F}$  permet d'obtenir des représentations tridimensionnelles de  $C\mathcal{K}$  et de ses sous-ensembles  $R_r\mathcal{K}$ . Tout sous-ensemble  $\mathcal{E}$  peut en effet être identifié à son image réciproque  $\mathcal{F}^{-1}(\mathcal{E})$ , qui est un domaine dans  $\mathbb{R}^3$ . Ainsi, l'enveloppe convexe s'identifie au tétraèdre  $\mathcal{T}'_3$ , et les déformations  $e_i$  dans  $\mathcal{K}$  s'identifient aux 4 sommets du tétraèdre. Il est vérifié à partir des expressions (3.45) que  $\{e_1, e_2\}$  et  $\{e_3, e_4\}$  sont les seules paires de variantes compatibles dans  $\mathcal{K}$ . Donc

$$R_1\mathcal{K} = [e_1, e_2] \cup [e_3, e_4].$$

Cet ensemble (de dimension 1, c'est-à-dire une courbe) est représenté sur la figure 3.9.

**Figure 3.9.**

Bornes par laminations de rang 1 pour un problème à 4 phases



L'ensemble  $R_2\mathcal{K}$  est formé des déformations de la forme

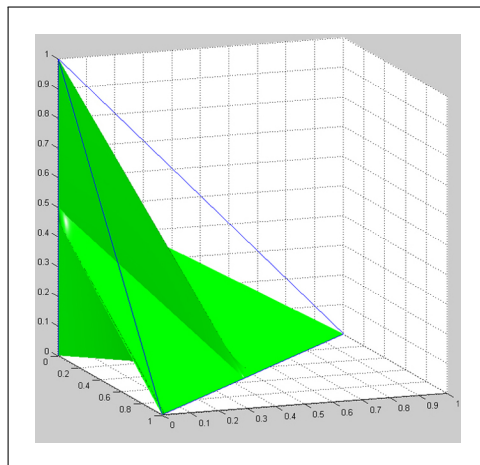
$$\theta(ae_1 + (1 - a)e_2) + (1 - \theta)(be_3 + (1 - b)e_4) \quad (3.48)$$

où  $(\theta, a, b)$  sont dans  $[0, 1]$  et tels que les déformations  $\{ae_1 + (1 - a)e_2, be_3 + (1 - b)e_4\}$  sont compatibles.

Cet ensemble  $R_2\mathcal{K}$  de dimension 2, c'est-à-dire une surface, est représenté figure 3.10.

**Figure 3.10.**

Bornes par laminations de rang 2 pour un problèmes à 4 phases



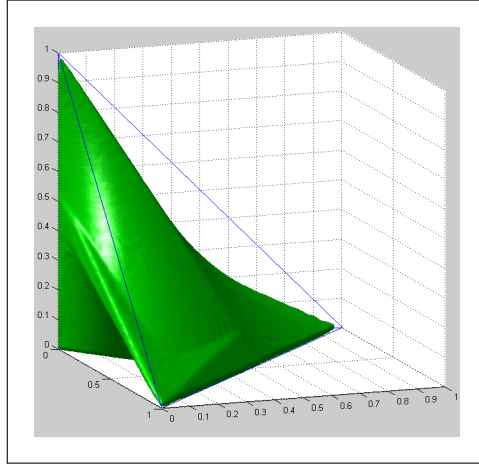
Pour  $a$  donné, il existe en effet au plus 2 valeurs de  $b$  tels que  $\{ae_1 + (1 - a)e_2, be_3 + (1 - b)e_4\}$  sont compatibles.

Les points de  $R_2\mathcal{K}$  sont donc paramétrés par les variables  $(b, \theta)$ , ce qui exprime le fait que  $R_2\mathcal{K}$  est une surface.

L'ensemble  $R_3\mathcal{K}$ , défini à partir de  $R_2\mathcal{K}$  via (3.43), est représenté figure 3.11. C'est un domaine de volume non nul.

**Figure 3.11.**

Bornes par laminations de rang 3 pour un problèmes à 4 phases



On a la chaîne d'inclusions

$$\mathcal{K} = R_0\mathcal{K} \subset R_1\mathcal{K} \subset \dots \subset R_r\mathcal{K} \subset Q\mathcal{K} \subset C\mathcal{K} \subset \text{vect}(\mathcal{K})$$

ce qui exprime le fait que  $R_n\mathcal{K}$  et  $C\mathcal{K}$  donnent un encadrement (au sens de l'inclusion d'ensembles) sur  $Q\mathcal{K}$ . Comme tous les ensembles considérés sont dans l'espace vectoriel  $\text{vect}(\mathcal{K})$ , la qualité de l'encadrement obtenu peut être estimée en comparant la mesure de ces ensembles dans  $\text{vect}(\mathcal{K})$ . Pour un ensemble  $\mathcal{E}$  donné dans  $\text{vect} \mathcal{K}$ , sa mesure  $|\mathcal{E}|$  est définie par

$$|\mathcal{E}| = \int_{e \in \text{vect}(\mathcal{K})} \chi_{\mathcal{E}}(e) de \tag{3.49}$$

où  $\chi_{\mathcal{E}}(e)$  est égal à 1 si  $e \in \mathcal{E}$ , et nul sinon. La mesure  $|\mathcal{E}|$  d'un ensemble donné  $\mathcal{E}$  est directement reliée au volume (dans  $\mathbb{R}^3$ ) de sa représentation tridimensionnelle  $\mathcal{F}^{-1}(\mathcal{E})$ . En effet, comme  $\mathcal{F}$  est affine,

$$|\mathcal{E}| = J \int_{\theta \in \mathcal{F}^{-1}(\mathcal{E})} d\theta \tag{3.50}$$

où  $J$  est le jacobien de  $\mathcal{F}$  et égal au produit mixte  $[e_1 - e_4, e_2 - e_4, e_3 - e_4]$ .

Par exemple, comme  $\mathcal{F}^{-1}(C\mathcal{K}) = \mathcal{T}'_3$  et  $\int_{\theta \in \mathcal{T}'_3} d\theta = 1/6$ , alors

$$|CK| = \frac{1}{6}[e_1 - e_4, e_2 - e_4, e_3 - e_4].$$

On a  $|R_1\mathcal{K}| = |R_2\mathcal{K}| = 0$  car  $R_1\mathcal{K}$  et  $R_2\mathcal{K}$  sont respectivement une courbe et une surface. Par contre,  $R_3\mathcal{K}$  a une mesure non nulle.

On pourrait espérer obtenir un ensemble plus grand en poursuivant le processus de lamination. Cependant, à la précision numérique des calculs près, on obtient que  $R_4\mathcal{K} = R_3\mathcal{K}$  (et par suite que  $R_r\mathcal{K} = R_3\mathcal{K}$  pour  $r \geq 3$ ).

Cette construction peut être réalisée pour tout problème à 4 phases (on trouvera d'autres exemples dans (Michaël PEIGNEY, 2013a)). Notons que si les déformations  $e_i$  sont toutes incompatibles entre elles, alors  $R_1\mathcal{K} = \mathcal{K}$  et par suite  $R_r\mathcal{K} = \mathcal{K}$  pour tout  $r \geq 1$ . Ceci n'implique pas que  $Q\mathcal{K} = \mathcal{K}$ . Par exemple, même si  $e_1, e_2, e_3$  sont incompatibles, il peut exister  $e \in [e_2, e_3]$  compatible avec  $e_1$ , auquel cas  $[e_2, e] \subset Q\mathcal{K}$  (BHATTACHARYA, FIROOZYE et al., 1994).

À l'inverse, si toutes les déformations  $e_1, e_2, e_3, e_4$  sont compatibles deux à deux, alors on peut montrer (BHATTACHARYA, 1993) que

$$R_3\mathcal{K} = Q\mathcal{K} = C\mathcal{K}.$$

Cette propriété n'est pas limitée au cas de 4 phases et s'étend au cas plus général  $\mathcal{K} = \{e_1, \dots, e_n\}$  : si les déformations  $e_i$  sont deux à deux compatibles, alors

$$R_{n-1}\mathcal{K} = Q\mathcal{K} = C\mathcal{K}. \tag{3.51}$$

Dans ce cas, l'ensemble des déformations minimisant  $W$  coïncide avec l'enveloppe convexe de  $\mathcal{K}$ . De plus, toute déformation dans  $Q\mathcal{K}$  est réalisée par une microstructure laminée de rang  $n - 1$ . Ceci n'exclut pas la formation d'autres microstructures.

### 3.8. Bornes inférieures non convexes sur $W$

La borne (3.24) peut être améliorée en utilisant la méthode de translation (LURIE et al., 1984; MURAT et al., 1985; MILTON, 2004). L'idée centrale est d'appliquer la démarche précédente en remplaçant  $w_i$  par  $w_i - u$  où  $u$  est une fonction à ce stade arbitraire. Tant que  $w_i - u$  est convexe pour tout  $i$ , les relations générales écrites au paragraphe précédent restent valables : on a

$$(W - U) = -(W - U)^* \tag{3.52}$$

où

$$(W - U) = \inf_{e \in A(\bar{e})} \frac{1}{|\Omega|} \int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x})(w_i - u)(e) d\omega \tag{3.53}$$

et

$$(W - U)^* = \inf_{\sigma | \operatorname{div} \sigma = 0} \frac{1}{|\Omega|} \int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x})(w_i - u)^*(\sigma) d\omega - \bar{e} : \bar{\sigma}. \quad (3.54)$$

Le raisonnement suit les mêmes lignes que précédemment : en se restreignant à des champs  $\sigma$  constants dans (3.54), on obtient une majoration sur  $(W - U)^*$ , dont on déduit - via la formule de Clapeyron (3.52) - une minoration sur  $(W - U)$ . Plus précisément :

$$\sup_{\bar{\sigma}} \left\{ \bar{e} : \bar{\sigma} - \sum_{i=0}^n \theta_i (w_i - u)^*(\bar{\sigma}) \right\} \leq (W - U). \quad (3.55)$$

Supposons que  $u$  est quasiconvexe, c'est-à-dire vérifie

$$\frac{1}{|\Omega|} \int_{\Omega} u(e) d\omega \geq u(\bar{e})$$

pour tout  $e \in A(\bar{e})$ . Alors

$$\int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x})(w_i - u)(e) d\omega \leq -|\Omega| u(\bar{e}) + \int_{\Omega} \sum_{i=0}^n \chi_i(\mathbf{x}) w_i(e) d\omega.$$

On obtient ainsi la relation

$$u(\bar{e}) + \frac{1}{|\Omega|} \sup_{\bar{\sigma}} \left\{ \bar{e} : \bar{\sigma} - \sum_{i=0}^n \theta_i (w_i - u)^*(\bar{\sigma}) \right\} \leq W(\bar{e}; \chi). \quad (3.56)$$

Le membre de gauche est une borne inférieure sur  $W(\bar{e}; \chi)$ .

Pour obtenir des bornes explicites, il faut préciser et bien choisir le potentiel  $u$  utilisé. Une voie explorée dans (Michaël PEIGNEY, 2009) est de considérer des fonctions  $u$  quadratiques, c'est-à-dire pouvant s'écrire sous la forme

$$u(e) = \frac{1}{2} e : \mathbf{K} : e$$

où  $\mathbf{K}$  est un tenseur symétrique d'ordre 4. Pour les fonctions quadratiques, on sait en particulier que la quasiconvexité est équivalente à la convexité de rang 1, condition plus facile à manipuler. Tant que  $\mathbf{L} - \mathbf{K}$  est défini positif, alors  $(w_i - u)^*(\bar{\sigma})$  est fini et donné par

$$(w_i - u)^*(\bar{\sigma}) = \frac{1}{2} (\bar{\sigma} + \mathbf{L} : e_i) : (\mathbf{L} - \mathbf{K})^{-1} : (\bar{\sigma} + \mathbf{L} : e_i) - \frac{1}{2} e_i : \mathbf{L} : e_i - m_i.$$

En remplaçant dans (3.56) puis en maximisant par rapport à  $\bar{\sigma}$  ;

$$W(\bar{e}; \theta) \geq CW(\bar{e}, \theta) + G(\theta, \mathbf{K}) \quad (3.57)$$

avec  $CW(\bar{e}, \theta)$  donné par (3.23) et

$$G(\theta, \mathbf{K}) = -\frac{1}{2} \left( \sum_{i=0}^n \theta_i e_i \right) : \mathbf{M}(\mathbf{K}) : \left( \sum_{i=0}^n \theta_i e_i \right) + \frac{1}{2} \sum_{i=0}^n \theta_i e_i : \mathbf{M}(\mathbf{K}) : e_i$$

où

$$\mathbf{M}(\mathbf{K}) = \mathbf{L} - \mathbf{L} : (\mathbf{L} - \mathbf{K})^{-1} : \mathbf{L}.$$

L'inégalité (3.57) vaut pour tout  $\mathbf{K}$  quasiconvexe tel que  $\mathbf{L} - \mathbf{K} > 0$ .

En considérant une famille  $\mathcal{C}$  de tenseurs  $\mathbf{K}$  quasiconvexes, alors

$$W(\bar{e}; \theta) \geq CW(\bar{e}, \theta) + G(\theta)$$

avec

$$G(\theta) = \sup_{\mathbf{K} \in \mathcal{C} | \mathbf{L} - \mathbf{K} > 0} -\frac{1}{2} \left( \sum_{i=0}^n \theta_i e_i \right) : \mathbf{M}(\mathbf{K}) : \left( \sum_{i=0}^n \theta_i e_i \right) + \frac{1}{2} \sum_{i=0}^n \theta_i e_i : \mathbf{M}(\mathbf{K}) : e_i. \quad (3.58)$$

Une telle famille  $\mathcal{C}$  peut être construite en considérant les fonctions de la forme  $e \mapsto -a : e^*$  où  $a$  est symétrique positif et  $e^*$  est le tenseur adjugué de  $e$ , donné en représentation matricielle (dans une base orthonormée) par

$$e_{ii}^* = e_{jj} e_{kk} - e_{jk}^2, \quad e_{jk}^* = (-1)^{i+j} (e_{ji} e_{ki} - e_{jk} e_{ii})$$

pour toute permutation  $\{i, j, k\}$  de  $\{1, 2, 3\}$ . Nous pouvons vérifier que la fonction quadratique  $e \mapsto -a : e^*$  est quasi-onvexe si  $a \geq 0$  (Michaël PEIGNEY, 2008). Soit  $\mathbf{K}(a)$  le tenseur symétrique d'ordre 4 tel que  $\frac{1}{2} e : \mathbf{K}(a) : e = -a : e^*$  pour tout  $e$ . En représentation matricielle,  $\mathbf{K}(a)$  est défini par les relations

$$\begin{aligned} K_{iijj} &= -a_{kk}, \quad K_{iijk} = K_{jkii} = \frac{1}{2} a_{jk}, \quad K_{ijij} = K_{jiij} = \frac{1}{2} a_{kk}, \\ K_{ijik} &= K_{jii k} = K_{ijk i} = K_{jiki} = -\frac{1}{4} a_{jk} \end{aligned} \quad (3.59)$$

pour toute permutation  $\{i, j, k\}$  de  $\{1, 2, 3\}$ . En utilisant la famille de fonctions quasiconvexes ainsi définie, l'expression (3.58) devient

$$G(\theta) = \sup_{\mathbf{a} \geq 0 | \mathbf{L} - \mathbf{K}(\mathbf{a}) \geq 0} \frac{1}{4} \sum_{i,j} \theta_i \theta_j (e_i - e_j) : \mathbf{M}(\mathbf{a}) : (e_i - e_j) \quad (3.60)$$

où  $\mathbf{M}(\mathbf{a}) = \mathbf{L} - \mathbf{L} : (\mathbf{L} - \mathbf{K}(\mathbf{a}))^{-1} : \mathbf{L}$ .

Nous pouvons trouver dans (Michaël PEIGNEY, 2013c; Michaël PEIGNEY, 2009) des illustrations de la borne énergétique (3.57).

Ici nous utilisons cette borne pour estimer l'ensemble des déformations à énergie macroscopique minimale. Pour  $T < T^{crit}$ , toute déformation  $\bar{e} \in Q\mathcal{K}$  peut s'écrire sous la forme  $\bar{e} = \sum_{i=1}^n \theta_i e_i$  où  $\theta \in \mathcal{T}$  vérifie  $G(\theta) = 0$ . Nous pouvons montrer



(Michaël PEIGNEY, 2013a) que

$$G(\boldsymbol{\theta}) = 0 \iff \sum_{i,j} \theta_i \theta_j (\mathbf{e}_i - \mathbf{e}_j)^* \leq 0.$$

Par conséquent,

$$Q\mathcal{K} \subset P\mathcal{K}$$

où

$$P\mathcal{K} = \left\{ \sum_{i=1}^n \mathbf{e}_i \mid \boldsymbol{\theta} \in \mathcal{T}; \sum_{i,j} \theta_i \theta_j (\mathbf{e}_i - \mathbf{e}_j)^* \leq 0 \right\}.$$

Cet ensemble est inclus dans l'enveloppe convexe  $C\mathcal{K}$  et fournit donc en général une borne plus précise que  $C\mathcal{K}$ .

### 3.9. Transformations cubiques-monocliniques

Nous nous intéressons dans cette section aux transformations cubiques-monocliniques, pour lesquelles il y a 12 variantes de martensite. Il y a deux types de transformations cubiques-monocliniques, listées dans les tables 3.2 et 3.3. Comme pour la table 3.1, les représentations matricielles sont relatives au repère orthonormé de la structure cubique de l'austénite.

Pour ces deux types de transformation, chaque variante est compatible avec seulement 7 des 11 autres. En particulier, les variantes ne sont pas deux à deux compatibles, et l'on peut montrer que les ensembles  $Q\mathcal{K}$  pour ces transformations ne sont pas convexes (BHATTACHARYA et KOHN, 1997 ; Michaël PEIGNEY, 2013a).

La structure de l'ensemble  $Q\mathcal{K}$  a jusqu'à présent été étudiée principalement à l'aide de bornes convexes. Puisque  $Q\mathcal{K}$  n'est pas convexe, de telles bornes sont nécessairement *strictes*. Par exemple, l'enveloppe convexe de  $\mathcal{K}$  a été utilisée comme borne supérieure pour étudier les déformations recouvrables dans les alliages NiTi (SHU et al., 1998). Une borne inférieure convexe a été proposée dans (BHATTACHARYA et KOHN, 1997) en observant que  $\mathbf{e}_{2i-1}$  and  $\mathbf{e}_{2i}$  sont compatibles pour  $i = 1, \dots, 6$ .

Il en découle que  $Q\mathcal{K}$  contient les déformations  $\mathbf{e}'_i = (\mathbf{e}_{2i-1} + \mathbf{e}_{2i})/2$ . Ces déformations  $\mathbf{e}'_i$  correspondent à la transformation cubique-orthorombique (voir table 3.1) : on sait que les tenseurs  $\mathbf{e}'_i$  sont deux à deux compatibles. Par conséquent,  $Q\mathcal{K}$  contient l'enveloppe convexe  $\mathcal{S}$  de  $\mathbf{e}'_1, \dots, \mathbf{e}'_6$ , donnée par les expressions (3.29). En d'autres termes, l'ensemble  $\mathcal{S}$  défini par (3.29) est une borne inférieure convexe sur  $Q\mathcal{K}$ . Ceci prouve en particulier que  $Q\mathcal{K}$  est de dimension 5 pour les transformations monocliques.

Une représentation schématique des bornes convexes  $\mathcal{S}$  et  $C\mathcal{K}$  est donnée figure 3.12. Y sont représentées les 12 déformations  $\mathbf{e}_i$  comme des points cocycliques. Les déformations qui sont compatibles sont reliées par un segment (par souci de lisibilité, chaque déformation  $\mathbf{e}_i$  est représentée comme étant compatible avec seulement 2 des 11 autres). La borne inférieure  $\mathcal{S}$  (en vert) est

l'enveloppe convexe de 6 points particuliers obtenus comme milieux de segments reliant deux déformations compatibles. La borne supérieure  $CK$  est l'union des domaines verts et rouges.

**Table 3.2.**  
Transformation cubique-monoclinique-I

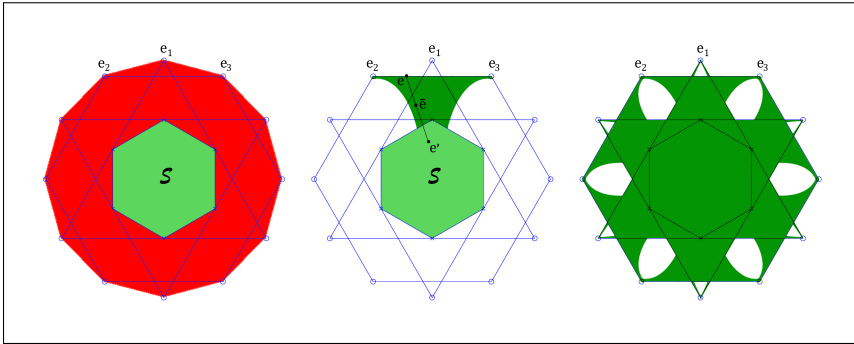
$e_1^I$	$e_2^I$	$e_3^I$	$e_4^I$
$\begin{pmatrix} \alpha & \delta & \epsilon \\ \delta & \alpha & \epsilon \\ \epsilon & \epsilon & \beta \end{pmatrix}$	$\begin{pmatrix} \alpha & \delta & -\epsilon \\ \delta & \alpha & -\epsilon \\ -\epsilon & -\epsilon & \beta \end{pmatrix}$	$\begin{pmatrix} \alpha & -\delta & -\epsilon \\ -\delta & \alpha & \epsilon \\ -\epsilon & \epsilon & \beta \end{pmatrix}$	$\begin{pmatrix} \alpha & -\delta & \epsilon \\ -\delta & \alpha & -\epsilon \\ \epsilon & -\epsilon & \beta \end{pmatrix}$
$e_5^I$	$e_6^I$	$e_7^I$	$e_8^I$
$\begin{pmatrix} \alpha & \epsilon & \delta \\ \epsilon & \beta & \epsilon \\ \delta & \epsilon & \alpha \end{pmatrix}$	$\begin{pmatrix} \alpha & -\epsilon & \delta \\ -\epsilon & \beta & -\epsilon \\ \delta & -\epsilon & \alpha \end{pmatrix}$	$\begin{pmatrix} \alpha & -\epsilon & -\delta \\ -\epsilon & \beta & \epsilon \\ -\delta & \epsilon & \alpha \end{pmatrix}$	$\begin{pmatrix} \alpha & \epsilon & -\delta \\ \epsilon & \beta & -\epsilon \\ -\delta & -\epsilon & \alpha \end{pmatrix}$
$e_9^I$	$e_{10}^I$	$e_{11}^I$	$e_{12}^I$
$\begin{pmatrix} \beta & \epsilon & \epsilon \\ \epsilon & \alpha & \delta \\ \epsilon & \delta & \alpha \end{pmatrix}$	$\begin{pmatrix} \beta & -\epsilon & -\epsilon \\ -\epsilon & \alpha & \delta \\ -\epsilon & \delta & \alpha \end{pmatrix}$	$\begin{pmatrix} \beta & -\epsilon & \epsilon \\ -\epsilon & \alpha & -\delta \\ \epsilon & -\delta & \alpha \end{pmatrix}$	$\begin{pmatrix} \beta & \epsilon & -\epsilon \\ \epsilon & \alpha & -\delta \\ -\epsilon & -\delta & \alpha \end{pmatrix}$

**Table 3.3.**  
Transformation cubique-monoclinique-II

$e_1^{II}$	$e_2^{II}$	$e_3^{II}$
$\begin{pmatrix} \alpha + \epsilon & \delta & 0 \\ \delta & \alpha - \epsilon & 0 \\ 0 & 0 & \beta \end{pmatrix}$	$\begin{pmatrix} \alpha - \epsilon & \delta & 0 \\ \delta & \alpha + \epsilon & 0 \\ 0 & 0 & \beta \end{pmatrix}$	$\begin{pmatrix} \alpha + \epsilon & -\delta & 0 \\ -\delta & \alpha - \epsilon & 0 \\ 0 & 0 & \beta \end{pmatrix}$
$e_4^{II}$	$e_5^{II}$	$e_6^{II}$
$\begin{pmatrix} \alpha - \epsilon & -\delta & 0 \\ -\delta & \alpha + \epsilon & 0 \\ 0 & 0 & \beta \end{pmatrix}$	$\begin{pmatrix} \alpha + \epsilon & 0 & \delta \\ 0 & \beta & 0 \\ \delta & 0 & \alpha - \epsilon \end{pmatrix}$	$\begin{pmatrix} \alpha - \epsilon & 0 & \delta \\ 0 & \beta & 0 \\ \delta & 0 & \alpha + \epsilon \end{pmatrix}$
$e_7^{II}$	$e_8^{II}$	$e_9^{II}$
$\begin{pmatrix} \alpha + \epsilon & 0 & -\delta \\ 0 & \beta & 0 \\ -\delta & 0 & \alpha - \epsilon \end{pmatrix}$	$\begin{pmatrix} \alpha - \epsilon & 0 & -\delta \\ 0 & \beta & 0 \\ -\delta & 0 & \alpha + \epsilon \end{pmatrix}$	$\begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha - \epsilon & \delta \\ 0 & \delta & \alpha + \epsilon \end{pmatrix}$
$e_{10}^{II}$	$e_{11}^{II}$	$e_{12}^{II}$
$\begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha + \epsilon & \delta \\ 0 & \delta & \alpha - \epsilon \end{pmatrix}$	$\begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha - \epsilon & -\delta \\ 0 & -\delta & \alpha + \epsilon \end{pmatrix}$	$\begin{pmatrix} \beta & 0 & 0 \\ 0 & \alpha + \epsilon & -\delta \\ 0 & -\delta & \alpha - \epsilon \end{pmatrix}$

**Figure 3.12.**

Représentation schématique des bornes pour les problèmes à 12 phases



L'écart peut être estimé entre les bornes convexes  $S$  et  $CK$  en comparant la mesure (dans  $\text{vect } \mathcal{K}$ ) de ces deux ensembles. La valeur exacte de  $|S|$  est calculable à partir de l'expression analytique (3.29) de  $S$ . En l'absence d'une expression exacte pour  $CK$ ,  $|CK|$  peut être estimé numériquement en utilisant un algorithme général de calcul d'enveloppe convexe (BARBER et al., 1996).

La table 3.4 reporte les valeurs  $|S|/|CK|$  obtenues pour différents matériaux. On observe que le rapport  $|S|/|CK|$  est relativement petit ( $< 0.18$ ), surtout pour la transformation monoclinique-I ( $< 0.07$ ). Ceci indique que l'écart entre les deux bornes  $S$  et  $CK$  est relativement grand. On cherche dans la suite à réduire cet écart en considérant des bornes non convexes sur  $QK$ .

**Table 3.4.**

 Mesures des différentes bornes (exprimées en fonction de  $|CK|$ )

matériau	$ S / CK $	$ S_1\mathcal{K} / CK $	$ S_2\mathcal{K} / CK $	$ PK / CK $
<i>monoclinique-I</i>				
Ni-49.75Ti	0.0645	0.543	0.824	0.994
CuZr	0.0543	0.547	0.865	0.998
<i>monoclinique-II</i>				
Cu-20Zn-12Ga	0.154	0.576	0.787	0.922
Cu-39.3Zn	0.131	0.539	0.761	0.924
Cu-15Zn-17Al	0.176	0.592	0.756	0.917
$\beta_1^I$ Cu-14Al-4Ni	0.0915	0.470	0.706	0.918

Pour construire une borne inférieure, on peut penser utiliser les ensembles  $R_r\mathcal{K}$  introduits précédemment. On se heurte en pratique à une difficulté de taille : d'après (3.44), il est nécessaire de prendre  $r \geq 5$  pour obtenir un ensemble  $R_r\mathcal{K}$  de mesure non nulle. Or, le coût de calcul augmente très rapidement avec le rang  $r$  de lamination. En présence de 12 déformations de transformation, il s'avère très difficile numériquement de dépasser  $r = 2$ . Une autre stratégie explorée

dans (Michaël PEIGNEY, 2013a) est de considérer des déformations compatibles entre  $S$  et  $R_r\mathcal{K}$ , c'est-à-dire de considérer l'ensemble  $S_r\mathcal{K}$  défini par

$$S_r\mathcal{K} = \bigcup_{\substack{\mathbf{e} \in R_r\mathcal{K}, \mathbf{e}' \in \mathcal{S} \\ \det(\mathbf{e} - \mathbf{e}') = 0}} [\mathbf{e}, \mathbf{e}']. \quad (3.61)$$

Cet ensemble contient  $S$  et a donc une mesure non nulle pour tout  $r \geq 1$ . On peut ainsi se contenter d'une faible valeur de  $r$  (typiquement  $r \leq 2$ ), permettant de mener à bien les calculs. Les résultats obtenus, reportés dans la table 3.4 confirment la pertinence de cette approche : la borne  $S_r\mathcal{K}$  améliore la borne  $S$  de façon significative. L'écart entre les meilleures bornes disponibles (à savoir  $S_2\mathcal{K}$  et  $PK$ ) se retrouve largement réduit par rapport au résultat obtenu avec les bornes convexes ( $S$  et  $CK$ ). Il en découle une meilleure estimation de l'ensemble  $Q\mathcal{K}$ . En comparant les résultats obtenus pour la transformation monoclinique-I avec ceux obtenus pour la transformation monoclinique-II, on constate que

$$\frac{|PK^{II}|}{|CK^{II}|} < \frac{|PK^I|}{|CK^I|},$$

Ce résultat s'interprète comme suit : pour la transformation monoclinique-I, l'ensemble  $PK^I$  est plus proche de l'enveloppe convexe  $\mathcal{K}$  que pour la transformation monoclinique-II. La borne  $S_2\mathcal{K}$  vérifie la même propriété. Ceci suggère que  $Q\mathcal{K}$  est plus proche de  $CK$  pour la transformation monoclinique-I.

Bien que le calcul de mesures soit instructif et permet d'avoir une vision globale, il ne caractérise pas complètement les différentes bornes introduites. Afin d'illustrer les bornes obtenues de façon différente, considérons comme en 3.5 les déformations  $\bar{\mathbf{e}}(\omega, \tau)$  de la forme 3.31. On pose

$$\begin{aligned} C\tau(\omega) &= \sup\{\tau | \bar{\mathbf{e}}(\omega, \tau) \in CK\} \quad , \quad Q\tau(\omega) = \sup\{\tau | \bar{\mathbf{e}}(\omega, \tau) \in Q\mathcal{K}\}, \\ P\tau(\omega) &= \sup\{\tau | \bar{\mathbf{e}}(\omega, \tau) \in PK\} \quad , \quad S_r\tau(\omega) = \sup\{\tau | \bar{\mathbf{e}}(\omega, \tau) \in S_r\mathcal{K}\}. \end{aligned}$$

La chaîne d'inclusions  $S \subset S_1\mathcal{K} \subset S_2\mathcal{K} \subset Q\mathcal{K} \subset PK \subset CK$  implique que

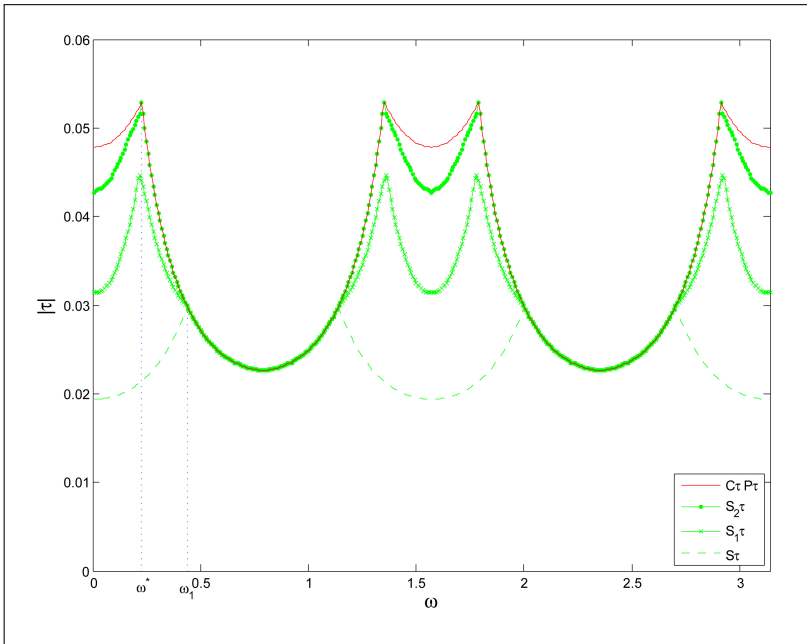
$$S\tau \leq S_1\tau \leq S_2\tau \leq Q\tau \leq P\tau \leq C\tau. \quad (3.62)$$

Ces fonctions sont représentées figure 3.13 pour Ni-49.75Ti (monoclinique-I martensite) et figure 3.14 pour l'alliage  $\beta'_1\text{Cu-14Al-4Ni}$  (monoclinique-II martensite).

Pour l'alliage  $\beta'_1\text{Cu-14Al-4Ni}$ , les fonctions  $P\tau$  et  $S_2\tau$  sont très proches, donnant ainsi une bonne estimation de la fonction  $Q\tau$  qui caractérise l'enveloppe quasi-convexe. L'écart entre  $P\tau$  and  $S_2\tau$  n'est pas aussi faible pour l'alliage Ni-49.75Ti, mais on observe néanmoins une amélioration significative par rapport à la borne convexe  $S\tau$ .

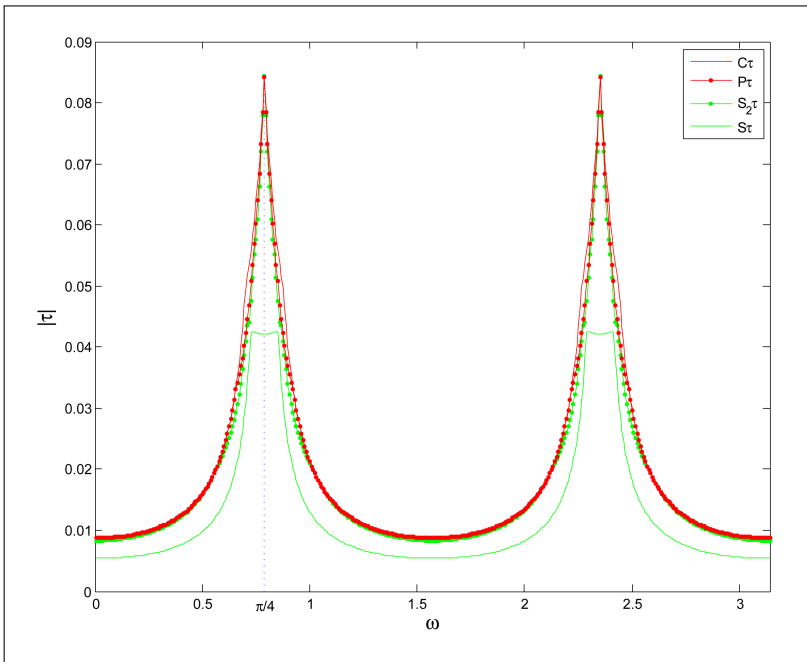
**Figure 3.13.**

Bornes sur les valeurs  $(\omega, \tau)$  telles que  $\bar{e}(\omega, \tau) \in Q\mathcal{K}^I$



**Figure 3.14.**

Bornes sur les valeurs  $(\omega, \tau)$  telles que  $\bar{e}(\omega, \tau) \in Q\mathcal{K}^{II}$



### 3.10. Conclusion

Comme il a été montré dans ce chapitre, l'étude du comportement des alliages à mémoire de forme repose sur un problème d'optimisation de formes. Celui-ci consiste à chercher les distributions spatiales des différentes phases qui minimisent l'énergie totale. Bien que la solution exacte reste en général hors de portée, on a présenté différentes bornes qui permettent, au moins dans certains cas, d'obtenir une estimation relativement précise des microstructures (voir par exemple (SHU et al., 1998) pour des comparaisons avec des résultats expérimentaux). Ces bornes sont susceptibles d'améliorations, et des progrès continuent d'être réalisés dans cette voie (CHENCHIAH et al., 2013).

Les microstructures auxquelles on s'est intéressé dans ce chapitre correspondent à une contrainte macroscopique nulle, mais on peut de la même façon étudier les microstructures associées à une contrainte macroscopique arbitraire (Michaël PEIGNEY, 2009 ; GOVINDJEE, HACKL et al., 2007 ; GOVINDJEE et MIEHE, 2001). Ceci permet notamment de suivre l'évolution des microstructures au cours d'un cycle de chargement mécanique. Notons également que l'approche présentée peut être étendue aux polycristaux ainsi qu'aux grandes déformations (Michaël PEIGNEY, 2008 ; Michaël PEIGNEY, 2013b ; Michaël PEIGNEY, 2016) mais l'analyse devient beaucoup plus complexe. Enfin, au delà du cadre des alliages à mémoire de forme, on notera qu'une problématique similaire se retrouve dans d'autres matériaux (ferroélectriques par exemple). Nous pouvons en particulier citer la découverte récente de nouveaux alliages qui offrent des perspectives nouvelles en matière de récupération d'énergie thermique (SONG et al., 2013).

### Bibliographie

- N.H. Andersen, Bente Lebech et Henning Friis Poulsen.** « The structural phase-diagram and oxygen equilibrium partial-pressure of  $\text{YBa}_2\text{Cu}_3\text{O}_6 + x$  studied by neutron powder diffraction and gas volumetry ». In : *Physica C : Superconductivity and its Application* 172 (1990), pages 31-42.
- C. Bradford Barber, David P. Dobkin, David P. Dobkin et Hannu Huhdanpaa.** « The Quickhull Algorithm for Convex Hulls ». In : *ACM Trans. Math. Softw.* 22.4 (déc. 1996), pages 469-483. ISSN : 0098-3500. DOI : 10.1145/235815.235821.
- Kaushik Bhattacharya.** « Comparison of the geometrically nonlinear and linear theories of martensitic transformations ». In : *Continuum Mechanics and Thermodynamics* 5 (1993), pages 205-242.
- Kaushik Bhattacharya.** *Microstructure and Martensite : Why it Forms and how it Gives Rise to the Shape-memory Effect*. Tome 2. Oxford University Press, 2003.
- Kaushik Bhattacharya.** « Wedge-like microstructure in martensites ». In : *Acta Metallurgica et Materialia* 39 (1991), pages 2431-2444.

- Kaushik Bhattacharya, Nikan B. Firoozye, Richard D. James et Robert V. Kohn.** « Restrictions on microstructure ». In : *Proceedings of the Royal Society of Edinburgh : Section A Mathematics* 124.05 (1994), pages 843-878.
- Kaushik Bhattacharya et Robert V. Kohn.** « Energy minimization and the recoverable strains in polycrystalline shape memory alloys ». In : *Arch. Rational Mech. Anal.* 139 (1997), pages 99-180.
- Isaac Vikram Chenchiah et Anja Schlömerkemper.** « Non-laminate microstructures in monoclinic-I martensite ». In : *Archive for Rational Mechanics and Analysis* 207.1 (2013), pages 39-74.
- Bernard Dacorogna.** *Direct Methods in the Calculus of Variations.* Springer, 2008.
- Bernard Dacorogna.** « Quasiconvexity and relaxation of non convex variational problems. » In : *J. Funct. Anal.* 46 (1982), pages 102-118.
- Sanjay Govindjee, Klaus Hackl et Richard Heinen.** « An upper bound to the free energy of mixing by twin-compatible lamination for n-variant martensitic phase transformations ». In : *Continuum Mechanics and Thermodynamics* 18.7-8 (2007), pages 443-453.
- Sanjay Govindjee et Christian Miehe.** « A multi-variant martensitic phase transformation model : formulation and numerical implementation ». In : *Comput. Mech. Appli. Mech. Eng.* 191 (2001), pages 215-238.
- Philippe Hannequart, Michaël Peigney, Jean-François Caron, Olivier Baverel et Emmanuel Viglino.** « The Potential of Shape Memory Alloys in Deployable Systems—A Design and Experimental Approach ». In : *Humanizing Digital Reality.* Springer Singapore, sept. 2017, pages 237-246. DOI : 10.1007/978-981-10-6611-5\_21.
- Robert V. Kohn.** « Relaxation of a double-well energy ». In : *Continuum Mechanics and Thermodynamics* 3 (1991), pages 193-236.
- Konstantin A. Lurie et Andrej V. Cherkaev.** « Exact estimates of conductivity of composites formed by two isotropically conducting media taken in prescribed proportion ». In : *Proceedings of the Royal Society of Edinburgh : Section A Mathematics* 99 (1984), pages 71-87.
- Graeme W. Milton.** *The theory of composites.* Cambridge University Press, 2004.
- François Murat et Luc Tartar.** « Calcul des variations et homogénéisation ». In : *Les méthodes de l'homogénéisation : théorie et applications en physique.* 1985, pages 319-369.
- Michaël Peigney.** « A non-convex lower bound on the effective free energy of polycrystalline shape memory alloys ». In : *Journal of the Mechanics and Physics of Solids* 57 (2009), pages 970-986.
- Michaël Peigney.** « Improved bounds on the energy-minimizing strains in martensitic polycrystals ». In : *Continuum Mechanics and Thermodynamics* 28 (2016), pages 923-946.

- Michaël Peigney.** « On the energy-minimizing strains in martensitic microstructures Part 2 : Geometrically linear theory ». In : *Journal of the Mechanics and Physics of Solids* (2013).
- Michaël Peigney.** « On the energy-minimizing strains in martensitic microstructures-Part 1 : Geometrically nonlinear theory ». In : *Journal of the Mechanics and Physics of Solids* (2013).
- Michaël Peigney.** « Recoverable strains in composite shape-memory alloys ». In : *Journal of the Mechanics and Physics of Solids* 56 (2008), pages 360-375.
- Michaël Peigney.** « Stress-Free Strains in Martensitic Microstructures ». In : *Materials Science Forum*. Tome 738. Trans Tech Publ. 2013, pages 10-14.
- Michaël Peigney et Jean-Philippe Seguin.** « An incremental variational approach to coupled thermo-mechanical problems in anelastic solids. Application to shape-memory alloys ». In : *International Journal of Solids and Structures* 50.24 (2013), pages 4043-4054.
- Michaël Peigney, Jean-Philippe Seguin et Eveline Hervé-Luanco.** « Numerical simulation of shape memory alloys structures using interior-point methods ». In : *International Journal of Solids and Structures* 48 (2011), pages 2791-2799.
- Jean Salençon.** *Mécanique des Milieux Continus*. Éditions de l'École Polytechnique, 2007.
- Y.C. Shu et Kaushik Bhattacharya.** « The influence of texture on the shape-memory effects in polycrystals ». In : *Acta Mater.* 15 (1998), pages 5457-5473.
- Valery Smyshlyaev et J.R. Willis.** « On the relaxation of a three-well energy ». In : *Proc. R. Soc. Lond. A* 455 (1998), pages 779-814.
- Yintao Song, Xian Chen, Vivekanand Dabade, Thomas W. Shield et Richard D. James.** « Enhanced reversibility and unusual microstructure of a phase-transforming material ». In : *Nature* 502.7469 (2013), pages 85-88.



## Chapitre 4

# Optimisation globale pour les contours actifs

Pierre CHARBONNIER<sup>1</sup>, Jean-Philippe TAREL<sup>2</sup>

*Résumé – Les contours actifs sont des courbes déformables que l'on vient positionner dans les images pour y capturer des structures d'intérêt : on parle de segmentation d'images. La plupart du temps, cet ajustement est formulé comme l'optimisation d'une fonctionnelle d'énergie, caractérisée par la présence de nombreux minima locaux, correspondant à des solutions peu pertinentes.*

*Dans ce chapitre, nous passons en revue les principaux modèles de contours actifs existants, puis nous décrivons des solutions récemment développées, assurant la détermination de solutions globalement optimales. Il s'agit, d'une part, d'algorithmes de calcul de chemins optimaux et, d'autre part, de techniques de relaxation convexe. Les premiers, adaptés à la recherche de courbes optimales entre deux points, procèdent par propagation d'une distance géodésique et rétro-parcours. Les secondes, applicables à certaines formes de contours actifs orientés région, se positionnent dans un espace convexe, en cherchant une approximation de la fonction caractéristique des régions, et optimisent une fonctionnelle elle aussi convexe.*

### 4.1. Introduction

L'une des tâches récurrentes en analyse d'images est la *segmentation*, qui consiste à partitionner les images en régions correspondant à un fond et à un

---

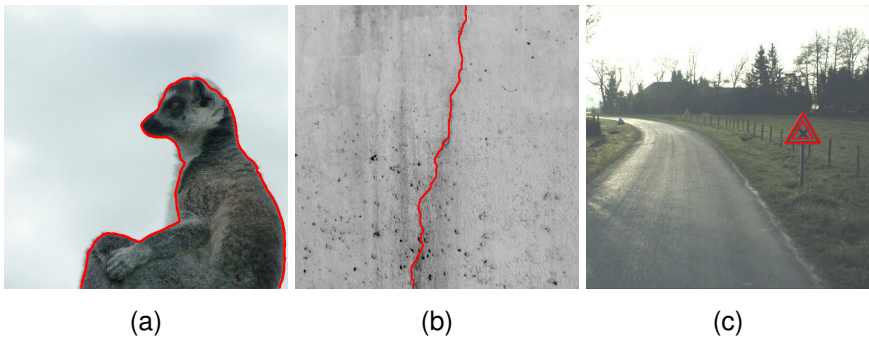
1. CEREMA/ENDSUM

2. IFSTTAR/COSYS

ou plusieurs objets d'intérêt. Bien qu'elle paraisse simple pour un opérateur humain, cette action n'est pas facile à automatiser, à cause de la variabilité d'aspect, parfois très complexe, des objets et du fait des perturbations, souvent importantes, des observations. Une réponse classique à ces difficultés est d'utiliser des modèles des objets ou de leurs contours. Pour cela, on utilise souvent des figures géométriques, des courbes ou des surfaces, dont la position, la forme, voire la topologie sont gouvernées par un nombre plus ou moins important de paramètres : on parle de *modèles déformables* (figure. 4.1). Certains d'entre eux sont spécifiques à des classes d'objets particulières, par exemple : des polygones pour modéliser des objets manufacturés. D'autres, moins contraints, sont adaptés à une grande variété d'applications.

**Figure 4.1.**

Exemples de modèles déformables. Une courbe permet de segmenter la silhouette d'un animal (a) ou suivre une fissure dans une image (b). Une combinaison de polygones est adaptée à la segmentation d'un objet manufacturé (c)



C'est le cas d'une des familles les plus connues de modèles déformables : les *contours actifs* ou *snakes* (KASS et al., 1988). Il s'agit de courbes (en analyse d'images 2D) ou de surface (en 3D), dont l'évolution est décrite par une équation aux dérivées partielles, souvent associée à l'optimisation d'une fonction ou *critère*. Cette dernière quantifie l'adéquation de la forme en évolution à l'objet recherché en termes de données image. De plus, elle pénalise les configurations *a priori* les moins adaptées à la résolution du problème. La segmentation peut alors être vue comme la recherche d'une forme optimale, au sens de ce critère.

Ce paradigme peut être décliné de nombreuses façons, ce qui a fait le succès des contours actifs, dont un grand nombre de variantes ont été proposées au cours des 25 dernières années. Cependant, il est apparu très tôt que la mise en œuvre algorithmique de l'équation d'évolution ne permettait qu'une optimisation locale de l'énergie associée à la courbe.

Il en résulte que celle-ci a un fort risque de rester « coincée » dans une configuration insatisfaisante, proche de son initialisation, surtout en présence de perturbations ou pour des images complexes. Cette sensibilité au bruit et à l'initialisation, le manque d'attractivité des modèles image, mais aussi l'impossibilité de gérer des changements de topologie de la courbe en cours

d'évolution ont rapidement été identifiés comme les inconvénients majeurs des contours actifs. Ce constat a amené de nombreux chercheurs à travailler à la définition d'approches plus performantes et plus optimales, avec des niveaux plus ou moins importants d'interaction avec l'utilisateur.

Dans la première partie de ce chapitre (§ 4.2) nous présentons de manière synthétique le modèle initial des contours actifs, ainsi que deux de ses principales évolutions : les contours actifs géodésiques et les contours actifs orientés région. Les premiers sont, comme leur nom l'indique, fondés sur l'optimisation d'une distance géodésique associée à la courbe en évolution. Les seconds recherchent une partition optimale de l'image en régions statistiquement homogènes.

Les deux dernières parties sont consacrées aux algorithmes permettant d'atteindre l'optimum global dans chacun des deux cas : nous verrons comment exploiter le principe de moindre action pour trouver la courbe géodésique optimale reliant deux points désignés par l'utilisateur (§ 4.3) et comment des techniques de relaxation convexe permettent d'atteindre le minimum global pour une famille particulière de contours actifs région (§ 4.4).

## 4.2. Les contours actifs

Dans cette partie, nous proposons d'abord quelques rappels sur le modèle initial des contours actifs, ou *snakes*, proposé par Kass *et al.* (KASS *et al.*, 1988) en 1988. Très générique, relativement simple à implanter, mais souffrant aussi d'un certain nombre de limitations, ce modèle a engendré une abondante littérature.

Nous présentons ensuite deux évolutions majeures de ce modèle, les contours actifs géodésiques et les contours actifs *région*, et renvoyons le lecteur à (CHARBONNIER, 2012) pour une synthèse bibliographique plus fournie.

### 4.2.1. Les contours actifs *classiques*

Le modèle des contours actifs est, dans le cas de la segmentation en 2D, une courbe paramétrique  $\Gamma : [a, b] \subset \mathbb{R} \rightarrow \mathbb{R}^2$ , ouverte ou fermée, à laquelle est associée une énergie à minimiser pour obtenir la segmentation :

$$E(\Gamma) = \alpha \int_a^b |\Gamma'(q)|^2 dq + \beta \int_a^b |\Gamma''(q)|^2 dq + \lambda \int_a^b g(|\nabla I(\Gamma(q))|) dq. \quad (4.1)$$

Les deux premiers termes mesurent l'élasticité et la rigidité de la courbe et leur minimisation tend donc à la régulariser. Le troisième terme mesure l'adéquation de la courbe à l'image. La fonction de *potentiel*,  $g$ , étant choisie décroissante, ce terme est minimum lorsque la courbe se situe sur les bords de l'objet à segmenter (là où le gradient de l'image,  $\nabla I$ , est le plus fort). Les coefficients  $\alpha$ ,  $\beta$  et  $\lambda$  sont, la plupart du temps, réglés par l'utilisateur.

Notons que la formulation initiale contenait des termes supplémentaires, permettant l'interaction avec un opérateur.

La plupart du temps, l'optimisation de la fonctionnelle est effectuée par une technique de descente de gradient. Cela conduit à une équation aux dérivées partielles, appelée *équation d'évolution* du contour actif :

$$\frac{\partial \Gamma(q)}{\partial t} = \underbrace{\alpha \Gamma''(q) - \beta \Gamma^{(4)}(q)}_{F_{INT}(\Gamma)} - \underbrace{\lambda \nabla g(|\nabla I(\Gamma(q))|)}_{F_{IMA}(\Gamma, I)}, \quad (4.2)$$

où  $\Gamma''$  et  $\Gamma^{(4)}$  représentent les dérivées spatiale seconde et quatrième, respectivement, de  $\Gamma$  et  $t$  est un temps artificiel. En discrétisant l'équation (4.2) en temps, on obtient pour chaque point de la courbe :

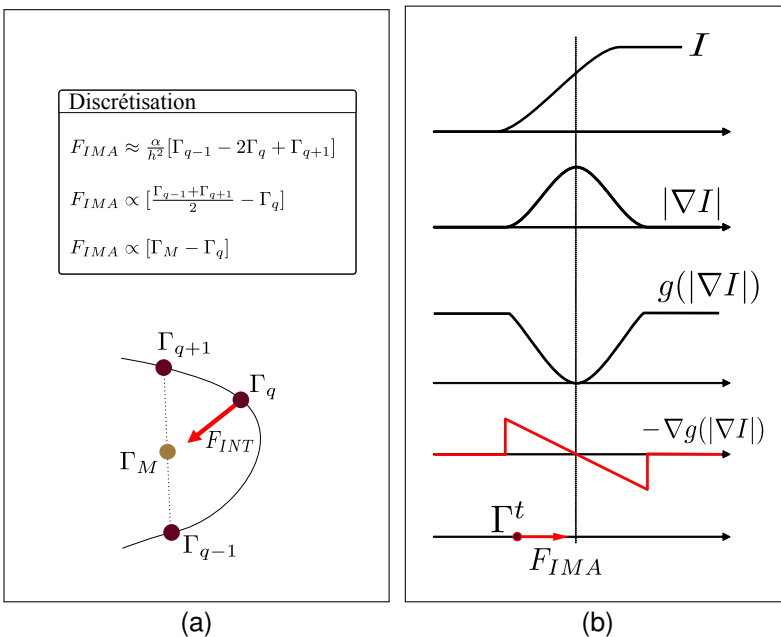
$$\Gamma^{t+1}(q) = \Gamma^t(q) + \delta t (F_{INT}(q) + F_{IMA}(q)). \quad (4.3)$$

L'évolution spatiale de chaque point de la courbe au cours du temps peut donc être vue comme la résultante de deux forces : une force interne, régularisante, ( $F_{INT}$ ) et une force image ( $F_{IMA}$ ).

Leur effet est illustré, dans un cas simplifié, sur la figure 4.2.

**Figure 4.2.**

(a) Illustration de la force interne  $F_{INT}$  dans le cas purement élastique, i.e.  $\beta = 0$  : chaque point de la courbe est attiré vers le point milieu de ses voisins (noté  $\Gamma_M$ ). (b) Illustration de la force image, selon un profil 1D perpendiculaire à un contour dans l'image : chaque point de la courbe est attiré vers le fond du *puits* de potentiel  $g$



Malgré les nécessaires précautions à prendre sur le choix des pas de discrétisation (spatial et temporel), le modèle originel est simple à mettre en œuvre. Par contre, il souffre d'un certain nombre de limitations, voire de défauts.

Ainsi, le terme élastique confère au contour actif un sens d'évolution préférentiel en contraction. La « force ballon » (force d'intensité constante, colinéaire à la normale au contour (TERZOPOULOS et al., 1988 ; L. D. COHEN, 1991)) permet de combattre l'élasticité naturelle de la courbe pour, par exemple, la faire évoluer en expansion.

De plus, comme on le comprend sur la figure 4.2 (a), le potentiel  $g$  est constant loin des contours et le terme image est donc inopérant. Des versions plus attractives, fondées sur un lissage de la force image par diffusion (Gradient Vector Flow, ou GVF, (XU et PRINCE, 1998)), ou bien sur l'exploitation de cartes de distance à des éléments de contours extraits préalablement (L. D. COHEN, 1991) ont été proposées.

Par ailleurs, le bruit présent dans les images provoque des puits de potentiels parasites, qui peuvent bloquer le contour actif loin de la solution recherchée. Pour améliorer la convergence des contours actifs fermés, des schémas algorithmiques alternatifs, multi-résolution (LEROY et al., 1996), incrémentiels (BERGER, 1991), ou encore fondés sur l'utilisation de deux snakes couplés (GUNN et al., 1997 ; VELASCO et al., 2001), ont été proposés.

Le schéma algorithmique de base a également été adapté pour gérer la fusion de courbes entre elles ainsi que leur scission (MCINERNEY et al., 1995 ; DURIKOVIC et al., 1995 ; DELINGETTE et al., 2000 ; PRECIOSO et al., 2002). Les mécanismes additionnels développés pour gérer ces changements de topologie offrent une souplesse accrue par rapport à l'initialisation des courbes. En particulier, la technique des *seed-snakes* (MCINERNEY et al., 1995), consistant à paver l'espace image de contours actifs capables de fusionner entre eux ou de disparaître s'ils ne rencontrent aucun objet d'intérêt, est aujourd'hui une stratégie d'initialisation très répandue.

Enfin, bien que des schémas de discrétisation le plus courant reste celui des différences finies, des alternatives sont parfois employées (éléments finis (PETLAND et al., 1991 ; L. D. COHEN et I. COHEN, 1993), polygones (FUA et al., 1990 ; CHESNAUD et al., 1999), *splines* (MENET et al., 1990 ; BRIGGER et al., 2000), descripteurs de Fourier (STAIB et al., 1992 ; LEROY et al., 1996 ; DUFRENOIS, 2000)). Dans tous les cas, le résultat obtenu demeure sensible à la paramétrisation de la courbe.

## 4.2.2. Les contours actifs géométriques et géodésiques

### 4.2.2.1. Équations d'évolution

Une approche alternative, issue de travaux de recherche sur la théorie de l'évolution de courbes, est apparue au début des années 1990. Il s'agit d'étudier l'évolution dans le temps d'une courbe fermée  $\Gamma$ , selon une équation aux dérivées

partielles (EDP) du type :

$$\frac{\partial \Gamma(s)}{\partial t} = F(s)\mathcal{N}, \quad (4.4)$$

où  $\mathcal{N}$  représente la normale unitaire (entrante) à la courbe.  $F$  est une fonction vitesse, fondée sur des quantités géométriques, indépendantes de la paramétrisation de la courbe.

Ainsi, lorsque l'amplitude de la force est une constante,  $\nu$ , on obtient une évolution selon la force ballon. Notons que celle-ci correspond, selon le signe de  $\nu$  à la minimisation ou à la maximisation de l'aire de la région définie par la courbe. Lorsque la force est d'amplitude proportionnelle à la courbure  $\kappa$ , la courbe évolue comme un élastique. Cette EDP correspond à la minimisation de la longueur de la courbe (on l'appelle *Euclidean shortening flow*). Ces deux forces fournissent donc naturellement le « moteur » d'un contour actif. Pour stopper l'évolution de la courbe lorsqu'elle atteint les bords des objets, il suffit de multiplier la vitesse par une fonction décroissante de la norme du gradient de l'intensité. On obtient donc :

$$\frac{\partial \Gamma}{\partial t} = g(|\nabla I|)(\nu + \kappa)\mathcal{N}. \quad (4.5)$$

Cette équation correspond aux contours actifs *géométriques*, proposés indépendamment dans (CASELLES, CATTÉ et al., 1993) et dans (MALLADI et al., 1995).

La fonction  $g$  ne s'annulant jamais complètement, on peut ajouter à cette équation un terme d'arrêt supplémentaire, attirant le contour vers le minimum du potentiel  $g$  associé aux contours des objets. On obtient ainsi l'EDP des contours actifs *géodésiques* (CASELLES, KIMMEL et al., 1997 ; KICHENASSAMY et al., 1996) :

$$\frac{\partial \Gamma}{\partial t} = g(|\nabla I|)(\nu + \kappa)\mathcal{N} - \langle \nabla g, \mathcal{N} \rangle \mathcal{N}. \quad (4.6)$$

Leur dénomination provient du fait que (4.6) correspond (pour  $\nu = 0$ ) à la minimisation d'une longueur géodésique de la courbe, la métrique utilisée dépendant de l'image *via* la fonction  $g$  :

$$L_g(\Gamma) = \int_0^1 g(|\nabla I(\Gamma(q))|)|\Gamma'(q)|dq = \int_0^{L(\Gamma)} g(|\nabla I(\Gamma(s))|)ds, \quad (4.7)$$

où  $L(\Gamma)$  est la longueur géométrique de la courbe et  $s$ , l'abscisse curviligne. On montre (CASELLES, KIMMEL et al., 1997 ; AUBERT et BLANC-FÉRAUD, 1998), que ce flot correspond également à la minimisation d'un critère de contour actif, sans contrainte de rigidité (i.e. dans le cas où  $\beta = 0$ ). De fait, une comparaison des équations d'évolution renforce cette impression de proximité entre les deux approches, comme remarqué dans (CHARBONNIER et CUISENAIRE, 1996) et (XU, PHAM et al., 2000 ; XU, ANTHONY et al., 2000). En posant  $\alpha = 1$ ,  $\beta = 0$  et en se plaçant en paramétrisation intrinsèque, ce qui entraîne  $\Gamma''(s) = \kappa(s)\mathcal{N}$ , l'équation

d'évolution du contour actif classique (4.2) s'écrit :

$$\frac{\partial \Gamma}{\partial t} = \kappa \mathcal{N} - \lambda \nabla g(\Gamma). \quad (4.8)$$

On constate, en comparant (4.6) et (4.8), que la différence tient en la multiplication de la force interne par la fonction  $g$ , ce qui tend à la relaxer à proximité des contours des objets, en la projection de la force image sur la normale, ce qui est gage de stabilité dans une implantation paramétrique (CHARBONNIER et CUISENAIRE, 1996) et en l'introduction de la force ballon, ce qui favorise la convergence.

Les contours actifs géodésiques apparaissent ainsi comme une version de contours actifs plus performante, à la fois par rapport au modèle classique et au modèle géométrique. C'est, certainement, l'une des explications du succès de cette approche. Celui-ci doit cependant beaucoup à l'algorithmique associée, la technique des *Level Sets*, qui permet une gestion transparente des changements de topologie et une indépendance vis-à-vis des problèmes de paramétrisation.

#### 4.2.2.2. Algorithme des Level Sets

L'algorithme d'évolution de courbe par lignes de niveaux ou *level-sets* est né, à la fin des années 1980, de travaux de recherche portant sur la simulation numérique en mécanique des fluides, et plus particulièrement, sur l'évolution d'interfaces (OSHER et SETHIAN, 1988).

Son principe se fonde sur une représentation intrinsèque, eulérienne, de la courbe en évolution. Celle-ci est considérée comme la ligne de niveau 0 d'une fonction *hôte*,  $\psi$ , fonction scalaire de la variable d'espace et suffisamment régulière. Classiquement, on choisit pour  $\psi$  la fonction distance signée : la région intérieure à  $\Gamma$ , notée  $\Omega_{int}$  correspond, par exemple, à des niveaux négatifs tandis que la région extérieure,  $\Omega_{ext}$ , reçoit des valeurs positives. Notons que dans ce cas,  $|\nabla \psi| = 1$  et que les quantités géométriques intervenant dans l'équation d'évolution (4.4) peuvent s'exprimer à partir de  $\psi$ . Par exemple, la courbure s'écrit :  $\kappa = \frac{\nabla \psi}{|\nabla \psi|}$ .

Par ailleurs, en dérivant la relation  $\psi(\Gamma) = 0$ , on montre facilement que si la courbe  $\Gamma$  évolue selon (4.4), alors :

$$\frac{\partial \psi}{\partial t} = F |\nabla \psi|. \quad (4.9)$$

Ainsi, l'algorithme d'évolution de courbe consiste à construire la fonction  $\psi$  à partir de la courbe initiale  $\Gamma(s, t = 0)$ , à la faire évoluer selon (4.9) jusqu'à convergence et à en extraire finalement la courbe de niveau 0.

Même si un certain nombre de précautions doivent être prises lors de sa mise en œuvre (SETHIAN, 1999 ; OSHER et FEDKIW, 2003), notamment pour éviter tout problème numérique, l'algorithme des *Level Sets* offre plusieurs intérêts, qui expliquent son très fort succès. D'une part, la représentation eulérienne

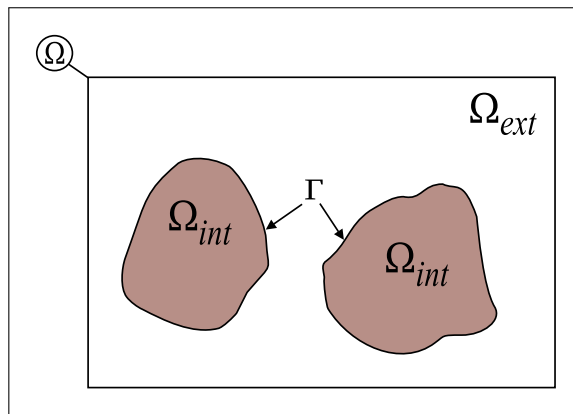
utilisée permet de s'affranchir des questions liées à la paramétrisation de la courbe, inhérentes aux représentations explicites, lagrangiennes. D'autre part, la méthode se généralise facilement au cas 3D. Enfin, et surtout, la topologie de la courbe de niveau 0 évolue librement au cours du temps, sans qu'il soit besoin de gérer directement les scissions ni les fusions.

#### 4.2.3. Les contours actifs *région*

L'un des défauts que présentent les modèles décrits précédemment est que la force image qui leur est associée ne dépend que d'informations évaluées le long de la courbe  $\Gamma$ . On parle d'approche *frontière* ou *contour*. Par ailleurs, ces informations sont le plus souvent reliées au gradient de l'image, dont le calcul numérique est sensible au bruit. Un concept alternatif, celui des contours actifs orientés région, est apparu vers le milieu des années 1990 (L. D. COHEN, BARDINET et al., 1993; RONFARD, 1994; ZHU et al., 1996), se développant surtout dans les années 2000, à la suite de (CHAN et VESE, 2001). L'idée est de formuler le problème de segmentation de l'image comme un problème de partitionnement. Dans le cas à deux régions (ou *phases*), la région intérieure à la courbe fermée  $\Gamma$  correspond à l'objet recherché, tandis que la région extérieure représente le fond, comme l'illustre la figure 4.3.

**Figure 4.3.**

Partitionnement du domaine image,  $\Omega$  en une région de fond,  $\Omega_{ext}$ , et une région correspondant à l'objet d'intérêt,  $\Omega_{int}$ . Le contour actif,  $\Gamma$ , marque la frontière entre les deux régions. D'après (FOULONNEAU, 2004)



Les régions, potentiellement disjointes, sont supposées homogènes et facilement discernables entre elles en termes de statistiques. Elles sont décrites soit par une fonction densité de probabilité, soit par des moments statistiques (moyenne, variance) ou des paramètres déduits comme l'entropie. Ces caractéristiques peuvent être supposées connues, issues d'une phase d'apprentissage, ou bien évaluées au fur et à mesure de l'évolution de la partition. À partir de ces *descripteurs* de régions (et, éventuellement, de termes *contour* (PARAGIOS et al., 1999)), on définit une fonctionnelle, dont la minimisation conduit à l'équation d'évolution du contour actif.



#### 4.2.3.1. Le modèle de Chan et Vese

Il existe de nombreuses façons de définir des fonctionnelles d'énergie pour les contours actifs région. La formulation la plus souvent citée a été proposée par Chan et Vese (CHAN et VESE, 2001) pour le cas à 2 régions, puis étendue au cas multi-phases dans (VESE et al., 2002).

Dans sa forme la plus générale, elle s'écrit :

$$E_{CV}(u_i, \Gamma) = \sum_{i=1}^{N_{reg}} \int_{\Omega_i} (u_i - I)^2 d\mathbf{x} + \nu |\Gamma|, \quad (4.10)$$

où  $\Omega_i$  note les  $N_{reg}$  régions délimitées par l'ensemble des courbes,  $\Gamma$ , dont la longueur totale est notée  $|\Gamma|$ . Le premier terme tend à imposer aux régions d'être aussi homogènes que possible en termes d'intensité, autour d'une valeur caractéristique  $u_i$ . Le second impose des frontières aussi courtes que possible et promeut donc la régularité des courbes. Notons que, si l'on fixe  $\Gamma$ , l'optimum de la fonctionnelle est atteint lorsque, pour tout  $i$ ,  $u_i$  correspond à  $\mu_i$ , la moyenne empirique de  $I$  dans la région  $\Omega_i$ . Par ailleurs, si on fixe les constantes  $u_i$ , l'évolution de la courbe permettant de minimiser  $E_{CV}$  obéit (dans le cas à  $N_{reg} = 2$  classes) à l'équation :

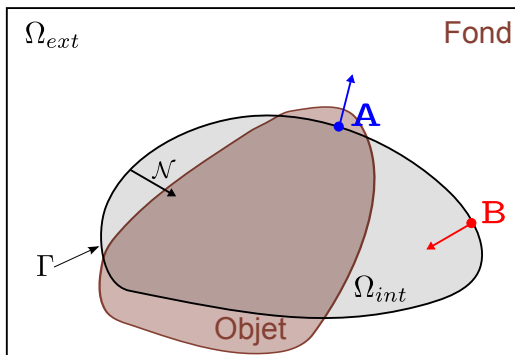
$$\frac{\partial \Gamma}{\partial t} = ((I - \mu_{int})^2 - (I - \mu_{ext})^2) \mathcal{N} + \nu \kappa \mathcal{N}, \quad (4.11)$$

où  $\mu_{int}$  et  $\mu_{ext}$  sont la moyenne de la région intérieure à la courbe et la moyenne du fond, respectivement.

La figure 4.4 illustre le fonctionnement du premier terme de (4.11), qui correspond à la force image s'appliquant en chaque point de la courbe en évolution.

#### Figure 4.4.

Interprétation de la force image  $((I - \mu_{int})^2 - (I - \mu_{ext})^2) \mathcal{N}$  en deux points du contour actif, dans le cas où  $\mathcal{N}$ , la normale à la courbe  $\Gamma$  est dirigée vers l'intérieur. L'intensité au point A est proche de  $\mu_{int}$  et la force (en bleu) est donc contraire à la normale. Au point B, l'intensité est proche de  $\mu_{ext}$  et la courbe a tendance à se contracter (flèche rouge)



Notons que si l'on considère que les deux régions ont une densité de probabilité gaussienne, de même variance et de moyennes respectives  $\mu_{int}$  et  $\mu_{ext}$ , l'équation d'évolution (4.11) est équivalente à :

$$\frac{\partial \Gamma}{\partial t} = (\log(p_{ext}(I)) - \log(p_{int}(I)))\mathcal{N} + \nu\kappa\mathcal{N}, \quad (4.12)$$

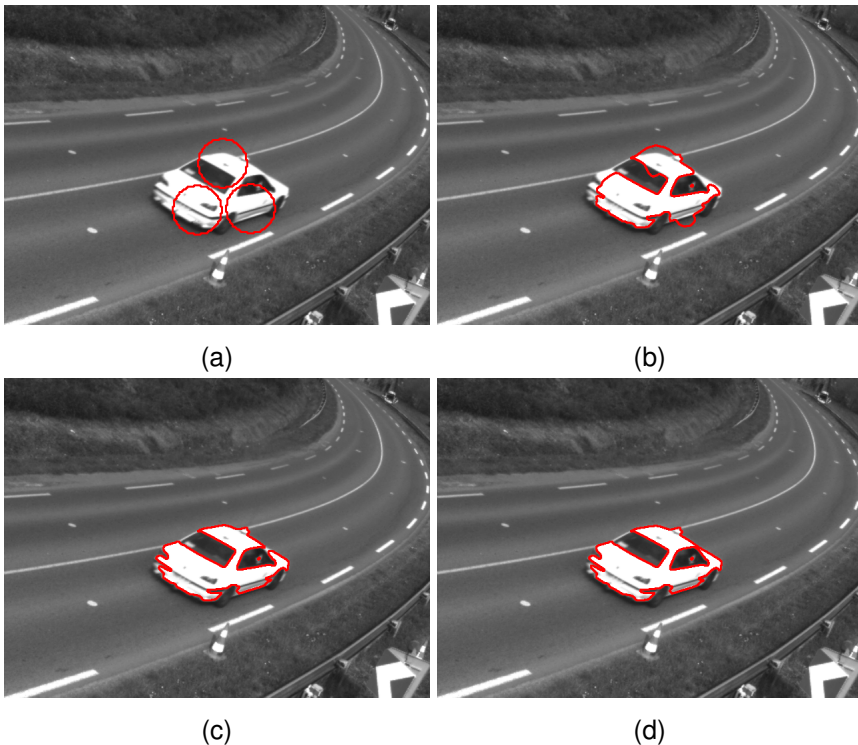
On retrouve ainsi, dans le modèle de Chan et Vese, un cas particulier de la notion de *compétition de régions* proposée par Ronfard (RONFARD, 1994) et formalisée dans (ZHU et al., 1996) : l'évolution de la courbe tend à englober un point dans la région dont il est statistiquement le plus proche.

L'algorithme d'optimisation proposé dans (CHAN et VESE, 2001 ; VESE et al., 2002) effectue de façon alternée le calcul des moyennes et l'évolution de la courbe selon (4.11), jusqu'à ce que la courbe n'évolue plus.

La figure 4.5 montre un exemple d'application de cette méthode sur une image en niveaux de gris.

**Figure 4.5.**

Exemple de segmentation par contour actif orienté région (CHARBONNIER, 2009) : modèle de Chan et Vese. Image originale et initialisation (a), résultat après 10 (b), 20 (c) et 31 itérations, résultat final (d). Image IFSTTAR Nantes. Le contour est épaissi pour une meilleure visualisation



#### 4.2.3.2. Autres modèles

Comme celle de Chan et Vese, beaucoup de fonctionnelles de contours actifs région favorisent l'homogénéité des régions formant la partition de l'image. L'approche alternative en reconnaissance des formes consiste à rechercher des régions aussi *dissemblables* entre elles que possible.

C'est, par exemple, ce que proposent Yezzi *et al.* dans (YEZZI JR et al., 1999), où la différence quadratique des moyennes (ou des variances) des régions est maximisée. Nous renvoyons le lecteur intéressé à (CREMERS, ROUSSON et al., 2007 ; CHARBONNIER, 2009) pour un aperçu plus complet des modèles de contours actifs région.

Notons que les paramètres statistiques des régions, qui interviennent dans la plupart des fonctionnelles dépendent de la position des frontières. Il y a lieu de tenir compte de ces dépendances lors de la dérivation des équations d'évolution. La notion de dérivée de domaine ou dérivée eulérienne, issue des travaux de Zolésio (SOKOLOWSKI et al., 1992 ; DELFOUR et al., 2001) et également employée dans l'estimation du flot optique (SCHNÖRR, 1992), offre un cadre théorique rigoureux pour effectuer ce type de calculs (AUBERT, BARLAUD et al., 2003).

#### 4.2.3.3. Bilan

Les contours actifs région ont un certain nombre de bonnes propriétés vis-à-vis de la tâche de segmentation. En premier lieu, ils prennent en compte l'information image d'une manière plus globale que les approches frontière, ce qui les rend moins sensibles à l'initialisation<sup>3</sup>. Par ailleurs, les contours actifs région n'impliquent pas forcément de dérivation de l'image, ce qui est gage de stabilité numérique. Enfin, ils offrent la possibilité de segmenter des régions sans frontière apparente (CHAN et VESE, 2001 ; J. KIM et al., 2005).

Au passif de ces méthodes, on notera que le cas des régions multiples (cas multi-phases) est beaucoup plus difficile à gérer que le cas de la segmentation binaire. De plus, même si la technique des *seed-snakes* permet d'améliorer les choses, la mise en œuvre de ces modèles repose sur une technique de descente de gradient. Les solutions obtenues correspondent donc généralement à des optima locaux des fonctionnelles, ce qui n'est pas totalement satisfaisant.

### 4.3. Optimisation globale par recherche de chemin minimal

Les approches que nous allons présenter dans ce paragraphes permettent de déterminer la courbe réalisant le minimum global d'une énergie de type contour actif géodésique. Comme nous l'avons fait remarquer à la section 4.2.2, de telles énergies peuvent également être vues comme des longueurs, dans une

3. Cette affirmation est toutefois à nuancer pour certains types d'images, comme on en rencontre en imagerie médicale, sujettes à de fortes variations spatiales d'intensité, ou pour des objets de nature hétérogène. Des approches région plus locales (LI et al., 2008 ; LANKTON et al., 2008 ; BROX et al., 2009) ont été développées par la suite pour gérer ces situations.

métrique dépendant de l'image. L'optimisation s'apparente donc à une recherche de chemin de longueur minimale. La solution globalement optimale peut être atteinte dans deux cas : celui des courbes ouvertes définies entre deux points,  $P_0$  et  $P_1$ , désignés par l'utilisateur et celui des courbes fermées englobant un point  $P$ . Outre le modèle énergétique employé, ces deux approches ont donc en commun la nécessité d'une interaction avec l'utilisateur.

### 4.3.1. Cas des courbes ouvertes

Dans (L. D. COHEN et KIMMEL, 1997), Cohen et Kimmel proposent de rechercher, dans l'image à segmenter, la courbe minimisant une énergie de type contour actif géodésique, très proche de (4.7) :

$$E'_g(\Gamma) = \alpha \int_0^{L(\Gamma)} |\Gamma'(s)|^2 ds + \int_0^{L(\Gamma)} g(\Gamma(s)) ds = \int_0^{L(\Gamma)} \tilde{g}(\Gamma(s)) ds = L'_g(\Gamma) \quad (4.13)$$

où  $s$  représente l'abscisse curviligne,  $g$  est le potentiel image habituel et  $\tilde{g} = \alpha + g$ . Le potentiel est, selon l'application considérée, une fonction de l'intensité (ex. : suivi de lignes, fissures, routes, vaisseaux sanguins) ou de son gradient le long de la courbe (ex. : suivi du bord d'objets), nous le notons  $g(\Gamma)$  pour simplifier. On remarque que le passage d'énergie à longueur géodésique est, ici, direct puisque  $|\Gamma'(s)| = 1$  en paramétrisation intrinsèque (i.e. par l'abscisse curviligne). Le scalaire  $\alpha$  joue le paramètre de régularisation : on peut montrer (L. D. COHEN et KIMMEL, 1997) que la courbure maximale que peut prendre la courbe solution est inversement proportionnelle à  $\alpha$ .

#### 4.3.1.1. Stratégie d'optimisation

Le calcul de la courbe minimisant (4.13) est réalisé selon la stratégie suivante : dans un premier temps, une *surface d'action minimale*  $U_0(x)$ , fonction scalaire de la variable d'espace, est calculée. Sa valeur au point  $x$  est l'énergie associée au chemin optimal, au sens de la distance géodésique  $L'_g$ , reliant  $x$  à  $P_0$ . Dans un second temps, lorsque  $U_0(P_1)$  est déterminée, un simple rétro-parcours par descente de gradient permet d'extraire le chemin optimal recherché. Nous pouvons également extraire des chemins multiples, reliant plusieurs points d'arrivée au même point de départ. Le lecteur intéressé pourra se référer à (L. D. COHEN et KIMMEL, 1997 ; DESCHAMPS, 2001) pour plus de détails.

On montre (voir, par exemple, (FOMEL, 1997 ; L. D. COHEN et KIMMEL, 1997)) que la surface d'action minimale obéit à l'équation eikonale<sup>4</sup> :

$$\tilde{g}(x) = |\nabla U_0(x)|. \quad (4.14)$$

avec  $U_0(P_0) = 0$ . Notons que si  $\tilde{g}$  est une constante, la surface d'action minimale correspond à une simple carte de distance euclidienne au point de départ. Plusieurs approches sont envisageables pour résoudre cette équation. La plus

4. Cette équation, due à Hamilton, est à la base de l'optique géométrique, d'où sa dénomination, proposée par Burns en 1895 (« eikonal » venant du mot grec « *εικων* » signifiant « image »). Pour plus de développements sur l'analogie avec l'optique géométrique, voir la thèse de T. Deschamps (DESCHAMPS, 2001).

immédiate consiste à la discrétiser et à appliquer un schéma itératif (ROUY et al., 1992). Cette approche a toutefois une complexité en  $O(N^2)$ , où  $N$  est le nombre de points de la grille.

La méthode usuelle, plus rapide, consiste à calculer  $U_0$  par propagation à partir du point de départ,  $P_0$  pour lequel la valeur de  $U_0$  est nulle. Cette idée apparaît déjà dans (QIN et al., 1992), mais c'est en faisant le lien avec l'algorithme  $A^*$  de Dijkstra (DIJKSTRA, 1959), utilisé pour la recherche de plus court chemin dans les graphes, que Tskitsiklis (TSITSIKLIS, 1995), puis (de manière indépendante) Sethian (SETHIAN, 1996) proposent des algorithmes réellement efficaces, de complexité  $O(N \log N)$ <sup>5</sup>.

#### 4.3.1.2. Algorithme $A^*$

L'algorithme  $A^*$  utilise le principe d'optimalité de Bellman, exploité en programmation dynamique : le chemin le plus court est propagé d'un point à ses voisins, par ajout de la contribution locale  $\tilde{g}$  minimale. Une valeur de  $U_0$  peut donc être affectée à un point dès lors que l'un de ses voisins est atteint. Naturellement, le point doit rester *vivant* (i.e. sa valeur de  $U_0$  doit pouvoir être revue à la baisse) jusqu'à ce qu'il devienne lui-même le minimum des points vivants. En effet, il est alors impossible de trouver un chemin géodésique plus court pour l'atteindre. L'utilisation d'une structure de tas ordonné pour extraire à chaque instant le point de valeur  $U_0$  minimale parmi les points vivants confère à l'algorithme une complexité en  $O(N \log N)$ . Les algorithmes de délinéation d'objets tels que le *Live Wire* (FALCÃO et al., 1998) ou les *Intelligent Scissors* (MORTENSEN et al., 1995 ; MORTENSEN et al., 1999) implantés dans des logiciels de traitement d'image comme Gimp utilisent, de manière interactive, le même genre de méthode que l'algorithme  $A^*$ .

#### 4.3.1.3. Algorithme Fast Marching

Nous pouvons remarquer que l'ensemble des points vivants se comporte comme un front, de valeur  $U_0$  à peu près constante, et progressant à partir de  $P_0$  à une vitesse spatiale inversement proportionnelle au potentiel local  $\tilde{g}$ . Cette observation est à la base de l'algorithme *Fast Marching* de Sethian (SETHIAN, 1996), qui considère le calcul de  $U_0$  comme le problème de propagation d'une ligne de niveau  $\mathcal{L}$  selon l'équation :

$$\frac{\partial \mathcal{L}}{\partial t} = \frac{1}{\tilde{g}} \mathcal{N}, \quad (4.15)$$

à partir d'un cercle infinitésimal centré en  $P_0$ . Notons que  $U_0$  peut s'interpréter comme le temps d'arrivée  $t$  du front  $\mathcal{L}$  à une distance euclidienne donnée (L. D. COHEN et KIMMEL, 1997), ce qui permet un parallèle avec l'algorithme des *Level Sets*. En effet,  $U_0$  jouant le rôle de fonction hôte, son évolution selon (4.9) s'écrit :  $\frac{\partial U_0}{\partial t} = \frac{1}{\tilde{g}} |\nabla U_0| = \frac{\partial t}{\partial t} = 1$ , et l'on retrouve l'équation eikonale. L'algorithme *Fast Marching* peut donc être vu comme une formulation stationnaire du problème

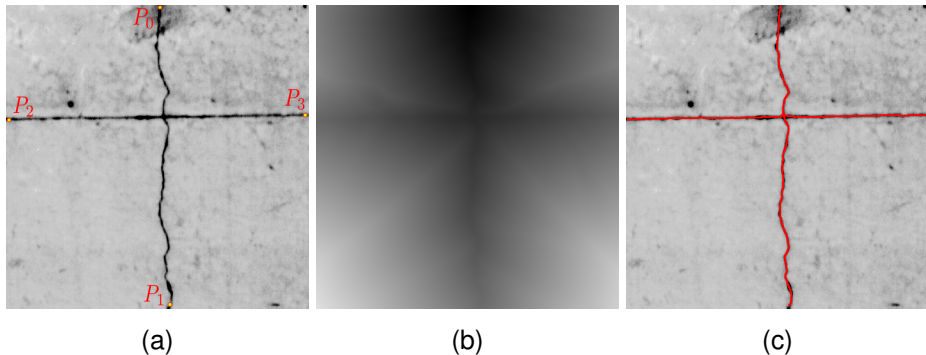
5. Voir également (HELMSEN et al., 1996).

d'évolution de courbe (SETHIAN, 1999), réservé au cas où la vitesse ne change pas de signe au cours du temps.

À la différence de l'algorithme  $A^*$ , dont il suit l'architecture, l'algorithme *Fast Marching* ne procède pas par incrémentation locale de  $U_0$ , mais par résolution de (4.14). Il est donc un peu plus coûteux, mais, en contrepartie, beaucoup plus précis, puisque la métrique sous-jacente est euclidienne et non de type *city-block*.

#### Figure 4.6.

Exemple de suivi de ligne de fissure par contour actif géodésique entre plusieurs extrémités. (a) Image originale et points désignés par l'utilisateur ; (b) surface d'action minimale  $U_0$  propagée depuis  $P_0$  ; (c) résultats de l'extraction obtenu par descente de gradient sur l'image (b) à partir des points  $P_1$ ,  $P_2$  et  $P_3$ . Le contour est épaissi pour une meilleure visualisation



La figure 4.6 montre un exemple d'application de cet algorithme dans le domaine du génie civil (suivi de ligne de fissure dans du béton)<sup>6</sup>. Dans ce cas, le potentiel image,  $g$ , est directement fonction du niveau de gris.

#### 4.3.1.4. Vitesse et précision

Simple à programmer et efficace, l'algorithme *Fast Marching* est très utilisé pour les applications d'analyse d'images. Dans certains domaines, cependant, une précision supérieure peut être intéressante. En effet, le schéma numérique le plus courant correspond à une discrétisation au premier ordre de l'opérateur de dérivation. Il s'ensuit que l'algorithme peine à générer de fortes courbures des lignes de niveau. Ainsi, dans le cadre de la propagation d'une distance euclidienne à partir d'un point, on observe que les premières courbes de niveau ne sont pas parfaitement circulaires. De plus, le coût  $\tilde{g}$  est considéré constant autour de chaque point de la grille. Cela signifie qu'entre deux points voisins, la vitesse d'évolution du front ne sera pas la même suivant le sens de propagation considéré. Hassouna *et al.* (HASSOUNA *et al.*, 2007), Danielsson *et al.* (DANIELSSON *et al.*, 2003) ou, plus récemment, Appia et Yezzi (APPIA *et al.*, 2013) ont proposé des schémas numériques plus précis et isotropes, demeurant efficaces d'un point de vue calculatoire.

6. Cet algorithme est mis en œuvre (CHARBONNIER, GUILLARD *et al.*, 1999 ; CHARBONNIER et MOLIARD, 2002 ; CHARBONNIER et MOLIARD, 2003) dans le logiciel de traitement d'image PICTURE, distribué par l'IFSTTAR jusqu'en 2016.

L'efficacité algorithmique est, en effet, un point primordial dans le contexte d'une application interactive. Des techniques traitant les points par groupe (*Group Marching* (S. KIM, 2001)), ou *via* une structure de gestion de priorités (YATZIV et al., 2006) permettent d'atteindre une complexité en  $O(N)$ , sans perte substantielle de précision. Enfin, une version abandonnant la propriété de causalité de la propagation au profit d'une parallélisation de l'algorithme a été proposée dans (JEONG et al., 2008), ce qui permet son implantation efficace sur processeur graphique.

#### 4.3.1.5. Extension à la détection à partir d'un point unique

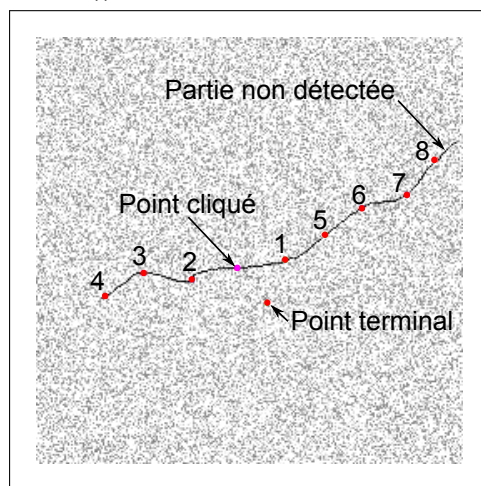
Il y a quelques années, Kaul *et al.* ont proposé une approche fondée sur la recherche de chemins minimaux, permettant de détecter des lignes à partir d'un seul point désigné par l'utilisateur (KAUL et al., 2012). L'idée, initialement développée dans (BENMANSOUR et al., 2009) vient de l'observation suivante : la propagation de la distance géodésique est plus rapide (spatialement) le long des lignes suivies que dans les régions homogènes.

Autrement dit, à partir d'un point situé le long de la ligne à détecter, le premier point atteint par le front à une distance euclidienne  $\lambda$  donnée du point de départ se situe également sur la ligne. Un rétro-parcours entre ce point et le point de départ permet la détection d'un premier linéament. La procédure se poursuit, de manière incrémentale, à partir de l'ensemble formé des deux points.

Par rapport à (BENMANSOUR et al., 2009), la méthode proposée dans (KAUL et al., 2012) résout le problème du test d'arrêt de la croissance de la courbe, en exploitant l'information de distance entre points successifs. Moyennant quelques modifications dans la gestion des tests d'arrêt, la méthode peut être étendue au cas des courbes ramifiées, ainsi qu'aux courbes fermées.

#### Figure 4.7.

Application de la méthode de Kaul *et al.* sur une image synthétique (d'après (KAUL et al., 2012))



La figure 4.7 montre un exemple d'application de cette méthode sur une image synthétique. Le point terminal est le premier point détecté hors de la courbe par le test proposé dans (KAUL et al., 2012).

Nous pouvons constater que la longueur de la courbe détectée diffère au plus de sa longueur réelle de la valeur de  $\lambda$  (égale à 30 pixels, dans cet exemple). Enfin, on remarquera que, si cette méthode apporte plusieurs améliorations par rapport à la méthode initiale de Cohen et Kimmel (L. D. COHEN et KIMMEL, 1997) (un seul point de départ, traitement automatique des courbes fermées et des ramifications), il n'est pas évident qu'elle demeure globalement optimale.

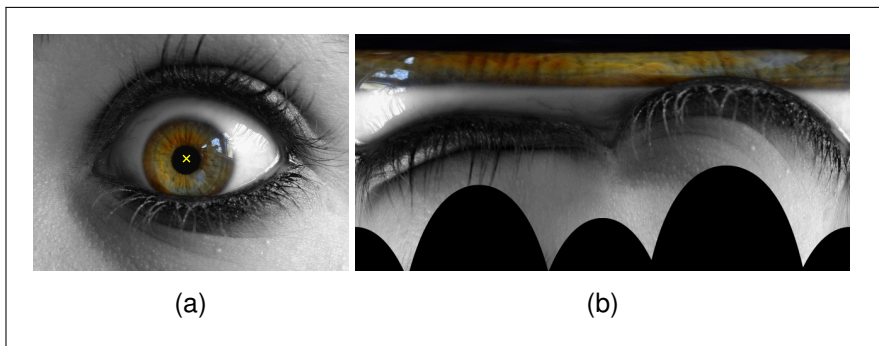
#### 4.3.2. Cas des courbes fermées

La méthode de Cohen et Kimmel (L. D. COHEN et KIMMEL, 1997) peut être étendue à la segmentation de contours fermés de plusieurs autres façons que celle proposée dans (KAUL et al., 2012). Le point  $P_0$  étant fixé, la première variante consiste à rechercher le long du contour, un point  $P_1$  tel qu'il existe deux courbes globalement optimales le reliant à  $P_0$ . Une méthode permettant de déterminer de tels points (qui sont des points-selles de  $U_0$ ) est proposée dans (L. D. COHEN et KIMMEL, 1997). Cette technique est également utilisée dans (L. D. COHEN, 2001) pour mettre en place un outil de fermeture automatique de contours par groupement perceptif (*perceptual grouping*).

Une autre façon d'étendre la méthode aux courbes fermées s'inspire de recherches initialement motivées par les applications d'imagerie panoramique. Dans cette approche, l'image est « dépliée » de manière radiale puis interpolée sur une grille rectangulaire (technique du *polar unwrapping*) à partir d'un point  $P_{int}$  désigné par l'utilisateur, voir figure 4.8. On transforme ainsi la recherche d'une courbe optimale fermée entourant le point  $P_{int}$  en un problème de recherche de chemin optimal *circulaire*<sup>7</sup> ou *Circular Shortest Path* (CSP) dans l'image dépliée.

**Figure 4.8.**

(a) Image originale; (b) image « dépliée » de manière radiale à partir du point marqué d'une croix jaune dans (a).



7. Un chemin circulaire (ici, horizontal) est  $2\pi$ -périodique : l'ordonnée de son extrémité à gauche de l'image est égale, éventuellement au pixel près, à celle de son extrémité droite.

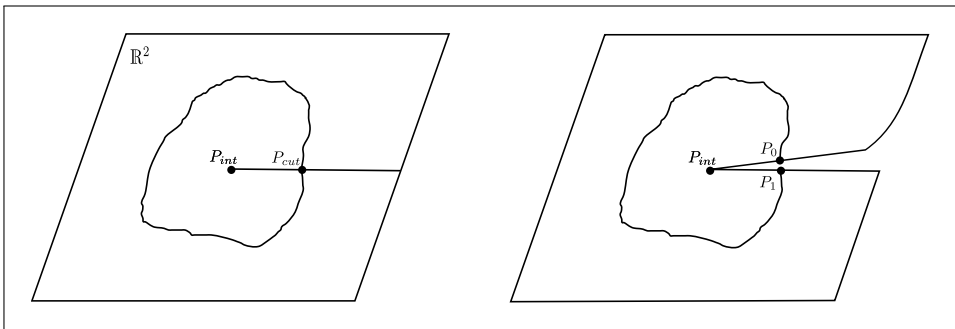


L'approche « naïve » du problème consiste à l'aborder de manière exhaustive, en recherchant l'ensemble des chemins optimaux reliant les points de la première colonne de l'image à ceux (au pixel près) de la dernière colonne. Cette approche est, évidemment, très coûteuse en temps de calcul. Dans (SUN et al., 2003), Sun et Pallottino définissent trois méthodes alternatives, plus rapides mais n'assurant pas que le chemin optimal résultant soit forcément circulaire. Des techniques de type *Branch and Bound* (APPLETON et SUN, 2003) ou recherche dichotomique (DE LA GORCE et al., 2006) ont été proposées par la suite pour résoudre le problème de façon exacte, avec de meilleures performances.

Dans (APPLETON et TALBOT, 2005), un principe légèrement différent est adopté : il consiste à couper le plan image le long d'une demi-droite issue du point  $P_{int}$ . À partir d'un point  $P_{cut}$  de cette demi-droite, des extrémités  $P_0$  et  $P_1$  peuvent alors être définies de part et d'autre de la découpe (figure. 4.9). Une recherche de chemin minimal par Fast Marching et rétro-parcours peut alors être effectuée à partir de ces points de départ et d'arrivée.

#### Figure 4.9.

Découpage du plan image dans le cas d'un objet convexe (d'après (APPLETON et TALBOT, 2005))



L'avantage de cette variante est qu'elle peut s'adapter au cas où la courbe, concave, traverse plusieurs fois le plan de coupe, moyennant une représentation plus complexe, hélicoïdale, du domaine image. Tout le problème revient alors à choisir le point  $P_{cut}$  optimal de la manière la plus efficace possible. Pour cela, Appleton et Talbot proposent une adaptation de la technique de recherche arborescente de (APPLETON et SUN, 2003).

Si la méthode d'Appleton et Talbot permet de segmenter des objets non convexes, on notera qu'elle ne gère pas directement les topologies complexes. Ainsi, pour segmenter des objets à plusieurs composantes connexes, il est nécessaire de positionner un point par composante. De plus, comme la méthode originale de Cohen et Kimmel pour les courbes ouvertes, l'énergie minimisée est une énergie de type frontière.

Ainsi, les algorithmes que nous venons de décrire ne permettent pas d'optimiser des fonctionnelles région, dont on sait pourtant qu'elles sont souvent plus performantes en segmentation.

## 4.4. Optimisation globale pour les contours actifs région

Quelle que soit la modélisation considérée, classique, géodésique ou région, la vitesse d'évolution des contours actifs n'est définie que sur la courbe en évolution. De plus, suivant son initialisation, le contour actif n'explore qu'une partie du domaine image. Ainsi, une courbe initialisée autour d'un objet comprenant un trou capture les frontières externes de l'objet mais ne peut pas segmenter le trou. Une illustration de ce problème est donnée sur la figure 4.10 (b-c), pour des images de cellules, segmentées à l'aide du modèle de Chan et Vese.

Pour remédier à ce problème d'optimum local, lié à l'impossibilité d'explorer complètement l'espace des solutions, une première tentative a été l'utilisation de la notion de dérivée topologique, issue de recherches sur l'optimisation de forme (DELFOUR et al., 2001).

Définie sur l'ensemble du domaine image, celle-ci mesure la sensibilité de l'énergie à une modification topologique des domaines étudiés, par introduction de trous. Une fois celui-ci calculé, on peut se servir du gradient topologique de la fonction d'énergie pour définir un algorithme de descente pour les contours actifs, comme dans (SHI, 2005) (en reconstruction tomographique) ou dans (HE et al., 2007) (en segmentation, selon le modèle de Chan et Vese). Cependant, l'emploi d'un algorithme de *Level sets*, où la vitesse n'est en principe définie que le long de l'interface, nécessite une adaptation (BURGER et al., 2004).

Ces premières tentatives n'ont, à notre connaissance, pas eu de suite, ces méthodes ayant rapidement été dépassées par l'essor des techniques de *convexification*.

Il s'agit de techniques variationnelles, issues de recherches sur le débruitage d'images binaires (NIKOLOVA, 2004), d'études théoriques sur la régularisation par variation totale (RUDIN et al., 1992 ; CHAN et ESEDOĞLU, 2005) et de travaux sur l'optimisation convexe (CHAMBOLLE, 2004). Ces investigations ont débouché sur des méthodes rapides, précises et assurant dans certains cas l'optimalité globale pour la segmentation à deux classes à partir de fonctionnelles orientées région (NIKOLOVA et al., 2006 ; BRESSON et al., 2007).

Nous nous focaliserons ici sur les approches variationnelles continues, mais il existe des méthodes d'un niveau équivalent dans le formalisme discret (ces travaux ont, d'ailleurs, bien souvent inspiré le développement des méthodes continues). C'est, en particulier, le cas des modèles de type *max flow - min cut*, associés aux algorithmes de découpe de graphe ou *graph cuts* (BOYKOV et al., 2001 ; APPLETON et TALBOT, 2006).

### 4.4.1. Relaxation convexe : principe général

Reprenons le problème de segmentation par contour actif région, selon la fonctionnelle de Chan et Vese (§ 4.2.3). On considère que l'image est une fonction binaire, formée d'une région intérieure  $\Omega_{int}$ , d'intensité  $\mu_{int}$ , et d'une région extérieure  $\Omega_{ext} = \Omega \setminus \Omega_{int}$ , d'intensité  $\mu_{ext}$ , séparées par une frontière  $\Gamma$ ,

de sorte que :

$$I(\mathbf{x}) = \mu_{int} \cdot \chi_{\Omega_{int}}(\mathbf{x}) + \mu_{ext} \cdot \chi_{\Omega_{ext}}(\mathbf{x}), \quad (4.16)$$

où  $\chi_{\Sigma}$  représente la fonction caractéristique, binaire, de la région  $\Sigma$ . Le problème consiste à trouver la courbe qui minimise la fonctionnelle :

$$E_{CV}(\Omega_{int}) = \int_{\Omega_{int}} (I - \mu_{int})^2 d\mathbf{x} + \int_{\Omega \setminus \Omega_{int}} (I - \mu_{ext})^2 d\mathbf{x} + \nu Per(\Omega_{int}), \quad (4.17)$$

où  $Per(\Omega_{int}) = |\Gamma|$  est la longueur de l'interface entre les régions intérieure et extérieure. Nous savons que ce problème n'a pas de solution unique car l'ensemble des régions binaires n'est pas un ensemble convexe. Toutes les techniques d'optimisation, en particulier, celle des *level-sets*, conduisent donc à des minima locaux. Il est montré que l'équation d'évolution de la courbe, éq. (4.11), se traduit en termes d'ensembles de niveaux par :

$$\frac{\partial \psi}{\partial t} = \delta(\psi) \left\{ ((I - \mu_{int})^2 - (I - \mu_{ext})^2) + \nu div \left( \frac{\nabla \psi}{|\nabla \psi|} \right) \right\}. \quad (4.18)$$

où, en général, nous remplaçons la fonction Dirac  $\delta$ , par une fonction continue,  $\delta_{\epsilon}$  de façon à régulariser le schéma numérique.

Comme cette dernière est à support étendu, des changements de signe de  $\psi$  peuvent apparaître plus ou moins loin de la courbe. Ceci permet, par exemple la capture de trous dans les objets à l'aide d'une courbe initialisée à l'extérieur de ceux-ci. Cette propriété va dans le sens d'une exploration plus complète du domaine image, nécessaire pour une optimisation globale. Nous pouvons pousser plus loin cette idée, en remarquant que, puisque la fonction  $\delta_{\epsilon}$  est positive, la supprimer ne change pas l'ensemble des points stationnaires de (4.18) et l'on peut donc appliquer, sur l'ensemble des courbes de niveau :

$$\frac{\partial \psi}{\partial t} = \left\{ ((I - \mu_{int})^2 - (I - \mu_{ext})^2) + \nu div \left( \frac{\nabla \psi}{|\nabla \psi|} \right) \right\}. \quad (4.19)$$

Or, cette équation d'évolution correspond à une minimisation par descente de gradient de :

$$E_{CEN}(\psi) = \int_{\Omega} |\nabla \psi| + \lambda \int_{\Omega} \{ (\mu_{int} - I(\mathbf{x}))^2 - (\mu_{ext} - I(\mathbf{x}))^2 \} \psi(\mathbf{x}) d\mathbf{x}, \quad (4.20)$$

qui est une fonctionnelle convexe par rapport à  $\psi$ . Cependant, on peut noter que cette énergie est homogène de degré 1 (elle est linéaire en  $\psi$  :  $E_{CEN}(\alpha \cdot \psi) = \alpha \cdot E_{CEN}(\psi)$ ) et n'a donc pas de minimiseur. Une rapide étude de l'équation d'évolution (4.19) associée permet de le comprendre<sup>8</sup>. Si on discrétise

8. Pour le justifier, on peut aussi remarquer qu'il existe une infinité de représentations d'une courbe par un ensemble de niveaux (GOLDSTEIN et al., 2009a)

cette équation en temps, l'évolution de la fonction hôte  $\psi$  est donnée par :

$$\psi^{t+1} = \psi^t + \delta t \left\{ ((I - \mu_{int})^2 - (I - \mu_{ext})^2) + \nu \operatorname{div} \left( \frac{\nabla \psi^t}{|\nabla \psi^t|} \right) \right\}. \quad (4.21)$$

Considérons le cas idéal où la courbe est parfaitement positionnée sur les frontières de l'objet. On a, par convention,  $\psi(\mathbf{x}) < 0$  pour un point  $\mathbf{x}$  appartenant à  $\Omega_{int}$ . Si l'image est effectivement binaire, on a également  $I(\mathbf{x}) \simeq \mu_{int}$ . Autrement dit, si l'on omet pour simplifier le terme de courbure, l'incrément appliqué à  $\psi$  est négatif. De même, pour un point  $\mathbf{x}$  situé à l'extérieur de l'objet, on a à la fois  $\psi(\mathbf{x}) > 0$  et  $I(\mathbf{x}) \simeq \mu_{ext}$  et l'incrément est positif. En d'autres termes, l'application répétée de (4.21) aura pour effet de rendre toujours plus négatives les valeurs de  $\psi$  à l'intérieur de l'objet et plus positives les valeurs de  $\psi$  à l'extérieur de celui-ci. Donc,  $\psi$  va tendre vers une carte binaire à valeurs  $\pm\infty$  selon l'appartenance à la région  $\Omega_{int}$ . Cet écueil peut être évité de manière simple, en bornant la fonction  $\psi$  : on choisit donc de remplacer  $\psi$  par une fonction  $0 \leq u(\mathbf{x}) \leq 1$  et on considère (NIKOLOVA et al., 2006) la minimisation de la fonctionnelle suivante :

$$E_{CEN}(u) = \int_{\Omega} |\nabla u| + \lambda \int_{\Omega} \{(\mu_{int} - I(\mathbf{x}))^2 - (\mu_{ext} - I(\mathbf{x}))^2\} u(\mathbf{x}) d\mathbf{x}, \quad (4.22)$$

sous la contrainte  $0 \leq u(\mathbf{x}) \leq 1$ .

Nous pouvons remarquer que non seulement la fonctionnelle (4.22) est convexe, mais que  $u$  appartient également à un espace convexe, plus précisément l'ensemble des fonctions à variations bornées sur  $[0, 1]$ . L'optimisation de  $E_{CEN}$  peut donc être conduite de manière globale. On remarque aussi que la convexité a été obtenue en remplaçant la recherche de la région  $\Omega_{int}$ , que l'on peut définir par sa fonction caractéristique binaire, par celle d'une fonction continue sur  $[0, 1]$ , donc en relâchant la contrainte de binarité. C'est ce qui donne le nom de *relaxation convexe*<sup>9</sup> à cette approche.

On montre alors, et c'est sans doute le résultat le plus important de (NIKOLOVA et al., 2006), que pour  $\mu_{int}, \mu_{ext} \in \mathbb{R}$  fixés, si  $u^*$  (tel que  $u^*(\mathbf{x}) \in [0, 1]$ ) minimise  $E_{CEN}$ , alors  $\Omega_{int}(\mu) = \{\mathbf{x} : u^*(\mathbf{x}) \geq \mu\}$  est un minimiseur global de l'énergie (4.17) pour presque tout  $\mu \in [0, 1]$ .

La preuve de ce théorème (NIKOLOVA et al., 2006), inspirée de travaux plus anciens (STRANG, 1982 ; STRANG, 1983), utilise notamment la formule de la co-aire, qui relie la variation totale de la fonction  $u$  à l'intégrale des longueurs de ses courbes de niveau :

$$\int_{\Omega} |\nabla u| = \int_0^1 \operatorname{Per}(\{\mathbf{x} : u(\mathbf{x}) > \mu\}) d\mu, \quad (4.23)$$

9. Cette notion est également exploitée, mais de manière différente, dans (CREMERS, SCHMIDT et al., 2008).

et l'identité (où  $\chi_{[a,b]}(\cdot)$  est la fonction caractéristique 1D de l'intervalle  $[a, b]$ ) :

$$u(\mathbf{x}) = \int_0^1 \chi_{[0, u(\mathbf{x})]}(\mu) d\mu, \quad (4.24)$$

pour obtenir :

$$E_{CEN}(u) = \int_0^1 E_{CV}(\Omega_{int}(\mu)) d\mu - C, \quad (4.25)$$

où  $C$  est une constante. Il s'ensuit que si  $u^*$  est un minimiseur de  $E_{CEN}$ , alors  $\Omega_{int}(\mu)$  doit être un minimiseur de  $E_{CV}$  pour toutes les valeurs de  $\mu$ , à un ensemble de mesure nulle près. Notons que ce théorème a été étendu pour toutes les valeurs de  $\mu$ , pour  $\mu \in [0, 1]$  dans (BERKELS, 2009).

En vertu de ce théorème de seuillage, le principe général de la méthode d'optimisation globale par relaxation convexe consiste à minimiser (4.22) sous contrainte  $u^*(\mathbf{x}) \in [0, 1]$ , puis à seuiller le résultat pour obtenir la partition souhaitée. Notons que n'importe quelle valeur de seuil peut être utilisée, d'après le théorème, ce qui est cohérent avec le fait que le schéma numérique associé a un effet binarisant, comme nous l'avons fait remarquer : en pratique les valeurs de  $u^*$  sont soit très proches de 0, soit très proches de 1 (NIKOLOVA et al., 2006).

#### 4.4.2. Algorithmes d'optimisation

Le point clef de la méthode est la minimisation sous contrainte de (4.22). Une descente de gradient est possible. C'est l'approche initialement retenue dans (NIKOLOVA et al., 2006). Le problème est tout d'abord transformé en remplaçant la contrainte  $u(\mathbf{x}) \in [0, 1]$  par l'utilisation d'un terme supplémentaire faisant intervenir une fonction « vallée », nulle sur  $[0, 1]$  et linéaire sur les bords de cet intervalle. Toutefois, le terme  $L^1$  qui apparaît dans le critère  $E_{CEN}$  nécessite une régularisation et l'utilisation d'un pas de temps faible (BRESSON et al., 2007). De plus, les méthodes régularisées ont le défaut de lisser les frontières dans l'image  $u^*$ , ce qui rend la segmentation sensible au choix du seuil.

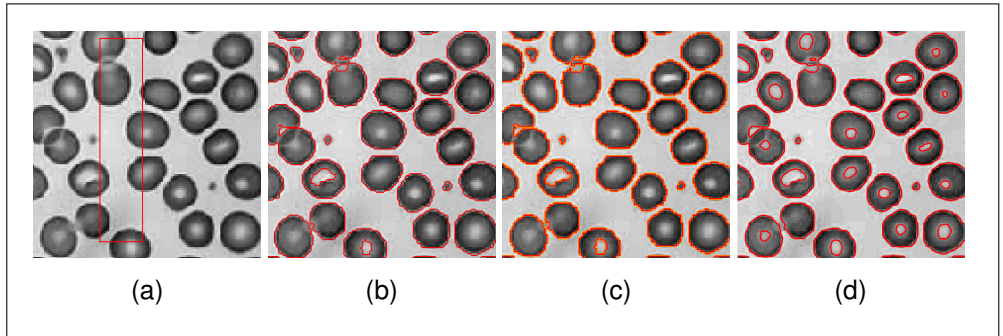
Des méthodes plus rapides et plus précises, fondées sur des formulations duales développées notamment pour les décompositions d'image en structure et texture (voir par exemple (CHAMBOLLE, 2004) ou (AUJOL et al., 2005)), peuvent être appliquées pour accélérer l'optimisation.

Plus récemment (GOLDSTEIN et al., 2009a) la méthode dite Split-Bregman a permis d'atteindre des temps de calcul comparables à ceux des *Graph Cuts*. Nous avons, à titre d'exemple, appliqué cette méthode à une image de cellule (figure 4.10). Les temps de calcul sont largement inférieurs à ceux des algorithmes *Level Sets*, même accélérés, que nous utilisons habituellement.

Par ailleurs, on constate que les zones blanches à l'intérieur de certaines cellules sont correctement segmentées, alors que les algorithmes *Level Sets* ne les capturent qu'à proximité de la courbe initiale.

**Figure 4.10.**

Initialisation pour les algorithmes Level Sets (a) et résultats de segmentation par contour actif orienté région (modèle de Chan et Vese) d'une image  $230 \times 230$  tirée de (YEZZI JR et al., 1999) (b,c,d). Algorithme levels sets (OSHER et SETHIAN, 1988), implantation C+Matlab (A. Foulonneau) : temps de calcul environ 12 secondes (b). Algorithme Level Sets rapides de Shi et Karl (SHI et KARL, 2005), implantation Matlab (G. Gaullier) : temps de calcul environ 4s (c). Algorithme de Goldstein *et al.* (GOLDSTEIN et al., 2009a), implantation C+Matlab (aimablement fournie par X.Bresson) : temps de calcul environ 0,04s (d). Moyennes des régions :  $\mu_{int} = 202$ ,  $\mu_{ext} = 87$  pour (b,c) et  $\mu_{int} = 205$ ,  $\mu_{ext} = 86$  (d).



#### 4.4.3. Extensions

La méthode de Chan *et al.* a été adaptée dans (BRESSON et al., 2007) en utilisant une norme pondérée de  $\nabla u$  en lieu et place de la norme  $L^1$ , reliant ainsi les approches de Chan et Vese et de Mumford et Shah aux contours actifs géodésiques. Par ailleurs, l'approche peut s'étendre à toute fonctionnelle de la forme :

$$E(u) = \int_{\Gamma} g(\mathbf{x}(s)) ds + \lambda \int_{\Omega_{int}} r_{int}(\mathbf{x}) d\mathbf{x} + \lambda \int_{\Omega \setminus \Omega_{int}} r_{ext}(\mathbf{x}) d\mathbf{x}. \quad (4.26)$$

Ainsi, elle est appliquée dans (HOUBOU et al., 2008) à un critère région de dissimilarité (maximisation de la J-divergence) et dans (NI et al., 2009), à un critère région d'homogénéité fondé sur l'utilisation d'histogrammes locaux. Dans (MORY et ARDON, 2007), une approche très voisine, appelée *fuzzy region competition*, est proposée. Elle consiste à remplacer dans (4.26) la région  $\Omega_{int}$  par une fonction d'appartenance *floue*  $u$ , non binaire. On obtient ainsi une expression convexe, très proche de (4.22). Dans (MORY, ARDON et THIRAN, 2007), ce principe est appliqué à des fonctionnelles fondées sur des modèles non paramétriques, par noyaux de Parzen, des densités de probabilité globales des régions. Le modèle est ensuite adapté à des mesures plus locales.

L'extension de ce type de méthodes au cas multi-phases est un sujet difficile et toujours d'actualité. La version proposée dans (POCK et al., 2009) semble s'approcher de l'optimum global. L'introduction de contraintes de forme dans

le formalisme de la relaxation convexe fait également l'objet de travaux de recherche (WERLBERGER et al., 2009).

#### 4.4.4. Limitations

Il est très important de noter que les approches décrites précédemment ne sont globales que vis-à-vis de l'optimisation de forme. En réalité, le modèle de Chan et Vese comporte trois inconnues, puisque les moyennes (ou, dans la formulation générale, les caractéristiques des régions) ne sont *a priori* pas connues. L'approche usuelle, dans la lignée de ce qui se pratiquait avec l'algorithme des *level sets*, est un schéma d'optimisation alternant ajustement de forme et estimation des paramètres.

Malheureusement, nous n'avons plus aucune garantie qu'une solution globale puisse être atteinte par ce type de schéma! Plusieurs exemples où deux solutions différentes sont obtenus à partir de deux initialisations différentes est montré dans (BROWN et al., 2011). Curieusement, ce problème a très peu été exploré ces dernières années. Nous pouvons citer les travaux de Strandmark *et al.* (STRANDMARK et al., 2009) incluant un algorithme de type *branch and bound* et, plus récemment, ceux de Brown *et al.* (BROWN et al., 2011). Ces derniers exploitent une technique de relaxation convexe proposée dans (GOLDSTEIN et al., 2009b) pour estimer simultanément  $u$  et les paramètres des régions. L'optimalité globale n'est pas assurée, mais un critère permettant de vérifier de combien la solution calculée s'en éloigne est proposé.

## 4.5. Conclusion

Les contours actifs sont une catégorie très populaire de modèles déformables et ont généré une littérature abondante et riche au cours des 25 dernières années. Ceci s'explique, d'une part, par l'aspect très générique de leur principe, qui les rend adaptés à un grand nombre d'applications. D'autre part, la difficulté du problème abordé et les relatives faiblesses des premiers modèles proposés ont également motivé un grand nombre de recherches. Enfin, ces approches sont nourries de travaux menés dans disciplines variées comme la physique, les méthodes numériques, la géométrie, les statistiques, ou encore la théorie de l'optimisation.

Il est, évidemment, impossible de dresser un état de l'art exhaustif dans un domaine aussi foisonnant et, dans la première partie de ce chapitre, nous nous sommes limités à un tour d'horizon des approches les plus connues. Nous pouvons toutefois remarquer que, parmi les principales problématiques rencontrées et les variantes ou améliorations proposées, la question de l'optimalité des solutions estimées est un sujet central, depuis l'origine de ces méthodes.

En effet, selon un canevas usuel en reconnaissance des formes, les contours actifs reposent sur la résolution d'un problème d'optimisation de paramètres ou, plus généralement, de formes. Ce type de problème est connu pour sa difficulté,

liée notamment à la présence d'optima locaux. Celle-ci peut être due à la non-convexité de l'espace dans lequel sont recherchées les solutions (par exemple, l'espace des fonctions caractéristiques binaires) ou bien à la non-convexité de la fonctionnelle minimisée, voire même, des deux aspects combinés. Dans ce chapitre, nous avons focalisé notre attention sur les principales approches ayant abouti à des outils aujourd'hui opérationnels. Plus précisément, nous avons présenté les techniques basées sur la recherche de chemins minimaux, qui offrent des outils optimaux pour les contours actifs orientés frontière, avec un certain besoin d'interactivité avec l'opérateur.

Nous avons ensuite expliqué le principe des techniques de relaxation convexe, qui permettent d'optimiser certaines fonctionnelles orientées région. Ces méthodes, souvent mises en concurrence avec les techniques de *graph cuts*, qui représentent leur pendant dans un formalisme discret, ont connu un essor important ces dernières années. Leur succès est dû à une certaine généralité du formalisme proposé, à la présence de résultats théoriques concernant leur optimalité. Il tient, surtout, à la disponibilité d'algorithmes d'optimisation efficaces fondés notamment sur la notion de dualité. À l'heure actuelle, les approches fondées sur l'optimisation convexe sont en pleine expansion dans le domaine de la segmentation.

Si l'on fait le bilan de ces méthodes, on se rend compte que l'on est progressivement passé au cours des années 1990 de représentations explicites, sous forme de courbes, à des représentations implicites des formes, comme les ensemble de niveaux ou, depuis le milieu des années 2000, les fonctions caractéristiques. D'un problème d'optimisation de courbe, on se ramène aujourd'hui à l'optimisation d'une fonction bidimensionnelle et l'on retrouve des fonctionnelles très proches de critères définis dans le domaine de la restauration d'images bruitées.

En fait, segmentation et restauration sont intimement liées, et les travaux de Mumford et Shah (MUMFORD et al., 1989) continuent d'inspirer ces deux domaines de recherches. En revenir à un problème de segmentation pourrait passer pour un retour en arrière : dans le cas discret, par exemple, on retrouve des fonctionnelles bien connues dans le domaine des champs Markoviens, très en vogue à partir de la fin des années 1980. Ceci étant, les outils théoriques et pratiques avec lesquels on revisite aujourd'hui ces méthodes ouvrent de nouvelles possibilités et permettent de définir des outils de segmentation innovants et efficaces.

Nous l'avons vu, les modèles et les algorithmes ont beaucoup progressé au cours de ces 20 ans et on dispose d'un panel d'exemples de fonctionnelles adaptées à des situations variées, et de plusieurs techniques optimales de segmentation. Les besoins en recherche méthodologique sont cependant loin d'être taris. En effet, tous les problèmes ne sont pas résolus, loin de là. Par exemple, nous avons noté que dans le cas des modèles région binaires, l'optimalité n'était assurée que par rapport à la forme, et non par rapport à l'ensemble « forme plus paramètres », ce qui est rarement souligné dans les publications. Très peu de travaux ont abordé



cette question. Un autre exemple est celui de la segmentation globale simultanée de plusieurs objets, et l'incorporation de contraintes de haut niveau portant sur la forme ou la topologie des objets. Dans ces domaines, bien des progrès restent à faire pour développer des modèles plus sophistiqués, en meilleure adéquation avec l'extrême variété des situations rencontrées dans les applications pratiques.

## Bibliographie

**Vikram Appia** et **Yezzi Anthony**. « Symmetric Fast Marching schemes for better numerical isotropy ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.9 (sept. 2013), pages 2298-2304.

**Ben Appleton** et **Changming Sun**. « Circular shortest paths by Branch and Bound ». In : *Pattern Recognition* 36.11 (2003), pages 2513-2520.

**Ben Appleton** et **Hugues Talbot**. « Globally Minimal Surfaces by Continuous Maximal Flows ». In : *IEEE Transactions On Pattern Analysis And Machine Intelligence* 28.1 (2006), pages 106-118.

**Ben Appleton** et **Hugues Talbot**. « Globally Optimal Geodesic Active Contours ». In : *Journal of Mathematical Imaging and Vision* 23.1 (2005), pages 67-86.

**Gilles Aubert**, **Michel Barlaud**, **Olivier Faugeras** et **Stéphanie Jehan-Besson**. « Image segmentation using active contours : calculus of variations or shape gradients ? » In : *SIAM, Journal on Applied Mathematics* 63.6 (sept. 2003). RR-INRIA 4483, juin 2002, pages 2128-2154.

**Gilles Aubert** et **Laure Blanc-Féraud**. *An elementary proof of the equivalence between 2D and 3D classical snakes and geodesic active contours*. Rapport de Recherche RR-3340. Sophia Antipolis : INRIA, jan. 1998.

**Jean-François Aujol** et **Antonin Chambolle**. « Dual Norms and Image Decomposition Models ». In : *International Journal of Computer Vision* 63.1 (2005), pages 85-104.

**Fethallah Benmansour** et **Laurent D. Cohen**. « Fast Object Segmentation by Growing Minimal Paths from a Single Point on 2D or 3D Images ». In : *Journal of Mathematical Imaging and Vision* 33.2 (fév. 2009), pages 209-221. ISSN : 0924-9907. DOI : 10.1007/s10851-008-0131-0.

**Marie-Odile Berger**. « Les contours actifs : modélisation, comportement, convergence ». Thèse de doctorat. Nancy : INPL, 1991.

**Benjamin Berkels**. « An unconstrained multiphase thresholding approach for image segmentation ». In : *Proceedings of the Second International Conference on Scale Space Methods and Variational Methods in Computer Vision (SSVM 2009)*. Tome 5567. Lecture Notes in Computer Science. 2009, pages 26-37.

**Yuri Boykov**, **Olga Veksler** et **Ramin Zabih**. « Fast Approximate Energy Minimization via Graph Cuts ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.11 (nov. 2001), pages 1222-1239. ISSN : 0162-8828. DOI : 10.1109/34.969114.

- Xavier Bresson, Selim Esedoğlu, Pierre Vanderghenst, Jean-Philippe Thiran et Stanley Osher.** « Fast Global Minimization of the Active Contour/Snake Model ». In : *Journal of Mathematical Imaging and Vision* 28.2 (2007), pages 151-167.
- P. Brigger, J. Hoeg et M. Unser.** « B-spline Snakes : A Flexible Tool for Parametric Contour Detection ». In : *IEEE Transactions on Image Processing* 9.9 (sept. 2000), pages 1484-1496. ISSN : 1057-7149. DOI : 10.1109/83.862624.
- Ethan S. Brown, Tony F. Chan et Xavier Bresson.** « Completely convex formulation of the Chan-Vese image segmentation model ». In : *International Journal of Computer Vision* 98.1 (2011), pages 103-121.
- Thomas Brox et Daniel Cremers.** « On local region models and a statistical interpretation of the piecewise smooth Mumford-Shah functional ». In : *International Journal of Computer Vision* 84.2 (août 2009), pages 184-193.
- Martin Burger, Benjamin Hackl et Wolfgang Ring.** « Incorporating Topological Derivatives into Level Set Methods ». In : *Journal of Computational Physics* 194.1 (fév. 2004), pages 344-362. ISSN : 0021-9991. DOI : 10.1016/j.jcp.2003.09.033.
- Vincent Caselles, Francine Catté, Tomeu Coll et Françoise Dibos.** « A geometric model for active contours in image processing ». In : *Numerische Mathematik* 66 (oct. 1993), pages 1-31. DOI : 10.1007/BF01385685.
- Vincent Caselles, Ron Kimmel et Guillermo Sapiro.** « Geodesic active contours ». In : *International Journal of Computer Vision* 22.1 (fév. 1997), pages 61-79.
- Antonin Chambolle.** « An Algorithm for Total Variation Minimization and Applications ». In : *Journal of Mathematical Imaging and Vision* 20.1-2 (jan. 2004), pages 89-97. ISSN : 0924-9907. DOI : 10.1023/B:JMIV.0000011325.36760.1e.
- Tony F. Chan et Selim Esedoğlu.** « Aspects of total variation regularized  $L^1$  function approximation ». In : *SIAM Journal on Applied Mathematics* 65.5 (2005), pages 1817-1837. DOI : 10.1137/040604297.
- Tony F. Chan et Luminita A. Vese.** « Active Contours Without Edges ». In : *Trans. Img. Proc.* 10.2 (fév. 2001), pages 266-277. ISSN : 1057-7149. DOI : 10.1109/83.902291.
- Pierre Charbonnier.** « Modèles de forme et d'apparence en traitement d'images ». Habilitation à Diriger des Recherches. Université de Strasbourg, sept. 2009.
- Pierre Charbonnier.** *Optimisation globale pour les contours actifs*. Rapport bibliographique OR 11W101. IFSTTAR, mar. 2012.
- Pierre Charbonnier et Olivier Cuisenaire.** *Une étude des contours actifs : modèles classique, géométrique et géodésique*. Rapport technique 163. Laboratoire TELE, Place du Levant 2, 1348 LOUVAIN-LA-NEUVE (Belgique) : Université Catholique de Louvain, juil. 1996.

- Pierre Charbonnier, Yves Guillard et Xavier Clady.** *Détection automatique de fissures sur ouvrages d'art par analyse d'images.* Rapport technique FAER 283138. LCPC, avr. 1999.
- Pierre Charbonnier et Jean-Marc Moliard.** « Calculs de chemins minimaux, suivi de fissures et autres applications ». In : *Journées des Sciences de l'Ingénieur du réseau des laboratoires des Ponts et Chaussées.* Sous la direction de **LCPC.** Actes des journées scientifiques du LCPC. Dourdan, France, déc. 2003, pages 201-206.
- Pierre Charbonnier et Jean-Marc Moliard.** *Implantation dans PICTURE d'un algorithme de segmentation d'images par minimisation de distance géodésique pour le suivi de bords ou de fissures.* Rapport technique FAER 282410, OR 11A025. LCPC, fév. 2002.
- Christophe Chesnaud, Philippe Réfrégier et Vlady Boulet.** « Statistical Region Snake-Based Segmentation Adapted to Different Physical Noise Models ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* 21.11 (nov. 1999), pages 1145-1157. ISSN : 0162-8828. DOI : 10.1109/34.809108.
- Laurent D. Cohen.** « Multiple Contour Finding and Perceptual Grouping using Minimal Paths ». In : *Journal of Mathematical Imaging and Vision* 14.3 (mai 2001), pages 225-236.
- Laurent D. Cohen.** « On active contour models and balloons ». In : *Computer Vision, Graphics, and Image Processing. Image Understanding* 53.2 (1991), pages 211-218.
- Laurent D. Cohen, Éric Bardinet et Nicholas Ayache.** *Surface reconstruction using active contour models.* Rapport de Recherche RR-1824. Sophia Antipolis : INRIA, fév. 1993.
- Laurent D. Cohen et I. Cohen.** « Finite-Element Methods for Active Contour Models and Balloons for 2-D and 3-D Images ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.11 (nov. 1993), pages 1131-1147.
- Laurent D. Cohen et Ron Kimmel.** « Global Minimum for Active Contour Models : A Minimal Path Approach ». In : *International Journal of Computer Vision* 24.1 (1997), pages 57-78.
- Daniel Cremers, Mikael Rousson et Rachid Deriche.** « A Review of Statistical Approaches to Level Set Segmentation : Integrating Color, Texture, Motion and Shape ». In : *Int. J. Comput. Vision* 72.2 (avr. 2007), pages 195-215. ISSN : 0920-5691. DOI : 10.1007/s11263-006-8711-1.
- Daniel Cremers, Frank R. Schmidt et Frank Barthel.** « Shape Priors in Variational Image Segmentation : Convexity, Lipschitz Continuity and Globally Optimal Solutions ». In : *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Anchorage, Alaska, juin 2008. DOI : 10.1109/CVPR.2008.4587446.

- Per-Erik Danielsson** et **Qingfen Lin**. « A Modified Fast Marching Method ». In : *Proceedings of the 13th Scandinavian Conference on Image Analysis. SCIA'03*. Halmstad, Sweden : Springer-Verlag, 2003, pages 1154-1161. ISBN : 3-540-40601-8.
- Martin de La Gorce** et **Nikos Paragios**. « Fast Dichotomic Multiple Search Algorithm for Shortest Circular Path ». In : *Proc. 18th International Conference on Pattern Recognition (ICPR'06)*. Tome 2. 2006, pages 403-406.
- Michel C. Delfour** et **Jean-Paul Zolésio**. *Shape and geometries : analysis, differential calculus and optimization*. Advances in design and control. SIAM, 2001.
- Hervé Delingette** et **Montagnat Johan**. *Topology and Shape Constraints on Parametric Active Contours*. Rapport de Recherche RR-3880. Sophia Antipolis : INRIA, 2000.
- Thomas Deschamps**. « Extraction de Courbes et Surfaces par Methodes de Chemins Minimaux et Ensembles de Niveaux. Applications en Imagerie Medicale 3D ». Olivier FAUGERAS Philippe CINQUIN Ron KIMMEL Françoise DIBOS Wiro NIESSEN Nicolas ROUGON. Theses. Université Paris Dauphine - Paris IX, déc. 2001.
- Edsger W. Dijkstra**. « A Note on Two Problems in Connexion with Graphs ». In : *Numer. Math.* 1.1 (déc. 1959), pages 269-271. ISSN : 0029-599X. DOI : 10.1007/BF01386390.
- Franck Dufrenois**. « Étude d'un modèle déformable de Fourier pour la segmentation et le suivi d'objets 2D et 3D ». In : *Traitement du Signal* 17.2 (2000), pages 153-178.
- Roman Durikovic**, **Kazufumi Kaneda** et **Hideo Yamashita**. « Dynamic contour : a texture approach and contour operations ». In : *The Visual Computer* 11.6 (1995), pages 277-289.
- Alexandre A. Falcão**, **Jayaral K. Udupa**, **Supun Samarasekera**, **Shoba Sharma**, **Bruce Elliot Hirsch** et **Roberto de A. Lotufo**. « User-steered image segmentation paradigms : Live-wire and live-lane ». In : *Graphical Models and Image Processing* 60.4 (juil. 1998), pages 233-260.
- Sergey Fomel**. *A variational formulation of the fast marching eikonal solver*. Rapport Technique 95. Université de Standford : Stanford Exploration Project, sept. 1997.
- Alban Foulonneau**. « Une contribution à l'introduction de contraintes géométriques dans les contours actifs orientés région ». Thèse de doctorat. Strasbourg : Université Louis Pasteur - Strasbourg I, déc. 2004.
- Pascal Fua** et **Yvan G. Leclerc**. « Model Driven Edge Detection ». In : *Machine Vision and Applications* 3 (1990), pages 45-56.
- Tom Goldstein**, **Xavier Bresson** et **Stanley Osher**. *Geometric Applications of the Split Bregman Method : Segmentation and Surface Reconstruction*. Rapport de Recherche CAM09-06. UCLA, fév. 2009.

- Tom Goldstein, Xavier Bresson et Stanley Osher.** *Global minimization of Markov random fields with applications to optical flow*. Rapport de Recherche CAM09-77. UCLA, fév. 2009.
- Steve R. Gunn et Mark S. Nixon.** « A Robust Snake Implementation ; A Dual Active Contour ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* 19.1 (jan. 1997), pages 63-68. ISSN : 0162-8828. DOI : 10.1109/34.566812.
- Sabry M. Hassouna et Aly A. Farag.** « Multistencils Fast Marching Methods : A highly accurate solution to the Eikonal equation on cartesian domains ». In : *IEEE Transactions On Pattern Analysis And Machine Intelligence* 29.9 (sept. 2007), pages 1563-1574.
- Lin He et Stanley Osher.** « Solving the Chan-Vese Model by a Multiphase Level Set Algorithm Based on the Topological Derivative ». In : *Proceedings of the 1st International Conference on Scale Space and Variational Methods in Computer Vision*. SSVM'07. Ischia, Italy : Springer-Verlag, 2007, pages 777-788. ISBN : 978-3-540-72822-1.
- John Helmsen, Elbridge Puckett, Phillip Colella et Milo Dorr.** « Two new methods for simulating photolithography development in 3D ». In : 2726 (juin 1996), pages 253-261. DOI : 10.1117/12.240959.
- Nawal Houhou, Jean-Philippe Thiran et Xavier Bresson.** « Fast texture segmentation model based on the shape operator and active contour ». In : *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2008)*. Anchorage, USA, 2008, pages 1-8.
- Won-Ki Jeong et Ross Whitaker.** « A Fast Iterative Method for Eikonal Equations ». In : *SIAM Journal of Scientific Computing* 30.5 (juil. 2008), pages 2512-2534. DOI : 10.1.1.117.8943.
- Michael Kass, Andrew Witkin et Demetri Terzopoulos.** « Snakes : active contour models ». In : *International Journal of Computer Vision* 1.4 (jan. 1988), pages 321-331. DOI : 10.1007/BF00133570.
- Vivek Kaul, Anthony Yezzi et Yichang Tsai.** « Detecting curves with unknown endpoints and arbitrary topology using minimal paths ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.10 (oct. 2012), pages 1952-1965. DOI : 10.1007/BF00379537.
- Satyanad Kichenassamy, Arun Kumar, Peter Olver, Allen Tannenbaum et Anthony Jr Yezzi.** « Conformal curvature flows : From phase transitions to active vision ». In : *Archive for Rational Mechanics and Analysis* 134 (sept. 1996), pages 275-301.
- Junmo Kim, John W. Fisher, Anthony Yezzi, Mujdat Çetin et Alan S. Willsky.** « A nonparametric statistical method for image segmentation using information theory and curve evolution ». In : *IEEE Transactions on Image Processing* 14.10 (oct. 2005), pages 1486-1502. DOI : 10.1109/TIP.2005.854442.
- Seongjai Kim.** « An  $O(N)$  level set method for eikonal equation ». In : *SIAM Journal on Scientific Computing* 22.6 (2001), pages 2178-2193. DOI : 10.1137/S1064827500367130.

- Shawn Lankton et Allen Tannenbaum.** « Localizing region-based active contours ». In : *IEEE Transactions on Image Processing* 17.11 (nov. 2008), pages 1-11.
- Bertrand Leroy, Isabelle L. Herlin et Laurent D. Cohen.** « Multi-resolution algorithms for active contour models ». In : *Proc. 12th International Conference on Analysis and Optimization of Systems : Images, Wavelets and PDE's (ICAOS'96)*. Tome 219/1996. Lecture Notes in Control and Information Sciences. Paris : Springer, juin 1996, pages 58-65. DOI : 10.1007/3-540-76076-8\_117.
- Chunming Li, Chiu-Yen Kao, J. C. Gore et Zhaohua Ding.** « Minimization of Region-Scalable Fitting Energy for Image Segmentation ». In : *Trans. Img. Proc.* 17.10 (oct. 2008), pages 1940-1949. ISSN : 1057-7149. DOI : 10.1109/TIP.2008.2002304.
- R. Malladi, James A. Sethian et B. Vemuri.** « Shape modeling with front propagation : a level set approach ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.2 (fév. 1995), pages 158-175.
- Tim McInerney et Demetri Terzopoulos.** « Topologically Adaptable Snakes ». In : *Proceedings of the Fifth International Conference on Computer Vision. ICCV '95*. Washington, DC, USA : IEEE Computer Society, 1995, pages 840-. ISBN : 0-8186-7042-8.
- Sylvie Menet, Philippe Saint-Marc et Gerard Medioni.** « B-snakes : Implementation and application to stereo ». In : *Proc. 3rd International Conference on Computer Vision (ICCV'90)*. 1990, pages 720-726.
- Eric N. Mortensen et William A. Barrett.** « Intelligent Scissors for Image Composition ». In : *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '95*. New York, NY, USA : ACM, 1995, pages 191-198. ISBN : 0-89791-701-4. DOI : 10.1145/218380.218442.
- Eric N. Mortensen et William A. Barrett.** « Toboggan-Based Intelligent Scissors with a Four-Parameter Edge Model ». In : *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Tome 2. Ft. Collins, USA, juin 1999, pages 2452-2458. DOI : 10.1109/CVPR.1999.784720.
- Benoit Mory et Roberto Ardon.** « Fuzzy region competition : a convex two-phase segmentation framework ». In : *Proceedings of the 1st international conference on Scale space and variational methods in computer vision. SSVM'07*. Ischia, Italie, 2007, pages 214-226.
- Benoit Mory, Roberto Ardon et Jean-Philippe Thiran.** « Variational Segmentation using Fuzzy Region Competition and Local Non-Parametric Probability Density Functions ». In : *IEEE 11th International Conference on Computer Vision, ICCV*. Rio de Janeiro, Brésil, oct. 2007, pages 1-8.
- David Mumford et Jayant Shah.** « Optimal approximations by piecewise smooth functions and associated variational problems ». In : *Communications on Pure and Applied Mathematics* 42.5 (1989), pages 577-685. ISSN : 00103640. DOI : 10.1002/cpa.3160420503.

- K. Ni, Xavier Bresson, Tony F. Chan et Selim Esedoğlu.** « Local Histogram Based Segmentation Using the Wasserstein Distance ». In : *International Journal of Computer Vision* 84.1 (août 2009), pages 97-111.
- Mila Nikolova.** « A variational approach to remove outliers and impulse noise ». In : *Journal of Mathematical Imaging and Vision* 20 (2004), pages 99-120.
- Mila Nikolova, Selim Esedoğlu et Tony F. Chan.** « Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models ». In : *SIAM Journal on Applied Mathematics* 66.5 (2006). (UCLA CAM Report 04-54, 2004), pages 1632-1648.
- Stanley Osher et R. Fedkiw.** *Level set methods and dynamic implicit surfaces.* Applied mathematical science. New York, USA : Springer, 2003. ISBN : 0-387-95482-1.
- Stanley Osher et James A. Sethian.** « Fronts propagating with curvature-dependant speed : algorithms based on Hamilton-Jacobi formulations ». In : *Journal of Computational Physics* 79.1 (nov. 1988), pages 12-49.
- Nikos Paragios et Rachid Deriche.** « Geodesic Active Regions for Supervised Texture Segmentation ». In : *Proceedings of the International Conference on Computer Vision - Volume 2 - Volume 2.* ICCV '99. Washington, DC, USA : IEEE Computer Society, 1999, pages 926-. ISBN : 0-7695-0164-8.
- Alex Petland et Bradley Horowitz.** « Recovery of Nonrigid Motion and Structure ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* 13.7 (juil. 1991), pages 730-742. ISSN : 0162-8828. DOI : 10.1109/34.85661.
- Thomas Pock, Antonin Chambolle, Daniel Cremers et Horst Bischof.** « A Convex Relaxation Approach for Computing Minimal Partitions ». In : *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).* 2009, pages 810-817.
- Frederic Precioso et Michel Barlaud.** « B-Spline active contour with handling of topology changes for fast video segmentation ». In : *Eurasip Journal on Applied Signal Processing, special issue : image analysis for multimedia interactive services - PART II* 2002.6 (juin 2002), pages 555-560. DOI : 10.1155/s1110865702203121.
- Fuhao Qin, Yi Luo, Kim B. Olsen, Wenying Cai et Gerard Schuster.** « Finite-difference solution of the eikonal equation along expanding wavefronts ». In : *Geophysics* 57.3 (1992), pages 478-487. DOI : 10.1190/1.1443263.
- Rémi Ronfard.** « Region-based Strategies for Active Contour Models ». In : *Int. J. Comput. Vision* 13.2 (oct. 1994), pages 229-251. ISSN : 0920-5691. DOI : 10.1007/BF01427153.
- Elisabeth Rouy et Agnès Tourin.** « A Viscosity Solutions Approach to Shape-from-shading ». In : *SIAM J. Numer. Anal.* 29.3 (juin 1992), pages 867-884. ISSN : 0036-1429. DOI : 10.1137/0729053.

- Leonid I. Rudin, Stanley Osher et Emad Fatemi.** « Nonlinear total variation based noise removal algorithms ». In : *Physica D : Nonlinear Phenomena* 60.1 (1992), pages 259-268. ISSN : 0167-2789. DOI : 10.1016/0167-2789(92)90242-F.
- Christoph Schnörr.** « Computation of Discontinuous Optical Flow by Domain Decomposition and Shape Optimization ». In : *Int. J. Comput. Vision* 8.2 (août 1992), pages 153-165. ISSN : 0920-5691. DOI : 10.1007/BF00127172.
- James A. Sethian.** « A fast marching level set method for monotonically advancing fronts ». In : *Proc. Nat. Acad. Sci.* 93.4 (1996), pages 1591-1595.
- James A. Sethian.** *Level set methods and fast marching methods : evolving interfaces in computational geometry, fluid mechanics, computer vision and material sciences.* Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, 1999.
- Yonggang Shi.** « Object-Based Dynamic Imaging With Level Set Methods ». Thèse de doctorat. Université de Boston, USA, 2005.
- Yonggang Shi et W. C. Karl.** « A fast level set method without solving PDEs ». In : *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Tome 2. Mar. 2005, pages 97-100.
- Jan Sokolowski et Jean-Paul Zolésio.** *Introduction to shape optimization : shape sensitivity analysis.* Tome 16. Springer Series in Computational Mathematics. Berlin, Heidelberg, New York : Springer Verlag, juil. 1992.
- Lawrence H. Staib et James S. Duncan.** « Boundary Finding with Parametrically Deformable Models ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* 14.11 (nov. 1992), pages 1061-1075. ISSN : 0162-8828. DOI : 10.1109/34.166621.
- Petter Strandmark, Frederik Kahl et Niels Chr. Overgaard.** « Optimizing parametric total variation models. » In : *IEEE International Conference on Computer Vision*. 2009, pages 2240-2247. DOI : 10.1109/ICCV.2009.5459464.
- Gilbert Strang.** «  $L^1$  and  $L^\infty$  and approximation of vector fields in the plane ». In : *Nonlinear Partial Differential Equations in Applied Science*. Sous la direction de **H. Fujita, P. Lax et G. Strang**. Tome 5. Lecture Notes in Num. Appl. Anal. 1982, pages 273-288. DOI : 10.1016/S0304-0208(08)72097-8.
- Gilbert Strang.** « Maximal Flow Through a Domain ». In : *Math. Program.* 26.2 (juin 1983), pages 123-143. ISSN : 0025-5610. DOI : 10.1007/BF02592050.
- Changming Sun et Stefano Pallottino.** « Circular Shortest Path in images ». In : *Pattern Recognition* 36.3 (mar. 2003), pages 709-719. DOI : 10.1016/S0031-3203(02)00085-7.
- Demetri Terzopoulos, Andrew Witkin et Michael Kass.** « Constraints on deformable models : recovering 3D shape and nongrid motion ». In : *Artificial Intelligence* 36.1 (1988), pages 91-123.
- John N. Tsitsiklis.** « Efficient algorithms for globally optimal trajectories ». In : *IEEE Transactions on Automatic Control* 40.9 (sept. 1995), pages 1528-1538.



- Fernandio A. Velasco** et **L. Marroquin José**. « Robust parametric active contours : the Sandwich Snakes ». In : *Machine Vision and Applications* 12 (2001), pages 238-242. DOI : 10.1007/s001380050143.
- Luminita A. Vese** et **Tony F. Chan**. « A multiphase level set framework for image segmentation using the Mumford and Shah model ». In : *International Journal of Computer Vision* 50.3 (2002), pages 271-293.
- Manuel Werlberger, Thomas Pock, Markus Unger** et **Horst Bischof**. « A Variational Model for Interactive Shape Prior Segmentation and Real-Time Tracking ». In : *Proceedings of the Second International Conference on Scale Space and Variational Methods in Computer Vision*. SSVM '09. Voss, Norway : Springer-Verlag, 2009, pages 200-211. ISBN : 978-3-642-02255-5. DOI : 10.1007/978-3-642-02256-2\_17.
- Chenyang Xu, Yezzi Anthony** et **Jerry L. Prince**. « On the relationship between parametric and geometric active contours ». In : *Proc. 34th Asilomar Conference on Signal, Systems and Computers*. Oct. 2000, pages 483-489.
- Chenyang Xu, Dzung L. Pham** et **Jerry L. Prince**. « Medical Image Segmentation Using Deformable Models ». In : *Handbook of Medical Imaging – Volume 2 : Medical Image Processing and Analysis*. Tome PM80. SPIE Press, juin 2000. Chapitre 3, pages 129-174.
- Chenyang Xu** et **Jerry L. Prince**. « Snakes, Shapes, and Gradient Vector Flow ». In : *IEEE Transactions on Image Processing* 7.3 (mar. 1998), pages 359-369.
- Liron Yatziv, Albero Bartesaghi** et **Guillermo Sapiro**. « O(N) implementation of the fast marching algorithm ». In : *Journal of computational physics* 212 (2006), pages 393-399. DOI : 10.1016/j.jcp.2005.08.005.
- Anthony Yezzi Jr, Andy Tsai** et **Alan Willsky**. « A Statistical Approach to Snakes for Bimodal and Trimodal Imagery ». In : *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*. ICCV '99. Washington, DC, USA : IEEE Computer Society, 1999, pages 898-. ISBN : 0-7695-0164-8.
- Song Chun Zhu** et **Alan Yuille**. « Region Competition : Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* 18.9 (sept. 1996), pages 884-900. ISSN : 0162-8828. DOI : 10.1109/34.537343.



## Chapitre 5

# Optimisation de fonctions pseudo-booléennes

*Laurent CARAFFA<sup>1</sup>, Jean-Philippe TAREL<sup>1</sup>, Mathias PAGET<sup>1</sup>*

*Résumé – Une approche pour faire du traitement d'image consiste à poser les problèmes comme la minimisation d'une énergie sur l'espace des images qui sont représentées par des fonctions 2D. L'optimisation de ce type d'énergie passe par le développement de schémas numériques et donc par la nécessaire discrétisation de l'espace des fonctions choisies et de l'énergie utilisée. Les images étant en pratique représentées de façon discrétisée, une autre approche consiste à poser les problèmes comme la minimisation d'une énergie directement dans le domaine discret. Cela conduit généralement à introduire une représentation de l'image sous la forme d'un graphe afin de pouvoir modéliser les interactions entre voisins.*

*Avec cette approche, nous disposons du cadre théorique de l'optimisation quadratique pseudo-booléenne (QPBO) dans le cas où les variables sont binaires et de certaines extensions aux cas non-binaires. Dans ce chapitre, les principaux résultats obtenus dans ce cadre théorique QPBO sont présentés de façon succincte afin d'introduire les outils disponibles actuellement.*

*Enfin, l'utilisation de ces outils est illustrée sur le problème de la reconstruction 3D à partir de paires stéréoscopiques, mais ils peuvent s'appliquer à bien d'autres problèmes au delà du traitement d'image.*

---

1. IFSTTAR

## 5.1. Introduction

En traitement d'images et dans d'autres domaines, de nombreux problèmes conduisent à la recherche d'un optimum global d'une énergie multi-labels. La solution recherchée est alors représentée par un ensemble de variables indexées par les sommets d'un graphe et prenant leurs valeurs parmi un ensemble de labels (entiers). Soit  $L = \{0, \dots, m\}$  l'ensemble des labels, par exemple  $\{0, 1\}$  pour une segmentation binaire,  $\{0, \dots, 255\}$  pour la restauration d'une image 8 bits,  $\{0, \dots, m\}$  pour un problème de reconstruction 3D à partir de paires stéréoscopiques avec  $m$  la disparité maximale admissible. Dans ces exemples, le graphe associé a pour sommets les pixels et les arrêtes codent les voisinages entre pixels.

Soit  $n$  le nombre de variables et de sommets,  $f : L^n \rightarrow \mathbb{R}$  est une fonction déduite du modèle du problème à traiter, où chaque configuration  $\mathbf{x}$  de  $L^n$  correspond une valeur d'énergie :

$$f(\mathbf{x}) = \sum_{u \in C_1} \vartheta_u(L_u) + \sum_{u,v \in C_2} \vartheta_{uv}(L_u, L_v) + \sum_{u,v,w \in C_3} \vartheta_{uvw}(L_u, L_v, L_w) + \dots \quad (5.1)$$

où  $C_k$  est l'ensemble des sous-ensembles de  $k$  sommets et  $\vartheta$  est la fonction de coût de chaque sous-ensemble. Lorsque le modèle dérive d'un modèle probabiliste, les solutions les plus probables ont les énergies les plus faibles, on essaiera donc de résoudre le problème de minimisation suivant :

$$\min_{\mathbf{x} \in L^n} f(\mathbf{x}). \quad (5.2)$$

La recherche des labels qui minimisent (5.1) est généralement complexe. Le développement de modèles nécessite donc la connaissance des techniques d'optimisation disponibles.

En optimisation, il est primordial de savoir s'il est possible de trouver, et dans quels délais, le minimum global d'une fonction. Si on peut trouver le minimum global d'une énergie, il n'y a pas d'ambiguïté sur le résultat de l'optimisation, et l'on peut évaluer si le modèle proposé est bien en adéquation avec le problème à résoudre. Inversement, lorsque seulement un minimum local est obtenu, on ne sait pas si les erreurs viennent de ce que l'on n'a pas atteint le minimum global de la fonction, ou si c'est le modèle qui est incorrect.

Lorsque le nombre de configurations possibles n'est pas important, il est assez simple de trouver le minimum de l'énergie. Une recherche exhaustive sur l'ensemble des configurations est suffisante pour permettre de trouver celle qui minimise l'énergie. Mais quand il existe des interactions locales entre les variables, le nombre de configurations possibles devient très important avec le nombre de variables. Par exemple, considérons le cas de l'optimisation d'un champ de Markov binaire sur une image  $512 \times 512$  avec prise en compte des

voisinages deux à deux ; il existe alors  $2^{512 \times 512}$  configurations possibles. Sachant que les images actuelles sont constituées de plusieurs millions de pixels et que la cardinalité de l'ensemble des labels peut être beaucoup plus importante (256 pour du débruitage, ou encore plus pour la stéréovision avec une grande base), il est donc de toute évidence impossible de parcourir l'ensemble des solutions. Des algorithmes spécialisés ont donc dû être conçus pour optimiser ce type de problème avec un grand espace de recherche.

En optimisation continue, lorsqu'une fonction est convexe, tous les minima locaux sont globaux. De façon semblable à l'optimisation continue, il existe en optimisation discrète une classe de fonctions qui permet, malgré une croissance exponentielle de l'espace de recherche, de trouver un minimum global en un temps polynomial. Ce sont les fonctions dites *sous-modulaires*, qu'il est important de distinguer des fonctions non sous-modulaires. Une des méthodes les plus utilisées pour optimiser les fonctions sous-modulaires est connue sous le nom de coupe de graphe (*graph-cuts*), voir par exemple (KOLMOGOROV et al., 2002).

Toutefois, l'espace des fonctions sous-modulaires est assez restreint. De plus, au delà d'une certaine cardinalité des tailles des voisinages, pouvoir dire d'une fonction qu'elle est sous-modulaire s'avère être un problème *NP-Difficile*. Il arrive aussi fréquemment que l'on veuille optimiser des fonctions non sous-modulaires, il est donc utile dans ce cas, de disposer de méthodes spécifiques d'optimisation même si les résultats obtenus sont partiels. Compte tenu de l'importance de ce type d'optimisation dans différents champs d'application, de nombreux algorithmes ont vu le jour.

L'une des approches les plus efficaces pour optimiser le type d'énergies (5.2) est celle fondée sur l'optimisation de fonctions pseudo-booléennes. L'étude de l'optimisation des fonctions pseudo-booléennes date des années 50. Observées initialement dans la théorie des jeux, ce fut l'une des principales motivations de leur étude en recherche opérationnelle. Par la suite, la découverte de la présence récurrente de fonctions pseudo-booléennes dans un spectre très large d'applications et de domaines a fortement influé sur l'évolution et le gain d'intérêt de ce domaine.

Dans le cas des labels binaires, on nomme  $\mathbf{B} = \{0, 1\}$  l'ensemble des labels. Une fonction  $f : \mathbf{B}^n \rightarrow \mathbb{R}$  est dite pseudo-booléenne. Elle peut être écrite de façon unique comme un polynôme de  $n$  variables de la façon suivante :

$$f(x_1, \dots, x_n) = c_0 + \sum_{i=1}^n c_i x_i + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j + \sum_{1 \leq i < j < k \leq n} c_{ijk} x_i x_j x_k + \dots \quad (5.3)$$

Dans le cas binaire, l'énergie (5.1) peut être réécrite comme une fonction pseudo-booléenne, voir la partie 5.2.2. Dans le cas non binaire, le passage à des fonctions pseudo-booléennes est décrit dans la partie 5.3.2.

Nous avons choisi d'introduire dans la partie 5.2 l'optimisation de fonctions pseudo-booléennes, en suivant la présentation faite dans (BOROS et Peter L

HAMMER, 2002). Cette partie est donc une synthèse partielle de cet article d'introduction sur l'état de l'art de l'optimisation de fonctions pseudo-bouliennes. L'ensemble des théorèmes que nous présentons est limité à une partie de cet article.

Certaines notions sont détaillées par des exemples complémentaires à ceux de l'article. Cette présentation est assez différente de celle proposée habituellement en coupe de graphe, même si on retrouve comme un cas particulier la coupe de graphe habituelle comme expliqué en partie 5.2.7. La partie suivante 5.3 présente des algorithmes pour l'optimisation approchée de fonctions multi-labels fondées sur la théorie de l'optimisation de fonctions pseudo-bouliennes. Enfin, dans la partie 5.4, trois exemples d'utilisation de l'optimisation de fonctions pseudo-bouliennes sont discutés afin d'expliquer comment utiliser en pratique ce type d'optimisation.

## 5.2. Optimisation de fonctions pseudo-bouliennes

### 5.2.1. Notations

Dans la suite, on nomme  $\mathbf{V} = \{1, 2, \dots, n\}$  l'ensemble des indices de 1 à  $n$ , où  $n$  est le nombre de variables.  $\mathbf{x} = (x_1, \dots, x_n)$  est un vecteur binaire et par définition  $\bar{x}_i = 1 - x_i$  est le complément pour  $i \in \mathbf{V}$ . On note  $\mathbf{X} = \{x_1, \bar{x}_1, \dots, x_n, \bar{x}_n\}$  l'ensemble de ces symboles.

### 5.2.2. Représentations

On dispose d'au moins deux représentations pour la même fonction pseudo-boulienne.

#### 5.2.2.1. Forme multilinéaire polynomiale unique

$$f(\mathbf{x}) = \sum_{S \subseteq \mathbf{V}} c_S \prod_{j \in S} x_j \quad (5.4)$$

On appelle degré de  $f$  la taille du plus grand ensemble  $S \subseteq \mathbf{V}$  pour lequel  $c_S \neq 0$ . Cette représentation est utile pour étudier certaines propriétés de  $f$ .

#### 5.2.2.2. Posiforme

Une fonction pseudo-boulienne peut toujours être représentée par un polynôme à coefficients positifs, au terme constant près, ou *posiforme* de la forme suivante :

$$\phi(\mathbf{x}) = a_\emptyset + \sum_{T \subseteq \mathbf{X}} a_T \prod_{u \in T} u \quad (5.5)$$

où  $a_T \geq 0$ . Pour représenter une fonction pseudo-boulienne par une posiforme, une façon de procéder consiste à remplacer le premier élément  $x$  par  $1 - \bar{x}$  dans chaque terme où  $c_S < 0$  dans la forme multilinéaire unique. Cette dernière opération est nommée complément.

**Table 5.1.**

Table de vérité de la fonction pseudo-booléenne de l'exemple 5.1. Les minimums de  $f_1$  sont en gras

	<b>0 0</b> 0 0 <b>0 0 0 0</b> <b>1 1</b> 1 1 1 1 1
<b>x</b>	<b>0 0</b> 0 0 <b>1 1 1 1</b> 0 0 0 0 <b>1 1</b> 1 1
	<b>0 0</b> 1 1 <b>0 0</b> 1 <b>1 0</b> 0 1 1 0 0 1 1
	<b>0 1</b> 0 1 <b>0 1</b> 0 1 0 1 0 1 0 1 0 1
$f_1(\mathbf{x})$	<b>0 0</b> 13 4 <b>0 0</b> 9 <b>0 5</b> 5 14 5 5 12 14 12

Il existe plusieurs posiformes pour une même fonction pseudo-booléenne, donc pour une même table de vérité. On notera  $\mathcal{P}(f)$  la famille des posiformes représentant la même fonction  $f$ .

**Exemple 5.1.** Ces deux posiformes :

$$\phi_{11}(\mathbf{x}) = 5x_1 + 4\bar{x}_1\bar{x}_2x_3 + 7x_1x_2x_4 + 9x_3\bar{x}_4$$

$$\phi_{12}(\mathbf{x}) = x_1 + 4x_1x_2 + 4x_1\bar{x}_2\bar{x}_3 + 7x_1x_2x_4 + 4\bar{x}_2x_3 + 9x_3\bar{x}_4$$

ont la même forme multilinéaire polynomiale unique :

$$f_1(\mathbf{x}) = 5x_1 + 13x_3 - 4x_1x_3 - 4x_2x_3 - 9x_3x_4 + 4x_1x_2x_3 + 7x_1x_2x_4.$$

En effet, ces trois fonctions pseudo-booléennes vérifient la même table de vérité, table. 5.1.

### 5.2.3. Propriétés des fonctions pseudo-booléennes

La structure de posiforme permet de vérifier certaines propriétés. En effet, une posiforme ne possédant aucun terme négatif, la valeur de la posiforme ne peut être inférieure à 0, à la constante près. Dans un sens, minimiser une posiforme équivaut à essayer d'annuler le plus de termes possible. Il en découle que minimiser une posiforme est équivalent au problème d'optimisation de satisfabilité booléenne maximale *MAX-SAT* (Maximum Satisfiability problem), qui est *NP-complet*. Sous l'hypothèse  $P \neq NP$ , on est alors seulement capable de trouver en un temps polynomial une affectation partielle des variables binaires qui est dans l'*optimum* ((BOROS et Peter L HAMMER, 2002), p.18).

On appelle affectation partielle, un vecteur binaire  $\mathbf{y} \in \mathbf{B}^S$  correspondant à un sous-ensemble  $S \subseteq \mathbf{V}$ . De plus, pour un sous-ensemble  $S \subseteq \mathbf{V}$  d'indices et un vecteur  $\mathbf{y} \in \mathbf{B}^S$ , on note par  $\mathbf{x}[S] \in \mathbf{B}^S$  le vecteur correspondant aux indices dans  $S$ , c'est-à-dire  $\mathbf{x}[S] = (x_i | i \in S)$ . Pour une affectation partielle  $\mathbf{y} \in \mathbf{B}^S$  et pour un

vecteur  $\mathbf{x} \in \mathbf{B}^n$ , on définit l'échange de  $\mathbf{x}$  comme le vecteur binaire  $\mathbf{z}$  par :

$$z_j = \begin{cases} x_j & \text{si } j \notin S \\ y_j & \text{si } j \in S \end{cases} \quad (5.6)$$

et on le note par  $\mathbf{z} = \mathbf{x}[S \leftarrow \mathbf{y}]$ . Par exemple, si  $n = 5$ ,  $S = \{1, 2, 5\}$  et  $\mathbf{y}$  est l'affectation partielle  $y_1 = 1, y_2 = 0$  et  $y_5 = 1$  alors l'affectation de  $\mathbf{x} = (1, 1, 1, 0, 0)$  par  $\mathbf{y}$  va être le vecteur  $\mathbf{z} = (1, 0, 1, 0, 1)$ .

Soit une fonction pseudo-booléenne  $f$ , un vecteur  $\mathbf{x} \in \mathbf{B}^n$  et une affectation partielle  $\mathbf{y} \in \mathbf{B}^n$  pour un sous-ensemble  $S \subseteq \mathbf{V}$ , alors on a les définitions suivantes.

**Persistence forte** Pour  $f$  aux valeurs de  $\mathbf{y}$ , si pour tout  $\mathbf{x} \in \text{Argmin}_{\mathbf{B}^n}(f)$ , alors  $\mathbf{x}[S] = \mathbf{y}$ . En d'autres termes, on dit qu'une affectation est persistante forte si l'ensemble des valeurs de l'affectation  $\mathbf{y}$  appartient à tous les minima globaux de la fonction  $f$ .

**Persistence faible** Pour  $f$  aux valeurs de  $\mathbf{y}$ , si pour tout  $\mathbf{x} \in \text{Argmin}_{\mathbf{B}^n}(f)$ , alors  $\mathbf{x}[S \leftarrow \mathbf{y}] \in \text{Argmin}_{\mathbf{B}^n}(f)$ . On dit qu'une affectation est persistante faible si toutes les valeurs de  $\mathbf{x}$  appartenant à tous les minima globaux de la fonction  $f$  le sont encore après affectation.

**Exemple 5.1.** Pour illustrer les deux notions précédentes, reprenons la fonction  $f_1$  de l'exemple 5.1 et sa posiforme  $\phi_{11}$ . On considère l'affectation partielle  $\mathbf{y}^* = (0) \in \mathbf{B}^{\{1\}}$ . Nous pouvons constater avec la table 5.1 que pour tous les minima globaux,  $x_1 = (0)$ . L'affectation  $\mathbf{y}$  est donc persistante forte.

Soit  $\mathbf{z}^* = (0, 1) \in \mathbf{B}^{\{1,2\}}$  une autre affectation partielle. Les solutions  $\mathbf{x} = (0, 0, 0, 0)$  et  $\mathbf{x} = (0, 0, 0, 1)$  appartiennent au minimum global de la fonction  $f_1$ , pourtant leur deuxième valeur  $x_2$  est différente de 1. L'affectation n'est donc pas persistante forte. Néanmoins, l'application de l'affectation  $\mathbf{z}^*$  sur  $\mathbf{x} = (0, 0, 0, 0)$  et de  $\mathbf{x} = (0, 0, 0, 1)$  donne  $\mathbf{x} = (0, 1, 0, 0)$  et  $\mathbf{x} = (0, 1, 0, 1)$ . Ces deux vecteurs appartiennent bien au minimum global de la fonction donc cette affectation est persistante faible.

#### 5.2.4. Fonction pseudo-booléenne quadratique

L'optimisation d'une fonction pseudo-booléenne quadratique est un problème abondamment étudié. L'une des principales propriétés de ce type d'optimisation est qu'il est possible de trouver en un temps polynomial une borne inférieure aux valeurs de la fonction appelée la *roof duality*. Une fois cette borne inférieure atteinte, il est possible d'en déduire une affectation partielle ayant comme propriété d'être persistante forte.

On nomme  $\mathcal{F}_2$  la famille des fonctions pseudo-booléennes quadratiques. Une fonction pseudo-booléenne quadratique peut être représentée par sa forme



multilinéaire polynomiale unique :

$$f(\mathbf{x}) = c_0 + \sum_{i=1}^n c_i x_i + \sum_{1 \leq i < j \leq n} c_{ij} x_i x_j \quad (5.7)$$

ou par une posiforme quadratique de la forme suivante :

$$\phi(\mathbf{x}) = a_0 + \sum_{u \in \mathbf{X}} a_u u + \sum_{u, v \in \mathbf{X}, u \neq v} a_{uv} uv \quad (5.8)$$

où  $a_u \geq 0$  et  $a_{uv} \geq 0$ . Généralement,  $a_0$  est nommé *terme constant*, les termes de degré 1 *terme linéaire*, et ceux de degré 2 *terme quadratique*. Comme dans (BOROS et Peter L HAMMER, 2002), le terme constant  $a_0$  de la posiforme  $\phi$  est noté  $C(\phi)$ . On notera aussi  $\mathcal{P}_2(f)$  la famille des posiformes quadratiques représentant la même fonction  $f$ .

### 5.2.5. Roof duality

Trouver le minimum d'une fonction pseudo-boulienne quadratique quelconque est un problème *NP-complet*, mais on peut calculer une borne inférieure aux valeurs de la fonction appelée *roof duality* en un temps polynomial. Une façon intuitive d'interpréter la *roof duality* est de trouver quelle est la posiforme  $\phi$  dans un ensemble de fonctions pseudo-bouliennes vérifiant la même table de vérité, dont le terme constant  $C(\phi)$  est le plus grand.

En effet, les coefficients d'une posiforme étant tous positifs, le terme constant maximum de  $\phi$  est une borne inférieure de  $f$ .

Ce problème a été étudié en optimisation de fonctions pseudo-bouliennes quadratiques et il existe plusieurs techniques permettant de calculer cette borne inférieure de la fonction. Elle est notée :

- $M_2(f)$  quand elle est obtenue par majorisation,
- $C_2(f)$  par la complémentation,
- $L_2(f)$  par la linéarisation,
- autres quand on utilise le Lagrangien, *paved duality*, etc.

Un résultat important est que toutes ces bornes sont les mêmes et égales à  $\max_{\phi \in \mathcal{P}_2(f)} C(\phi)$ .

Les méthodes pour les obtenir sont donc équivalentes.

**Théorème 5.1.** Pour toutes fonctions pseudo-bouliennes quadratiques  $f \in \mathcal{F}_2$ , alors (voir (BOROS et Peter L HAMMER, 2002)) :

$$M_2(f) = C_2(f) = L_2(f) \leq \min_{\mathbf{x} \in \mathbf{B}^n} f(\mathbf{x}).$$

Une fois la *roof duality* atteinte (c'est à dire après avoir transformé la posiforme d'origine en une posiforme avec la plus grande constante possible) la propriété suivante permet d'extraire un sous-ensemble persistant fort.

**Théorème 5.2. (Persistance forte)** Soit une fonction pseudo-booléenne quadratique  $f \in \mathcal{F}_2$ , soit  $\phi \in \mathcal{P}_2(f)$  une posiforme la représentant tel que  $C(\phi) = C_2(\phi)$ , si  $a_u > 0$  pour des labels  $u \in \mathbf{X}$ , alors  $u = 0$  dans tous les vecteurs binaires  $\mathbf{x} \in \text{Argmin}(f)$  minimisant  $f$  (voir (BOROS et Peter L HAMMER, 2002)).

Autrement dit, s'il reste un terme linéaire dans la formule obtenue après avoir atteint la *roof duality*, alors chaque affectation annulant le terme linéaire appartiendra à tous les minima globaux.

### 5.2.6. Réduction

Trouver la *roof duality* est un problème de classe polynomiale. En effet, elle peut être obtenue en résolvant un problème de programmation linéaire. Comme précédemment indiqué, plusieurs méthodes existent pour calculer cette borne. Dans cette partie, nous allons détailler une méthode fondée sur la théorie des graphes. Cette méthode est très efficace car fondée sur des algorithmes de la théorie des graphes très étudiés et optimisés. Ces algorithmes sont donc parmi les plus rapides pour les problèmes discrets.

L'algorithme est décomposé en plusieurs étapes : tout d'abord, la posiforme est traduite sous la forme d'un graphe dit induit. Ensuite, il faut effectuer la réduction de la fonction pseudo-booléenne pour obtenir la valeur de la *roof duality*. Cela se fait en pratique par poussage de flot sur le graphe induit et conduit à la détermination d'un graphe résiduel. Enfin, les affectations persistantes fortes sont obtenues à partir du parcours du graphe réduit.

#### 5.2.6.1. Construction du graphe induit

Soit une fonction pseudo-booléenne quadratique  $f \in \mathcal{F}_2$  donnée par la posiforme  $\phi \in \mathcal{P}_2(f)$  de forme (5.8). On associe à cette posiforme quadratique un graphe orienté  $G_\phi = (N, A)$  où l'ensemble de sommets est défini par  $N = \mathbf{X} \cup \{0, \bar{0}\}$ .

Il y a deux sommets par variable : le sommet  $x_i$ , noté  $i$  dans le graphe, représentant le cas  $x_i = 1$  et le sommet  $\bar{x}_i$ , noté  $\bar{i}$  dans le graphe, représentant le cas  $\bar{x}_i = 1$ , plus deux sommets  $x_0$  et  $\bar{x}_0$  représentant les cas  $x_0 = 1$  et  $\bar{x}_0 = 1$ . Ces deux sommets sont notés  $0$  et  $\bar{0}$ , respectivement, dans le graphe induit.

Les arcs sont définis à partir des termes de la posiforme. À chaque terme quadratique  $a_{uv}uv$  correspond deux arcs  $(\overrightarrow{u, \bar{v}})$  et  $(\overrightarrow{\bar{v}, u})$  avec chacun un poids  $\frac{1}{2}a_{uv}$ . À chaque terme linéaire  $a_uu$  correspond aussi deux arcs  $(\overrightarrow{u, \bar{x}_0})$  et  $(\overrightarrow{x_0, \bar{u}})$  avec chacun comme poids  $\frac{1}{2}a_u$  (voir (BOROS et Peter L HAMMER, 2002)).

**Exemple 5.2.** Soit la posiforme quadratique suivante :

$$\phi_2(\mathbf{x}) = 10x_1 + 8\bar{x}_1x_2 + 6\bar{x}_2x_3 + 4\bar{x}_3. \tag{5.9}$$

Le premier terme va induire deux arcs orientés  $(\overrightarrow{x_1, \bar{x}_0})$  et  $(\overrightarrow{x_0, \bar{x}_1})$  de poids 5.

Le deuxième terme va induire deux arcs orientés  $(\overrightarrow{\bar{x}_1, \bar{x}_2})$  et  $(\overrightarrow{x_2, x_1})$  de poids 4.

Le troisième terme va induire deux arcs orientés  $\overrightarrow{(\bar{x}_2, \bar{x}_3)}$  et  $\overrightarrow{(x_3, x_2)}$  de poids 3.

Le quatrième terme va induire deux arcs orientés  $\overrightarrow{(\bar{x}_3, \bar{x}_0)}$  et  $\overrightarrow{(x_0, x_3)}$  de poids 2.

Le graphe induit est montré dans la figure 5.1(a).

### 5.2.6.2. Calcul de la borne inférieure par flot maximum

Une fois le graphe induit construit, nous allons voir qu'il y a équivalence entre une somme alternée dans une équation pseudo-booléenne et un chemin augmentant le flot dans le graphe induit de la posiforme  $\phi$ . La caractéristique d'une somme alternée est qu'il est possible de faire apparaître, par une identité algébrique, une constante, et par conséquent, d'augmenter le terme constant  $C(\phi)$  de la posiforme pour atteindre la valeur de *roof duality*  $C_2(\phi)$ . Cette réduction est réalisée en pratique par un algorithme de poussage de flot sur le graphe induit grâce à l'équivalence entre somme alternée et chemin augmentant.

### 5.2.6.3. Équivalence entre somme alternée et chemin augmentant

Un flot possible dans un graphe  $G = (N, A)$  avec comme source  $x_0$  et comme puits  $\bar{x}_0$  est une application  $\zeta : A \rightarrow \mathbb{R}_+$  respectant les contraintes :

$$\zeta(u, v) \leq c_{u,v} \quad \text{pour tous les arcs } (\overrightarrow{u, v}) \in A, \text{ et} \quad (5.10)$$

$$\sum_{(\overrightarrow{u, v}) \in A} \zeta(u, v) = \sum_{(\overrightarrow{v, w}) \in A} \zeta(v, w) \quad \text{pour tous les sommets } v \in \mathbf{X}. \quad (5.11)$$

La première contrainte indique que le flot ne peut pas dépasser la capacité de chaque arc. La seconde contrainte impose que le flot total entrant soit égal au flot total sortant en chaque sommet.

Pour un graphe donné  $G = (N, A)$  et un flot possible  $\zeta$  dans ce graphe, on appelle graphe réduit  $G[\zeta] = (N, A^\zeta)$  avec comme capacités :

$$c_{uv}^\zeta = \begin{cases} c_{uv} - \zeta(u, v) & \text{pour } (\overrightarrow{u, v}) \in A \\ \zeta(u, v) & \text{pour } (\overrightarrow{v, u}) \in A \end{cases} \quad (5.12)$$

Quand  $u_1, u_2, \dots, u_k \in \mathbf{X}$ , on appelle somme alternée, la forme quadratique suivante :

$$u_1 + \bar{u}_1 u_2 + \bar{u}_2 u_3 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k. \quad (5.13)$$

On remarque que cette forme quadratique est bien homogène et alternée en introduisant  $\bar{x}_0$  et  $x_0$ . Une posiforme quadratique  $\phi$  contient une somme alternée (5.13) de poids  $\omega$  si nous avons  $a_{u_1} \geq \omega$ ,  $a_{\bar{u}_j u_{j+1}} \geq \omega$  pour  $j = 1, \dots, k-1$  et  $a_{\bar{u}_k} \geq \omega$  pour tous les coefficients correspondants de  $\phi$  (voir (BOROS et Peter L HAMMER, 2002)).

**Proposition 5.1.** L'identité suivante est vérifiée pour les sommes alternées :

$$u_1 + \bar{u}_1 u_2 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k = 1 + u_1 \bar{u}_2 + \dots + u_{k-1} \bar{u}_k. \quad (5.14)$$

Cela se démontre en changeant  $x_i$  en  $1 - \bar{x}_i$ . Avec cette identité, une posiforme quadratique  $\phi$  qui contient une somme alternée (5.13) de poids  $\omega$  peut être transformée en une posiforme équivalente ayant un terme constant plus important. Pour ce faire, il faut dans un premier temps réécrire  $\phi$  en :

$$\phi = \omega[u_1 + \bar{u}_1 u_2 + \bar{u}_2 u_3 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k] + \phi'.$$

Par construction,  $\phi'$  est aussi une posiforme quadratique. Par conséquent, en appliquant l'identité (5.14), on obtient :

$$\phi = \omega + \omega[u_1 + \bar{u}_1 u_2 + \bar{u}_2 u_3 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k] + \phi'.$$

En construisant le graphe induit d'une somme alternée suivant les règles précédentes, on voit qu'il y a une correspondance entre une somme alternée contenue dans une posiforme  $\phi$  de poids  $\omega$  et un chemin augmentant de capacité  $\omega$  dans le graphe induit de  $\phi$ . La proposition suivante peut donc être énoncée :

**Proposition 5.2.** On considère une posiforme  $\phi \in P_2(f)$  et un flot possible  $\zeta$  dans le graphe  $G = G_\phi$ . Alors  $x_0, u_1, \dots, u_k, \bar{x}_0$  est un chemin augmentant de capacité  $\omega > 0$  dans ce flot si et seulement si  $u_1 + \bar{u}_1 u_2 + \dots + \bar{u}_{k-1} u_k + \bar{u}_k$  est une somme alternée de poids  $\omega$  dans la posiforme  $\phi$ .

**Exemple 5.2.** Reprenons la posiforme  $\phi_2$  de l'exemple précédent. Son graphe induit est montré en figure 5.1(a). Nous allons illustrer l'équivalence entre la capacité d'un chemin augmentant le flot sur le graphe induit et le poids d'une somme alternée utilisée lors de la réduction de la posiforme.

Tous les coefficients de  $\phi_2$  sont supérieurs ou égaux à 4. On choisit donc  $\omega = 4$  et on effectue le regroupement suivant :

$$\phi_2(\mathbf{x}) = \underbrace{4}_{\omega} \underbrace{(x_1 + \bar{x}_1 x_2 + \bar{x}_2 x_3 + \bar{x}_3)}_{\text{somme alternée}} + \underbrace{6x_1 + 4\bar{x}_1 x_2 + 2\bar{x}_2 x_3}_{\text{forme résiduelle } \phi'} \quad (5.15)$$

$$= 4(1 + x_1 \bar{x}_2 + x_2 \bar{x}_3) + 6x_1 + 4\bar{x}_1 x_2 + 2\bar{x}_2 x_3 \quad (5.16)$$

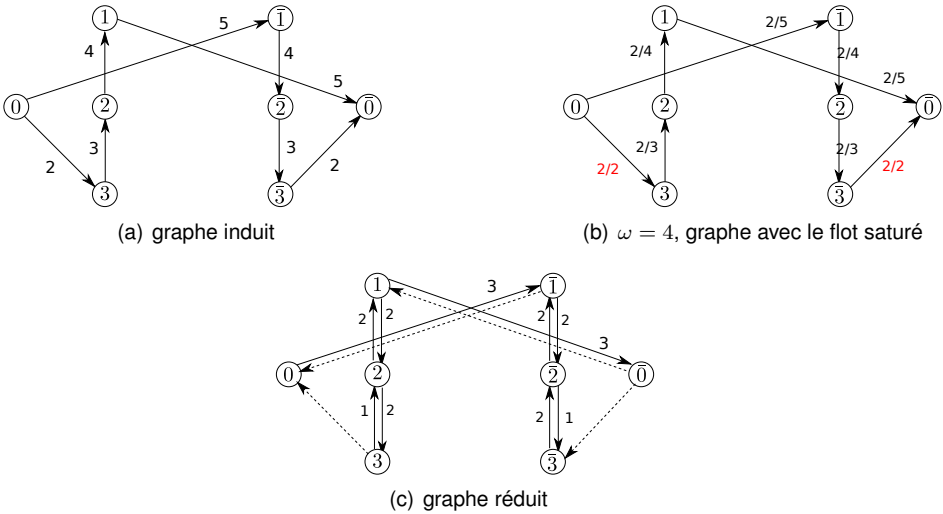
$$= 4 + \underbrace{6x_1 + 4x_1 \bar{x}_2 + 4\bar{x}_1 x_2 + 2\bar{x}_2 x_3 + 4x_2 \bar{x}_3}_{\text{forme réduite } \psi}. \quad (5.17)$$

En appliquant l'identité (5.14) sur la somme alternée qui est maintenant dans (5.15), un terme constant apparaît dans l'équation (5.16). Ceci a comme conséquence, une fois (5.16) développée, d'atteindre la valeur de *roof duality* dans ce cas.

Ces manipulations algébriques ont leur équivalent en terme de poussage de flot sur le graphe. En effet, à partir de la figure 5.1(b), on voit que le flot maximum jusqu'à saturation est obtenu en faisant passer un flot de 2 dans les deux chemins possibles entre 0 et  $\bar{0}$ . Ces deux chemins qui sont symétriques correspondent à la somme alternée dans l'équation (5.15). Nous pouvons aussi constater que  $\omega = 2 + 2 = 4$ . En construisant le graphe réduit avec (5.12) et en simplifiant

**Figure 5.1.**

Équivalence entre la réduction de la posiforme et le poussage de flot sur le graphe induit. (a) montre le graphe induit par la posiforme de l'exemple 5.2. (b) montre le poussage de flot maximum. (c) montre le graphe réduit



au niveau des arcs qui partent ou arrivent de  $x_0$  et  $\bar{x}_0$  (cela correspond à la réduction des termes linéaires par complémentation), on arrive au graphe réduit affiché en figure 5.1(c). On voit qu'il correspond exactement à la forme réduite  $\psi$  dans l'équation (5.17). Les arcs en pointillés ont une capacité nulle.

5.2.6.4. Lien entre le flot maximum et la valeur de roof duality

La réduction précédente doit être réalisée jusqu'à ce qu'il ne soit plus possible de factoriser de nouvelles sommes alternées. Alors, la propriété fondamentale de cette réduction est que la valeur *roof duality*  $C_2(f)$  est atteinte en sommant le terme constant de la posiforme initiale avec la valeur du flot maximum atteinte par l'algorithme de poussage de flot.

**Théorème 5.3.** Soit une fonction pseudo-boulienne quadratique  $f$ , et une posiforme quadratique la représentant  $\phi \in P_2(f)$ , et  $\zeta^*$  la valeur du flot maximum dans  $G_\phi$ , alors :

$$C_2(f) = C(\phi) + \zeta^*.$$

Après avoir calculé le flot maximum, et donc la valeur de *roof duality*, il est possible de trouver les affectations persistantes fortes. Une façon efficace est d'utiliser le graphe réduit.

**Théorème 5.4 (Persistance forte et flot maximum).** Soit  $\phi \in P_2(f)$  pour une fonction pseudo-boulienne quadratique  $f$ , soit  $\zeta^*$  un flot maximum dans  $G = G_\phi$ , soit  $S \in L$  l'ensemble des sommets dans  $G$  qui peuvent être atteints en partant de  $x_0$  par un chemin avec une capacité résiduelle positive, alors  $u(\mathbf{x}^*) = 1$  pour tout  $u \in S$  et pour tous les vecteurs  $\mathbf{x}^* \in \text{Argmin}(f)$ .

En d'autres termes, pour toutes variables pouvant être atteintes dans le graphe réduit en partant de la source  $x_0$  par des capacités strictement positives, leur affectation à 1 est persistante forte.

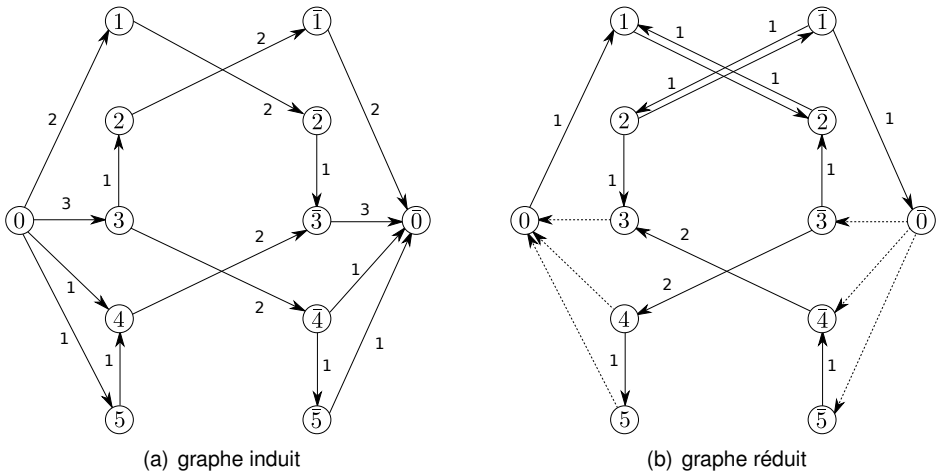
5.2.6.5. Récapitulatif

Pour résumer, les affectations persistantes fortes pour une fonction pseudo-booléenne quadratique  $f$  sont calculées de la façon suivante :

- Trouver une posiforme quadratique  $\phi$  de même table de vérité que  $f$ .
- Construire le graphe induit  $G_\phi$  de  $\phi$ .
- Calculer le flot maximum  $\zeta^*$  de  $G_\phi$ .
- Construire le graphe réduit  $G_\psi$  de  $G_\phi$  à partir du flot maximum  $\zeta^*$ .
- Pour tous les sommets pouvant être atteints à partir de la source  $x_0$ , affecter les variables correspondantes à 1.

Figure 5.2.

(a) montre le graphe induit de l'exemple 5.3 et (b) le graphe réduit



**Exemple 5.3.** La figure 5.2(a) montre le graphe induit. En observant la figure 5.2(a), on peut constater qu'il est possible de faire passer 6 flots de poids 1 sur chacun des chemins suivants :

$$x_0 \rightarrow x_1 \rightarrow \bar{x}_2 \rightarrow \bar{x}_3 \rightarrow \bar{x}_0 \text{ et } x_0 \rightarrow x_3 \rightarrow x_2 \rightarrow \bar{x}_1 \rightarrow \bar{x}_0, \tag{5.18}$$

$$x_0 \rightarrow x_3 \rightarrow \bar{x}_4 \rightarrow \bar{x}_0 \text{ et } x_0 \rightarrow x_4 \rightarrow \bar{x}_3 \rightarrow \bar{x}_0, \tag{5.19}$$

$$x_0 \rightarrow x_3 \rightarrow \bar{x}_4 \rightarrow \bar{x}_5 \rightarrow \bar{x}_0 \text{ et } x_0 \rightarrow x_5 \rightarrow x_4 \rightarrow \bar{x}_3 \rightarrow \bar{x}_0. \tag{5.20}$$

Soit la fonction pseudo-booléenne quadratique suivante :

$$f_3(\mathbf{x}) = 10 - 4x_1 - 4x_3 - 2x_4 + 4x_4x_2 - 2x_2x_3 + 4x_3x_4 - 2x_4x_5.$$

**Table 5.2.**

Table de vérité de la fonction pseudo-booléenne de l'exemple 5.3. Les minimums de  $f_3$  sont en gras

<b>x</b>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1
	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
$f_3(\mathbf{x})$	10	10	8	6	6	6	8	6	10	10	8	6	4	4	6
<b>x</b>	1	1	1	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	1	1	1	1	1	1	1
	0	0	0	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	1	1	1	1	1	1	1
	0	0	0	<b>0</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	0	0	0	0	1	1	1
	0	0	1	<b>1</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>1</b>	0	0	1	1	0	0	1
	0	1	0	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>	0	1	0	1	0	1	0
$f_3(\mathbf{x})$	6	6	4	<b>2</b>	<b>2</b>	<b>2</b>	4	<b>2</b>	10	10	8	6	4	4	6

En substituant avec l'identité  $x = 1 - \bar{x}$  chaque terme quadratique ayant un coefficient négatif, on a la posiforme suivante :

$$\phi_3(\mathbf{x}) = -4 + 4\bar{x}_1 + 6\bar{x}_3 + 2\bar{x}_4 + 2\bar{x}_5 + 4x_1x_2 + 2\bar{x}_2x_3 + 4x_3x_4 + 2\bar{x}_4x_5.$$

Ces flots pouvant être poussés séquentiellement, ce sont des chemins augmentants. Ces chemins augmentants correspondent respectivement aux sommes alternées suivantes :

$$\bar{x}_1 + x_1x_2 + \bar{x}_2x_3 + \bar{x}_3 \text{ et } \bar{x}_3 + x_3\bar{x}_2 + x_2x_1 + \bar{x}_1, \tag{5.21}$$

$$\bar{x}_3 + x_3x_4 + \bar{x}_4 \text{ et } \bar{x}_4 + x_4x_3 + \bar{x}_3, \tag{5.22}$$

$$\bar{x}_3 + x_3x_4 + \bar{x}_4x_5 + \bar{x}_5 \text{ et } \bar{x}_5 + x_5\bar{x}_4 + x_4x_3 + \bar{x}_3. \tag{5.23}$$

On remarque qu'il n'y a pas d'autre chemin augmentant dans le graphe Fig. 5.2(a). Le nombre de chemins augmentants maximum a été atteint, et la valeur du flot maximum est donc de  $\zeta^* = 6$ . Le graphe réduit peut être construit comme montré en figure 5.2(b). Le graphe réduit correspond à la posiforme quadratique suivante :

$$\psi_3(\mathbf{x}) = 2\bar{x}_1 + 2x_1x_2 + 2\bar{x}_1\bar{x}_2 + 2x_2\bar{x}_3 + 4\bar{x}_3\bar{x}_4 + 2x_4\bar{x}_5$$

qu'il est aussi possible de retrouver en faisant algébriquement la réduction sur la posiforme  $\phi_3$ .

Comme le flot maximum est  $\zeta^* = 6$ , on a :  $\phi_3(\mathbf{x}) = C(\phi_3) + \zeta^* + \psi_3(\mathbf{x}) = 2 + \psi_3(\mathbf{x})$ , d'où  $C_2(f_3) = 2$ . Le graphe réduit figure 5.2(b) a deux sommets directement atteignables à partir de la source : le sommet 1 et  $\bar{2}$ . D'après le théorème 5.4, les affectations  $x_1 = 1$  et  $x_2 = 0$  sont donc vérifiées pour chaque minimum global de  $f_3$ . Nous pouvons donc affecter ces valeurs à  $\psi_3(\mathbf{x})$ . Ainsi, minimiser  $f_3$  équivaut maintenant à minimiser  $\psi_3^*(\mathbf{x}) = 4\bar{x}_3\bar{x}_4 + 2x_4\bar{x}_5$ . Les affectations ( $\mathbf{x}$ ) qui annulent cette fonction sont persistantes faibles. Ceci se vérifie sur la table 5.2.

## 5.2.7. Cas sous-modulaire

### 5.2.7.1. Définition

Une fonction  $f$  sur les ensembles est sous-modulaire si :

$$f(X) + f(Y) \geq f(X \cup Y) + f(X \cap Y) \quad \forall X, Y \quad (5.24)$$

est respecté pour tous les sous-ensembles  $X, Y$ . Le pendant algébrique de cette définition, pour une fonction pseudo-booléenne  $f : \mathbf{B}^n \rightarrow \mathbb{R}$  quadratique, est :

$$\frac{\partial f}{\partial x_i \partial x_j}(\mathbf{x}) \leq 0. \quad (5.25)$$

Autrement dit, une fonction pseudo-booléenne quadratique est sous-modulaire si, sous sa forme multilinéaire polynomiale unique (5.7), ses dérivées secondes sont toutes négatives ou nulles. Cela implique que tous ses termes quadratiques ont un coefficient de signe négatif. Nous allons maintenant voir que trouver le minimum global d'une fonction sous-modulaire est aussi équivalent à la recherche de la coupe minimale d'un graphe, avec certaines simplifications par rapport au cas général de la *roof duality* vu précédemment.

### 5.2.7.2. Posiforme quadratique sous-modulaire

Quand une fonction est sous-modulaire, il est toujours possible de l'écrire comme une posiforme  $\phi$  sous la forme (à une constante près) :

$$\phi(\mathbf{x}) = \sum_{i \in P} a_i x_i + \sum_{j \in N} a_j \bar{x}_j + \sum_{1 \leq i \leq j \leq n} a_{ij} x_i \bar{x}_j \quad (5.26)$$

où  $P, N \subseteq \mathbf{V}$  et tous les coefficients  $a_i$  ( $i \in P \cup N$ ) et  $a_{ij}$  ( $1 \leq i \leq j \leq n$ ) sont positifs.

### 5.2.7.3. Construction du graphe induit

Comme expliqué dans la partie 5.2.6.1 précédente, il est possible de construire le graphe induit de la posiforme 5.26. Comme il n'y a pas de terme en  $x_i x_j$  ou  $\bar{x}_i \bar{x}_j$ , il n'y a pas d'arcs entre les sommets  $i$  et  $\bar{i}$  dans le graphe induit et par conséquent, le graphe réduit peut être partagé en deux graphes indépendants et symétriques l'un de l'autre. Cela implique qu'il est possible d'associer à  $\phi$  le graphe plus simple  $N_\phi = (V, A)$  construit de la façon suivante :

- l'ensemble des sommets est  $V = \{s, t\} \cup \mathbf{V}$ .



— l'ensemble des arcs est

$$A = \{ \overrightarrow{(s, j)} \mid j \in N \} \cup \{ \overrightarrow{(i, t)} \mid i \in P \} \cup \{ \overrightarrow{(i, j)} \mid 1 \leq i < j \leq n \}$$

où les capacités des arcs sont respectivement  $c_{sj} = a_j$  pour  $j \in N$ ,  $c_{i,t} = a_i$  pour  $i \in P$  et  $c_{ij} = a_{ij}$  pour  $1 \leq i < j \leq n$ .

En d'autres termes, l'ensemble des sommets  $V$  est composé de l'ensemble des indices  $\mathbf{V}$  ainsi qu'une source  $s$  et un puits  $t$ . On construit, pour chaque terme unitaire, un arc. Quand le terme est une variable  $x_i$ , un arc du sommet  $i$  au puits  $t$  est créé, avec comme poids le coefficient  $a_i$ . Quand le terme est le complément d'une variable  $\bar{x}_j$ , un arc de la source  $s$  au sommet  $j$  est créé, avec comme poids le coefficient  $a_j$ . Enfin, on construit pour chaque terme  $a_{ij}x_i\bar{x}_j$  un arc entre  $i$  et  $j$ . C'est ce type de graphe induit qui est construit plus habituellement dans les méthodes dites par coupe de graphe, voir par exemple (KOLMOGOROV et al., 2002).

#### 5.2.7.4. Minimisation par coupe de graphe

Il y a une correspondance entre une coupe qui sépare la source  $s$  et le puits  $t$  dans le graphe et le vecteur binaire  $\mathbf{x} \in \mathbf{B}^n \leftrightarrow S_{\mathbf{x}} \stackrel{\text{def}}{=} \{s\} \cup \{j \mid x_j = 1\}$ . Il est donc facile de voir qu'avec cette définition, on a pour tout  $\mathbf{x} \in \mathbf{B}^n$  :

$$\phi(\mathbf{x}) = \sum_{\substack{i \in S_{\mathbf{x}} \\ j \notin S_{\mathbf{x}}}} a_{i,j} \quad (5.27)$$

avec la convention que seuls les arcs traversants de  $s$  à  $t$  comptent. Un minimum de  $\phi(\mathbf{x})$  correspond à une coupe de poids minimal de  $N_{\phi}$ . Une coupe minimale peut être trouvée par une méthode de maximisation du flot à travers le graphe  $N_{\phi}$ . Cela permet donc de trouver une affectation de toutes les variables de  $\mathbf{x}$  qui atteint la valeur minimale de  $\phi$ .

**Exemple 5.4.** On cherche à minimiser la posiforme suivante :

$$\phi_4(\mathbf{x}) = 3\bar{x}_1 + 6x_1 + 4\bar{x}_2 + x_2 + 2x_2\bar{x}_1. \quad (5.28)$$

Il est aisé de vérifier que cette posiforme est sous-modulaire.

En se plaçant dans le cas général, avec les règles de la partie 5.2.6.1, on peut construire le graphe induit montré en figure 5.3(a). Ce graphe se décompose clairement en deux graphes indépendants. En suivant les règles précédentes, on construit alternativement le graphe induit simplifié dans le cas sous-modulaire, comme montré en figure 5.3(b).

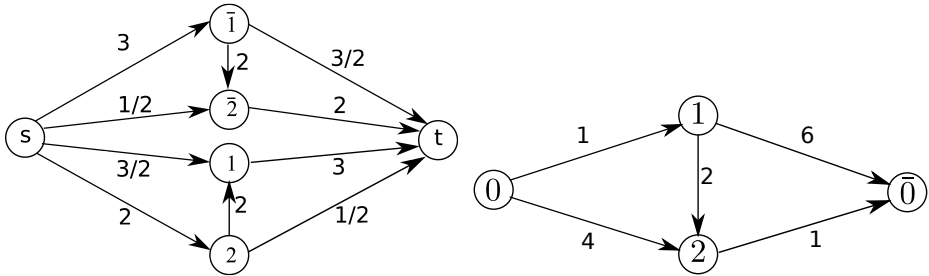
La figure 5.3(c) montre les quatre coupes possibles. Elles sont respectivement de poids 7, 6, 10 et 7. Par exemple pour la coupe  $c_2$  de la figure 5.3(c), son poids est de  $3 + 2 + 1 = 6$  car les arcs  $(s, 1)$ ,  $(1, 2)$   $(2, t)$  sont coupés, tous en allant de la source au puits.

Pour la coupe  $c_3$ , son poids est  $4 + 6 = 10$  car seuls les arcs  $(s, 2)$  et  $(1, t)$  sont coupés en allant de la source au puits. Les capacités des 4 coupes correspondent bien aux valeurs de la table de vérité de  $\phi_4$  en table 5.3.

En conséquence, la capacité de la coupe minimale dans le graphe induit figure 5.3(b)) est égale à la valeur minimum de la fonction  $\phi_4$  dans (5.28).

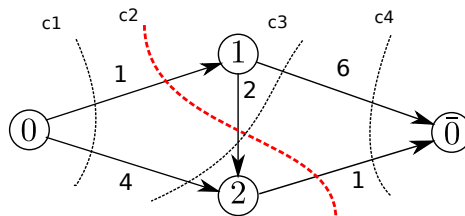
**Figure 5.3.**

Graphes induit dans le cas sous-modulaire quadratique de l'exemple 5.4



(a) Graphe induit dans le cas général

(b) Graphe induit dans le cas sous-modulaire



(c) Coupes possibles de poids  $w(c_1) = 7$ ,  $w(c_2) = 6$ ,  $w(c_3) = 10$  et  $w(c_4) = 7$ .

**Table 5.3.**

Table de vérité de la fonction pseudo-booléenne de l'exemple 5.4. Le minimum est en gras

<b>x</b>	<b>0 0</b> 1 1
	0 <b>1 0</b> 1
$f_4(\mathbf{x})$	7 <b>6</b> 10 7

La coupe minimale peut être obtenue en cherchant le flot maximum. Dans cet exemple, on voit que l'on peut faire passer un flot de valeur 2 de  $s$  à  $t$  en passant par les sommets 2 puis 1, un flot de 1 en passant par le sommet 2 et un flot de 3 en passant par le sommet 1. On vérifie donc que les 3 arcs coupés par la coupe  $c_2$  sont bien saturés par le flot maximum.

Avec la coupe minimum, seule le sommet 2 est du côté de la source  $s$ , donc  $x_2 = 1$ . Inversement, le sommet 1 étant du côté du puits  $t$ ,  $x_1 = 0$ . On retrouve bien la solution qui atteint la valeur minimum de  $\phi_4$  dans la table 5.3. Nous pouvons vérifier que l'on trouve le même résultat en procédant sur le graphe induit dans le cas général.

### 5.2.8. Optimisation de fonctions pseudo-booléennes de degré supérieur ou égal à 3

Il est difficile d'optimiser une fonction pseudo-booléenne de degré supérieur ou égal à 3, à cause de la présence de minima locaux, surtout, dans le cas non sous-modulaire. Pour optimiser de telles fonctions, deux approches ont néanmoins été proposées. La première approche consiste à décomposer une fonction de degré quelconque en fonction pseudo-booléenne quadratique possédant le même minimum et à l'optimiser avec la méthode par *roof duality*. La seconde approche est de généraliser directement le principe de *roof duality* au cas de degré strictement supérieur à deux (KAHL et al., 2011).

Ceci est difficile. Nous allons maintenant seulement résumer la méthode proposée dans (ISHIKAWA, 2009) qui est fondée sur le principe de substitution.

#### 5.2.8.1. Substitution

Une première méthode pour décomposer une fonction pseudo-booléenne d'ordre  $n$  à l'ordre 2 a été introduite en 1975 (ROSENBERG, 1975). Elle consiste à ajouter une variable auxiliaire afin de décomposer les termes de grands ordres en plusieurs termes d'ordres inférieurs. Cette méthode est itérative et fonctionne pour un degré quelconque.

Cette méthode de substitution qui permet, en remplaçant le produit de deux variables par une variable auxiliaire, de réduire le degré maximum d'une fonction pseudo-booléenne. On considère ces deux équivalences :

$$xy = z \iff xy - 2xz - 2yz + 3z = 0 \quad (5.29)$$

$$xy \neq z \iff xy - 2xz - 2yz + 3z > 0 \quad (5.30)$$

Les deux équivalences peuvent être facilement vérifiées en faisant les tables de vérité.

Il est donc possible, dans une fonction pseudo-booléenne, de remplacer le produit  $xy$  par  $z$  dans un terme de degré supérieur à deux, et d'ajouter l'expression  $xy - 2xz - 2yz + 3z$  comme un terme de pénalité, avec un grand coefficient multiplicatif. Alors, d'après (ROSENBERG, 1975), le vecteur minimisant la nouvelle équation permet d'obtenir le vecteur minimisant l'équation de départ en omettant les variables introduites en plus, ce qui se traduit par :

$$\min_{w,x,y} wxy = \min_{w,x,y,z} wz + \mu(xy - 2xz - 2yz + 3z) \quad (5.31)$$

où  $w, x, y$  sont les variables de départ et  $z$  est la variable ajoutée.  $\mu$  est le coefficient multiplicatif. Il doit être assez grand, pour cela, il peut être initialisé

**Table 5.4.**

Tables de vérité des fonctions pseudo-booléennes de l'exemple 5.5. Le minimum est en gras

<b>x</b>	0 0 0 0 1 1 1 1
	0 0 1 1 0 0 1 1
	0 1 0 1 0 1 0 1
$f_5(\mathbf{x})$	0 0 0 0 0 0 0 -1

<b>x</b>	0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1
	0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1
	0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1
	0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1
$f'_5(\mathbf{x})$	0 18 0 15 0 6 0 3 0 6 0 3 8 2 8 -1

avec  $\mu = 1 + \sum_{S \in \mathbf{v}} |c_S|$ . Cette étape de substitution peut être itérée jusqu'à ce que le degré maximum du résultat soit égal à deux.

**Exemple 5.5.** Soit la fonction pseudo-booléenne :

$$f_5(x_1, x_2, x_3) = 2x_1x_2 - 3x_1x_2x_3. \tag{5.32}$$

En substituant  $x_1x_2$  par une nouvelle variable  $x_4$  avec  $\mu = 1 + 2 + 3 = 6$ , la fonction (5.32) se réécrit en :

$$\begin{aligned} f'_5(x_1, x_2, x_3, x_4) &= 2x_1x_2 - 3 \underbrace{x_1x_2}_{x_4} x_3 + 6(x_1x_2 - 2x_1x_4 - 2x_2x_4 + 3x_4) \\ &= 18x_4 + 8x_1x_2 - 3x_3x_4 - 12x_1x_4 - 12x_2x_4. \end{aligned} \tag{5.33}$$

En observant la table de vérité 5.4 de la forme originale  $f_5$  (5.32) et de la forme quadratique  $f'_5$  (5.33), on constate que le minimum de  $f'_5$  est atteint pour le vecteur  $\mathbf{x} = (1111)$ . Si on omet la variable ajoutée  $x_4$  lors de la décomposition, le vecteur des variables d'origine est alors  $\mathbf{x} = (111)$  et il permet bien d'atteindre le minimum global de  $f_5$ , comme on le vérifie facilement dans la table de vérité.

### 5.2.8.2. Autres substitutions

Le principal défaut de la méthode par substitution précédente, est que, pour chaque substitution, un terme non sous-modulaire est ajouté. Au fil des itérations, le nombre de termes non sous-modulaires augmente donc. Or, le nombre d'affectations persistantes fortes lors du calcul de la *roof duality* dépend directement du nombre et des coefficients des termes quadratiques non sous-modulaires dans la forme multilinéaire polynomiale unique. Comme constaté dans (GALLAGHER et al., 2011), plus il y a de termes non sous-modulaire, et plus leurs coefficients sont importants, et moins il y a d'affectations fortes lors du calcul de la *roof duality*.

Après ce constat, de nouvelles méthodes ont été proposées afin de minimiser l'amplitude et le nombre des termes non sous-modulaire, donc de coefficient positifs, lors de la décomposition d'un degré quelconque en termes de degré quadratique (ISHIKAWA, 2009).

En effet, plusieurs types de décompositions sont donc possibles pour un même terme (KOLMOGOROV et al., 2002 ; ROSENBERG, 1975 ; ISHIKAWA, 2009).

Ces méthodes ont permis d'augmenter considérablement le nombre de persistances fortes lors du calcul de la *roof duality*, ce qui a permis de beaucoup mieux optimiser les problèmes d'origine. Néanmoins, le choix de la bonne substitution n'est pas simple. Dans (GALLAGHER et al., 2011), il est donc proposé, pour chaque terme, de choisir entre différents types de décompositions. Pour cela, une technique d'inférence est proposée pour permettre de choisir le type de décomposition en minimisant soit, le nombre d'arcs non sous-modulaires, soit l'amplitude des termes non sous-modulaires dans la fonction pseudo-booléenne finale. Cette méthode a l'avantage de tirer parti de chaque type de substitution et d'augmenter le nombre d'affectations persistantes fortes.

D'autres méthodes ont par la suite été proposées pour améliorer le taux d'affectation et la rapidité de la réduction. En effet, il est possible de décomposer directement une fonction de degré quelconque en un degré quadratique. Dans (ISHIKAWA, 2011), la méthode par substitution proposée fonctionne quel que soit le degré de la fonction.

Une autre méthode est proposée dans (FIX et al., 2011) par coupe de graphe.

## 5.3. Extension au cas multi-labels

### 5.3.1. Optimum global

Une des propriétés les plus remarquables est la possibilité de trouver en temps polynomial une solution qui minimise globalement une fonction pseudo-booléenne sous-modulaire. Cette propriété a pu être étendue au cas multi-labels lorsque la fonction ou l'énergie (5.1) contient au maximum des termes à deux variables et que la fonction réelle associée à ces termes est convexe (ISHIKAWA, 2003).

### 5.3.2. Stratégie de fusion

L'optimisation multi-labels d'une fonction discrète peut être approchée comme une succession de sous-problèmes binaires. En effet, si nous prenons deux vecteurs de labels, il est possible de trouver grâce aux méthodes d'optimisation binaires précédentes, quelle est la combinaison des composantes de ces deux vecteurs qui minimise la fonction. Ce procédé est nommé une *fusion* (LEMPITSKY et al., 2010).

La stratégie la plus courante est de considérer le vecteur  $\mathbf{L}^t$  des labels de la solution courante et un nouveau vecteur de labels  $\mathbf{L}^p$ . L'objectif de la fusion est de trouver la meilleure sélection des labels  $\mathbf{L}^{t+1}$ , c'est à dire celle pour laquelle l'énergie (5.1) est la plus faible. Pour chaque variable, un choix booléen doit être réalisé. Afin de combiner les vecteurs  $\mathbf{L}^t$  et  $\mathbf{L}^p$ , la combinaison  $\mathbf{L}^b(B)$  est définie par la fonction booléenne auxiliaire suivante :

$$\mathbf{L}(\mathbf{x}) = \bar{\mathbf{x}} \cdot \mathbf{L}^t + \mathbf{x} \cdot \mathbf{L}^p \quad (5.34)$$

où  $\mathbf{x}$  est un vecteur de variables booléennes et où le produit noté avec un point est terme à terme. Il y a donc une variable binaire  $x_i$  par site  $i$ . Quand  $x_i = 0$ , le label courant  $L_i^t$  est conservé pour le site  $i$ . Au contraire, quand  $x_i = 1$ , c'est le label proposé  $L_i^p$  qui est sélectionné.

Une fois que le meilleur  $\mathbf{x}^*$  est obtenu par optimisation quadratique pseudo-bouléenne, la meilleure combinaison de labels devient la solution courante :

$$\mathbf{L}^{t+1} = \mathbf{L}(\mathbf{x}^*). \quad (5.35)$$

Pour trouver  $\mathbf{x}^*$ , on doit minimiser la fonction suivante :

$$\phi(\mathbf{x}) = f(\mathbf{L}(\mathbf{x})). \quad (5.36)$$

Soit en utilisant la définition de  $f$  à partir de (5.1) :

$$\begin{aligned} \phi(\mathbf{x}) &= \sum_{u \in C_1} \vartheta_u(L_u^t) \bar{x}_u + \vartheta_u(L_u^p) x_u \\ &+ \sum_{u, v \in C_2} \vartheta_{uv}(L_u^t, L_v^t) \bar{x}_u \bar{x}_v + \vartheta_{uv}(L_u^t, L_v^p) \bar{x}_u x_v \\ &\quad + \vartheta_{uv}(L_u^p, L_v^t) x_u \bar{x}_v + \vartheta_{uv}(L_u^p, L_v^p) x_u x_v \\ &+ \sum_{u, v, w \in C_3} \dots \end{aligned} \quad (5.37)$$

Si la fonction pseudo-bouléenne (5.37) est sous-modulaire, alors toutes les variables de  $\mathbf{x}$  pourront être déterminées à cette étape. Si elle n'est pas sous-modulaire, seul un sous-ensemble de variable sera déterminé, dans un premier temps, en utilisant la *roof duality*. Dans un deuxième temps, les affectations persistantes faibles seront calculées en minimisant l'équation réduite. En cas

d'indétermination, plusieurs stratégies sont possibles, la plus simple est de garder la solution courante.

Si la fonction (5.37) a un degré supérieur à 2, elle peut être décomposée en une forme de degré quadratique, et ensuite optimisée comme vu précédemment.

### 5.3.3. Stratégie d'exploration

Dans la méthode itérative précédente, il faut choisir à chaque itération une proposition intéressante à fusionner. Ce choix est particulièrement critique car il détermine l'espace des solutions qui peuvent être explorées durant l'optimisation. Même si l'étape de fusion binaire trouve un minimum global car la fonction à minimiser est sous-modulaire, cela n'implique en rien que l'itération des fusions arrive à converger vers un minimum même local de la fonction multi-labels. En effet, c'est seulement un minimum local dans l'espace des solutions explorées qui sera obtenu, d'où l'importance de la stratégie d'exploration. La façon de choisir les propositions va aussi déterminer si les sous-problèmes binaires sont sous-modulaires ou pas.

#### 5.3.3.1. $\alpha$ -expansion

C'est la stratégie d'exploration dites par  $\alpha$ -expansion qui semble la plus efficace jusqu'à présent. La méthode par  $\alpha$ -expansion consiste à effectuer une succession de fusions, avec comme proposition  $\mathbf{L}^p$  un vecteur de valeurs identiques. Les itérations sont effectuées de façon cyclique sur les différentes valeurs de labels possibles jusqu'à stabilisation (BOYKOV et al., 2001).

On montre, comme par exemple dans (PAGET et al., 2015), qu'avec l' $\alpha$ -expansion, les sous-problèmes seront toujours sous-modulaires dès que l'énergie (5.1) contient au maximum des termes à deux variables et que la fonction réelle associée à ces termes est symétrique et concave pour les valeurs positives. Ainsi, chaque sous-problème binaire peut être optimisé de façon globale par une recherche de coupe de graphe minimale, ce qui permet d'assigner la totalité des labels (voir la partie 5.2.7.2). Si le minimum global de chaque sous-problème binaire est trouvé lors de la fusion, c'est seulement un minimum local de l'énergie multi-labels qui sera obtenu dans l'espace de recherche des  $\alpha$ -expansions, nommé  $\alpha$ -expansion move (BOYKOV et al., 2001).

Il faut noter que l'algorithme  $\alpha$ -expansion original n'est pas introduit comme nous venons de le faire avec l'aide de la programmation pseudo-booléenne. Cette façon de l'introduire permet de dériver une implantation originale et plus facile à programmer que la méthode  $\alpha$ -expansion d'origine.

#### 5.3.3.2. $\beta$ -jump

Une autre stratégie d'exploration dites par  $\beta$ -jump est possible même si elle n'est pas aussi efficace en pratique. La méthode par  $\beta$ -jump consiste à fusionner la solution courante avec une proposition qui est la solution courante décalée d'un même incrément des labels. L'intérêt du  $\beta$ -jump réside dans la condition suffisante que doit vérifier l'énergie pour conduire à des sous-problèmes sous-modulaires. Cette condition suffisante est que l'énergie contient au maximum des termes

à deux variables et que la fonction réelle associée à ces termes est convexe, voir (PAGET et al., 2015).

### 5.3.3.3. Alternée

Les conditions sur l'énergie avec  $\alpha$ -expansion et  $\beta$ -jump étant assez complémentaires, il est possible d'alterner les deux méthodes afin de traiter des énergies qui ne vérifient aucune de ces deux conditions, comme cela a été proposé dans (PAGET et al., 2015). Alternier  $\alpha$ -expansion et  $\beta$ -jump apparaît comme une approche alternative à l'utilisation de la *roof duality*.

**Exemple 5.6.** Soit l'énergie suivante :

$$f_6(l_1, l_2, l_3) = (l_1 - 2)^2 + (l_2 - 3)^2 + (l_3 - 2)^2 + \lambda|l_1 - l_2| + \lambda|l_2 - l_3|. \quad (5.38)$$

Cette énergie correspond à un problème de débruitage d'une image à  $3 \times 1$  pixels de valeurs d'intensité  $(2, 3, 2)$  avec un terme de régularisation de type  $L_1$  (ou variation totale) entre paires de voisins connexes. Les intensités prennent leurs valeurs ou labels dans l'intervalle des entiers  $[0, 4]$ . On cherche à optimiser cette énergie par  $\alpha$ -expansion à partir de la solution initiale  $\mathbf{L}^0 = (1, 3, 1)$ . Commençons par faire une expansion avec  $\alpha = 2$ . La solution proposée est donc  $\mathbf{L}^p = (2, 2, 2)$ . On constate que la combinaison booléenne auxiliaire peut être écrite comme  $\mathbf{L}(\mathbf{x}) = (2 - \bar{x}_1, 2 + \bar{x}_2, 2 - \bar{x}_3)$ . Le sous-problème binaire est donc après substitution de  $\mathbf{L}(\mathbf{x})$  dans  $f_6$  :

$$\begin{aligned} f_6(\mathbf{L}(\mathbf{x})) &= (\bar{x}_1)^2 + (\bar{x}_2 - 1)^2 + (\bar{x}_3)^2 + \lambda|\bar{x}_1 + \bar{x}_2| + \lambda|\bar{x}_2 + \bar{x}_3| \\ &= \bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \lambda(\bar{x}_1 + \bar{x}_2 + \bar{x}_2 + \bar{x}_3) \\ &= 1 + (1 + \lambda)\bar{x}_1 + (2\lambda - 1)\bar{x}_2 + (1 + \lambda)\bar{x}_3 \end{aligned}$$

après simplifications en utilisant le fait que le carré d'un variable binaire est égale à la variable et que la valeur absolue sur une valeur positive s'élimine. On déduit de la dernière expression que le minimum du sous-problème binaire est atteint pour  $\mathbf{x}^* = (1, 1, 1)$ . Le vecteur de labels sélectionné est donc  $(2, 2, 2)$ . En effectuant les autres étapes de l' $\alpha$ -expansion, on verra que cette solution est stable. C'est donc la solution finale qui sera obtenue par  $\alpha$ -expansion. Elle est d'énergie 1 et dans ce cas là, elle est de valeur minimale pour  $f_6$ .

## 5.4. Exemples d'utilisation

Afin d'illustrer l'intérêt des méthodes d'optimisation présentées dans ce chapitre, nous allons décrire trois applications, leur modèle, et la façon de les optimiser. Le premier exemple est celui de la segmentation binaire d'une image à mettre en relation avec les méthodes décrites dans le chapitre précédent. Ce problème est binaire et sous-modulaire. Le deuxième exemple est celui de la reconstruction 3D à partir de paires stéréoscopiques. Le problème est multi-labels et conduit à des sous-problèmes sous-modulaires. Enfin, le troisième et dernier exemple est le débruitage. Le problème est multi-labels et conduit à des problèmes non



sous-modulaires. D'autres exemples sont décrits, par exemple, dans (CARAFFA et TAREL, 2013a; CARAFFA et TAREL, 2013b; CARAFFA, 2013).

#### 5.4.1. Segmentation binaire d'une image

Le problème de la segmentation binaire d'une image a été traité dans le chapitre précédent avec une approche continue. Il est aussi possible de traiter de ce problème de façon discrète par coupe de graphe.

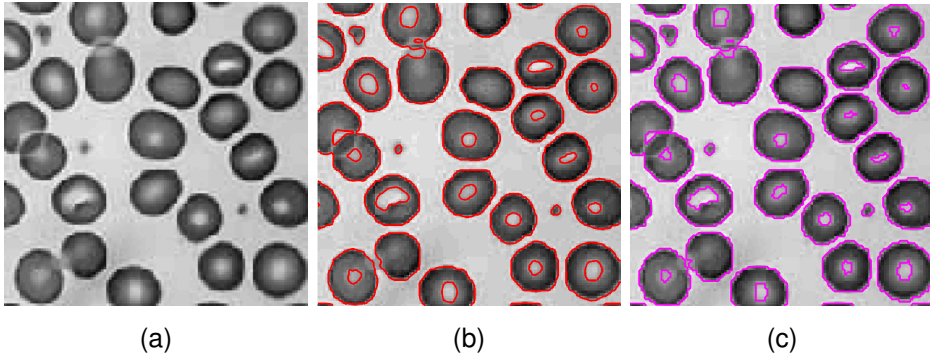
Ainsi, pour l'image  $I_{ij}, (i, j) \in \Omega$ , utilisée dans la figure 4.10, nous savons qu'elle peut-être segmentée en deux régions d'intensité moyenne  $\mu_{ext} = 86$  et  $\mu_{int} = 205$ . L'énergie discrète à minimiser est donc à partir de (4.17) :

$$f_7(\mathbf{x}) = \sum_{(i,j) \in \Omega} (I_{ij} - \mu_{int})^2 x_{ij} + (I_{ij} - \mu_{ext})^2 \bar{x}_{ij} + \nu |x_{ij} - x_{i+1,j}| + \nu |x_{ij} - x_{i,j+1}|. \quad (5.39)$$

Du fait de l'égalité binaire  $|x - y| = x\bar{y} + y\bar{x}$ , la fonction précédente est sous-modulaire et peut être minimisée globalement. Le résultat est montré en figure 5.4(c) et comparé au résultat de la méthode dite Split-Bergman qui donne aussi la solution au minimum global (figure 5.4(b)).

#### Figure 5.4.

L'image de la figure 4.10(a) est segmentée avec une méthode d'optimisation continue dite de Split-Bregman (b) et par un méthode discrète par coupe de graphe (c). Les mêmes valeurs de moyennes de région sont utilisés :  $\mu_{int} = 205$  et  $\mu_{ext} = 86$ . Les deux résultats sont assez proches. Le temps de calcul par coupe de graphe est de 40ms ( $\nu = 800$ )



#### 5.4.2. Reconstruction stéréoscopique

La reconstruction à partir de paires stéréoscopique consiste à retrouver le modèle 3D de la scène à partir de deux vues dont l'une est décalée horizontalement sur le plan focal par rapport à l'autre. Grâce à cette disposition, dite à géométrie épipolaire rectifiée, la projection d'un point de la scène 3D dans les deux caméras se fera sur la même ligne horizontale.

Si l'on connaît cette projection dans les deux images, alors, il est possible, par triangulation, de retrouver la profondeur. La différence de colonne d'une

projection entre les deux images s'appelle la disparité. L'ensemble des disparités d'une image est appelée la carte de disparité.

Supposons que nous avons deux images rectifiées  $I^g$  et  $I^d$  de la même scène. La reconstruction stéréoscopique consiste à estimer la carte de disparité  $D$  à partir des deux images. L'ensemble des pixels est noté  $\Omega$  et est indicé par le couple d'indices  $(i, j)$ .

L'énergie du modèle dans le cas le plus simple est la suivante :

$$f_s(D) = \sum_{(i,j) \in \Omega} (I_{ij}^g - I_{i(j+D_{ij})}^d)^2 + \nu |D_{ij} - D_{i(j+1)}| + \nu |D_{ij} - D_{(i+1)j}|. \quad (5.40)$$

Le premier terme de l'énergie représente le terme d'attache aux données, c'est à dire l'appariement entre l'image gauche  $I^g$  et l'image droite  $I^d$ . C'est simplement la différence de leurs intensités sachant la disparité. Quand l'intensité de l'image gauche est proche de celle de l'image droite, le score est faible.

Inversement, quand les deux intensités sont différentes, le score est fort. Le second terme de l'énergie est un terme de régularisation dit  $L_1$  ou variation totale qui a pour but de favoriser l'égalité des disparités entre pixels voisins.

Nous cherchons à optimiser l'énergie  $f_s$  par rapport à la variable discrète  $D$ . L'avantage de cette énergie est son caractère sous-modulaire. En effet, le terme d'attache aux données est unaire, c'est à dire que chaque terme est fonction d'une seule variable de  $D_{ij}$ . Le terme de régularisation est fonction de deux variables, paire et concave pour les valeurs positives. En conséquence, l'énergie (5.40) conduit toujours à des sous-problèmes sous-modulaires quand on utilise  $\alpha$ -expansion. Le terme de régularisation est aussi convexe, donc l'énergie (5.40) conduit toujours à des sous-problèmes sous-modulaires quand on utilise  $\beta$ -jump.

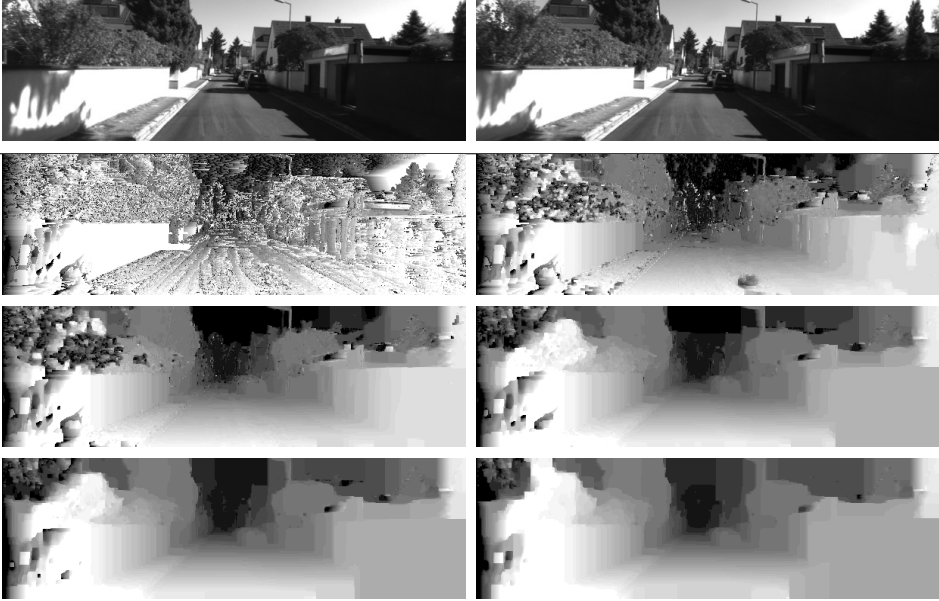
La figure 5.5 montre la solution minimale de l'énergie (5.40) avec l'algorithme  $\alpha$ -expansion pour différentes valeurs de  $\nu$  sur une paire stéréoscopique de la base *KITTI Vision Benchmark Suite*<sup>2</sup>. Nous pouvons voir que lorsque  $\nu = 0$ , c'est à dire sans régularisation, le résultat n'est pas bon à cause d'ambiguïtés. Avec  $\nu = 1$ , la carte de disparité obtenue est plus lisse. Il reste cependant quelques zones comme le bas de la chaussée mal reconstruites. Avec  $\nu = 4$ , la chaussée est plus lisse et donc mieux reconstruite.

Avec  $\nu = 16$ , la chaussée au loin est encore mieux reconstruite, cependant, un plan fronto-parallèle s'est formé au niveau de bord droit dû au fort poids du terme de régularisation. Ce phénomène est encore plus important avec  $\nu = 32$  et  $\nu = 64$ . Pour cette scène,  $\nu = 4$  semble donc être un bon compromis.

2. <http://www.cvlibs.net/datasets/kitti/>

**Figure 5.5.**

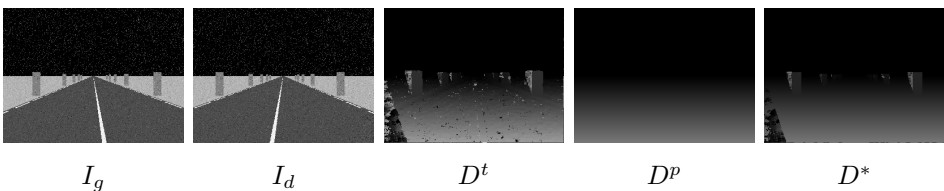
Résultat de reconstruction par  $\alpha$ -expansion sur (5.40) en fonction de la valeur de  $\nu$ . Première ligne : l'image gauche et droite de la scène. Seconde ligne :  $\nu = 0$  et  $\nu = 1$ . Troisième ligne :  $\nu = 4$  et  $\nu = 16$ . Quatrième ligne :  $\nu = 32$  et  $\nu = 64$



Pour pallier au problème de la reconstruction de la chaussée, il est possible de faire une étape de fusion entre la carte de disparité obtenue par minimisation de l'énergie (5.40) et une carte de disparité consistant au plan de la route. Le résultat de cette fusion est montré sur une autre paire d'images plus faciles à comprendre en figure 5.6. Nous constatons que les profondeurs sur les objets fronto-parallèles sont bien préservées et que la route est mieux reconstruite.

**Figure 5.6.**

Reconstruction raffinée par fusion avec un modèle a priori de la route. De gauche à droite : l'image gauche et droite, la carte de disparité obtenue par la minimisation de (5.40), la carte de disparité de la route et le résultat de la fusion



### 5.4.3. Débruitage d'une image

Le problème du débruitage d'une image consiste à enlever le bruit en chaque pixel. Cela peut se faire par lissage des zones uniformes en préservant les bords importants.

Étant donné une image bruitée  $I$ , et dans le cas d'un bruit supposé additif, indépendant et gaussien, le problème du débruitage peut être posé comme la minimisation de l'énergie suivante par rapport à la variable  $D$  qui représente l'image débruitée :

$$f_g(D) = \sum_{(i,j) \in \Omega} (I_{ij} - D_{ij})^2 + \lambda g(|D_{ij} - D_{i(j+1)}|) + \lambda g(|D_{ij} - D_{(i+1)j}|) \quad (5.41)$$

où  $g$  est une fonction croissante sur les valeurs positives. Le premier terme est le terme d'attache aux données qui empêche l'image  $D$  de s'éloigner trop de l'image bruitée  $I$ . Les deux derniers termes sont des termes de régularisation qui force  $D$  à être lisse. Lorsque  $g$  est l'identité, on retrouve à nouveau la régularisation  $L_1$  ou variation totale.

Si ce choix est assez efficace en terme de lissage, il a le défaut de faire apparaître des marches d'escaliers sur les zones où l'intensité varie de façon assez lente comme par exemple sur la joue de Lena dans la figure 5.7.

**Figure 5.7.**

Image originale (a) et un zoom (b). La même image bruitée par un bruit gaussien (c) et son zoom (d). L'image débruitée avec une régularisation  $L_1$  (e) et son zoom (f). Enfin, l'image débruitée avec une régularisation  $L_1$  modifiée par une petite cuvette en zéro (g) et son zoom (h). Le débruitage avec  $L_1$  est obtenu par  $\alpha$ -expansion et par une méthode alternant  $\alpha$ -expansion et  $\beta$ -jump dans le cas  $L_1$  modifiée



Il est donc préférable d'utiliser pour  $g$  une fonction  $L_1$  en modifiant la forme en zéro pour avoir une petite cuvette quadratique au lieu d'un angle droit. Cela permet d'obtenir des dégradés plus lisses comme illustré en figure 5.7. La

fonction n'étant plus concave, il faut utiliser la méthode alternée pour résoudre chaque sous-problème binaire.

## 5.5. Conclusion

Le cadre de l'optimisation des fonctions pseudo-booléennes a fait apparaître l'importance de la distinction entre fonction non sous-modulaire et fonction sous-modulaire. Dans ce dernier cas, des algorithmes en temps polynomial permettent de trouver un minimum global de l'énergie par exemple par coupe de graphe.

Mais, l'ensemble des fonctions sous-modulaires est assez restreint. En effet, dans le cadre binaire, seuls les énergies avec des termes d'ordres au maximum 2 à coefficients négatifs sont sous-modulaires et il est alors possible de les optimiser globalement en un temps polynomial. Les fonctions binaires avec des termes d'ordre supérieur à 2 ne sont généralement pas sous-modulaires. Il existe des techniques qui permettent de transformer une fonction pseudo-booléenne quelconque en une fonction pseudo-booléenne quadratique partageant le même minimum global. Mais ces techniques de transformation entraînent l'apparition de termes non sous-modulaires. Il n'est donc plus possible d'appliquer les algorithmes d'optimisation globale sous-modulaire.

Étonnamment, et malgré le caractère non sous-modulaire de la fonction à minimiser, certaines affectations de labels, appartenant au minimum global, peuvent néanmoins être trouvées en un temps polynomial. En effet, la structure éparsée de certains graphes permet de détecter, pour un sous-ensemble de variables, si l'affectation d'un label pour une variable donnée appartient au minimum global ou non. Cette propriété est importante car elle permet, même si le minimum global d'une énergie ne peut pas être atteint complètement du fait de sa non sous-modularité, d'identifier un sous-ensemble de variables pour lesquelles on peut trouver l'affectation appartenant au minimum global. La méthode pour le faire est la *roof duality*.

Pour passer à l'optimisation multi-labels, une heuristique est de décomposer le problème en une succession de sous-problèmes binaires, nommés fusions, où il faut choisir entre une solution courante, et une proposition de label. Chaque étape de fusion sélectionne une solution globale de l'énergie du sous-problème binaire, si elle est sous-modulaire. Pour autant, en itérant les fusions, il n'y a pas de garantie de converger vers un minimum local.

Enfin, nous avons vu que l'optimisation de fonctions pseudo-booléennes quadratiques apporte un cadre assez large pour optimiser les énergies rencontrées en traitement d'image et peut être aussi utilisée dans d'autres contextes.

Il faut noter que cette présentation succincte ne représente qu'une petite partie de l'optimisation des fonctions pseudo-booléennes. Pour aller plus loin, l'article (BOROS, Peter L. HAMMER et al., 2006) propose de nouvelles méthodes pour l'optimisation de fonctions pseudo-booléennes quadratiques tel que le *probing* qui permet d'améliorer le calcul de la *roof duality*. Autre exemple, la

thèse (STRANDMARK, 2012), soutenue en 2012, propose un état de l'art complet sur l'optimisation discrète appliquée au domaine de la vision par ordinateur.

### 5.5.1. Récapitulatif

Pour résumer, les points clés à retenir de l'optimisation des fonctions pseudo-booléennes sont :

- N'importe quelle fonction pseudo-booléenne peut être transformée en une posiforme.
  - Les affectations vérifiant la persistance forte avec la *roof duality* peuvent être trouvées en un temps polynomial par poussage de flot.
  - Les affectations persistantes faibles peuvent être trouvées en résolvant la posiforme réduite.
- Si la fonction pseudo-booléenne est sous-modulaire, alors on peut trouver un minimum global en un temps polynomial.
- Une fonction pseudo-booléenne quelconque peut toujours être réduite en une posiforme quadratique en ajoutant des variables supplémentaires, et ayant le même minimum global.
  - Il existe plusieurs réductions possibles.
  - La réduction entraîne l'apparition de termes non sous-modulaires.
  - Le nombre et les coefficients des termes non sous-modulaires dépendent directement de la réduction utilisée.
  - Le nombre et les coefficients des termes non sous-modulaires influent sur le nombre d'affectations persistantes fortes.
- Nous pouvons décomposer un problème multi-labels en une série de sous-problèmes binaires en utilisant la fusion et diverses stratégies d'exploration ( $\alpha$ -expansion,  $\beta$ -jump, alternée), mais on n'obtient alors pas forcément un minimum global.

### 5.5.2. Outils libres

Plusieurs implantations d'algorithmes sous licence libre sont disponibles pour résoudre les différents problèmes abordés dans ce chapitre :

- $\alpha$ -expansion : <http://vision.csd.uwo.ca/code/>
- Optimisation de fonction pseudo-booléenne quadratique : <http://pub.ist.ac.at/~vnk/software.html>
- Réduction de fonction pseudo-booléennes quelconques en quadratique : <http://www.f.waseda.jp/hfs/software.html>

## Bibliographie

**Endre Boros** et **Peter L Hammer**. « Pseudo-boolean optimization ». In : *Discrete applied mathematics* 123.1 (2002), pages 155-225. DOI : 10.1016/S0166-218X(01)00341-9.

**Endre Boros**, **Peter L. Hammer** et **Gabriel Tavares**. *Preprocessing of unconstrained quadratic binary optimization*. Rapport technique. Rutgers University, 2006.

- Yuri Boykov, Olga Veksler et Ramin Zabih.** « Fast Approximate Energy Minimization via Graph Cuts ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23 (11 nov. 2001), pages 1222-1239. DOI : 10.1109/34.969114.
- Laurent Caraffa.** « Reconstruction 3D à partir de paires stéréoscopiques en conditions dégradées ». Thèse de doctorat. Université Pierre et Marie Curie, 2013.
- Laurent Caraffa et Jean-Philippe Tarel.** « Markov Random Field Model for Single Image Defogging ». In : *Proceedings of IEEE Intelligent Vehicle Symposium (IV'2013)*. Gold Coast, Australia, juin 2013, pages 994-999.
- Laurent Caraffa et Jean-Philippe Tarel.** « Stereo Reconstruction and Contrast Restoration in Daytime Fog ». In : *Proceedings of Asian Conference on Computer Vision (ACCV'12)*. Tome IV. LNCS. Daejeon, Korea : Springer, 2013, pages 13-25.
- Alexander Fix, Aritanan Gruber, Endre Boros et Ramin Zabih.** « A graph cut algorithm for higher-order Markov Random Fields ». In : *Proceedings of the 2011 International Conference on Computer Vision. ICCV '11*. Washington, DC, USA : IEEE Computer Society, 2011, pages 1020-1027. ISBN : 978-1-4577-1101-5. DOI : 10.1109/ICCV.2011.6126347.
- Andrew C. Gallagher, Dhruv Batra et Devi Parikh.** « Inference for order reduction in Markov random fields ». In : *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. CVPR '11*. Washington, DC, USA : IEEE Computer Society, 2011, pages 1857-1864. ISBN : 978-1-4577-0394-2. DOI : 10.1109/CVPR.2011.5995452.
- Hiroshi Ishikawa.** « Exact optimization for Markov random fields with convex priors ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.10 (2003), pages 1333-1336. DOI : 10.1109/TPAMI.2003.1233908.
- Hiroshi Ishikawa.** « Higher-order clique reduction in binary graph cut ». In : *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*. IEEE, 2009, pages 2993-3000. ISBN : 978-1-4244-3992-8. DOI : 10.1109/CVPRW.2009.5206689.
- Hiroshi Ishikawa.** « Transformation of General Binary MRF Minimization to the First-Order Case ». In : *IEEE Transactions On Pattern Analysis And Machine Intelligence* 33.6 (2011), pages 1234-1249. DOI : 10.1109/TPAMI.2010.91.
- Fredrik Kahl et Petter Strandmark.** « Generalized roof duality for pseudo-boolean optimization ». In : *Proceedings of the 2011 International Conference on Computer Vision. ICCV '11*. Washington, DC, USA : IEEE Computer Society, 2011, pages 255-262. ISBN : 978-1-4577-1101-5. DOI : 10.1109/ICCV.2011.6126250.
- Vladimir Kolmogorov et Ramin Zabih.** « What Energy Functions Can Be Minimized via Graph Cuts ? » In : *Proceedings of the 7th European Conference on Computer Vision-Part III. ECCV '02*. London, UK, UK : Springer-Verlag, 2002, pages 65-81. ISBN : 3-540-43746-0.

**Victor Lempitsky, Carsten Rother, Stefan Roth et Andrew Blake.** « Fusion Moves for Markov Random Field Optimization ». In : *IEEE Transactions On Pattern Analysis And Machine Intelligence* 32.8 (août 2010), pages 1392-1405. ISSN : 0162-8828. DOI : 10.1109/TPAMI.2009.143.

**Mathias Paget, Jean-Philippe Tarel et Laurent Caraffa.** « Extending  $\alpha$ -expansion to a larger set of regularization functions ». In : *Proceedings of International Conference on Image Processing (ICIP'15)*. <http://perso.lcpc.fr/tarel.jean-philippe/publis/icip15.html>. Quebec City, Canada, 2015.

**Ivo G. Rosenberg.** « Reduction of bivalent maximization to the quadratic case. » In : *Cahiers du Centre d'Etudes de Recherche Operationnelle* (1975), 17 :7174.

**Petter Strandmark.** « Discrete Optimization in Early Vision ». Thèse de doctorat. Centre for Mathematical Sciences LTH, Lund University, Sweden, 2012.



# Conclusion

Les différents chapitres qui composent cet ouvrage ont permis de mettre en évidence comment des problèmes d'optimisation de forme apparaissent naturellement dans différents domaines d'application tels que la conception de structures (chapitres 1-2), l'analyse d'images (chapitres 4-5), ou l'étude du comportement de matériaux à changement de phase (chapitre 3). Des traits communs émergent des différentes problématiques qui ont été abordées. Ainsi, les passages entre formulation discrète et formulation continues ont souvent été discutées : la formulation la plus naturelle du problème est souvent de nature discrète, au sens où elle porte sur des quantités à valeurs dans  $\{0, 1\}$  (ou  $\{0, n\}$ ). En conception de structures, la quantité mise en jeu est de nature binaire et correspond à la présence ou non de matière au point  $x$ . En analyse d'images, la quantité mise en jeu est le niveau de gris en chaque pixel. Pour l'étude des transformations de phase, la quantité mise en jeu est à valeurs dans  $\{0, n\}$  et indique la phase présente au point  $x$ .

Afin de faciliter la résolution, certaines des approches présentées consistent à passer de la formulation discrète à une formulation continue, c'est-à-dire à une formulation où la quantité mise en jeu peut varier continûment dans un intervalle (par exemple). C'est le cas de la relaxation convexe présentée au chapitre 4 ou encore des méthodes d'optimisation topologique présentées au chapitre 1. Pour certains des problèmes étudiés, la quantité continue ainsi introduite bénéficie d'une interprétation physique claire (par exemple, en conception de structures, la densité locale de matière). Notons que ce passage du discret vers le continu n'est pas systématiquement nécessaire : les méthodes de l'optimisation quadratique pseudo-booléenne (chapitre 5) portent directement sur la formulation discrète. Signalons également que les passages entre formulation discrète et continue se font dans les deux sens. Ainsi, pour les problèmes d'optimisation de structure (chapitre 1 par exemple), une étape de "discrétisation" est effectuée sur la solution du problème continu afin de générer une forme constructible.

Une autre caractéristique commune des problèmes abordés est qu'ils font apparaître des problèmes de minimisation "mal posés", au sens où la fonction à minimiser (qui s'interprète souvent à une énergie) possèdent de nombreux minima locaux. Les minima globaux (qui sont ceux recherchés) et les minima locaux ne coïncident que dans quelques cas particuliers.

En optimisation quadratique pseudo-booléenne (chapitre 5), cette situation particulière correspond au cas d'une fonction énergie sous-modulaire. Dans l'étude des matériaux à changement de phase, elle correspond au cas d'une énergie convexe (ce qui est vérifié seulement lorsque les phases sont géométriquement compatibles deux à deux, au sens explicité au chapitre 3). Ces situations très particulières ne couvrent pas l'éventail des problèmes rencontrés en pratique. Aussi, la multiplicité des minima locaux demeure une difficulté dans la plupart des problèmes traités, ce qui se traduit par une dépendance de la solution obtenue vis-à-vis de l'estimation initiale choisie pour démarrer la résolution.

Au travers de différents domaines d'application, plusieurs méthodes ont été abordées : optimisation topologique, algorithmes génétiques, level-sets, optimisation quadratique pseudo-booléenne pour ne citer que quelques exemples. Il convient de noter que ces méthodes ne sont pas restreintes au domaine d'application dans lequel elles ont été présentées dans cet ouvrage. De par la similitude déjà soulignée des problèmes mathématiques sous-jacents, les méthodes numériques utilisées en traitement d'images (par exemple) pourraient probablement être pertinentes en conception de structures. De fait, les méthodes de type level-sets (décrites au chapitre 5) sont actuellement de plus en plus utilisées en conception de structures. Il pourrait en aller de même d'autres méthodes.

Notons finalement que les progrès constants en outils de calcul (matériels et logiciels) et de traitement de données ouvrent de nouveaux horizons. Il en va de même des techniques de fabrication additives, actuellement en plein essor, qui offrent un moyen de construire les formes optimales issues d'un calcul de conception de structures, sans fortes contraintes de fabricabilité.

# Liste des figures

1.1. Chargement mécanique pour le problème de dimensionnement .	13
1.2. Recherche d'un pont optimal. Positionnement du problème . . .	19
1.3. Recherche d'un pont optimal. Solution relaxée $\rho$ obtenue par la méthode d'homogénéisation . . . . .	20
1.4. Recherche d'un pont optimal. Solution $\tilde{\rho}$ obtenue par homogénéisation, après filtrage . . . . .	20
1.5. Optimisation d'une poutre console . . . . .	22
1.6. Solutions obtenues par la méthode SIMP, pour deux états initiaux différents. Maillage à 4586 éléments . . . . .	22
1.7. Solutions obtenues par la méthode d'homogénéisation, avant (gauche) et après (droite) filtrage. Maillage à 4586 éléments . . .	22
1.8. Solutions obtenues par la méthode d'homogénéisation, avant (gauche) et après (droite) filtrage. Maillage à 8201 éléments . . .	23
1.9. Solutions obtenues par la méthode d'homogénéisation, avant (gauche) et après (droite) filtrage. Maillage à 12913 éléments . .	23
1.10. Dimensionnement robuste d'une poutre en flexion . . . . .	28
1.11. Représentation de la fonction $a(r)$ (gauche) et d'une suite maximisante (droite) pour un problème de dimensionnement robuste . . . . .	36
1.12. Chargement maximal en fonction du niveau d'incertitude sur le module d'Young . . . . .	37
1.13. États initiaux (notés $E_0, E_1, E_2$ ) utilisés . . . . .	39
1.14. Distributions critiques obtenues pour chacun des 3 états initiaux $E_0, E_1, E_2$ . . . . .	39
1.15. Distributions critiques obtenues pour chacun des 3 états initiaux $E_0, E_1, E_2$ (problème régularisé) . . . . .	41

1.16. Calcul bidimensionnel par éléments finis . . . . .	42
2.1. Représentation symbolique du domaine $\Omega$ ainsi que d'un objet plongé dans $\Omega$ de contour $\Gamma$ . . . . .	47
2.2. Courbes d'absorption des 9 matériaux retenus pour l'optimisation	50
2.3. Exemple d'optimisation avec un seul critère . . . . .	58
2.4. Exemple d'optimisation avec 2 critères . . . . .	59
2.5. Schéma récapitulatif d'une optimisation globale, une fois les paramètres d'environnement fixés . . . . .	63
2.6. Résultat de l'optimisation globale en pondérant l'indicateur ACOU à 100% . . . . .	64
2.7. Résultat de l'optimisation globale en pondérant l'indicateur ENV à 100% . . . . .	65
2.8. Résultat de l'optimisation globale en pondérant l'indicateur COST à 100% . . . . .	65
3.1. Transformation cubique-tétraogonale . . . . .	70
3.2. Transformation tétraogonale-orthorombique . . . . .	71
3.3. Interprétation de l'effet mémoire de forme en termes de changement de phase ( $T$ désigne la température) . . . . .	72
3.4. Répartition géométrique des phases dans le domaine $\Omega$ (ici carré).	76
3.5. Exemple de déformations (à gauche la configuration avant déformation, à droite la configuration après déformations) . . . . .	83
3.6. Valeurs de $(\omega, \tau)$ telles que $\bar{e}(\omega, \tau) \in Q\mathcal{K}$ ( CuAlNi) . . . . .	83
3.7. Structure laminée . . . . .	84
3.8. Structure laminée de rang 2 . . . . .	87
3.9. Bornes par laminations de rang 1 pour un problème à 4 phases .	90
3.10. Bornes par laminations de rang 2 pour un problèmes à 4 phases	90
3.11. Bornes par laminations de rang 3 pour un problèmes à 4 phases	91
3.12. Représentation schématique des bornes pour les problèmes à 12 phases . . . . .	97
3.13. Bornes sur les valeurs $(\omega, \tau)$ telles que $\bar{e}(\omega, \tau) \in Q\mathcal{K}^I$ . . . . .	99
3.14. Bornes sur les valeurs $(\omega, \tau)$ telles que $\bar{e}(\omega, \tau) \in Q\mathcal{K}^{II}$ . . . . .	99
4.1. Exemples de modèles déformables. . . . .	104

4.2. (a) Illustration de la force interne $F_{INT}$ dans le cas purement élastique. (b) Illustration de la force image. . . . .	106
4.3. Partitionnement du domaine image . . . . .	110
4.4. Interprétation de la force image . . . . .	111
4.5. Exemple de segmentation par contour actif orienté-region. . . . .	112
4.6. Exemple de suivi de ligne de fissure par contour actif géodésique entre plusieurs extrémités. . . . .	116
4.7. Application de la méthode de Kaul <i>et al.</i> sur une image synthétique	117
4.8. (a) Image originale; (b) image « dépliée » de manière radiale à partir du point marqué d'une croix jaune dans (a). . . . .	118
4.9. Découpage du plan image dans le cas d'un objet convexe (d'après (APPLETON et TALBOT, 2005)) . . . . .	119
4.10. Initialisation pour les algorithmes Level Sets . . . . .	124
5.1. Équivalence entre la réduction de la posiforme et le poussage de flot sur le graphe induit. . . . .	147
5.2. (a) montre le graphe induit de l'exemple 5.3 et (b) le graphe réduit	148
5.3. Graphes induits dans le cas sous-modulaire . . . . .	152
5.4. Méthodes de segmentation binaire . . . . .	159
5.5. Résultat de reconstruction par $\alpha$ -expansion sur (5.40) en fonction de la valeur de $\nu$ . . . . .	161
5.6. Reconstruction raffinée par fusion avec un modèle a priori de la route. . . . .	161
5.7. Méthodes de débruitage . . . . .	162



# Liste des tables

1.1. Résultats numériques ( $r = 1.5$ ) . . . . .	39
1.2. Valeur maximale de la contrainte de Von Mises . . . . .	42
2.1. Exemple de coût de construction et de démolition en fonction de la hauteur de l'écran . . . . .	55
2.2. Exemple de coût de maintenance des matériaux utilisés . . . . .	55
2.3. Critères environnementaux liés aux matériaux considérés . . . . .	56
2.4. Description de 4 familles d'écran . . . . .	57
2.5. Seuils de performance à atteindre en termes de gain relatif pour les trois critères acoustiques . . . . .	57
2.6. Nombre de générations de populations concernant l'optimisation	59
2.8. Système de notation pour l'indicateur ACOU . . . . .	61
2.9. Système de notation pour l'indicateur ENV . . . . .	61
2.10. Système de notation pour l'indicateur COST . . . . .	61
2.11. Possibilités considérées pour l'optimisation extrinsèque . . . . .	62
3.1. Transformation cubique-orthorombique . . . . .	81
3.2. Transformation cubique-monoclinique-I . . . . .	96
3.3. Transformation cubique-monoclinique-II . . . . .	96
3.4. Mesures des différentes bornes (exprimées en fonction de $ CK $ )	97
5.1. Table de vérité de la fonction pseudo-booléenne de l'exemple 5.1	141
5.2. Table de vérité de la fonction pseudo-booléenne de l'exemple 5.3	149
5.3. Table de vérité de la fonction pseudo-booléenne de l'exemple 5.4	152
5.4. Tables de vérité des fonctions pseudo-booléennes de l'exemple 5.5.	154

# Fiche bibliographique

<b>Collection</b> Ouvrages scientifiques		
<b>ISSN</b> 2558-3018	<b>ISBN</b> Papier 978-2-85782-743-6 PDF 978-2-85782-744-3	<b>Réf.</b> OSI3
<b>Titre</b> Optimisation de formes en sciences de l'ingénieur		
<b>Sous-titre</b> Méthodes et applications		
<b>Cordinateur</b> Michaël PEIGNEY <b>Auteurs</b> Michael PEIGNEY, Christophe HEINKELE, Thomas LEISSING, Jérôme DEFRANCE, Pierre CHARBONNIER, Jean-Philippe TAREL, Laurent CARAFFA, Mathias PAGET		
<b>Date de publication</b> Novembre 2018	<b>Langue</b> Français	
<b>Résumé</b> Ce recueil donne un aperçu des méthodes d'optimisation de formes et de leurs applications en sciences de l'ingénieur, en mettant l'accent sur quelques contributions récentes de l'IFSTTAR, du CEREMA, et du CSTB. Les applications présentées concernent des thématiques physiques autour des matériaux et structures ainsi que des thématiques en traitement d'images numériques. Au fil de cet ouvrage, nous mettons en évidence certains liens et similitudes qui existent entre des problématiques au premier abord très différentes. Cet ouvrage a été rédigé avec un souci pédagogique constant pour être accessible au lecteur non spécialiste.		
<b>Mots clés</b> optimisation de formes – éco-conception de stuctures – traitement d'images – matériaux à changement de phase		
<b>Nbre de pages</b> 175 p.	<b>Prix</b> Papier Gratuit PDF Gratuit	



# Publication data form

<b>Collection</b> Scientific book		
<b>ISSN</b> 2558-3018	<b>ISBN</b> Paper 978-2-85782-743-6 PDF 978-2-85782-744-3	<b>Ref.</b> OSI3
<b>Title</b> Shape optimization in engineering sciences		
<b>Subtitle</b> Methods and applications		
<b>Editor</b> Michaël PEIGNEY		
<b>Author(s)</b> Michaël PEIGNEY, Christophe HEINKELE, Thomas LEISSING, Jérôme DEFRANCE, Pierre CHARBONNIER, Jean-Philippe TAREL, Laurent CARAFFA, Mathias PAGET		
<b>Publication date</b> November 2018	<b>Language</b> French	
<b>Abstract</b> This collective book aims at presenting some general methods used in shape optimization together with relevant applications in engineering science. Special emphasis is put on recent research results achieved by IFSTTAR, CEREMA and CSTB in the field. The applications presented cover material sciences, structural design, and numerical images processing. There exist some deep connections between those seemingly far apart applications, as will be shown throughout the book. This book has been intended to be accessible to non-specialist readers wanting to learn about shape optimization.		
<b>Keywords</b> shape optimization – eco-design of structures – numerical images processing – phase transformation		
<b>Page number</b> 175 p.	<b>Price</b> Paper Free PDF Free	



**C**e recueil donne un aperçu des méthodes d'optimisation de formes et de leurs applications en sciences de l'ingénieur, en mettant l'accent sur quelques contributions récentes de l'IFSTTAR, du CEREMA et du CSTB. Les applications présentées concernent des thématiques physiques autour des matériaux et structures ainsi que des thématiques en traitement d'images numériques. Au fil de cet ouvrage, nous mettons en évidence certains liens et similitudes qui existent entre des problématiques au premier abord très différentes. Cet ouvrage a été rédigé avec un souci pédagogique constant pour être accessible au lecteur non spécialiste.

**T**his collective book aims at presenting some general methods used in shape optimization together with relevant applications in engineering science. Special emphasis is put on recent research results achieved by IFSTTAR, CEREMA and CSTB in the field. The applications presented cover material sciences, structural design, and numerical images processing. There exist some deep connections between those seemingly far apart applications, as will be shown throughout the book. This book has been intended to be accessible to non-specialist readers wanting to learn about shape optimization.

*Michaël PEIGNEY est ingénieur en chef des Ponts, Eaux et Forêts et chercheur au laboratoire Navier. Il enseigne à l'École Polytechnique, à l'École des Ponts ParisTech et à l'ENSTA. Ses thèmes de recherche concernent la mécanique des solides, notamment le développement de méthodes numériques en mécanique non linéaire, les approches multi-échelles et l'étude des transformations de phase.*

Illustration de la couverture : résultats d'un calcul numérique d'optimisation de formes pour un pont bidimensionnel (Source IFSTTAR)



**IFSTTAR**

**LES COLLECTIONS DE L'IFSTTAR**



ISBN : 978-2-85782-744-3  
ISSN : 2558-3018  
Réf : OSI3  
Novembre 2018